TITLE: MEF2C controls segment-specific gene regulatory networks that direct heart tube morphogenesis

AUTHOR LIST: Jonathon M. Muncie-Vasic[1], Tanvi Sinha[2], Alexander P. Clark[3], Emily F. Brower[1], Jeffrey J. Saucerman[3,4], Brian L. Black[2,5], Benoit G. Bruneau[1,6,7,8,9,*]

AFFILIATIONS: [1]Gladstone Institutes, San Francisco, CA; [2]Cardiovascular Research Institute, University of California, San Francisco, CA; [3]Department of Biomedical Engineering, University of Virginia, Charlottesville, VA; [4]Robert M. Berne Cardiovascular Research Center, University of Virginia, Charlottesville, VA; [5]Department of Biochemistry and Biophysics, University of California, San Francisco, CA; [6]Roddenberry Center for Stem Cell Biology and Medicine, Gladstone Institutes, San Francisco, CA; [7]Department of Pediatrics, University of California, San Francisco, CA; [8]Institute for Human Genetics, University of California, San Francisco, CA; [9]Eli and Edythe Broad Center for Regeneration Medicine and Stem Cell Research, University of California, San Francisco, CA; *Author for correspondence (benoit.bruneau@gladstone.ucsf.edu)

ABSTRACT

The gene regulatory networks (GRNs) that control early heart formation are beginning to be understood, but lineage-specific GRNs remain largely undefined. We investigated networks controlled by the vital transcription factor MEF2C, with a time course of single-nucleus RNA- and ATAC-sequencing in wild-type and *Mef2c*-null embryos. We identified a "posteriorized" cardiac gene signature and chromatin landscape in the absence of MEF2C. Integrating our multiomics data in a deep learning-based model, we constructed developmental trajectories for each of the outflow tract, ventricular, and inflow tract segments, and alterations of these in *Mef2c*-null embryos. We computationally identified segment-specific MEF2C-dependent enhancers, with activity in the developing zebrafish heart. Finally, using inferred GRNs we discovered that the *Mef2c*-null heart malformations are partly driven by increased activity of the nuclear hormone receptor NR2F2. Our results delineate lineage-specific GRNs in the early heart tube and provide a generalizable framework for dissecting transcriptional networks governing developmental processes.

A recurrent observation in developmental biology is that the same transcription factor (TF) can play different roles depending on the cells or developmental stage in which it is expressed. The cell type- or stage-specific activity of a TF is dependent on a number of factors, including its own expression level, the availability of potential co-factors, the presence or absence of other TFs that may act co-operatively or antagonistically, and the dynamic state of the chromatin landscape in which the TF is acting[1–4]. Together, these phenomena can be conceptualized as transcriptional or gene regulatory networks (GRNs) that can be used to understand the role of a particular TF within a given developmental context[5]. However, for most TFs, the precise GRNs underlying their pleiotropic activity in distinct cell types and at different developmental timepoints remain undefined.

The developing heart presents an excellent model for exploring this question. Several TFs critical for cardiac development have been identified and their expression patterns are well-defined[6–8]. Moreover, lineage tracing studies have revealed distinct progenitor cell types that contribute to the various segments of the linear heart tube[7–11] (Fig. 1a). To explore in detail how a specific TF can have distinct roles within different parts of a developing organ, we focused on a key cardiac transcription factor, MEF2C[12], and its role in each of the developing heart tube segments. In mice, loss of MEF2C leads to cardiac defects and embryonic lethality by mid-gestation at embryonic day (E) 10.5[13]. Interestingly, although MEF2C is expressed in both first heart field (FHF) and second heart field (SHF) progenitors in the cardiac crescent at E7.75, and throughout the developing heart tube from E8.5 to E9, the loss of MEF2C causes distinct defects in different segments of the heart tube (Fig. 1b-c, Extended Data Fig. 1a). At E9, the outflow tract, which gives rise to the aorta and pulmonary trunk, is severely hypoplastic and only a single hypoplastic ventricle has formed. By contrast, the inflow tract, precursor to the atria, appears to be expanded and improperly patterned (Fig. 1c). These segment-specific defects suggest that MEF2C plays distinct regulatory roles in the outflow tract, ventricles, and inflow tract.

To dissect the segment-specific regulation of heart tube morphogenesis by MEF2C, we performed combined single-nucleus RNA-seq and ATAC-seq (snRNA-seq and snATAC-seq) on wild-type (WT) and *Mef2c* knockout (KO) mouse embryos at key stages throughout heart tube development. Taking advantage of the ability to match the

transcriptomic and chromatin accessibility data for every cell in the dataset, we applied a suite of complementary bioinformatics tools to construct GRNs for cardiac progenitor lineages, precisely defined the role of MEF2C in regulating gene expression and chromatin accessibility in the developing heart tube, identified novel MEF2C-dependent enhancers, and discovered a genetic interaction that underlies the inflow tract malformation in *Mef2c* KO embryos (Fig. 1d).

RESULTS

Loss of MEF2C reduces expression of key cardiomyocyte genes across all heart tube segments and alters expression of anterior/posterior markers

To assess the role of MEF2C in controlling gene expression in the developing heart tube, we performed snRNA-seq on stage-matched WT and *Mef2c* KO embryos at E7.75, E8.5, and E9 (n = 2 embryos for each genotype at each stage, Extended Data Fig. 1b). All embryos carried the *Smarcd3*-F6-eGFP transgene reporter to mark cardiac progenitors[14]. At E7.75, we harvested the entire embryos, whereas at E8.5 and E9 we removed the headfolds and posterior trunk via microdissection to increase the proportion of cardiac progenitors captured for sequencing. We created separate Seurat objects[15] for the embryos profiled at each timepoint (E7.75, E8.5, and E9). Following pre-processing, quality control, normalization with scTransform[16], and dimensionality reduction, we performed unbiased Louvain clustering and plotted the data in uniform manifold approximation and projection (UMAP) space. We then used the expression of known marker genes to identify the cell types represented by each cluster (Extended Data Fig. 2a-f, Supplementary Table 1).

To resolve specific cardiomyocyte subtypes, we subset and re-clustered cell types of interest, including cardiac progenitors, cardiomyocytes, and related mesoderm cell types (Fig. 2a-c, Extended Data Fig. 2g-i, Supplementary Table 2). By doing so, we were able to identify clusters representing FHF and early differentiating cardiomyocytes (CMs/FHF), SHF, and juxtacardiac field (JCF) progenitors[17] at E7.75 (Fig. 2a). At E8.5, we could distinguish inflow tract cardiomyocytes (IFT-CMs), ventricular cardiomyocytes (V-CMs), and outflow tract cardiomyocytes (OFT-CMs) (Fig. 2b). At E9 we identified the IFT-CMs more specifically as either atrial or atrioventricular canal (A- or AVC-CMs) (Fig.

2c). We next performed differential gene expression analysis between *Mef2c* KO and WT cells within each of these cell types (Fig. 2d-f, Supplementary Table 3). Overall, more genes were down-regulated than up-regulated in the *Mef2c* KO relative to WT embryos, consistent with the notion that MEF2C is an activator of the myogenic transcriptional program[18]. Among the top differentially expressed genes (DEGs), we found several genes that have been associated with high confidence to congenital heart defects[19], including *Myh6*, *Myh7*, *Actc1*, and *Foxp1* (Extended Data Fig. 2j). Furthermore, nearly all these top DEGs were identified as the nearest genes to MEF2C-occupied peaks in a previously published MEF2C ChIP-seq dataset of embryonic hearts at E14[20], suggesting that these genes are direct targets of MEF2C in the developing heart.

Closer examination of our DEG data (Supplementary Table 3) revealed two categories of DEGs. The first category consisted of key cardiomyocyte genes that were substantially down-regulated across all three heart tube segments in the *Mef2c* KO embryos, including genes encoding sarcomeric proteins such as *Tnnt2* and *Ttn,* the cardiac transcription factor *Nkx2-5*, and the natriuretic peptide *Nppa* (Fig. 2g-i). These data confirm the importance of MEF2C in early cardiomyocyte fate specification, as a direct activator of contractile protein gene expression and by regulating additional TFs and functional proteins necessary for proper cardiomyocyte development. The second category of DEGs consisted of genes involved in anterior/posterior patterning of the heart tube. Specifically, genes such as *Tbx5*, *Gata4*, and *Wnt2,* which are normally expressed only in the posterior IFT segment of the heart tube, were not only up-regulated in the IFT-CMs, but their expression was also expanded into the V-CMs in *Mef2c* KO (Fig. 2g-i). Additionally, expression of the anterior OFT-specific gene *Tdgf1* was completely lost in *Mef2c* KO (Fig. 2h-i), as previously demonstrated[21]. Consistent with this apparent posteriorization of gene expression in the heart tube, we also observed a larger proportion of A-CMs and AVC-CMs in *Mef2c* KO embryos compared to WT at E9, with a concomitant reduction in the proportion of V-CMs and OFT-CMs (Extended Data Fig. 2k). We validated the observed changes in gene expression with whole-mount embryo fluorescence *in situ* hybridization, confirming that *Mef2c* KO leads to broad reduction in the cardiomyocyte transcriptional program and a notable posteriorization of the developing heart tube (Fig. 2j).

108

109 **MEF2C regulates chromatin accessibility broadly throughout the heart tube and in**
110 **a segment-specific manner**

111 To investigate how MEF2C regulates the observed segment-specific gene expression
112 changes, we integrated our chromatin accessibility and gene expression datasets using
113 ArchR[22]. We performed dimensionality reduction using ArchR's latent semantic index
114 (LSI) implementation on the snRNA-seq and snATAC-seq modalities and then plotted the
115 cells in UMAP space using the combined dimensions (Extended Data Fig. 3a-c). Matching
116 cell barcodes in the snRNA-seq and snATAC-seq libraries allowed us to transfer the cell
117 type labels from the Seurat objects to the integrated ArchR datasets. As we did with the
118 gene expression data, we subset the cardiac cell types and associated mesoderm in
119 order to better resolve cardiomyocyte subtypes. Finally, we created pseudobulk groups
120 consisting of WT and *Mef2c* KO cells for each given cell type label in the subset datasets
121 and called peaks representing accessible regions using MACS2[23].

122 Our initial analyses revealed few changes to chromatin accessibility at E7.75
123 (Extended Data Fig. 3d). This was not entirely unexpected, given that there is no
124 discernable phenotype in the *Mef2c* KO embryos at this timepoint, and agrees with the
125 relatively few changes in gene expression we observed in CMs/FHF, SHF, and JCF cells
126 at E7.75 (Fig. 2d). Thus, we focused on the E8.5 dataset to identify the earliest substantial
127 changes in chromatin accessibility induced by the loss of MEF2C (Fig. 3a). Using our
128 pseudobulked data, we identified differentially accessible regions (DARs) of chromatin
129 between *Mef2c* KO and WT cells in each of the subset cell types and found that, as
130 expected, cardiomyocytes displayed the largest numbers of DARs (Fig. 3b). We also
131 found that in each of the CM subtypes (IFT-CMs, V-CMs, and OFT-CMs), *Mef2c* KO CMs
132 lost more DARs than they gained (Fig. 3b), and these lost DARs were highly enriched for
133 MEF2 binding motifs (Extended Data Fig. 3e). These data further reinforce MEF2C's role
134 as an activator of the contractile gene transcriptional program. Interestingly, when we
135 examined the lost DARs in *Mef2c* KO embryos, we found 675 regions that demonstrated
136 lost accessibility in two or more heart tube segments, and 2,549 regions that lost
137 accessibility in a segment-specific manner (Fig. 3c). The former are likely regions involved
138 in the regulation of general cardiomyocyte genes, while the latter may control segment-

specific gene regulation. By contrast, the gained DARs in *Mef2c* KO embryos were almost exclusively unique to each heart tube segment (Fig. 3c). These segment-specific DARs are interesting because they provide evidence that distinct transcriptional networks exist for each of the three heart tube segments and that each segment-specific network is uniquely altered by the loss of MEF2C.

To identify the DARs most likely to represent *bona fide* regulatory elements, we performed a peak-to-gene linkage (P2G) analysis, which identifies statistically significant correlations between the accessibility of individual peaks and the expression of potential target genes within 250 kilobases of each peak. We found 817 P2G links in the OFT-CMs, 1,201 in the V-CMs, and 2,257 in the IFT-CMs. We then intersected the P2G links with the MEF2C-dependent DEGs at each timepoint (Supplementary Table 4) and plotted each P2G link according to the fold change of the peak's accessibility, the fold change of linked gene's differential expression, and the correlation strength of the P2G link (Fig. 3d).

These P2G DAR-DEG correlation plots reveal a number of interesting observations. Genes such as *Nkx2-5*, *Myh6*, and *Myh7*, whose expression decreased throughout the heart tube in *Mef2c* KO embryos, are linked to DARs in all three segments (Fig. 3d). In particular, the *Myh6*/*Myh7* locus contains multiple regions that have P2G links and clearly altered accessibility in OFT-CMs, V-CMs, and IFT-CMs (Fig. 3e). This suggests that these DARs are likely to be regulatory elements that control expression of *Myh6* and *Myh7* and are sensitive to the expression of *Mef2c* in each segment of the heart tube. By contrast, there are also P2G links between peaks with lost accessibility and decreased gene expression that are unique to the individual segments, such as links to *Wnt11* in the OFT-CMs, *Nppa* in the V-CMs, and *Cacna1c* in the IFT-CMs (Fig. 3d). Furthermore, examining the links between peaks with gained accessibility and genes with increased expression in *Mef2c* KO reveals additional evidence for the posteriorized transcriptional program we previously described. Namely, P2G links to *Tbx5* in the V-CMs, as well as *Gata4* and *Wnt2* in the IFT-CMs (Fig. 3d and 3f). These P2G links represent potential regulatory elements that are activated or over-activated in response to the loss of MEF2C and, at least in part, promote the increased posteriorization of the heart tube.

169 **Each heart tube segment exhibits a distinct MEF2C-depedent developmental**
170 **trajectory**
171 Our multiomic datasets containing precisely staged embryos at different developmental
172 timepoints provided us the opportunity to dissect how lineage trajectories unfold in WT
173 embryos and are altered by loss of MEF2C. To this end, we applied multimodal Models
174 for Integrated Regulatory Analysis (MIRA)[24], which uses deep learning and probabilistic
175 graphical models to construct sets of "topics" that represent either gene expression or
176 peak accessibility. These topics can then be used to plot cells in UMAP space, assign cell
177 identities, perform pseudotime analyses, and construct cell state trees (Fig. 4a).
178 We trained topic models using cardiac progenitors and cardiomyocyte subtypes
179 from both genotypes at each of our three timepoints (E7.75: CMs/FHF, SHF, JCF; E8.5:
180 aSHF, pSHF, OFT-CMs, V-CMs, IFT-CMs; E9: OFT-CMs, V-CMs, A-CMs, AVC-CMs) and
181 then used those models to construct joint representation UMAPs of the outflow tract,
182 ventricular, and inflow tract lineages (Fig. 4b, 4f, 4j). We next calculated pseudotime
183 values for cells within the UMAPs and used the results to construct cell state trajectories
184 for each of the three lineages (Fig. 4c-d, 4g-h, 4k-l).
185 We observed clear differences in the fate specification of *Mef2c* KO cells compared
186 to WT cells, as represented by branch points in the trajectories (Fig. 4d, 4h, 4l). Most
187 interestingly, we noticed that the structure of these trajectories, along with the manner
188 and timing by which *Mef2c* KO cells diverged from WT cells within the trajectories, was
189 different for all three lineages. For instance, within the outflow tract lineage, *Mef2c* KO
190 and WT cells proceeded similarly up until E8.5, and then diverged as development
191 proceeded to E9 (Fig. 4d). Within the ventricular lineage, there was no branch point.
192 Instead, *Mef2c* KO V-CMs from both E8.5 and E9 were located earlier in the trajectory
193 than even E8.5 WT V-CMs (Fig. 4h). This indicates that loss of MEF2C in V-CMs resulted
194 in a marked delay or termination of normal development rather than a divergence to an
195 alternative cell state. The situation was most complex for the inflow tract lineage, in which
196 there were two points of divergence between *Mef2c* KO and WT cells (Fig. 4l). First, a
197 subset of *Mef2c* KO cells branched away from the WT trajectory as they differentiated
198 from posterior SHF (pSHF) progenitors to IFT-CMs at E8.5. A second set of *Mef2c* KO
199 cells proceeded along the WT trajectory to E9, before ultimately diverging from WT cells

200 at the “terminal” AVC-CM and A-CM states. This complex branching behavior may reflect
201 the notion that both FHF and pSHF progenitors contribute to the IFT[11] and suggests that
202 MEF2C likely controls distinct GRNs within these two IFT progenitor subpopulations.
203 Together, these data reinforce our working hypothesis that MEF2C regulates
204 development of the heart tube in a segment-specific manner via distinct regulatory
205 networks.

