Heart tube morphogenesis is regulated by segment-specific gene regulatory networks controlled by MEF2C

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,*]

[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 transcription factor MEF2C plays a critical role in the development of the linear heart tube, but the specific transcriptional networks controlled by MEF2C remain largely undefined. To address this, we performed combined single-nucleus RNA- and ATAC-sequencing on wild type and MEF2C-null embryos at distinct stages of development. We identified a broadly "posteriorized" cardiac gene signature and chromatin landscape throughout the heart tube in the absence of MEF2C. By integrating our gene expression and chromatin accessibility data in a deep-learning based model, we were able to construct developmental trajectories for each of the outflow tract, ventricular, and inflow tract lineages and determined how each of these segment-specific trajectories were distinctly altered in the MEF2C-null embryos. We computationally identified potential segment-specific MEF2C-dependent enhancers, and from these candidates, identified novel enhancers with activity in the developing heart tube using transgenesis in zebrafish. Finally, using inferred gene regulatory networks we discovered a genetic interaction between *Mef2c* and the atrial nuclear hormone receptor *Nr2f2*, revealing that the MEF2C-null heart malformations are partly driven by a transcriptional network with increased NR2F2 activity. These studies not only provide a rich description of the genomic regulation of early heart tube development, but provide a generalizable framework for using genetic mutants to dissect the transcriptional networks that govern developmental processes.

INTRODUCTION

Transcription factors (TFs) regulate gene expression to direct cell fate decisions and control developmental processes. Over the course of embryonic development, TFs are typically expressed in specific spatiotemporal patterns that dictate the cell and tissue types that they regulate[1–3]. The activity of a given 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[4–7]. 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[8]. These GRNs can be particularly useful for untangling the complexities of gene regulation and answering open questions in the field of developmental biology, such as a how a single TF may have different activity within its expression domain.

The developing linear heart tube presents an excellent model for exploring how a single TF may play distinct roles within different regions of a developing tissue. The TFs that are critical for cardiac development have been clearly identified and their expression patterns are well-defined[1,9–11]. Moreover, lineage tracing studies and subsequent molecular analyses have revealed the distinct progenitor cell types that contribute to the various segments of the linear heart tube (Fig. 1A). Cardiac progenitors originate in the mesoderm that arises during gastrulation[12–14]. By E7.75 in the mouse, this cardiogenic mesoderm forms a morphologically distinct ridge of cells underneath the early head folds termed the cardiac crescent[15]. The cells that make up the protruding rostral aspect of the crescent arrive to this position first and are termed the first heart field (FHF), which later gives rise to the left ventricle and atria[1,10,11,16,17]. Cells located more posterior and medial to the FHF in the cardiac crescent are termed the second heart field (SHF) and primarily produce the outflow tract and right ventricle, along with contributing to the atria[10,11,16–18].

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. In mice, loss of MEF2C leads to cardiac defects that are embryonic lethal by mid-gestation at embryonic day E10.5[19]. Interestingly, although MEF2C is expressed throughout the

entire cardiac crescent (FHF and SHF) and developing heart tube from E7.75 to E9 (Fig. 1B, Fig. S1A), the loss of MEF2C causes distinct defects in different segments of the heart tube. At E9, the outflow tract is severely hypoplastic and only a single hypoplastic ventricle has formed, but by contrast, the inflow tract 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.

In order to dissect the segment-specific regulation of heart tube morphogenesis by MEF2C, we used the 10x Multiome platform to perform combined single-nucleus RNA-seq and ATAC-seq (snRNA-seq and snATAT-seq) on wild-type (WT) and MEF2C knockout (KO) mouse embryos at key stages throughout heart tube development. Specifically, we collected two WT and two MEF2C KO stage-matched embryos at E7.75, E8.5, and E9 (Fig. S1B). All embryos carried the *Smarcd3*-F6-eGFP transgene reporter to mark cardiac progenitors[20]. At E7.75, we harvested the entire embryos, whereas, at E8.5 and E9 we removed the headfolds and posterior trunk via microdissection in order to increase the proportion of cardiac progenitors captured in the 10x Multiome assay. By 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 molecular basis of the expanded inflow tract phenotype in MEF2C KO embryos (Fig. 1D).

RESULTS

Loss of MEF2C leads to broad changes in gene expression throughout the heart tube, including mis-regulation of key cardiomyocyte genes and anterior/posterior markers

To begin examining how MEF2C regulates the morphogenesis of specific heart tube segments, we used Seurat[21] to determine the changes in gene expression that result from MEF2C KO. We began by creating separate Seurat objects for the embryos profiled at each timepoint (E7.75, E8.5, and E9). Following pre-processing, quality control,

82 normalization with scTransform[22], and dimensionality reduction, we performed unbiased
83 Louvain clustering and plotted the data in uniform manifold approximation and projection
84 (UMAP) space. We then used the expression of marker genes to identify the cell types
85 represented by each cluster (Fig. S2A-F, Table S1A-C). In order to resolve specific
86 cardiomyocyte subtypes, we subset and re-clustered cell types of interest, including
87 cardiac progenitors, cardiomyocytes, and related mesoderm cell types (Fig. 2A-C, Fig.
88 S2G-I, Table S2A-C). By doing so, we were able to identify clusters representing FHF and
89 early differentiating cardiomyocytes (CMs/FHF), SHF, and juxtacardiac field[23] (JCF)
90 progenitors at E7.75 (Fig. 2A). Similarly, at E8.5 we could distinguish inflow tract
91 cardiomyocytes (IFT-CMs), ventricular cardiomyocytes (V-CMs), and outflow tract
92 cardiomyocytes (OFT-CMs) (Fig. 2B). At E9 we could more specifically identify the IFT-
93 CMs as either atrial or atrioventricular canal (A- or AVC-CMs) (Fig. 2C). Having identified
94 our cell types of interest, we next performed differential gene expression analysis
95 between MEF2C KO and WT cells within each of these cell types (Fig. 2D-F, Table S3A-
96 J). Generally speaking, we found larger numbers of genes down-regulated than up-
97 regulated with loss of MEF2C, agreeing with the notion of MEF2C as an activator of the
98 myogenic transcriptional program[24]. Among the top differentially expressed genes
99 (DEGs) within each cell type of interest, we found a number of genes that have been
100 associated with high confidence to congenital heart defects[25], including *Myh6*, *Myh7*,
101 *Actc1*, and *Foxp1* (Fig. S2J). Furthermore, nearly all of these top DEGs were identified
102 as the nearest gene to peaks identified in a previously published MEF2C ChIP-seq
103 dataset of embryonic hearts at E14[26], demonstrating that these genes are indeed likely
104 to be direct targets of MEF2C regulation in the developing heart.

105 By more closely examining the results of our differential gene expression analysis
106 (Table S3A-J), two categories of DEGs emerged. First, we observed substantial down-
107 regulation of key cardiomyocyte genes across all three heart tube segments. This
108 included cardiac transcription factors such as *Nkx2.5*, *Tbx5*, and *Gata4*, as well as
109 sarcomeric proteins including *Tnnt2* and *Ttn* (Fig. 2G-I). This result further emphasizes
110 the importance of MEF2C in early cardiomyocyte fate specification, both by directly
111 driving expression of contractile proteins and by promoting additional transcription factors
112 involved in regulating the cardiomyocyte transcriptional program. The second category of

113 DEGs we identified were genes involved in anterior/posterior patterning of the heart tube.
114 Specifically, genes such as *Tbx5*, *Gata4*, and *Wnt2* that are normally expressed only in
115 the posterior inflow tract segment of the heart tube were not only up-regulated in the IFT-
116 CMs, but their expression also expanded into the V-CMs in MEF2C KO (Fig. 2G-I).
117 Additionally, expression of the anterior outflow tract-specific gene *Tdgf1* was completely
118 lost in MEF2C KO (Fig. 2H-I), as previously demonstrated[27]. Finally, we validated these
119 changes in gene expression with whole-mount embryo fluorescence *in situ* hybridization,
120 confirming that MEF2C KO leads to broad reduction in the cardiomyocyte transcriptional
121 program and leads to a notable "posteriorization" of the developing heart tube (Fig. 2J).
122

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

125 In order to better understand how MEF2C regulates the observed segment-specific gene
126 expression changes, we next used ArchR[28] to integrate both the gene expression and
127 chromatin accessibility data. We performed dimensionality reduction using ArchR's latent
128 semantic index (LSI) implementation on the snRNA-seq and snATAC-seq modalities, and
129 then plotted the cells in UMAP space using the combined dimensions (Fig. S3A-C).
130 Because individual cells have matching barcodes in the 10x Multiome snRNA-seq and
131 snATAC-seq libraries, we simply transferred the cell type labels from our Seurat objects
132 to our ArchR datasets. As we did with the gene expression data, we subset the cardiac
133 cell types and associated mesoderm in order to better resolve cardiomyocyte subtypes.
134 Finally, we created pseudobulk groups consisting of WT and MEF2C KO cells for each
135 given cell type label in these subset datasets and called peaks in the accessibility data
136 using MACS2[29].