206 We gained further insights regarding the segment-specific regulation of heart tube
207 development by examining the top 200 genes and the top ranked TF binding motifs
208 enriched in the gene expression and chromatin accessibility topics (Supplementary Table
209 5). As WT cells diverged from *Mef2c* KO cells in the outflow tract lineage, their gene
210 expression topics became enriched for canonical outflow tract markers *Tdgf1* and
211 *Wnt11*[21,25], and their accessibility topics for MEF2 and FOS binding motifs (Fig. 4e). By
212 contrast, the *Mef2c* KO cells were marked by gene expression topics enriched for Notch
213 and Wnt signaling components such as *Dlk1* and *Fzd1*, and transcription factors *Meis1*
214 and *Hand2,* and exhibited enrichment of an accessibility topic containing ISL, NKX, and
215 TEAD binding motifs (Fig. 4e), suggesting these cells may retain a more progenitor-like
216 state in the absence of MEF2C. The gene expression and accessibility topics enriched
217 along the ventricular lineage trajectory supports our interpretation that the *Mef2c* KO cells
218 experience a substantial delay or termination of their normal differentiation. Notably,
219 topics that were enriched for CM genes (e.g. *Myl2*, *Tnnt1*) and cardiac TF binding motifs
220 (e.g. TBX, MEF2, NKX) marked the terminal end of the differentiation trajectory consisting
221 of only WT cells (Fig. 4i). Interestingly, the mid-point of the ventricular trajectory containing
222 E8.5 and E9 *Mef2c* KO cells is marked by an accessibility topic enriched for the nuclear
223 receptor NR2F2 binding motif, a transcription factor that is restricted to the IFT in WT
224 embryos[26] (Fig. 4i). In the complex branching inflow tract trajectory, terminal WT cells
225 were enriched for topics containing expected atrial genes, such as *Angpt1* and *Nr2f1*, and
226 binding motifs for TBX5 and SNAI, among others (Fig. 4m, Extended Data Fig. 4a-b). By
227 contrast, the *Mef2c* KO cells streamed into one of two terminal branches enriched for
228 genes such as *Wnt6, Jarid2, Tbx2,* and *Zeb1* and binding motifs for ERF, GATA, and
229 TEAD factors (Fig. 4m, Extended Data Fig. 4a-b). Together, these data illustrate the
230 lineage-specific changes in gene expression and chromatin accessibility that accompany

heart tube development and the distinct impact of the loss of MEF2C on these lineage-specific programs.

Candidate regulatory elements with MEF2C-dependent chromatin accessibility display enhancer activity in zebrafish

Our integrated chromatin accessibility data (Fig. 3) and developmental trajectory analyses with MIRA (Fig. 4) highlighted changes in chromatin accessibility during heart tube development that were MEF2C-dependent and segment-specific. We hypothesized that these regions contain regulatory elements that are necessary for driving segment-specific gene expression. To identify enhancers, we applied a series of filtering criteria to our integrated multiomic data and selected 12 candidates of the highest priority to screen based on the Log2FC value of peak accessibility and linkage to genes of interest (Fig. 5a, Supplementary Table 6). We enumerated these candidate regions with the prefix "MVEB," after the initials of authors Muncie-Vasic and Brower. Overlaying our chromatin accessibility data at these regions with published ChIP-seq datasets[20,27] confirmed that each of our candidates had clear and strong occupancy of both MEF2C and H3K27ac, a histone post-translational modification that marks active enhancers (Fig. 5b). Of note, one of our ventricular-specific candidates, MVEB8, was located within the *Nppa*/*Nppb* super enhancer cluster[28]. Furthermore, two of the candidates we identified from differential accessibility in IFT-CMs overlapped with regions found in the VISTA enhancer database[29], with demonstrated enhancer activity in the hearts of E11.5 mouse embryos (Fig. 5c-d).

To determine whether our candidate regions are *bona fide* enhancers with activity in early heart tube development, we employed Tol2 transgenesis in zebrafish[30,31], a relatively high-throughput *in vivo* screening system (Fig. 5e). We reasoned that because *Mef2c* has orthologs in zebrafish (*mef2ca* and *mef2cb*), and because there is a high degree of overlap between the key cardiac transcriptional regulators in both species[32], there should be concordance between the activity of our candidate regions in mouse and zebrafish. Indeed, specific regulatory elements have been shown to exhibit activity in the heart of both species[33–35].

Upon screening, seven of eleven candidates displayed enhancer activity in developing zebrafish hearts (Fig. 5f, Extended Data Fig. 5a-b), with one of our candidates

being excluded due to a highly repetitive “GT” sequence. The two candidates that overlapped with regions found in the VISTA enhancer database (MVEB1, MVEB2) exhibited activity in the heart by 72 hours post-fertilization (hpf) (Fig. 5f). Additionally, two candidates from our IFT-CM data (MVEB3 and MVEB6), two from our V-CM data (MVEB7 and MVEB8), and one from our OFT-CM data (MVEB12) demonstrated robust and specific activity in the zebrafish heart at 72 hpf (Fig. 5f, Extended Data Fig. 5a-b). To our knowledge, four of these latter candidates (MVEB3, MVEB6, MVEB7, and MVEB12) have not been previously reported and represent novel, MEF2C-dependent enhancers with activity in the developing heart. Interestingly, we also observed a range in the timing of enhancer activity in the embryonic zebrafish heart. Candidates MVEB2 and MVEB6 were clearly detectable as early as 24 hpf, while the others were not active until 72 hpf (Fig. 5g-h). This variation in the onset of enhancer activity could result from differences in the timing of accessibility of these regulatory elements, different sensitivity to MEF2C levels, or cooperation with other TFs or transcriptional regulators.

Loss of MEF2C induces an overactive posteriorized gene regulatory network driven by NR2F2 and GATA4

Our data indicate that loss of MEF2C causes a posteriorized gene signature (Fig. 2h-j) and chromatin accessibility landscape (Fig. 3d, 3f) in the developing heart tube. Moreover, as previously described[36], the expanded but aberrantly formed inflow tract that forms in *Mef2c* KO embryos (Fig. 1c) indicates that loss of MEF2C leads to a more complex heart tube phenotype than a simple failure of cardiac differentiation and growth. To determine what transcriptional regulators may drive the inflow tract malformation, we performed TF binding motif enrichment analysis on the regions of chromatin that gain accessibility in the *Mef2c* KO IFT-CMs at E8.5. We found that Nuclear Receptor (NR) and GATA motifs were among the most enriched (Fig. 6a). From this observation, we hypothesized that MEF2C co-binds certain regulatory elements with NR or GATA factors in the WT condition, and that these NR and GATA factors become free to bind different regulatory elements with loss of MEF2C. Consistent with this hypothesis, we found that nearly half of the sites of accessible chromatin that mark WT IFT-CMs and contain an NR or GATA

292 motif also contain a MEF2 motif, whereas the vast majority of sites that gain accessibility
293 in *Mef2c* KO and contain an NR or GATA motif do not (Fig. 6b, Extended Data Fig. 6a).
294 These data are particularly interesting given the importance of GATA factors, and
295 in particular GATA4, in cardiac development[37] and the knowledge that nuclear receptor
296 subfamily 2 group F member 2 (NR2F2, also known as COUP-TFII) is a necessary driver
297 of atrial development[26,38]. Using publicly available ChIP-seq datasets[38,39], we found
298 NR2F2 and GATA4 occupancy at a number of the regions of chromatin that gained
299 accessibility in the *Mef2c* KO IFT-CMs at E8.5 (Extended Data Fig. 6b-c), suggesting that
300 these DARs could indeed be sites of potential NR2F2 and GATA4 regulation. To
301 determine whether altered NR2F2 and GATA4 activity may drive the posteriorization of
302 the heart tube, we next asked if NR2F2 and GATA4 target genes were upregulated in the
303 inflow tract of *Mef2c* KO embryos. Using publicly available datasets to identify the target
304 genes of each TF[40,41], we calculated the odds ratio (OR) of finding NR2F2 or GATA4
305 target genes compared to non-target genes amongst our up-regulated DEGs at E8.5. We
306 found strong evidence that NR2F2 (OR = 1.52, p = 1.79e-4) and GATA4 (OR = 4.23, p <
307 2.2e-16) targets were enriched amongst the upregulated genes (Fig. 6c, Extended Data
308 Fig. 6d). Next, we took the list of NR2F2 and GATA4 target genes that were upregulated
309 (Supplementary Table 7) and calculated a module score[42] for the expression of these
310 genes in our dataset. In agreement with our working model that these NR2F2 and GATA4
311 targets are driving the posteriorized heart tube phenotype in *Mef2c* KO, we found the
312 NR2F2 and GATA4 module scores were increased in all segments of the heart tube at
313 E8.5 (Fig. 6d, Extended Data Fig. 6e).
314 To understand how altered NR2F2 and GATA4 activity upon loss of MEF2C affects
315 the entire transcriptional network, we used CellOracle[43] to construct GRNs for the WT
316 and *Mef2c* KO IFT-CMs at E8.5 and E9 (Fig. 6e). In brief, CellOracle uses integrated gene
317 expression and chromatin accessibility data to infer interactions between TFs and target
318 genes based on the accessibility of a given TF's binding motif near a potential target gene
319 and the correlation of expression between the TF and the potential target. To validate that
320 these inferred CellOracle networks effectively captured the GRNs that govern inflow tract
321 development, we simulated the knockout of *Mef2c in silico* in the E8.5 WT IFT-CMs
322 network and compared the predicted changes in gene expression to the measured

323 changes captured in our E9 dataset (Fig. 6f-g). We found that the networks correctly
324 predicted the gene expression changes with 74% accuracy, even with a relatively
325 permissive DEGs cutoff of Log2FC >= 0.75 (Fig. 6g). As expected, the prediction
326 accuracy increased with more stringent Log2FC cutoffs, and vice versa (Supplementary
327 Table 8).

328 We next visualized the subnetworks of 12 core cardiac TFs and the top 100 DEGs
329 in the WT and *Mef2c* KO inflow tracts (Fig. 6h). These subnetworks revealed that loss of
330 MEF2C causes a network-wide reorganization of gene regulation. First, we observed that
331 MEF2C may be critical for coordinating cooperative regulation by multiple cardiac TFs.
332 For instance, in the WT network, MEF2C, MEF2A, and TBX20 all have an inferred
333 activating effect on *Ttn* expression. However, in the *Mef2c* KO network, both MEF2A and
334 TBX20 lose their activating interaction with *Ttn.* Similarly, in the WT network, MEF2C,
335 MEF2A, and NKX2-5 are all inferred activators of *Myh6*, but in the *Mef2c* KO network,
336 MEF2A and NKX2-5 are no longer predicted to activate *Myh6* expression. These data
337 suggest that MEF2C is required for cooperative activation with other cardiac TFs in the
338 developing heart tube, similar to what was shown to occur during cardiomyocyte
339 reprogramming *in vitro*[44]. Additionally, we found that loss of MEF2C may alter the scope
340 and strength of interactions between other cardiac TFs and their gene targets. For
341 example, in the WT network, GATA4 and MEIS2 exhibit competitive regulation of *Fbxo32*,
342 which encodes a muscle-specific ubiquitin-E3 ligase that is critical for sarcomeric function
343 and associated with dilated cardiomyopathy[45,46]. This competitive relationship is
344 maintained in the *Mef2c* KO network, but the interactions are strengthened in both
345 directions. Finally, we observed that the activity of NR2F2 is substantially altered in the
346 *Mef2c* KO network, with a larger number of gene targets and apparent strengthening of
347 both activating and inhibitory interactions. Most notably, the activating interaction between
348 NR2F2 and *Angpt*1 expression is much stronger in the *Mef2c* KO network than in the WT
349 network. Interestingly, the activation of *Nr2f2* itself by TBX5 is lost upon *Mef2c* KO.

350 Next, we further narrowed in on the direct interactions between NR2F2 and GATA4
351 and their target genes in the WT and *Mef2c* KO inflow tract networks (Fig. 6i, Extended
352 Data Fig. 6f). From these analyses we observe similar behavior for both TFs –
353 approximately half of the WT interactions are maintained in the *Mef2c* KO, while nearly

twice as many new direct interactions are established. This is consistent with our working model that in response to loss of MEF2C, aberrant NR2F2 and GATA4 activity regulate additional genes that drive the inflow tract phenotype in the posterior of the heart tube.

Reduction of NR2F2 dosage partially rescues the heart tube phenotype in *Mef2c* KO embryos

Given our observation that loss of MEF2C causes aberrant NR2F2 activity in the IFT-CM GRNs, we theorized that reducing the dosage of NR2F2 may rescue the inflow tract phenotype in *Mef2c* KO embryos. To test this possibility, we generated *Mef2c*$^{+/-}$;*Nr2f2*$^{+/-}$ male mice and crossed them to *Mef2c*$^{+/-}$ females. Although *Nr2f2*$^{+/-}$ embryos have no apparent heart tube phenotype at E9.5 (Fig. 6j, Extended Data Fig. 7a), pups of this genotype were recovered at lower numbers than expected Mendelian ratios[26], suggesting that *Nr2f2* is a dosage-sensitive gene. We observed a notable partial rescue in ventricle and inflow tract development at E9.5 (18-24 somites) in the *Mef2c*$^{-/-}$;*Nr2f2*$^{+/-}$ embryos compared to *Mef2c*$^{-/-}$ (Fig. 6j-k), although the extent of the rescue varied (Extended Data Fig. 7b-c). The *Mef2c* KO single-ventricle phenotype persisted in *Mef2c*$^{-/-}$;*Nr2f2*$^{+/-}$ embryos, but these ventricles were larger and more expanded than in *Mef2c*$^{-/-}$ *embryos.* Moreover, the posterior structures of the heart tube in *Mef2c*$^{-/-}$;*Nr2f2*$^{+/-}$ embryos more closely resembled WT inflow tracts with expanded and looping prospective atria, compared to the completely disrupted inflow tract in *Mef2c*$^{-/-}$ embryos (Fig. 6j-k). This partial rescue of the ventricle and inflow tract in *Mef2c*$^{-/-}$;*Nr2f2*$^{+/-}$ embryos is evidence that the *Mef2c* KO phenotype in the posterior heart tube is driven, at least in part, by an altered gene regulatory network with increased and aberrant NR2F2 activity.

DISCUSSION

In this study we demonstrated that MEF2C controls complex, segment-specific GRNs that regulate gene expression and chromatin accessibility in the developing linear heart tube. We found that loss of MEF2C causes a posteriorization of the heart tube and revealed novel segment-specific regulatory elements with enhancer activity. These results address the long-standing question of how a given TF can play distinct regulatory roles in different cell types and at particular developmental stages. For instance, decades of work have

385 established that FHF progenitors give rise to the left ventricle and atria, while the SHF is
386 subdivided into the aSHF, which produces the outflow tract and right ventricle, and the
387 pSHF, which contributes to the atria[7–11]. Although these progenitor cell types have
388 different molecular signatures and give rise to individual structures with specific
389 morphology and function, they largely rely on a shared set of cardiac TFs to regulate their
390 development. Our data illustrate that these core cardiac TFs operate in unique GRNs for
391 each of the inflow tract, ventricular, and outflow tract lineages in the developing heart
392 tube. We found that distinct chromatin landscapes and regulatory elements are important
393 determinants for setting up these GRNs, but other embryological phenomena are likely
394 to be involved, such as morphogen signaling from differing neighbor cell types,
395 heterochronic timing of differentiation[47], and mechanical phenomena introduced by tissue
396 morphogenesis and the initiation of cardiomyocyte beating[48–50].

397 Using the GRNs inferred from our multiomic data, we identified a genetic
398 interaction between *Mef2c* and *Nr2f2* in the inflow tract of the developing heart tube.
399 Previous work from our lab showed that TF occupancy at specific loci is highly dependent
400 upon both TF-TF and TF-DNA affinities, and that TF-TF interactions not only serve to
401 regulate gene expression, but prevent the redistribution of one or both TFs to exogenous
402 loci[51]. The results we present here lend further support to those concepts. By removing
403 MEF2C, the set of possible TF-TF interactions is limited and causes altered activity of
404 partner TFs, including NR2F2 and GATA4. Our comparison of the WT and *Mef2c* KO
405 GRNs revealed that the transcriptional networks regulating early heart development are
406 highly sensitive to TF perturbations, with loss of MEF2C causing a reorganization of heart
407 tube GRNs that is not limited to direct MEF2C connections, and reduction of NR2F2
408 dosage capable of partially rescuing the heart tube malformations. Thus, these GRNs
409 may be implemented as a powerful tool for predicting the network-wide consequences of
410 genetic mutations that underlie CHDs, potentially leading to unexpected molecular targets
411 for intervention and correction.

412 Our approach provides a generalizable framework for integrating snRNA-seq and
413 snATAC-seq data to build GRNs that provide specific and comprehensive models of gene
414 regulation. By constructing these GRNs using both WT and mutant data, we were able to
415 immediately test predictions made from the WT networks and validated the inferred

regulatory interactions. These tools also allow specific networks to be built for particular cell or tissue types, provided there are enough cells in the group of interest to generate robust inferences. Moreover, the experimental mutant data revealed how the overall networks were altered – both in ways that would and would not be predicted from the WT networks. These unexpected differences help to illuminate indirect or transactivating gene regulation. Discoveries from these networks lead to many new hypotheses about regulatory elements, TF interactions, DNA binding, and chromatin organization. As these hypotheses are experimentally tested, the GRNs can be refined, leading to higher-fidelity models with greater predictive power. Such higher-fidelity networks will yield important new insights into the transcriptional regulation of developmental processes and will have the ability to reveal precisely how these processes are altered in the context of disease-causing mutations.

428 **MATERIALS AND METHODS**

429 **Mouse Models**

430 All mouse studies were performed in strict compliance with the UCSF Institutional Animal
431 Care and Use Committee. Mice were housed in a standard 12-hour light/dark animal
432 husbandry barrier facility at the Gladstone Institutes. The *Mef2c*$^{+/-}$ allele used was the
433 same as the original generated by Lin et al.[13]. The *Smarcd3*-F6-eGFP reporter allele was
434 generated in our lab and described previously[14]. The *COUP-TFII lacZ* knock-in allele
435 (abbreviated here as *Nr2f2*$^{+/-}$) was obtained from the Mutant Mouse Resource & Research
436 Centers and was originally generated by Takamoto et al.[52]. The *Mef2c*$^{+/-}$ and *Nr2f2*$^{+/-}$ lines
437 were maintained in the C57BL6/J background (Jackson Laboratory #664). The *Smarcd3*-
438 F6-eGFP allele was crossed into our *Mef2c*$^{+/-}$ line from a mixed background of C57BL6/J
439 (Jackson Laboratory #664) and CD-1 (Charles River #022).

440

441 **Timed Embryo Dissections and Genotyping**

442 To achieve timed matings, male and female mice were housed together in the evening
443 and pregnancy was assessed by vaginal plug the following morning. Gestational stage
444 was determined relative to noon on the day of plug detection, defined as day E0.5.
445 Embryos were dissected and, at later stages when yolk was present, also de-yolked, in
446 ice-cold PBS (Life Technologies, 14190250) with 1% fetal bovine serum (FBS; Thermo
447 Fisher Scientific, 10439016) on ice. The posterior trunk was removed by microdissection
448 for all embryos at E8.5 or later that were collected for imaging (IF or FISH via HCR) to
449 ensure unobstructed views of the heart tube. DNA for genotyping was extracted using
450 QuickExtract DNA Extraction Solution (Lucigen, QE09050) from harvested yolk sac
451 tissue, if available, or else from a micro-dissected piece of the extra-embryonic anterior
452 proximal region. For embryos collected for 10x Multiome library construction, rapid
453 genotyping was performed while embryos remained on ice using Phire Green Hot Start II
454 PCR Master Mix (Thermo Fisher Scientific, F126S), according to manufacturer's
455 protocols. For all other experiments, genotyping was performed after embryo fixation
456 using GoTaq Green Master Mix (VWR, PAM7122), according to manufacturer's protocols.
457 PCR products were run on a 1% agarose gel (Thermo Fisher Scientific, BP164-500) with
458 Biotium GelRed Nucleic Acid Gel Stain (Thermo Fisher Scientific, NC9594719) used for

459 DNA detection. Primers used for genotyping were ordered from Integrated DNA
460 Technologies (IDT). For genotyping the *Mef2c* allele, a three-primer mix was used (P1:
461 ACTGGCTGAGAGTTGTACCCAC, P2: GCGTGATTTGCTAGATAGTGGTAGAC, P3:
462 ATGTGGAATGTGTGCGAGGC), resulting in a single 314 bp band for WT, a single 242
463 bp band for *Mef2c*$^{-/-}$, and both bands for *Mef2c*$^{+/-}$. For genotyping the *Nr2f2* allele, a three-
464 primer mix was used (P4: CATCCGGGATATGTTACTGTCCGG, P5:
465 TGGGGAAGCTAAGTGTTGATGTGATTCC, P6: GCCGTGGGTTTCAATATTGGCTTC),
466 resulting in a single 786 bp band for WT, a single ~1600 bp band for *Nr2f2*$^{-/-}$, and both
467 bands for *Nr2f2*$^{+/-}$.
468
469 **Whole-Mount Embryo Immunostaining and Imaging**
470 Embryos collected for immunostaining were fixed for 1 h at room temperature in 4%
471 paraformaldehyde (PFA) freshly diluted from 16% weight/volume PFA aqueous solution
472 (Thermo Fisher Scientific, 043368-9M) in PBS (Life Technologies, 14190250). Embryos
473 were stored at 4°C for up to 1 month prior to immunostaining. To permeabilize and block
474 non-specific antigens, embryos were incubated at 37°C for 2 h with gentle rotation in a
475 blocking buffer consisting of PBS (Life Technologies, 14190250), 5% normal donkey
476 serum (Sigma Aldrich, S30-M), and 0.8% Triton X-100 (Sigma Aldrich, X-100-5ML). This
477 was then removed and replaced with primary antibodies diluted in the same blocking
478 buffer and incubated overnight at 37°C with gentle rotation. The next day, primary
479 antibody solution was removed and embryos were washed 3x 30 min in the same blocking
480 buffer at 37°C with gentle rotation. Secondary antibodies and DAPI (Abcam, ab228549,
481 used at 1 µg/mL) were added in a blocking buffer consisting of PBS (Life Technologies,
482 14190250), 5% normal donkey serum (Sigma Aldrich, S30-M), and 0.4% Triton X-100
483 (Sigma Aldrich, X-100-5ML), and incubated for 3-4 h at 37°C with gentle rotation.
484 Secondary antibody solution was removed and embryos were washed 3x 30 min in PBS
485 (Life Technologies, 14190250) with 0.1% Tween-20 (Thermo Fisher Scientific,
486 J20605.AP) at 37°C with gentle rotation. Embryos were imaged in PBS using an upright
487 epifluorescence microscope (Leica MZFLIII, Leica DFC 3000G, Lumen Dynamics XCite
488 120LED) and acquisition software LASX (Leica). Primary antibodies used were sheep
489 polyclonal MEF2C (R&D Systems, AF6786, used at 1:250) and rabbit polyclonal cardiac

490 Troponin T (Thermo Fisher Scientific, 15513-1-AP, used at 1:250). Secondary antibodies
491 used were Donkey anti-Sheep IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa
492 Fluor™ 647 (Thermo Fisher Scientific, A-21448, used at 1:1000) and Donkey anti-Rabbit
493 IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 555 (Thermo
494 Fisher Scientific, A-31572, used at 1:1000).
495
496 **Whole-Mount Embryo Fluorescence *In Situ* Hybridization via Hybridization Chain**
497 **Reaction and Imaging**
498 The protocol for *in situ* hybridization via hybridization chain reaction (HCR™) was adapted
499 from the optimized manufacturer's protocol for whole-mount mouse embryos (Molecular
500 Instruments). Embryos were fixed overnight at 4°C in 4% PFA freshly diluted from 16%
501 weight/volume PFA aqueous solution (Thermo Fisher Scientific, 043368-9M) in PBS (Life
502 Technologies, 14190250). Embryos were then washed 3x 5 min in PBS (Life
503 Technologies, 14190250) with 0.1% Tween-20 (Thermo Fisher Scientific, J20605.AP),
504 abbreviated PBST, and then processed through 10 min serial incubations in a dehydration
505 series of 25%, 50%, 75%, and 100% methanol solutions in PBST on ice. Embryos were
506 stored in 100% methanol at −20°C for up to 6 months prior to *in situ* hybridization.
507 Embryos were then rehydrated through 10 min serial incubations in a series of 75%, 50%,
508 25%, and 0% methanol solutions in PBST on ice. Embryos were then brought to room
509 temperature and permeabilized by 10 min incubation with 10 µg/mL Proteinase K (Sigma
510 Aldrich, 70663-5) in PBST, washed 2x 5 min in PBST at room temperature, post-fixed in
511 4% PFA freshly diluted from 16% weight/volume PFA aqueous solution (Thermo Fisher
512 Scientific, 043368-9M) in PBS (Life Technologies, 14190250), and then washed 3x 5 min
513 in PBST at room temperature. Embryos were pre-hybridized in Probe Hybridization Buffer
514 (Molecular Instruments) for 5 min at room temperature followed by an additional 30 min
515 at 37°C with fresh buffer. Single or multiplexed hybridization probes were then diluted in
516 pre-warmed Probe Hybridization Buffer (Molecular Instruments) and embryos were
517 hybridized for 48 h at 37°C. Embryos were then washed 4x 15 min with pre-warmed Probe
518 Wash Buffer (Molecular Instruments) at 37°C, followed by 2x 5 min with 5xSSCT (diluted
519 in nuclease-free water from 20xSSCT, Thermo Fisher Scientific, 15557044) at room
520 temperature. During washes, amplification hairpins were snap-cooled by heating to 95°C

521 for 90 s and then allowed to cool to room temperature in the dark. Embryos were pre-
522 amplified in Amplification Buffer (Molecular Instruments) for 5 min at room temperature.
523 Snap-cooled hairpins and DAPI (Abcam, ab228549, used at 1 µg/mL) were diluted in
524 Amplification Buffer (Molecular Instruments) and embryos were incubated in this solution
525 overnight in the dark at room temperature. The following day, embryos were washed
526 through a series of 5xSSCT washes (diluted in nuclease-free water from 20xSSCT,
527 Thermo Fisher Scientific, 15557044) in the dark at room temperature (2x 5 min, 2x 30
528 min, 1x 5 min). Embryos were imaged in PBS using an upright epifluorescence
529 microscope (Leica MZFLIII, Leica DFC 3000G, Lumen Dynamics XCite 120LED) and
530 acquisition software LASX (Leica). To maximize the utility of collected embryos, probes
531 were then stripped by overnight incubation in 80% formamide in water (Thermo Fisher
532 Scientific, BP228100) at room temperature, followed by 3x 5 min washes in PBST at room
533 temperature. Embryos could then be probed for additional targets, beginning the protocol
534 again at the pre-hybridization step. Probes used in this study were ordered from the
535 Molecular Instruments catalog, when possible, or custom-designed and manufactured by
536 Molecular Instruments. Hybridization probes used were: *Tnnt2*-B3 (used at 4 nM), *Ttn*-B2
537 (used at 4 nM), *Nppa*-B2 (used at 20 nM), *Nkx2.5*-B1 (used at 20 nM), *Tbx5*-B1 (used at
538 20 nM), *Gata4*-B1 (used at 20 nM), *Wnt2*-B2 (used at 20 nM), *Tdgf1*-B2 (used at 20 nM).
539 Amplification probes used were all used at 60 nM: B1-Alexa Fluor™ 647, B2-Alexa
540 Fluor™ 546, B2-Alexa Fluor™ 647, B3-Alexa Fluor™ 488.
541
542 **Embryo Preparation for 10x Multiome snRNA-seq and snATAC-seq**
543 For E7.75 embryos, whole embryos were dissected and harvested for single-nucleus
544 library generation. For E8.5 and E9 embryos, the head folds and posterior trunk were
545 removed by microdissection prior to harvesting for library generation to enrich the relative
546 capture of cardiac cell types. Two embryos were collected per genotype per embryonic
547 stage. Following dissection and rapid PCR genotyping, embryo samples selected for the
548 multiome experiment were incubated in 200 µl TrypLE Select (Thermo Fisher Scientific,
549 12563-011) for 5 min at 37°C, triturated gently by pipetting up and down, and then
550 incubated an additional 3 min at 37°C. The dissociated cell suspension was quenched
551 with 600 µl of PBS (Life Technologies, 14190250) with 1% FBS (Thermo Fisher Scientific,

552 10439016), singularized by passage through the 35 µm mesh of a 5 mL Falcon™ Round-
553 Bottom Polystyrene Test Tube with Cell Strainer Snap Cap (Thermo Fisher Scientific, 08-
554 771-23), pelleted by centrifugation at 300 g for 5 min at 4°C, and resuspended in 50 µl of
555 PBS (Life Technologies, 14190250) with 0.04% bovine serum albumin (BSA, Sigma
556 Aldrich, A1595). At this stage, the manufacturer's protocol for Nuclei Isolation for Single
557 Cell Multiome ATAC + Gene Expression Sequencing, Appendix, Low Cell Input Nuclei
558 Isolation (10x Genomics, CG000365 Rev B) was followed exactly to prepare nuclei for
559 library generation.
560
561 **Generation of snRNA-seq and snATAC-seq Libraries and Sequencing**
562 Libraries for snRNA-seq and snATAC-seq were prepared according to the manufacturer's
563 protocol for Chromium Next GEM Single Cell Multiome ATAC + Gene Expression (10x
564 Genomics, CG000338 Rev B), using the 10x Genomics Chromium controller, Chromium
565 Next GEM Single Cell Multiome ATAC + Gene Expression Reagent Bundle (10x
566 Genomics, PN-1000283), Chromium Next GEM Chip J Single Cell Kit (10x Genomics,
567 PN-1000234), Single Index Kit N Set A (10x Genomics, PN-1000212), and Dual Index Kit
568 TT Set A (10x Genomics, PN-1000215). Nuclei isolated from multiome samples were
569 processed directly into the snATAC-seq transposition reaction and then captured in 10x
570 Genomics GEMs via the Chromium Controller. GEMs were stored at -80°C for up to 4
571 weeks, allowing for collection of replicate embryos of each genotype at each development
572 stage. All subsequent library preparation steps were performed together for all embryos
573 of a given developmental stage to reduce the likelihood of batch artifacts. A targeted
574 maximum recovery of 10,000 cells per sample were loaded onto the 10x Genomics
575 Chromium controller instrument and each sample was indexed with a unique sample
576 identifier (Single Index Kit N Set A for snATAC-seq libraries, Dual Index Kit TT Set A for
577 snRNA-seq libraries). Final libraries were quality-controlled using an Agilent Bioanlyzer
578 instrument with Agilent High Sensitivity DNA Kit (Agilent, 5067-4626). The DNA
579 concentration of each library was measured using KAPA Library Quantification qPCR Kit
580 (Thermo Fisher Scientific, 50-196-5234), and then libraries were pooled and sequenced
581 on NovaSeq6000 S4 lanes (Illumina). All snRNA-seq libraries were sequenced to a depth
582 of at least 20,000 mean raw aligned reads per cell and snATAC-seq libraries were

583 sequenced to a depth of at least 35,000 mean raw aligned read pairs per cell (most
584 >60,000).
585
586 **Processing of Raw Sequencing Data**
587 Raw sequencing reads were processed using the 10x Genomics Cell Ranger Arc v2.0.0
588 pipeline. Reads were demultiplexed using cellranger-arc mkfastq and aligned with
589 cellranger-arc count to the mm10 reference genome containing an additional sequence
590 for eGFP to map reads of the *Smarcd3*-F6-eGFP reporter transcript.
591
592 **Seurat Analysis of snRNA-seq Data**
593 Gene expression data were analyzed using Seurat v4.3.0[15] in R. Datasets for each
594 developmental stage (E7.75, E8.5, E9) were analyzed as separate Seurat objects.
595 Counts matrices from each sample that were output by the Cell Ranger Arc pipeline were
596 used as inputs to the Read10X() and CreateSeuratObject() functions. Data were filtered
597 for quality control based on number of detected genes, unique molecular identifier (UMI)
598 counts, percent ribosomal transcripts, and percent mitochondrial transcripts.
599 SCTransform v2[16] was used to normalize and scale data, with regressions performed with
600 respect to percentage of ribosomal transcripts and percentage of mitochondrial
601 transcripts. Principal component analysis (PCA) was performed, followed by batch
602 correction with Harmony[53], and cells were clustered and visualized as UMAP embeddings
603 based on the top 30 principal components using FindNeighbors(), FindClusters(),
604 RunUMAP(), and DimPlot(). Cell types were annotated using the FindAllMarkers()
605 function with Wilcoxon rank-sum test (min.pct = 0.25, logfc threshold = 0.25, p_val_adj <
606 0.05) to identify cluster-specific marker genes (Supplementary Table 1, Extended Data
607 Fig. 2d-f). Cardiac progenitors, cardiomyocytes, and related mesoderm cell types were
608 subset into new Seurat objects. PCA analysis, batch correction, visualization, and
609 clustering were then performed on these new objects as described above. Cell types in
610 the subset objects were annotated using the FindAllMarkers() function with Wilcoxon
611 rank-sum test (min.pct = 0.25, logfc threshold = 0.5, p_val_adj < 0.05) to identify cluster-
612 specific marker genes that enabled more specific cell type labeling than in the full datasets
613 (Supplementary Table 2, Extended Data Fig. 2g-i). Differential gene expression was