137 Our initial analyses revealed few changes to chromatin accessibility at E7.75 (Fig.
138 S3D). This was not entirely unexpected, given that there is no discernable phenotype in
139 the MEF2C KO embryos at this timepoint, and agrees with the relatively few changes in
140 gene expression we observed in CMs/FHF, SHF, and JCF cells at E7.75 (Fig. 2D). Thus,
141 we chose to focus on the E8.5 dataset in order to identify the earliest substantial changes
142 in chromatin accessibility that lead to the MEF2C KO phenotype (Fig. 3A). Using our
143 pseudobulked data, we identified differentially accessible regions (DARs) of chromatin

144 between MEF2C KO and WT cells in each of the subset cell types and found that, as
145 expected, the most changes occurred in the cardiomyocytes (Fig. 3B). We also found that
146 in each of the IFT-CMs, V-CMs, and OFT-CMs, there were more lost DARs than gained
147 DARs in MEF2C KO (Fig. 3B), and that these lost DARs were highly enriched for MEF2C
148 binding motifs (Fig. S3E). These data further reinforce MEF2C's role as an activator of
149 the cardiac transcriptional program. Interestingly, when we examined the lost DARs in
150 MEF2C KO embryos, we found that there were 675 regions that demonstrated lost
151 accessibility in two or more heart tube segments, and 2,549 regions that lost accessibility
152 in a segment-specific manner (Fig. 3C). This likely reflects two sets of DARs with different
153 biological functions: i) regions that are involved in the regulation of general cardiomyocyte
154 genes, and ii) regions that are involved in segment-specific gene regulation. By contrast,
155 the gained DARs in MEF2C KO were almost exclusively unique to each heart tube
156 segment (Fig. 3C). This is interesting because it provides evidence that distinct
157 transcriptional networks exist for each of the three heart tube segments, and thus, each
158 network is uniquely affected by loss of MEF2C.

159 We performed peak-to-gene linkage (P2G) analysis in order to determine which of
160 the identified differentially accessible regions were most likely to be *bona fide* regulatory
161 elements. Peak-to-gene linkage identifies statistical correlation between the accessibility
162 of a given peak and the expression of potential target genes within 250 kilobases of the
163 peak. This revealed 817 P2G links in the OFT-CMs, 1,201 in the V-CMs, and 2,257 in the
164 IFT-CMs. To further refine these results, we intersected the P2G links with our differential
165 gene expression analysis in Seurat to identify which P2G links are connected to genes
166 that exhibit altered expression in MEF2C KO (Table S4A-C). We then visualized this data
167 by plotting each P2G link according to the Log2 fold change values from the differential
168 peak accessibility and differential gene expression tests, and colored each point by the
169 correlation strength of the P2G link (Fig. 3D).

170 These DAR-P2G correlation plots reveal a number of interesting observations.
171 Genes such as *Nkx2.5*, *Myh6*, and *Myh7*, which have decreased expression throughout
172 the heart tube in MEF2C KO, are also linked to the accessibility of DARs in all three
173 segments (Fig. 3D). In particular, the *Myh6*/*Myh7* locus contains a number of regions that
174 have P2G links and clearly altered accessibility in OFT-CMs, V-CMs, and IFT-CMs (Fig.