614 performed between *Mef2c* KO and WT cell types of interest using the FindMarkers()
615 function with Wilcoxon rank-sum test (min.pct = 0.1, logfc threshold = 0.25, p_val_adj <
616 0.05). The DotPlot() and RidgePlot() functions was used to visualize gene expression
617 levels of differentially expressed genes across cell types of interest. The
618 AddModuleScore() function(42) was used to add module scores for up-regulated (*Mef2c*
619 KO-vs-WT) NR2F2 and GATA4 target genes and the RidgePlot() function was used to
620 visualize these module scores in cell types of interest.
621
622 **ArchR Analysis of Integrated snRNA-seq and snATAC-seq Data**
623 *Initial Dataset Processing*
624 Chromatin accessibility and integrated multiome data was analyzed using the ArchR
625 software package v1.0.2[22] in R. The fragments files output by the Cell Ranger Arc pipeline
626 were used as inupts to the createArrowFiles() function to create Arrow Files for each
627 sample. These Arrow Files were used to create individual ArchR projects for
628 developmental stage (E7.75, E8.5, E9). The gene expression data was added using the
629 import10xFeatureMatrix() and addGeneExpressionMatrix() functions, using the matching
630 cell barcodes to combine the snRNA-seq and snATAC-seq data for each cell. Data were
631 filtered for quality control based on the number of unique fragments, TSS Enrichment
632 score, expression UMI counts, and number of detected genes. Doublets were filtered
633 using the addDoubletScores() and filterDoublets() functions. Dimensionality reduction
634 was performed using ArchR's iterative latent semantic index (LSI) implementation
635 individually for the snATAC-seq data (varFeatures = 25,000, dimsToUse = 1:30) and
636 snRNA-seq data (varFeatures = 5,000, dimsToUse = 1:30). Then, the reduced
637 dimensions were combined using addCombinedDims() and batch correction was
638 performed on the combined dimensions using Harmony[53]. Cells were then embedded in
639 UMAP space with addUMAP(), clustered using addClusters(), and visualized using
640 plotEmbedding(). Cell types were annotated by transferring metadata labels from the
641 previously analyzed Seurat objects. The getMarkerFeatures() function with Wilcoxon
642 rank-sum test (FDR <= 0.01, Log2FC >= 0.5) was used to identify cluster-specific marker
643 genes to confirm cell type labels. Cardiac progenitors, cardiomyocytes, and related
644 mesoderm cell types were subset into new ArchR projects. Dimensionality reduction,

645 batch correction, clustering, and visualization were performed as described above. Cell
646 types in the subset projects were annotated by transferring metadata labels from the
647 previously analyzed subset Seurat objects. The getMarkerFeatures() function with
648 Wilcoxon rank-sum test (FDR <= 0.01, Log2FC >= 0.5) was used to identify cluster-
649 specific marker genes to confirm subset cell type labels. Due to the small number of *Mef2c*
650 KO OFT-CMs in the E8.5 dataset, an ArchR project with combined E8.5 and E9 samples
651 was created and processed as described above for downstream analyses/comparisons
652 of OFT-CMs.
653
654 *Peak Calling, Differential Peak Testing, and Motif Enrichment*
655 Cells were pseudobulked into replicates based on their cell type and genotype labels (e.g.
656 WT V-CMs), with two pseudobulked replicates created corresponding to the two embryos
657 collected for each genotype at each developmental stage. Replicates were created using
658 the addGroupCoverages() function (minCells = 20, maxCells = 500, sampleRatio = 0.8).
659 Peaks were then called using addReproduciblePeakSet() with MACS2(23)
660 (reproducibility = 2, peaksPerCell = 500, minCells = 20). Differential peak accessibility
661 testing was performed using getMarkerFeatures() with Wilcoxon rank-sum test (FDR <=
662 0.15, Log2FC >= 0.5 or Log2FC <= -0.5) for each cell type of interest with *Mef2c* KO cells
663 used as the group of interest and WT cells used as the background. Transcription factor
664 binding motifs were annotated using addMotifAnnotations() with the Vierstra motif
665 archetypes[54] and motif enrichments for regions of gained or lost accessibility were
666 calculated using peakAnnoEnrichment() with cutoffs of FDR <= 0.15 and Log2FC <= -0.5
667 or Log2FC >= 0.5.
668
669 *Peak2Gene Linkages and Browser Track Visualization*
670 Peak2Gene Linkages were identified using the addPeak2GeneLinks() and
671 getPeak2GeneLinks() functions. Code written in this study was used to filter Peak2Gene
672 links based on presence of a MEF2C motif and linkage to specific genes (i.e. DEGs).
673 Peak2Gene links were visualized using the plotBrowserTrack() function. The
674 getGroupBW() function was used to export the accessibility coverage data in bigwig

format for visualization along with published ChIP-seq data using the IGV software package[55].

MIRA Analysis of Integrated snRNA-seq and snATAC-seq Data

Preparing Merged Datasets

We used the Multimodal models for Integrated Regulatory Analysis (MIRA)[24] python package to examine dynamic changes in gene expression and chromatin accessibility in WT and *Mef2c* KO embryos across the time course of developmental stages that we collected (E7.75, E8.5, E9). We first created a merged Seurat object and ArchR project containing only our cardiac cell types of interest. We then used sceasy[56] to convert our Seurat object and the peak matrix from the ArchR project to AnnData formats, which were used as inputs for the MIRA analyses.

Topic Modeling

The first step of the MIRA pipeline is to train and tune topic models for both the gene expression and chromatin accessibility data. Model training and tuning was performed with GPU support on nodes of the Wynton HPC Co-Op cluster at UCSF. For our gene expression topic model, we used the 4,088 most variable genes (min_disp = 0.1) as the "exogenous" gene set and top 1,959 variable genes (dispersion threshold > 0.7) as the "endogenous" gene set for model training. We then set the minimum and maximum learning rate bounds, tuned the model (iters = 64, max_topics = 55), and MIRA's tuner selected the best model that was used for all subsequent analyses, which contained 33 gene expression topics. To aid in the interpretation of results, we exported the top 200 genes for each of the 33 gene expression topics. We then repeated these steps to train and tune the accessibility topic model, using all features (i.e. there are no "exogenous" or "endogenous" peak sets), which resulted in a model containing 17 accessibility topics. To aid in the interpretation of these topics, we performed TF binding motif enrichment analysis on each topic and exported the TF motif enrichment scores (top_quantile = 0.2) for each of the 17 accessibility topics.

Pseudotime Analysis and Lineage Trajectories

706 For each of the heart tube segment lineages that we analyzed (OFT, V, IFT), we created
707 separate subsetted AnnData objects containing only the cell types that contribute to each
708 lineage. We generated UMAP embeddings for each lineage based on joint
709 representations of gene expression and accessibility topics. We then performed
710 pseudotime analysis by calculating the diffusion map, selecting the appropriate number
711 of diffusion map components, and generated a K nearest neighbor graph based on these
712 diffusion map components. To create pseudotime trajectories, we randomly selected a
713 WT cell in the earliest progenitor cell type of each lineage to serve as the initial cell, and
714 then manually selected terminal cells from the appropriate cell types based on the
715 structure of each trajectory (for OFT lineage: E9 WT OFT-CM and E9 KO OFT-CM; for V
716 lineage: E9 WT V-CM; for IFT lineage: E8.5 KO IFT-CM, E9 WT A-CM, E9 WT AVC-CM,
717 E9 KO A-CM, E9 KO AVC-CM). Given these initial and terminal cells, along with the KNN
718 pseudotime graph, the probability of reaching each terminal state was calculated for every
719 cell and these probabilities we parsed to create the bifurcating tree lineage trajectory plots
720 (stream plots). Gene expression and accessibility topics of interest were then plotted
721 along these stream plots to interpret the dynamic changes in gene expression and
722 chromatin accessibility that occur in WT and *Mef2c* KO cells as they differentiate through
723 each lineage.
724
725 **Cloning of Reporter Constructions for Zebrafish Transgenesis**
726 To identify candidate enhancer regions to clone, we utilized our E8.5 ArchR dataset
727 containing integrated chromatin accessibility and gene expression data. We identified all
728 peaks that lost accessibility in the *Mef2c* KO (3,224), then filtered to the peaks that
729 demonstrated altered accessibility in only one of the three heart tube segments (2,549),
730 and that contained a MEF2 binding motif (1,059). Next, we utilized ArchR's Peak2Gene
731 functionality and filtered down to peaks correlated with DEGs in the *Mef2c* KOs (130) and
732 were not located in promoter regions (119). Finally, from this list of 119 candidates, we
733 selected 12 of the highest priority to screen based on the Log2FC value of the change in
734 peak accessibility and linkage to genes of interest.
735 To generate the Tg(*MVEB:egfp*)^{sfc26} zebrafish lines, 650-900 nucleotide fragments
736 containing the candidate enhancers were cloned by PCR from mouse genomic DNA.

737 Candidate nucleotide sequences and the primers used to amplify them are listed in
738 Supplementary Table 6. The resulting fragments were cloned into the Tol2 transgenic
739 vector E1b-eGFP-Tol2[30] (RRID: Addgene_37845) using the In-Fusion Cloning system
740 (Takara Bio). All plasmid sequences were confirmed by long-read sequencing before
741 zebrafish injection (Plasmidsaurus). We were unable to successfully clone one of our
742 twelve candidates (MVEB10) due to a highly repetitive "GT" sequence, so our final screen
743 consisted of eleven candidate enhancers.
744
745 **Zebrafish Transgenesis Injections, Screening, and Imaging**
746 All zebrafish experiments were reviewed and approved by the UCSF Institutional Animal
747 Care and Use Committee and were performed in accordance with the Public Health
748 Service Policy on the Humane Care and Use of Laboratory Animals. Zebrafish were
749 raised under standard laboratory conditions at 28°C. The outbred wildtype zebrafish EKW
750 (*Danio rerio-Ekwill*) and Tg(*cmlc2:mCherry*)^{s890} lines have been described previously[57].
751 One-cell zebrafish embryos obtained from Tg(*cmlc2:mCherry*)^{s890} and wildtype EKW
752 (Ekwill strain) parents were injected with 1nl of injection mix containing 15pg of candidate-
753 enhancer-eGFP and 15pg of ß-actin-BFP reporter plasmids along with 30pg of *Tol2*
754 mRNA. The ß -actin promoter drives expression of BFP strongly in somite muscles and
755 was used as a positive injection control. As a negative control, we also injected and
756 screened embryos with an empty reporter construct containing only the minimal E1b
757 promoter and eGFP. Injected embryos were screened at 24 and 72 hours post-fertilization
758 (hpf) for eGFP and BFP expression. Representative embryos were imaged for eGFP
759 expression as a readout of candidate enhancer activity and for mCherry expression as a
760 myocardial marker. Enhancer activity was scored as a percentage of embryos expressing
761 eGFP in the heart compared to the total number of BFP-positive embryos. A minimum of
762 at least 50 BFP+ embryos were screened per construct, except for MVEB1, for which only
763 17 BFP+ embryos were obtained. The exact numbers screened for each construct are
764 displayed in Extended Data Fig. 5b.
765
766 **Comparative Analyses with Published Datasets**

767 NR2F2 and GATA4 target genes were downloaded from the Harmonizome[40] and TFLink[41]
768 databases, respectively. Fisher's exact test was used to determine the Odd's Ratios and
769 p-values for identifying NR2F2 and GATA4 targets among genes upregulated in *Mef2c*
770 KO IFT-CMs at E8.5. Publicly available ChIP-seq datasets were downloaded from the
771 Gene Expression Omnibus (GEO) in wig or bigwig format for MEF2C (GSE124008)[20],
772 H3K27ac (GSE52386)[27], GATA4 (GSE52123)[39], and NR2F2 (GSE46498)[38]. Datasets
773 provided in wig format were converted to bigwig format using the Galaxy[58] web tool
774 wigToBigWig. CrossMap[59] was used together with chain files from the UCSC Genome
775 Browser[60] to lift genome annotations from mm9 to mm10 reference genome assemblies,
776 and vice versa. For the browser tracks in Fig. 5b-c, the MEF2C and H3K27ac ChIP-seq
777 bigwigs were lifted from mm9 to mm10, the Galaxy[58] web tool bigwigAverage was used
778 to generate a single average MEF2C and H3K27ac occupancy track from the individual
779 replicate tracks, and the results were displayed along with our pseudobulked snATAC-
780 seq data using the IGV software package[55]. For the heat maps displayed in Extended
781 Data Fig. 6b-c, we first created Browser Extensible Data (BED) format files for the gained
782 DARs in *Mef2c* KO IFT-CMs containing NR or GATA motifs. We then lifted these BED
783 regions from mm10 to mm9 and used the Galaxy[58] web tools computeMatrix and
784 plotHeatmap with the downloaded NR2F2 and GATA4 ChIP-seq data to visualize NR2F2
785 and GATA4 occupancy at the DARs. At the computeMatrix step, missing data was
786 converted to zeroes. At the plotHeatmap step, kmeans clustering was used to generate
787 two clusters of DARs (effectively "occupied" and "not occupied") and these were sorted
788 by decreasing mean value prior to being plotted.
789
790 **Gene Regulatory Network Inference and Analysis**
791 We used CellOracle[43] to infer GRNs for WT and *Mef2c* KO IFT-CMs at E8.5 and E9. The
792 GRN inference process included the following steps:
793
794 *Base GRN Construction*
795 We used OFT-CM, V-CM, IFT-CM, and A-CM snATAC-seq data to construct base GRNs
796 that define the accessible TF binding motifs near genes for each of the following
797 conditions: E8.5-WT, E8.5-KO, E9-WT, and E9-KO. Pooling the OFT-CM, V-CM, and A-

CM/IFT-CMs at each timepoint mitigated artifacts caused by data sparsity. The binarized PeakMatrix output from ArchR was used as an input to Cicero's[61] run_cicero() function (with default parameters) to compute co-accessibility scores among peak pairs within 500 kb. Output scores were from -1 to 1, where higher scores indicate stronger co-accessibility. Each dataset generated over 20 million peak pairs for co-accessibility comparisons. We annotated transcription start sites (TSSs) with their corresponding genes for accessible peaks using the motif_analysis.get_tss_info() function in CellOracle. Unlabeled peaks were considered potential regulatory elements. Subsequently, we applied motif_analysis.integrate_tss_peak_with_cicero() to link unlabeled peaks with genes, filtering for peak-to-gene pairs with co-accessibility scores > 0.8. TF binding motifs within each 500 bp peak were identified using the motif_analysis.TFinfo() function in CellOracle, which converts peak IDs into DNA sequences using the mm10 reference genome. TF motif scanning used the gimme.vertebrate.v5.0[62] library to produce a set of TF motifs present in accessible regions within 500kb of each gene. Finally, differentially accessible chromatin regions in KO-vs-WT were determined for each timepoint (FDR < 0.01) and inaccessible regions were pruned from the base GRNs. The TFs that bind to motifs present in regions that are co-accessible with and within 500 kb of the TSS of each gene are inferred to have the potential to regulate that gene.

GRN Inference

Using CellOracle, we constructed GRNs in which edge strengths are quantified based on the TF-to-gene correlation, constrained by the accessible regions defined in the base GRN. We inferred GRNs for E8.5 WT, E8.5 KO, E9 WT, and E9 KO IFT-CMs, using the computed base GRNs to determine permissible TF-to-gene links. GRN inference used the top 5,000 differentially expressed genes in the snRNA-seq dataset. We used the CellOracle oracle.get_links() function, with alpha set to 1,000 and otherwise default parameters, fitting linear models with L2 weight regularization to derive coefficients relating TF expression to target gene expression. The Bagging Ridge model was used to fit coefficients, providing posterior distributions for each TF-to-gene relationship. Positive coefficients indicated activating, while negative coefficients indicated inhibiting

828 relationships. We then filtered the inferred network to include 12,000 edges with the
829 largest magnitude coefficients.
830
831 *Simulation of Mef2c Knockout Effects*
832 To predict the effect of *Mef2c* KO on gene expression, we identified genes directly
833 regulated by MEF2C in the E8.5 WT IFT-CM network. We predicted that reduced
834 expression of *Mef2c* in the network would downregulate genes directly activated by
835 MEF2C and upregulate genes directly inhibited by MEF2C in the E8.5 WT IFT-CM
836 network. The simulation results were compared against differentially expressed genes in
837 the E9 snRNA-seq IFT-CM experimental data (KO-vs-WT). Simulations were performed
838 and prediction accuracy tested with varying Log2FC cutoffs for differentially expressed
839 genes, as summarized in Supplementary Table 8.
840
841 *Network Visualization*
842 Whole-network visualizations were generated using Gephi[63], applying the Yifan Hu
843 layout, with node sizes scaled by betweenness centrality. PyVis[64] was used to visualize
844 subnetworks that included manually selected TFs (*Mef2a, Mef2c, Nr2f2, Tbx5, Gata4,*
845 *Meis2, Plagl1, Isl1, Pitx2, Tbx20, Hand2, Nkx2-5*) and the top 100 differentially expressed
846 genes (ranked in ascending order by p value) identified by comparing the E8.5 WT and
847 *Mef2c* KO IFT-CM snRNA-seq experimental data. Visualizations for direct interactions
848 between TFs (*Nr2f2* and *Gata4*) and targets in the E8.5 WT and *Mef2c* KO IFT-CMs were
849 developed using Python's NetworkX package[65], with a circular layout and edge widths
850 scaled by coefficient values.
851
852 **Phenotypic Analysis of *Mef2c*$^{-/-}$ and *Mef2c*$^{-/-}$; *Nr2f2*$^{+/-}$ Embryos**
853 We conducted comparative phenotypic analysis of *Mef2c*$^{-/-}$ and *Mef2c*$^{-/-}$; *Nr2f2*$^{+/-}$ embryos
854 by performing immunofluorescent staining of cardiac Troponin T and analyzing the
855 apparent heart tube phenotypes, initially unblinded to embryo genotype. We determined
856 that the heart tube phenotypes ranged in apparent severity and fell into roughly three
857 categories, with Category 1 being the most severely malformed and Category 3 the least

858 severe. We then blinded ourselves to the embryo genotypes and re-examined the
859 embryos, placing each into the appropriate category of severity.

ACKNOWLEDGEMENTS

J.M.M-V. is grateful for assistance from Dr. Kavitha Rao in learning Seurat, as well as from Drs Ryan Corces and Christina Theodoris for assistance with ArchR and MIRA, respectively. J.M.M-V. is also grateful to all his colleagues in the Bruneau Lab for many insightful discussions, important feedback on directions of the project, and help with experiments. We thank the Gladstone Genomics Core for use and maintenance of the Agilent Bioanalyzer and 10x Genomics Chromium Controller. Portions of this work were performed on the Wynton HPC Co-Op cluster, which is supported by UCSF research faculty and UCSF institutional funds. The authors wish to thank the UCSF Wynton team for their ongoing technical support of the Wynton environment.

AUTHOR CONTRIBUTIONS

J.M.M-V. and B.G.B. conceived and designed the experiments presented in this study. J.M.M-V. performed embryo dissections, whole-mount embryo immunostaining, whole-mount embryo FISH via HCR, embryo imaging, preparation of sequencing libraries, and data analysis via Seurat, ArchR, and MIRA. E.F.B. performed mouse colony maintenance and genotyping. J.M.M-V. and E.F.B. performed cloning of reporter constructs for zebrafish transgenesis. T.S. performed zebrafish transgenesis injections and imaging. A.P.C. performed GRN analyses using CellOracle and newly written code. B.G.B., B.L.B., and J.J.S. supervised and advised. J.M.M-V. prepared the figures and wrote the manuscript with input from the co-authors.

DECLARATION OF INTERESTS

B.G.B. is a founder, shareholder, and advisor of Tenaya Therapeutics and an advisor for Silver Creek Pharmaceuticals. The work presented here is not related to the interests of these commercial entities.

FUNDING

This work was funded by grants from the National Heart, Lung, and Blood Institute (NHLBI) of the National Institutes of Health (F32HL162450 to J.M.M-V.; T32HL007284 to A.P.C.; R01HL160665 and R01HL162925 to J.J.S.; R01HL177462 to B.L.B.;

R01HL114948 and R01HL155906 to B.G.B.). The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This research was also supported by the Roddenberry Foundation (B.G.B.), Additional Ventures (B.G.B.), and the Younger Family Fund (B.G.B.). J.M.M-V. was also supported by a postdoctoral fellowship from the American Heart Association together with The Children's Heart Foundation (Grant ID: 24POST1191660, DOI: https://doi.org/10.58275/AHA.24POST1191660.pc.gr.190926). Sequencing performed at the UCSF CAT was supported by UCSF PBBR, RRP IMIA, and NIH 1S10OD028511-01 grants. This work was also supported by an NIH/NCRR grant (C06 RR018928) to the J. David Gladstone Institutes.

DATA AND CODE AVAILABILITY

Raw and processed data for snRNA-seq and snATAC-seq datasets has been deposited in the NCBI Gene Expression Omnibus (GEO) under accession number GSE280587. Code used for multiomics analyses, GRNs, and plot generation is available in the GitHub repository associated with this manuscript: https://github.com/jmmuncie/Muncie-Vasic_2025.

908 **FIGURE CAPTIONS**

909 **Figure 1: MEF2C is expressed throughout the developing heart tube and its loss**
910 **causes segment-specific defects.**

911 a) Schematic of cardiac progenitors and their contributions to linear heart tube
912 development from cardiac crescent (E7.75) to looped heart tube (E9) stage. b)
913 Immunofluorescent staining of MEF2C (cyan) and cardiac Troponin T (cTnT, magenta) in
914 E7.75, E8.5, and E9 WT embryos. c) Representative images of WT and *Mef2c* KO
915 embryos at E7.75, E8.5 and E9. Cardiac progenitors are marked by the *Smarcd3*-F6-
916 eGFP reporter transgene (green). BF, brightfield. d) Schematic of the methodology and
917 biological insights presented in the current study. Elements of this panel were created in
918 BioRender. B, B. (2024) https://BioRender.com/i72e213. Scale bars = 200 μm. FHF, first
919 heart field; aSHF, anterior second heart field; pSHF, posterior second heart field; LV, left
920 ventricle; RV, right ventricle; V, ventricle; IFT, inflow tract; OFT, outflow tract.

921

922 **Figure 2: Loss of MEF2C reduces expression of key cardiomyocyte genes across**
923 **all heart tube segments and alters expression of anterior/posterior markers.**

924 a-c) UMAPs of snRNA-seq data for cardiac progenitors, cardiomyocytes, and related
925 mesoderm subtypes from E7.75 (a), E8.5 (b), and E9 (c) embryos labeled by cell type
926 (left) and genotype/sample ID (right). d-f) Bar plots displaying the number of up-regulated
927 and down-regulated genes in *Mef2c* KO relative to WT in cell types of interest at E7.75
928 (d), E8.5 (e), and E9 (f). g-i) Dot plots displaying expression of key CM genes and
929 anterior/posterior (A/P) markers at E7.75 (g), E8.5 (h), and E9 (i). j) Fluorescence *in situ*
930 hybridization of key CM genes and A/P markers in E8.5-E9 WT and *Mef2c* KO embryos.
931 Note the reduced expression of CM genes *Tnnt2*, *Ttn*, *Nppa*, and *Nkx2-5*, the expanded
932 expression of posterior markers *Tbx5*, *Gata4*, and *Wnt2*, and the loss of anterior OFT
933 marker *Tdgf1* in *Mef2c* KO embryos compared to WT. Scale bars = 200 μm. CMs,
934 cardiomyocytes; FHF, first heart field; SHF, second heart field; JCF, juxtacardiac field;
935 CrM, cranial mesoderm; PrxM, paraxial mesoderm; LPM, lateral plate mesoderm; SoM,
936 somitic mesoderm; NMPs, neuromesodermal progenitors; KPs, kidney progenitors; ExM,
937 extraembryonic mesoderm; HSCs, hematopoietic stem cells; V-CMs, ventricular
938 cardiomyocytes; IFT-CMs, inflow tract cardiomyocytes; OFT-CMs, outflow tract

939 cardiomyocytes; aSHF, anterior second heart field; pSHF, posterior second heart field;
940 PostM, posterior mesoderm; PhM, pharyngeal mesoderm; MixM, mixed mesoderm; A-
941 CMs, atrial cardiomyocytes; AVC-CMs, atrioventricular canal cardiomyocytes; Pe,
942 proepicardium; VP, venous pole; *, Genes known to be associated with CHDs[19]; #, Direct
943 targets of MEF2C based on MEF2C ChIP-seq data[20].
944
945 **Figure 3: MEF2C regulates chromatin accessibility broadly throughout the heart**
946 **tube and in a segment-specific manner.**
947 a) UMAP of integrated snRNA-seq and snATAC-seq data for cardiac progenitors,
948 cardiomyocytes, and related mesoderm subtypes at E8.5 labeled by cell types
949 determined from snRNA-seq clustering (left) and genotype/sample ID (right). b) Bar plots
950 displaying the number of gained and lost DARs in *Mef2c* KO relative to WT cell types of
951 interest at E8.5. c) Venn diagrams displaying unique and overlapping DARs in the three
952 heart tube segments. d) Scatter plots displaying the relationship between the Log2 fold
953 change (Log2FC) values of the DEG and DAR analyses for all identified Peak2Gene
954 (P2G) links in the three heart tube segments. Dots are colored by the P2G correlation
955 score. e-f) Genome browser tracks displaying snATAC-seq accessibility profiles at the
956 *Myh6*/*Myh7* (e) and *Wnt2* (f) loci for the indicated pseudobulked cell types. DARs are
957 highlighted and the peaks are indicated by red bars (*Mef2c* KO-vs-WT IFT-CMs). Loops
958 indicate P2G links, colored by the P2G correlation score. V-CMs, ventricular
959 cardiomyocytes; IFT-CMs, inflow tract cardiomyocytes; OFT-CMs, outflow tract
960 cardiomyocytes; aSHF, anterior second heart field; pSHF, posterior second heart field;
961 LPM, lateral plate mesoderm; PostM, posterior mesoderm; PhM, pharyngeal mesoderm;
962 NA, cells not available in the snRNA-seq dataset; DAR, differentially accessible region;
963 DEG, differentially expressed gene.
964
965 **Figure 4: Each heart tube segment exhibits a distinct MEF2C-depedent**
966 **developmental trajectory.**
967 a) Schematic of multimodal Models for Integrated Regulatory Analysis (MIRA)[24] pipeline.
968 b) UMAP of outflow tract lineage cells plotted by MIRA topic models labeled by cell type
969 (top) and genotype (bottom). c) Pseudotime plot of outflow tract lineage cells. d) Lineage

trajectory stream plots for outflow tract lineage cells labeled by cell type (top) and genotype (bottom). e) Stream plots displaying the flow of gene expression (top) and chromatin accessibility (bottom) topics in the outflow tract lineage cells. Examples of the top genes and TF binding motifs for dynamic topics are labeled. f-i) Same as (b-e), but for ventricular lineage cells. j-m) Same as (b-e) and (f-i), but for inflow tract lineage cells.

Figure 5: Candidate regulatory elements with MEF2C-dependent chromatin accessibility display enhancer activity in zebrafish.

a) Schematic of selection process to identify candidate MEFC-dependent enhancers from integrated snRNA-seq and snATAC-seq data. b) Genome browser tracks displaying snATAC-seq accessibility profiles, MEF2C ChIP-seq occupancy profiles[20], and H3K27ac ChIP-seq occupancy profiles[27] at example loci containing IFT-specific (left), V-specific (middle), and OFT-specific (right) candidate enhancers (yellow highlights). c) Genome browser tracks displaying snATAC-seq accessibility profiles, MEF2C ChIP-seq occupancy profiles[20], and H3K27ac ChIP-seq occupancy profiles[27] at the *Myh6*/*Myh7* locus, which contains two candidate enhancers with IFT-specific altered accessibility (yellow highlights, MVEB1 and MVEB2) that overlap with regions found in the VISTA Enhancer Browser database[29]. d) Images of E11.5 mouse embryos from the Vista Enhancer Browser database[29] demonstrating positive enhancer activity of regions that overlap with candidates MVEB1 and MVEB2. e) Schematic of the Tol2 transgenesis assay used to screen candidate enhancers in zebrafish. Elements of this panel were created in BioRender. B, B. (2024) https://BioRender.com/h83r503. f) Representative ventral view images of Tg(*cmlc2:mCherry*) zebrafish embryos at 72 hours post-fertilization (hpf) injected with candidate enhancers that demonstrated positive activity in the heart. Boxed area in the representative brightfield image (left) indicates the anatomical region of interest captured in the fluorescent images. g) Schematic representation of the observed onset of enhancer activity for candidate enhancers that demonstrated positive activity in the heart. h) Representative ventral view images of Tg(*cmlc2:mCherry*) zebrafish embryos at 24 and 72 hpf injected with *MVEB2:eGFP* or *MVEB6:eGFP* reporter constructs. Scale bars = 100 μm.

1001 **Figure 6: Loss of MEF2C induces an overactive posteriorized gene regulatory**
1002 **network that is partially rescued by reduced NR2F2 dosage.**
1003 a) TF binding motif enrichment analysis for gained DARs in *Mef2c* KO relative to WT IFT-
1004 CMs at E8.5. b) Pie charts showing the proportion of lost or gained DARs (*Mef2c* KO-vs-
1005 WT) containing NR and MEF2C motifs or only NR motifs. c) Odds ratio analysis for NR2F2
1006 target genes[40] amongst DEGs up-regulated in *Mef2c* KO IFT-CMs. p-value calculated
1007 using Fisher's exact test. d) Ridge plot displaying module scores for up-regulated NR2F2
1008 targets in *Mef2c* KO and WT heart tube segments. e) Inferred GRNs constructed for
1009 *Mef2c* KO and WT IFT-CMs at E8.5 and E9. Boxed region of E8.5 WT GRN is shown at
1010 higher magnification in (f). f) Schematic of *in silico* simulated *Mef2c* KO in the E8.5 WT
1011 GRN. g) Results of GRN validation displaying the high accuracy (74%) of predicted
1012 relative to measured gene expression changes at E9. h) Visualizations of subnetworks
1013 consisting of 12 cardiac TFs and the top 100 DEGs within the WT and *Mef2c* KO E8.5
1014 IFT-CM GRNs. i) Visualization of direct NR2F2 interactions in the WT and *Mef2c* KO E8.5
1015 IFT-CM GRNs. Direct interactions that occur upon *Mef2c* KO are highlighted. #, mis-
1016 regulated DEG in *Mef2c* KO IFT-CMs at E8.5; *, Direct target of NR2F2[40]. j-k)
1017 Immunofluorescent staining of cardiac Troponin T (cTnT, green) in representative E9.5
1018 embryos (18-24 somites) collected from *Mef2c*$^{+/-}$;*Nr2f2*$^{+/-}$ to *Mef2c*$^{+/-}$ crosses. Boxed
1019 regions in (j) are shown at higher magnification in (k). Arrows point to the ventricle, which
1020 is expanded in *Mef2c*$^{-/-}$;*Nr2f2*$^{+/-}$ embryos compared to *Mef2c*$^{-/-}$ embryos. Asterisks mark
1021 the atria, which are better developed and have undergone more looping in *Mef2c*$^{-/-}$
1022 ;*Nr2f2*$^{+/-}$ embryos compared to *Mef2c*$^{-/-}$ embryos. n=5 *Mef2c*$^{-/-}$ and n=7 *Mef2c*$^{-/-}$;*Nr2f2*$^{+/-}$
1023 embryos from 7 independent litters. Scale bars = 200 μm

1024 **REFERENCES**

1025 1. Hafner A, Boettiger A. The spatial organization of transcriptional control. Nat Rev Genet.
1026 2023;24(1):53–68.

1027 2. Wagh K, Stavreva DA, Upadhyaya A, Hager GL. Transcription Factor Dynamics: One
1028 Molecule at a Time. Annu Rev Cell Dev Biol. 2023;39(1):277–305.

1029 3. Spitz F, Furlong EEM. Transcription factors: from enhancer binding to developmental control.
1030 Nat Rev Genet. 2012;13(9):613–26.

1031 4. Lee TI, Young RA. Transcriptional Regulation and Its Misregulation in Disease. Cell.
1032 2013;152(6):1237–51.

1033 5. Mercatelli D, Scalambra L, Triboli L, Ray F, Giorgi FM. Gene regulatory network inference
1034 resources: A practical overview. Biochim Biophys Acta (BBA) - Gene Regul Mech.
1035 2020;1863(6):194430.

1036 6. Bruneau BG. Signaling and Transcriptional Networks in Heart Development and
1037 Regeneration. Cold Spring Harb Perspect Biol. 2013;5(3):a008292.

1038 7. Kelly RG, Buckingham ME, Moorman AF. Heart Fields and Cardiac Morphogenesis. Cold
1039 Spring Harb Perspect Med. 2014;4(10):a015750.

1040 8. Evans SM, Yelon D, Conlon FL, Kirby ML. Myocardial Lineage Development. Circ Res.
1041 2010;107(12):1428–44.

1042 9. Lescroart F, Chabab S, Lin X, Rulands S, Paulissen C, Rodolosse A, et al. Early lineage
1043 restriction in temporally distinct populations of Mesp1 progenitors during mammalian heart
1044 development. Nat Cell Biol. 2014;16(9):829–40.

1045 10. Devine WP, Wythe JD, George M, Koshiba-Takeuchi K, Bruneau BG. Early patterning and
1046 specification of cardiac progenitors in gastrulating mesoderm. eLife. 2014;3:e03848.

1047 11. Meilhac SM, Esner M, Kelly RG, Nicolas JF, Buckingham ME. The Clonal Origin of
1048 Myocardial Cells in Different Regions of the Embryonic Mouse Heart. Dev Cell. 2004;6(5):685–
1049 98.

1050 12. Martin JF, Schwarz JJ, Olson EN. Myocyte enhancer factor (MEF) 2C: a tissue-restricted
1051 member of the MEF-2 family of transcription factors. Proc Natl Acad Sci. 1993;90(11):5282–6.

1052 13. Lin Q, Schwarz J, Bucana C, Olson EN. Control of Mouse Cardiac Morphogenesis and
1053 Myogenesis by Transcription Factor MEF2C. Science. 1997;276(5317):1404–7.

1054 14. Krup AL, Winchester SAB, Ranade SS, Agrawal A, Devine WP, Sinha T, et al. A Mesp1-
1055 dependent developmental breakpoint in transcriptional and epigenomic specification of early
1056 cardiac precursors. Development. 2023;150(9):dev201229.

1057 15. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis
1058 of multimodal single-cell data. Cell. 2021;184(13):3573-3587.e29.

1059 16. Choudhary S, Satija R. Comparison and evaluation of statistical error models for scRNA-seq.
1060 Genome Biol. 2022;23(1):27.

1061 17. Tyser RCV, Ibarra-Soria X, McDole K, Jayaram SA, Godwin J, Brand TAH van den, et al.
1062 Characterization of a common progenitor pool of the epicardium and myocardium. Science.
1063 2021;371(6533).

1064 18. Black BL, Olson EN. Transcriptional control of muscle development by myocyte enhancer
1065 factor-2 (MEF2) proteins. Annu Rev Cell Dev Biol. 1998;14(1):167–96.

1066 19. Yang A, Alankarage D, Cuny H, Ip EKK, Almog M, Lu J, et al. CHDgene: A Curated
1067 Database for Congenital Heart Disease Genes. Circ: Genom Precis Med. 2022;15(3):e003539.

1068 20. Akerberg BN, Gu F, VanDusen NJ, Zhang X, Dong R, Li K, et al. A reference map of
1069 murine cardiac transcription factor chromatin occupancy identifies dynamic and conserved
1070 enhancers. Nat Commun. 2019;10(1):4907.

1071 21. Barnes RM, Harris IS, Jaehnig EJ, Sauls K, Sinha T, Rojas A, et al. MEF2C regulates
1072 outflow tract alignment and transcriptional control of Tdgf1. Development. 2016;143(5):774–9.

1073 22. Granja JM, Corces MR, Pierce SE, Bagdatli ST, Choudhry H, Chang HY, et al. ArchR is a
1074 scalable software package for integrative single-cell chromatin accessibility analysis. Nat Genet.
1075 2021;53(3):403–11.

1076 23. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based
1077 Analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.

1078 24. Lynch AW, Theodoris CV, Long HW, Brown M, Liu XS, Meyer CA. MIRA: joint
1079 regulatory modeling of multimodal expression and chromatin accessibility in single cells. Nat
1080 Methods. 2022;19(9):1097–108.

1081 25. Zhou W, Lin L, Majumdar A, Li X, Zhang X, Liu W, et al. Modulation of morphogenesis by
1082 noncanonical Wnt signaling requires ATF/CREB family–mediated transcriptional activation of
1083 TGFβ2. Nat Genet. 2007;39(10):1225–34.

1084 26. Pereira FA, Qiu Y, Zhou G, Tsai MJ, Tsai SY. The orphan nuclear receptor COUP-TFII is
1085 required for angiogenesis and heart development. Genes Dev. 1999;13(8):1037–49.

1086 27. Nord AS, Blow MJ, Attanasio C, Akiyama JA, Holt A, Hosseini R, et al. Rapid and
1087 Pervasive Changes in Genome-wide Enhancer Usage during Mammalian Development. Cell.
1088 2013;155(7):1521–31.

1089 28. Man JCK, Duijvenboden K van, Krijger PHL, Hooijkaas IB, Made I van der, Vries C de G
1090 de, et al. Genetic Dissection of a Super Enhancer Controlling the Nppa-Nppb Cluster in the
1091 Heart. Circ Res. 2021;128(1):115–29.

1092 29. Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA Enhancer Browser—a database
1093 of tissue-specific human enhancers. Nucleic Acids Res. 2007;35(suppl_1):D88–92.

1094 30. Birnbaum RY, Clowney EJ, Agamy O, Kim MJ, Zhao J, Yamanaka T, et al. Coding exons
1095 function as tissue-specific enhancers of nearby genes. Genome Res. 2012;22(6):1059–68.

1096 31. Fisher S, Grice EA, Vinton RM, Bessling SL, Urasaki A, Kawakami K, et al. Evaluating the
1097 biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in
1098 zebrafish. Nat Protoc. 2006;1(3):1297–305.

1099 32. Hinits Y, Pan L, Walker C, Dowd J, Moens CB, Hughes SM. Zebrafish Mef2ca and Mef2cb
1100 are essential for both first and second heart field cardiomyocyte differentiation. Dev Biol.
1101 2012;369(2):199–210.

1102 33. Zlatanova I, Sun F, Wu RS, Chen X, Lau BH, Colombier P, et al. An injury-responsive
1103 mmp14b enhancer is required for heart regeneration. Sci Adv. 2023;9(48):eadh5313.

1104 34. Yan R, Cigliola V, Oonk KA, Petrover Z, DeLuca S, Wolfson DW, et al. An enhancer-based
1105 gene-therapy strategy for spatiotemporal control of cargoes during tissue repair. Cell Stem Cell.
1106 2023;30(1):96-111.e6.

1107 35. Kang J, Hu J, Karra R, Dickson AL, Tornini VA, Nachtrab G, et al. Modulation of tissue
1108 repair by regeneration enhancer elements. Nature. 2016;532(7598):201–6.

1109 36. Verzi MP, McCulley DJ, Val SD, Dodou E, Black BL. The right ventricle, outflow tract, and
1110 ventricular septum comprise a restricted expression domain within the secondary/anterior heart
1111 field. Dev Biol. 2005;287(1):134–45.

1112 37. Tremblay M, Sanchez-Ferras O, Bouchard M. GATA transcription factors in development
1113 and disease. Development. 2018;145(20):dev164384.

1114 38. Wu S pin, Cheng CM, Lanz RB, Wang T, Respress JL, Ather S, et al. Atrial Identity Is
1115 Determined by a COUP-TFII Regulatory Network. Dev Cell. 2013;25(4):417–26.

1116 39. He A, Gu F, Hu Y, Ma Q, Ye LY, Akiyama JA, et al. Dynamic GATA4 enhancers shape the
1117 chromatin landscape central to heart development and disease. Nat Commun. 2014;5(1):4907.

1118 40. Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG, et al.
1119 The harmonizome: a collection of processed datasets gathered to serve and mine knowledge
1120 about genes and proteins. Database. 2016;2016(2016):baw100.

1121 41. Liska O, Bohár B, Hidas A, Korcsmáros T, Papp B, Fazekas D, et al. TFLink: an integrated
1122 gateway to access transcription factor–target gene interactions for multiple species. Database.
1123 2022;2022:baac083.

1124 42. Tirosh I, Izar B, Prakadan SM, II MHW, Treacy D, Trombetta JJ, et al. Dissecting the
1125 multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science.
1126 2016;352(6282):189–96.

1127 43. Kamimoto K, Stringa B, Hoffmann CM, Jindal K, Solnica-Krezel L, Morris SA. Dissecting
1128 cell identity via network inference and in silico gene perturbation. Nature. 2023;614(7949):742–
1129 51.

1130 44. Stone NR, Gifford CA, Thomas R, Pratt KJB, Samse-Knapp K, Mohamed TMA, et al.
1131 Context-Specific Transcription Factor Functions Regulate Epigenomic and Transcriptional
1132 Dynamics during Cardiac Reprogramming. Cell Stem Cell. 2019;25(1):87-102.e9.

1133 45. Al-Yacoub N, Shaheen R, Awad SM, Kunhi M, Dzimiri N, Nguyen HC, et al. FBXO32,
1134 encoding a member of the SCF complex, is mutated in dilated cardiomyopathy. Genome Biol.
1135 2016;17(1):2.

1136 46. Ghasemi S, Mahdavi M, Maleki M, Salahshourifar I, Kalayinia S. A novel likely pathogenic
1137 variant in the FBXO32 gene associated with dilated cardiomyopathy according to whole-exome
1138 sequencing. BMC Méd Genom. 2022;15(1):234.

1139 47. Rowton M, Perez-Cervantes C, Hur S, Jacobs-Li J, Lu E, Deng N, et al. Hedgehog signaling
1140 activates a mammalian heterochronic gene regulatory network controlling differentiation timing
1141 across lineages. Dev Cell. 2022;57(18):2181-2203.e9.

1142 48. Tyser RC, Miranda AM, Chen C mun, Davidson SM, Srinivas S, Riley PR. Calcium
1143 handling precedes cardiac differentiation to initiate the first heartbeat. eLife. 2016;5:e17113.

1144 49. Barriga EH, Franze K, Charras G, Mayor R. Tissue stiffening coordinates morphogenesis by
1145 triggering collective cell migration in vivo. Nature. 2018;554(7693):523–7.

1146 50. Caldarelli P, Chamolly A, Villedieu A, Alegria-Prévot O, Phan C, Gros J, et al. Self-
1147 organized tissue mechanics underlie embryonic regulation. Nature. 2024;1–8.

1148 51. Luna-Zurita L, Stirnimann CU, Glatt S, Kaynak BL, Thomas S, Baudin F, et al. Complex
1149 Interdependence Regulates Heterotypic Transcription Factor Distribution and Coordinates
1150 Cardiogenesis. Cell. 2016;164(5):999–1014.

1151 52. Takamoto N, You LR, Moses K, Chiang C, Zimmer WE, Schwartz RJ, et al. COUP-TFII is
1152 essential for radial and anteroposterior patterning of the stomach. Development.
1153 2005;132(9):2179–89.

1154 53. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and
1155 accurate integration of single-cell data with Harmony. Nat Methods. 2019;16(12):1289–96.

1156 54. Vierstra J, Lazar J, Sandstrom R, Halow J, Lee K, Bates D, et al. Global reference mapping
1157 of human transcription factor footprints. Nature. 2020;583(7818):729–36.

1158 55. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al.
1159 Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24–6.

1160 56. Cakir B, Prete M, Huang N, Dongen S van, Pir P, Kiselev VY. Comparison of visualization
1161 tools for single-cell RNAseq data. NAR Genom Bioinform. 2020;2(3):lqaa052.

1162 57. Zhang R, Han P, Yang H, Ouyang K, Lee D, Lin YF, et al. In vivo cardiac reprogramming
1163 contributes to zebrafish heart regeneration. Nature. 2013;498(7455):497–501.

1164 58. Community TG, Abueg LAL, Afgan E, Allart O, Awan AH, Bacon WA, et al. The Galaxy
1165 platform for accessible, reproducible, and collaborative data analyses: 2024 update. Nucleic
1166 Acids Res. 2024;52(W1):W83–94.

1167 59. Zhao H, Sun Z, Wang J, Huang H, Kocher JP, Wang L. CrossMap: a versatile tool for
1168 coordinate conversion between genome assemblies. Bioinformatics. 2014;30(7):1006–7.

1169 60. Raney BJ, Barber GP, Benet-Pagès A, Casper J, Clawson H, Cline MS, et al. The UCSC
1170 Genome Browser database: 2024 update. Nucleic Acids Res. 2023;52(D1):D1082–8.

1171 61. Pliner HA, Packer JS, McFaline-Figueroa JL, Cusanovich DA, Daza RM, Aghamirzaie D, et
1172 al. Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility
1173 Data. Mol Cell. 2018;71(5):858-871.e8.

1174 62. Bruse N, Heeringen SJ van. GimmeMotifs: an analysis framework for transcription factor
1175 motif analysis. bioRxiv. 2018;474403.

1176 63. Bastian M, Heymann S, Jacomy M. Gephi: An Open Source Software for Exploring and
1177 Manipulating Networks. ICWSM. 2009;3:361–2.

1178 64. Perrone G, Unpingco J, Lu H minn. Network visualizations with Pyvis and VisJS. arXiv.
1179 2020;

1180 65. Hagberg AA, Schult DA, Swart PJ. Exploring Network Structure, Dynamics, and Function
1181 using NetworkX. In: Varoquaux G, Vaught T, Millman J, editors. Proceedings of the 7th Python
1182 in Science Conference (SciPy 2008). 2008. p. 11–6.

Figure 1

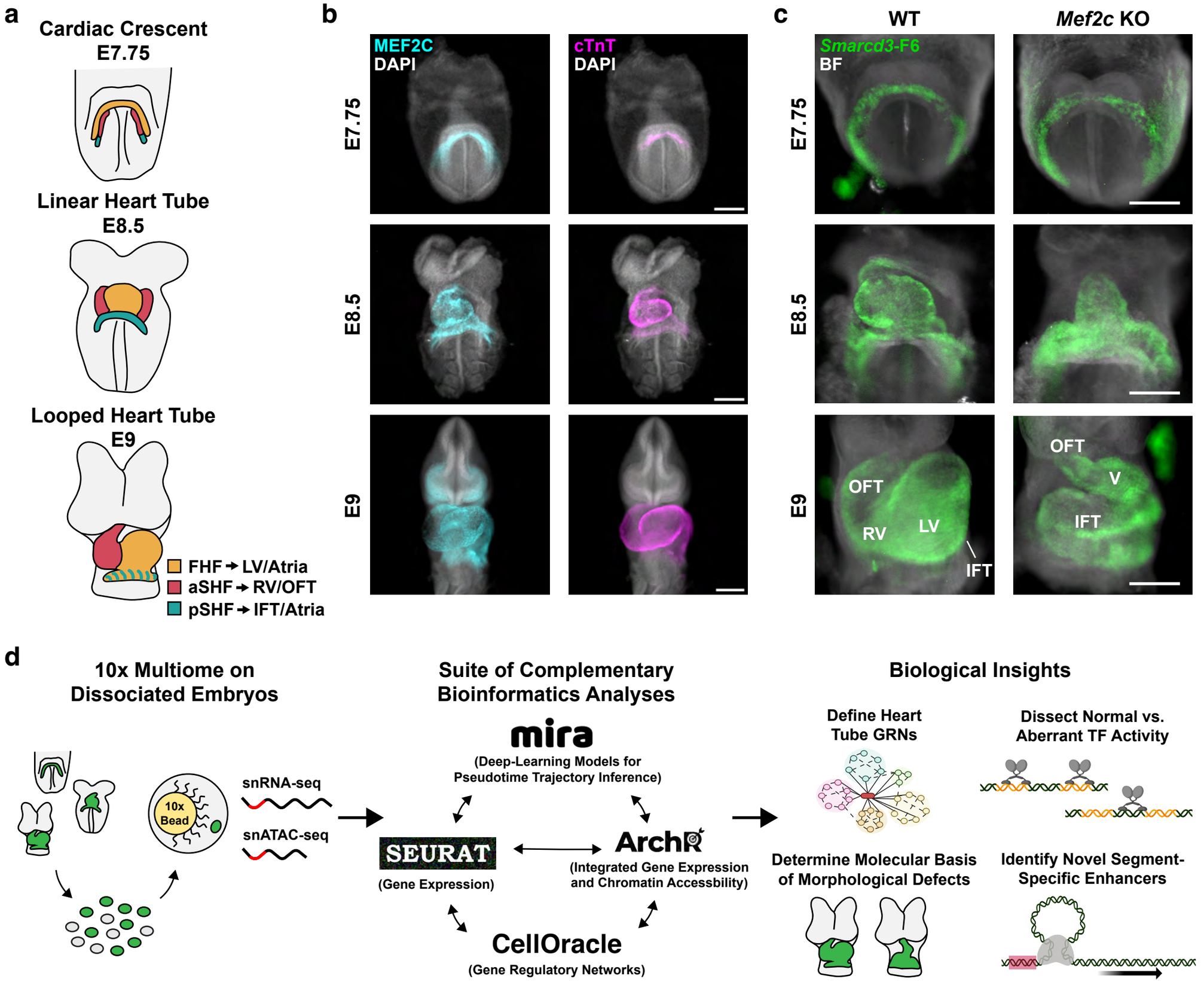

a
Cardiac Crescent
E7.75
Linear Heart Tube
E8.5
Looped Heart Tube
E9
FHF → LV/Atria
aSHF → RV/OFT
pSHF → IFT/Atria
b
MEF2C
DAPI
cTnT
DAPI
E7.75
E8.5
E9
c
WT
Mef2c KO
Smarcd3-F6
BF
E7.75
E8.5
E9
OFT
RV
LV
IFT
OFT
V
IFT
d
10x Multiome on
Dissociated Embryos
10x
Bead
snRNA-seq
snATAC-seq
Suite of Complementary
Bioinformatics Analyses
mira
(Deep-Learning Models for
Pseudotime Trajectory Inference)
SEURAT
(Gene Expression)
ArchR
(Integrated Gene Expression
and Chromatin Accessbility)
CellOracle
(Gene Regulatory Networks)
Biological Insights
Define Heart
Tube GRNs
Dissect Normal vs.
Aberrant TF Activity
Determine Molecular Basis
of Morphological Defects
Identify Novel Segment-
Specific Enhancers

Figure 2

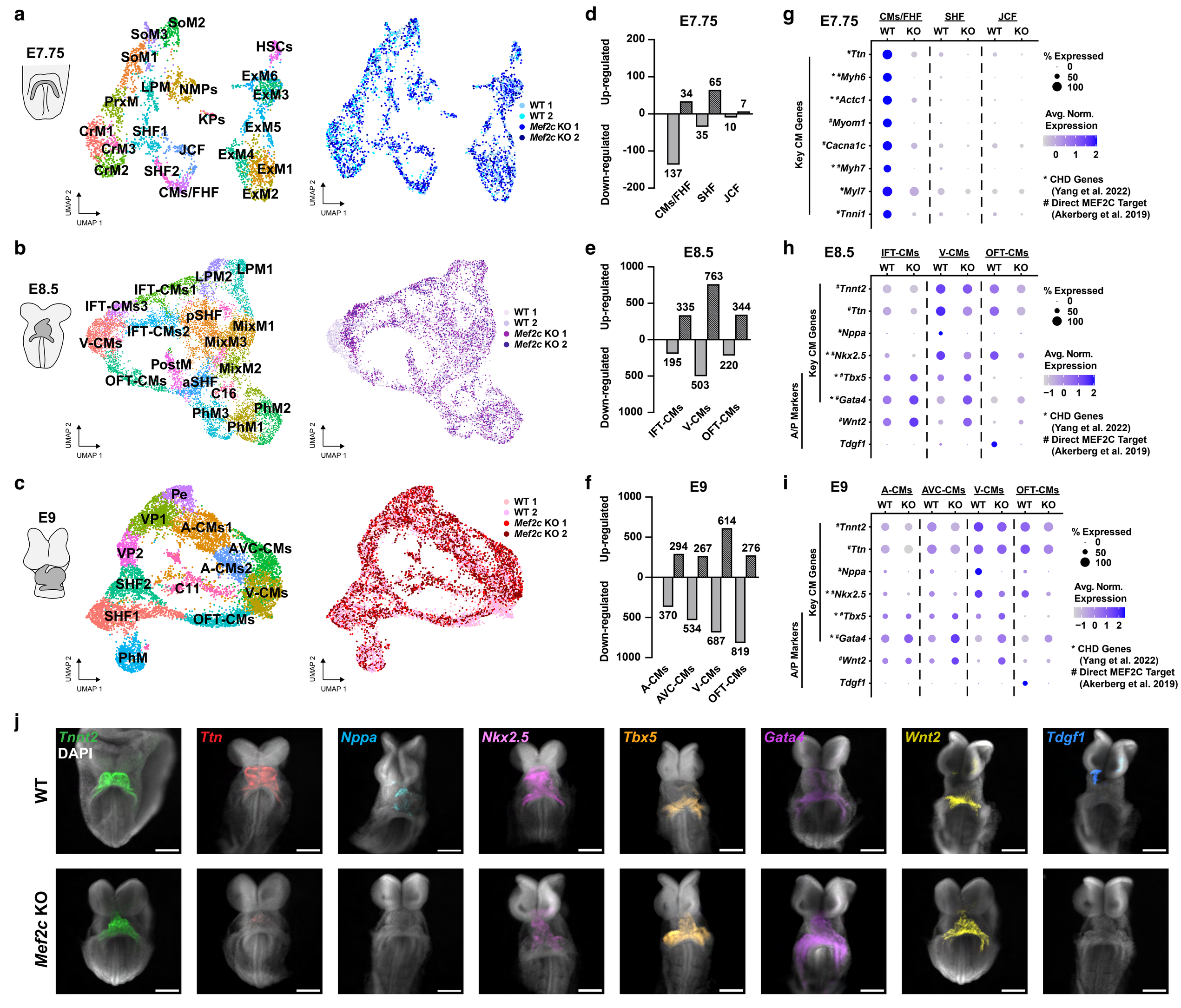

a
E7.75
SoM2
SoM3
SoM1
HSCs
ExM6
ExM3
LPM
NMPs
PrxM
KPs
ExM5
CrM1
SHF1
CrM3
JCF
ExM4
CrM2
SHF2
ExM1
CMs/FHF
ExM2
UMAP 2
UMAP 1
WT 1
WT 2
Mef2c KO 1
Mef2c KO 2
b
E8.5
LPM2
LPM1
IFT-CMs1
IFT-CMs3
pSHF
MixM1
IFT-CMs2
V-CMs
MixM3
PostM
MixM2
OFT-CMs
aSHF
C16
PhM2
PhM3
PhM1
c
E9
Pe
VP1
A-CMs1
AVC-CMs
VP2
A-CMs2
SHF2
C11
V-CMs
SHF1
OFT-CMs
PhM
d
E7.75
Up-regulated
Down-regulated
200
100
0
-100
-200
34
65
7
137
35
10
CMs/FHF
SHF
JCF
e
E8.5
1000
500
0
500
1000
335
763
344
195
503
220
IFT-CMs
V-CMs
OFT-CMs
f
E9
1000
500
0
500
1000
294
267
614
276
370
534
687
819
A-CMs
AVC-CMs
V-CMs
OFT-CMs
g
E7.75
CMs/FHF
SHF
JCF
WT
KO
Key CM Genes
#Ttn
*#Myh6
*#Actc1
#Myom1
#Cacna1c
*#Myh7
#Myl7
#Tnni1
% Expressed
0
50
100
Avg. Norm. Expression
0
1
2
* CHD Genes (Yang et al. 2022)
# Direct MEF2C Target (Akerberg et al. 2019)
h
E8.5
IFT-CMs
V-CMs
OFT-CMs
A/P Markers
#Tnnt2
#Ttn
#Nppa
*#Nkx2.5
*#Tbx5
*#Gata4
#Wnt2
Tdgf1
−1
0
1
2
i
E9
A-CMs
AVC-CMs
V-CMs
OFT-CMs
j
Tnnt2
DAPI
Ttn
Nppa
Nkx2.5
Tbx5
Gata4
Wnt2
Tdgf1
WT
Mef2c KO

Figure 3

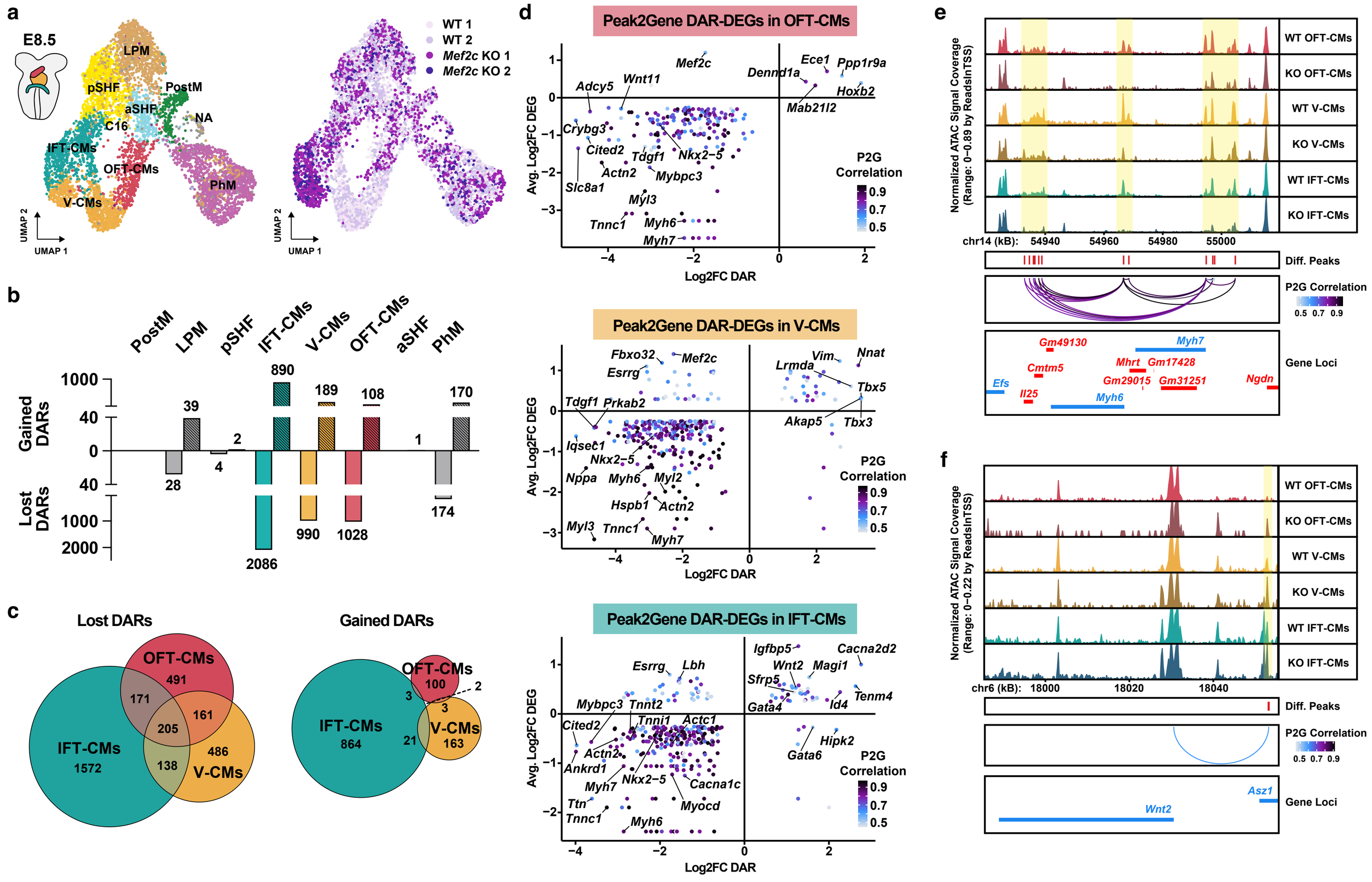

a
E8.5
LPM
pSHF
PostM
aSHF
C16
NA
IFT-CMs
OFT-CMs
PhM
V-CMs
UMAP 2
UMAP 1
WT 1
WT 2
Mef2c KO 1
Mef2c KO 2
b
PostM
LPM
pSHF
IFT-CMs
V-CMs
OFT-CMs
aSHF
PhM
Gained DARs
Lost DARs
1000
40
0
40
1000
2000
39
28
4
2
890
2086
189
990
108
1028
1
170
174
c
Lost DARs
OFT-CMs
491
171
161
205
IFT-CMs
1572
138
486
V-CMs
Gained DARs
OFT-CMs
100
3
2
3
IFT-CMs
864
21
V-CMs
163
d
Peak2Gene DAR-DEGs in OFT-CMs
Avg. Log2FC DEG
Log2FC DAR
Mef2c
Ece1
Ppp1r9a
Dennd1a
Hoxb2
Mab21l2
Adcy5
Wnt11
Crybg3
Cited2
Tdgf1
Nkx2−5
Actn2
Mybpc3
Slc8a1
Myl3
Tnnc1
Myh6
Myh7
P2G Correlation
0.9
0.7
0.5
Peak2Gene DAR-DEGs in V-CMs
Fbxo32
Mef2c
Esrrg
Vim
Nnat
Lrmda
Tbx5
Tdgf1
Prkab2
Akap5
Tbx3
Iqsec1
Nkx2−5
Nppa
Myh6
Myl2
Hspb1
Actn2
Myl3
Tnnc1
Myh7
Peak2Gene DAR-DEGs in IFT-CMs
Igfbp5
Cacna2d2
Esrrg
Lbh
Wnt2
Magi1
Sfrp5
Tenm4
Mybpc3
Tnnt2
Gata4
Id4
Cited2
Tnni1
Actc1
Hipk2
Actn2
Gata6
Ankrd1
Nkx2−5
Cacna1c
Myh7
Ttn
Myocd
Tnnc1
Myh6
e
Normalized ATAC Signal Coverage
(Range: 0−0.89 by ReadsInTSS)
WT OFT-CMs
KO OFT-CMs
WT V-CMs
KO V-CMs
WT IFT-CMs
KO IFT-CMs
chr14 (kB):
54940
54960
54980
55000
Diff. Peaks
P2G Correlation
0.5 0.7 0.9
Gene Loci
Gm49130
Myh7
Cmtm5
Mhrt
Gm17428
Efs
Gm29015
Gm31251
Ngdn
Il25
Myh6
f
Normalized ATAC Signal Coverage
(Range: 0−0.22 by ReadsInTSS)
WT OFT-CMs
KO OFT-CMs
WT V-CMs
KO V-CMs
WT IFT-CMs
KO IFT-CMs
chr6 (kB):
18000
18020
18040
Diff. Peaks
P2G Correlation
0.5 0.7 0.9
Gene Loci
Asz1
Wnt2

Figure 4

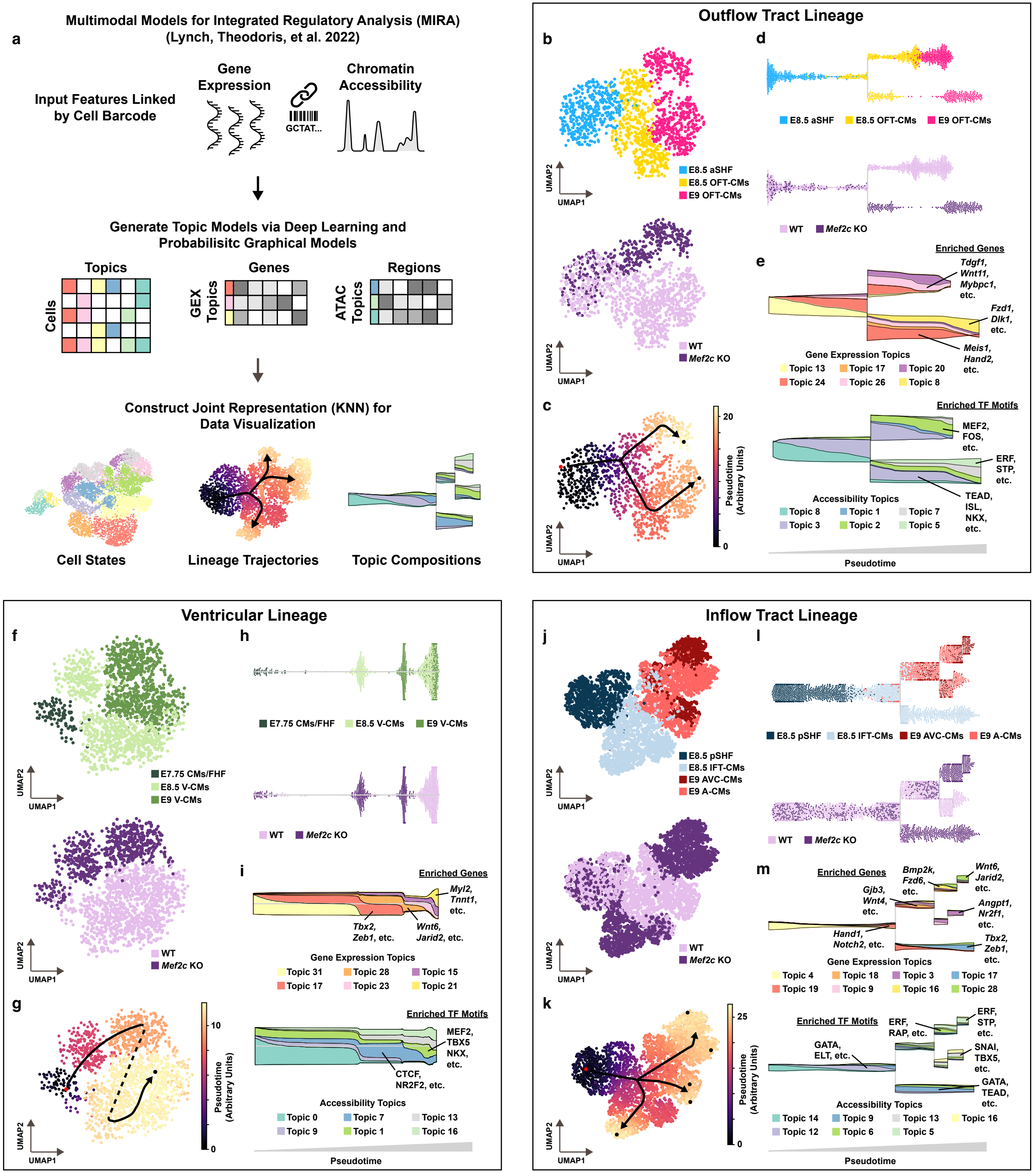

a
Multimodal Models for Integrated Regulatory Analysis (MIRA)
(Lynch, Theodoris, et al. 2022)
Input Features Linked by Cell Barcode
Gene Expression
Chromatin Accessibility
GCTAT...
Generate Topic Models via Deep Learning and Probabilisitc Graphical Models
Topics
Genes
Regions
Cells
GEX Topics
ATAC Topics
Construct Joint Representation (KNN) for Data Visualization
Cell States
Lineage Trajectories
Topic Compositions
Outflow Tract Lineage
b
UMAP2
UMAP1
E8.5 aSHF
E8.5 OFT-CMs
E9 OFT-CMs
WT
Mef2c KO
d
E8.5 aSHF
E8.5 OFT-CMs
E9 OFT-CMs
WT
Mef2c KO
e
Enriched Genes
Tdgf1, Wnt11, Mybpc1, etc.
Fzd1, Dlk1, etc.
Meis1, Hand2, etc.
Gene Expression Topics
Topic 13
Topic 17
Topic 20
Topic 24
Topic 26
Topic 8
c
20
0
Pseudotime (Arbitrary Units)
Enriched TF Motifs
MEF2, FOS, etc.
ERF, STP, etc.
TEAD, ISL, NKX, etc.
Accessibility Topics
Topic 8
Topic 1
Topic 7
Topic 3
Topic 2
Topic 5
Pseudotime
Ventricular Lineage
f
E7.75 CMs/FHF
E8.5 V-CMs
E9 V-CMs
h
E7.75 CMs/FHF
E8.5 V-CMs
E9 V-CMs
WT
Mef2c KO
i
Enriched Genes
Myl2, Tnnt1, etc.
Tbx2, Zeb1, etc.
Wnt6, Jarid2, etc.
Gene Expression Topics
Topic 31
Topic 28
Topic 15
Topic 17
Topic 23
Topic 21
g
10
0
Pseudotime (Arbitrary Units)
Enriched TF Motifs
MEF2, TBX5 NKX, etc.
CTCF, NR2F2, etc.
Accessibility Topics
Topic 0
Topic 7
Topic 13
Topic 9
Topic 1
Topic 16
Pseudotime
Inflow Tract Lineage
j
E8.5 pSHF
E8.5 IFT-CMs
E9 AVC-CMs
E9 A-CMs
l
E8.5 pSHF
E8.5 IFT-CMs
E9 AVC-CMs
E9 A-CMs
WT
Mef2c KO
m
Enriched Genes
Bmp2k, Fzd6, etc.
Wnt6, Jarid2, etc.
Gjb3, Wnt4, etc.
Angpt1, Nr2f1, etc.
Hand1, Notch2, etc.
Tbx2, Zeb1, etc.
Gene Expression Topics
Topic 4
Topic 18
Topic 3
Topic 17
Topic 19
Topic 9
Topic 16
Topic 28
k
25
0
Pseudotime (Arbitrary Units)
Enriched TF Motifs
ERF, RAP, etc.
ERF, STP, etc.
GATA, ELT, etc.
SNAI, TBX5, etc.
GATA, TEAD, etc.
Accessibility Topics
Topic 14
Topic 9
Topic 13
Topic 16
Topic 12
Topic 6
Topic 5
Pseudotime

Figure 5

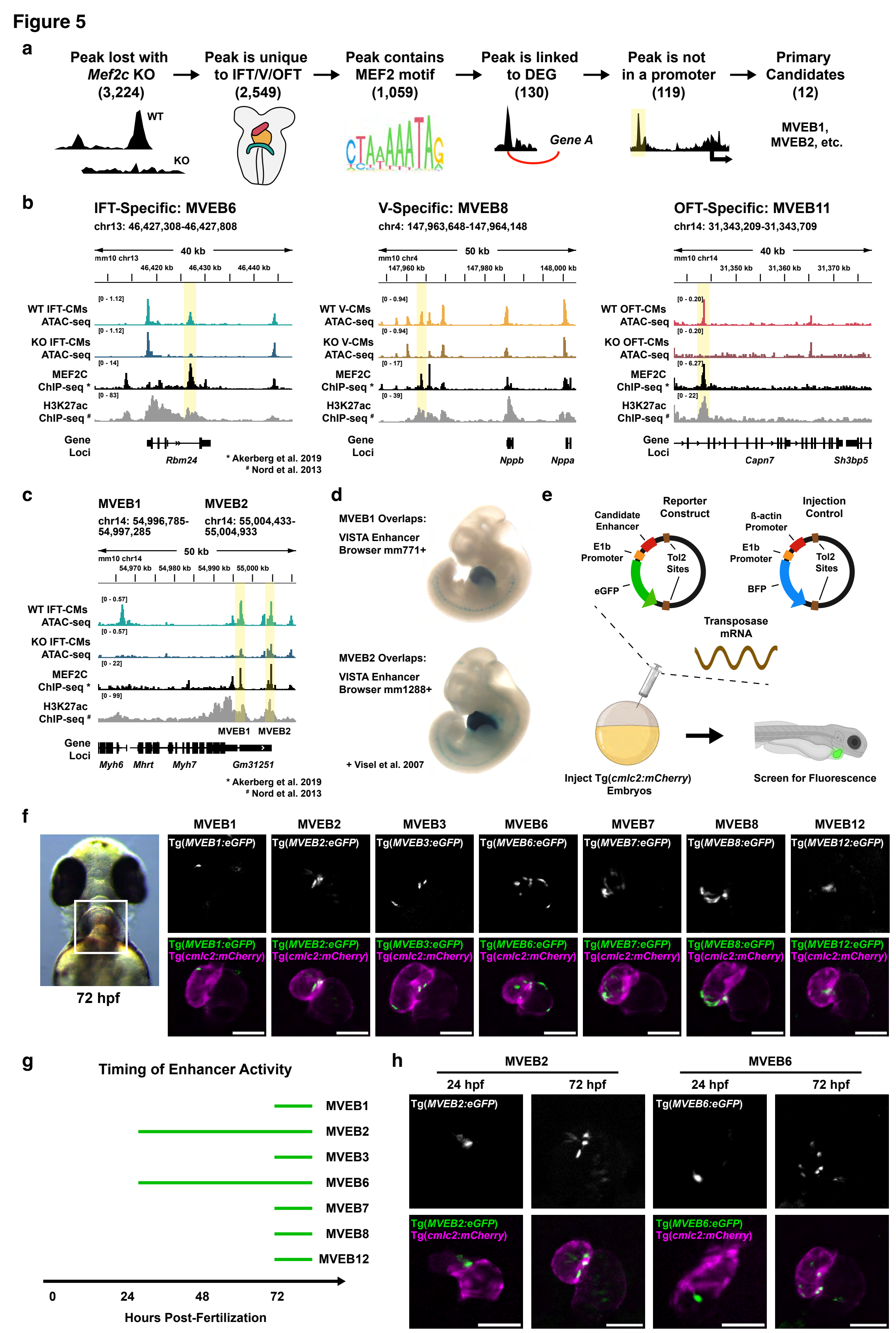
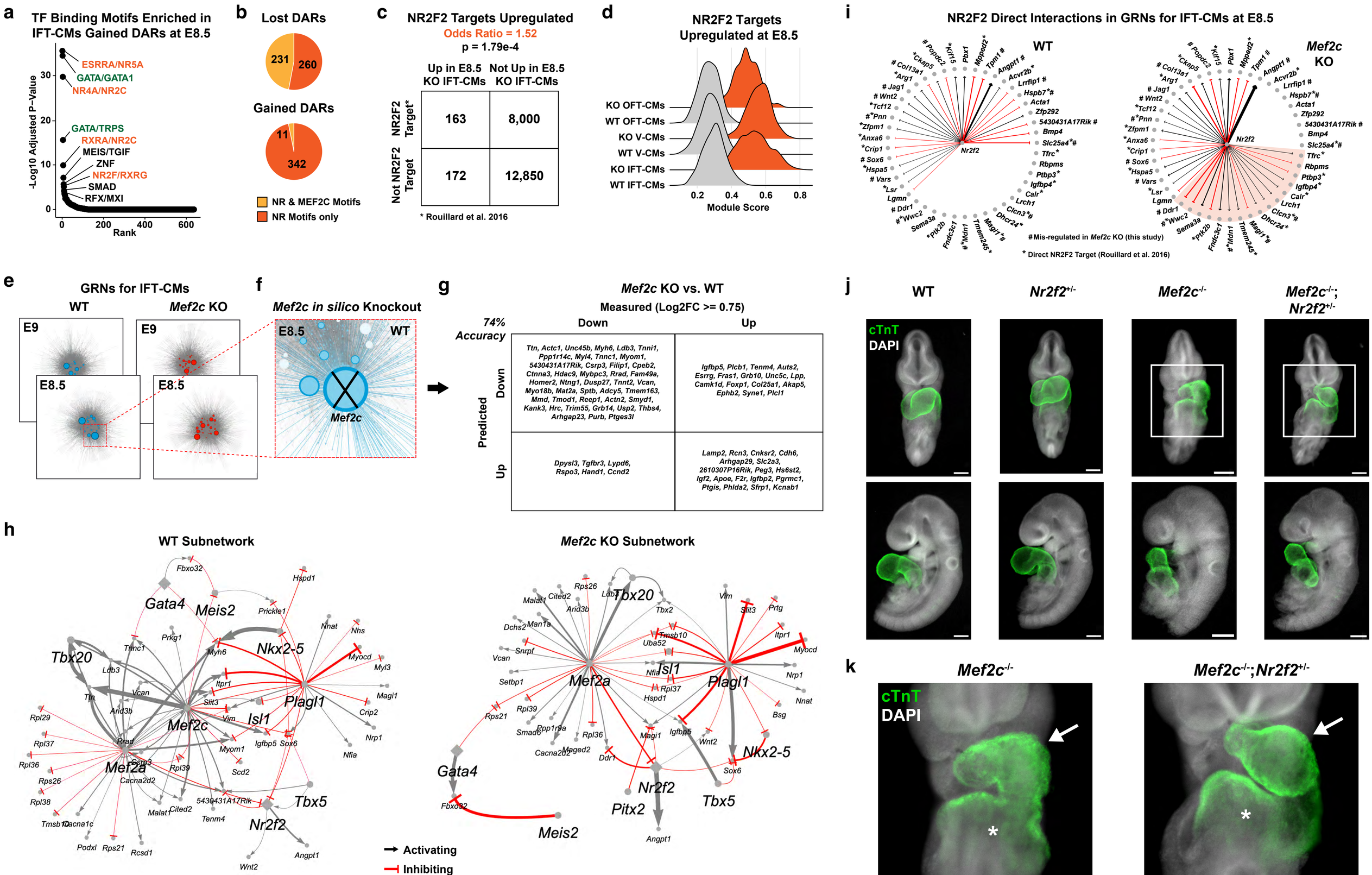

a
Peak lost with Mef2c KO (3,224)
Peak is unique to IFT/V/OFT (2,549)
Peak contains MEF2 motif (1,059)
Peak is linked to DEG (130)
Peak is not in a promoter (119)
Primary Candidates (12)
WT
KO
Gene A
MVEB1, MVEB2, etc.
b
IFT-Specific: MVEB6
chr13: 46,427,308-46,427,808
V-Specific: MVEB8
chr4: 147,963,648-147,964,148
OFT-Specific: MVEB11
chr14: 31,343,209-31,343,709
WT IFT-CMs ATAC-seq
KO IFT-CMs ATAC-seq
MEF2C ChIP-seq *
H3K27ac ChIP-seq #
WT V-CMs ATAC-seq
KO V-CMs ATAC-seq
WT OFT-CMs ATAC-seq
KO OFT-CMs ATAC-seq
Gene Loci
Rbm24
Nppb
Nppa
Capn7
Sh3bp5
* Akerberg et al. 2019
# Nord et al. 2013
c
MVEB1
chr14: 54,996,785-54,997,285
MVEB2
chr14: 55,004,433-55,004,933
50 kb
mm10 chr14
WT IFT-CMs ATAC-seq
KO IFT-CMs ATAC-seq
MEF2C ChIP-seq *
H3K27ac ChIP-seq #
Gene Loci
Myh6
Mhrt
Myh7
Gm31251
* Akerberg et al. 2019
# Nord et al. 2013
d
MVEB1 Overlaps:
VISTA Enhancer Browser mm771+
MVEB2 Overlaps:
VISTA Enhancer Browser mm1288+
+ Visel et al. 2007
e
Reporter Construct
Candidate Enhancer
E1b Promoter
eGFP
Tol2 Sites
Injection Control
ß-actin Promoter
BFP
Transposase mRNA
Inject Tg(cmlc2:mCherry) Embryos
Screen for Fluorescence
f
MVEB1
MVEB2
MVEB3
MVEB6
MVEB7
MVEB8
MVEB12
Tg(MVEB1:eGFP)
Tg(MVEB2:eGFP)
Tg(MVEB3:eGFP)
Tg(MVEB6:eGFP)
Tg(MVEB7:eGFP)
Tg(MVEB8:eGFP)
Tg(MVEB12:eGFP)
Tg(cmlc2:mCherry)
72 hpf
g
Timing of Enhancer Activity
MVEB1
MVEB2
MVEB3
MVEB6
MVEB7
MVEB8
MVEB12
0
24
48
72
Hours Post-Fertilization
h
MVEB2
MVEB6
24 hpf
72 hpf
Tg(MVEB2:eGFP)
Tg(MVEB6:eGFP)
Tg(cmlc2:mCherry)

Figure 6


a
TF Binding Motifs Enriched in IFT-CMs Gained DARs at E8.5
ESRRA/NR5A
GATA/GATA1
NR4A/NR2C
GATA/TRPS
RXRA/NR2C
MEIS/TGIF
ZNF
NR2F/RXRG
SMAD
RFX/MXI
-Log10 Adjusted P-Value
Rank
b
Lost DARs
231
260
Gained DARs
11
342
NR & MEF2C Motifs
NR Motifs only
c
NR2F2 Targets Upregulated
Odds Ratio = 1.52
p = 1.79e-4
Up in E8.5 KO IFT-CMs
Not Up in E8.5 KO IFT-CMs
NR2F2 Target*
Not NR2F2 Target
163
8,000
172
12,850
* Rouillard et al. 2016
d
NR2F2 Targets Upregulated at E8.5
KO OFT-CMs
WT OFT-CMs
KO V-CMs
WT V-CMs
KO IFT-CMs
WT IFT-CMs
Module Score
i
NR2F2 Direct Interactions in GRNs for IFT-CMs at E8.5
WT
Mef2c KO
Nr2f2
Mis-regulated in Mef2c KO (this study)
* Direct NR2F2 Target (Rouillard et al. 2016)
e
GRNs for IFT-CMs
WT
Mef2c KO
E9
E8.5
f
Mef2c in silico Knockout
E8.5
WT
Mef2c
g
Mef2c KO vs. WT
Measured (Log2FC >= 0.75)
74% Accuracy
Down
Up
Predicted
Down
Up
Ttn, Actc1, Unc45b, Myh6, Ldb3, Tnni1, Ppp1r14c, Myl4, Tnnc1, Myom1, 5430431A17Rik, Csrp3, Filip1, Cpeb2, Ctnna3, Hdac9, Mybpc3, Rrad, Fam49a, Homer2, Ntng1, Dusp27, Tnnt2, Vcan, Myo18b, Mat2a, Sptb, Adcy5, Tmem163, Mmd, Tmod1, Reep1, Actn2, Smyd1, Kank3, Hrc, Trim55, Grb14, Usp2, Thbs4, Arhgap23, Purb, Ptges3l
Igfbp5, Plcb1, Tenm4, Auts2, Esrrg, Fras1, Grb10, Unc5c, Lpp, Camk1d, Foxp1, Col25a1, Akap5, Ephb2, Syne1, Plcl1
Dpysl3, Tgfbr3, Lypd6, Rspo3, Hand1, Ccnd2
Lamp2, Rcn3, Cnksr2, Cdh6, Arhgap29, Slc2a3, 2610307P16Rik, Peg3, Hs6st2, Igf2, Apoe, F2r, Igfbp2, Pgrmc1, Ptgis, Phlda2, Sfrp1, Kcnab1
j
WT
Nr2f2+/-
Mef2c-/-
Mef2c-/-; Nr2f2+/-
cTnT
DAPI
h
WT Subnetwork
Mef2c KO Subnetwork
Gata4
Meis2
Tbx20
Nkx2-5
Plagl1
Isl1
Mef2c
Mef2a
Tbx5
Nr2f2
Pitx2
Activating
Inhibiting
k
Mef2c-/-
Mef2c-/-;Nr2f2+/-
cTnT
DAPI