175 3E). This suggests that these DARs are likely to be regulatory elements that drive
176 expression of these genes and are sensitive to the expression of MEF2C in the entire
177 heart tube. By contrast, there are also P2G links between peaks with lost accessibility
178 and decreased gene expression that are unique to the individual segments, such as links
179 to *Wnt11* in the OFT-CMs, *Nppa* in the V-CMs, and *Cacna1c* in the IFT-CMs (Fig. 3D).
180 Furthermore, examining the links between peaks with gained accessibility and genes with
181 increased expression in MEF2C KO reveals additional evidence for the increased
182 "posteriorized" transcriptional program we previously described. Namely, P2G links to
183 *Tbx5* in the V-CMs, as well as *Gata4* and *Wnt2* in the IFT-CMs (Fig. 3D and 3F). These
184 P2G links represent potential regulatory elements that are activated or over-activated in
185 response to the loss of MEF2C and, at least in part, drive the increased posteriorization
186 of the heart tube.
187
188 **Pseudotime analysis with MIRA reveals distinct MEF2C-depedent developmental**
189 **trajectories for each heart tube segment**
190 In order to fully take advantage of our multiomic dataset containing precisely staged
191 embryos at consecutive embryonic timepoints, we used multimodal Models for Integrated
192 Regulatory Analysis[30] (MIRA) to examine the altered temporal dynamics of gene
193 expression and chromatin accessibility in the developing heart tubes of WT and MEF2C
194 KO embryos. MIRA's probabilistic topic modeling borrows concepts that were developed
195 to identify abstract topics within natural language that give meaning to a collection of text,
196 and applies them to snRNA-seq and snATAC-seq. In this case, MIRA uses deep learning
197 and probabilistic graphical models to construct sets of "topics" that represent either genes
198 or accessibility peaks, and then uses these topics to define cell identities. Finally, MIRA
199 uses these topics to construct a joint representation that can be used to plot cells in UMAP
200 space, perform pseudotime analysis, and construct cell state trees (Fig. 4A).
201 We were particularly interested in using MIRA's ability to construct cell state trees
202 to understand how the lineage trajectories of OFT-CMs, V-CMs, and IFT-CMs were
203 altered due to the loss of MEF2C. We trained topic models using cardiac progenitors and
204 cardiomyocyte subtypes from both genotypes at each of our three timepoints (E7.75:
205 CMs/FHF, SHF, JCF; E8.5: aSHF, pSHF, OFT-CMs, V-CMs, IFT-CMs; E9: OFT-CMs, V-

206 CMs, A-CMs, AVC-CMs) and then used those models to construct joint representation
207 UMAPs of the outflow tract, ventricular, and inflow tract lineages (Fig. 4B, 4F, 4J). We
208 next calculated pseudotime values for cells within the UMAPs and used the results to
209 construct cell state trajectories for each of the three lineages (Fig. 4C-D, 4G-H, 4K-L).

210 By labeling the cells within these lineage trajectories by both cell type and
211 genotype, we can observe clear differences in the fate specification of MEF2C KO cells,
212 compared to WT cells, as represented by branch points in the trajectories (Fig. 4D, 4H,
213 4L). Most interestingly, we notice that the structure of these trajectories, along with the
214 manner and timing by which MEF2C KO cells diverge from WT cells within the
215 trajectories, is different for all three lineages. Specifically, within the outflow tract lineage,
216 MEF2C KO and WT cells proceed similarly up until E8.5, and then completely diverge as
217 development proceeds to E9 (Fig. 4D). Within the ventricular lineage, there is no branch
218 point. Instead, MEF2C KO V-CMs from both E8.5 and E9 are located earlier in the
219 trajectory than even E8.5 WT V-CMs (Fig. 4H). This indicates that the result of knocking
220 out MEF2C in V-CMs resembles a marked delay or termination of normal development
221 more so than a divergence to an alternative cell state. The situation is most complex for
222 the inflow tract lineage, in which there are two points of divergence between MEF2C KO
223 and WT cells (Fig. 4L). First, there is a subset of MEF2C KO cells that branch away from
224 the WT trajectory as they differentiate from pSHF progenitors to IFT-CMs at E8.5. A
225 second set of MEF2C KO cells proceed along the WT trajectory to E9, before ultimately
226 diverging from WT cells at the "terminal" AVC-CM and A-CM states. Together, these data
227 reinforce our working hypothesis that MEF2C regulates development of the heart tube in
228 a segment-specific manner via distinct regulatory networks.

229 We gained further insights regarding the segment-specific regulation of heart tube
230 development by examining the top 200 genes and the top ranked TF binding motifs
231 enriched in the gene expression and chromatin accessibility topics (Table S5). As WT
232 cells diverge from MEF2C KO cells in the outflow tract lineage, they are marked by gene
233 expression topics enriched for canonical outflow tract markers *Tdgf1* and *Wnt11*[27,31], and
234 their accessibility topics are marked by peaks enriched for MEF2 and FOS binding motifs
235 (Fig. 4E). Interestingly, the MEF2C KO cells are marked by expression of Notch and Wnt
236 signaling components such as *Dlk1* and *Fzd1*, and transcription factors *Meis1* and *Hand2*

237 (Fig. 4E). The accessibility topic enriched for ISL, NKX, and TEAD is maintained to a
238 larger extent in the outflow tract of MEF2C KO cells (Fig. 4E), suggesting these cells may
239 retain a more progenitor-like state in the absence of MEF2C. The gene expression and
240 accessibility topics enriched along the ventricular lineage trajectory supports our
241 interpretation that the MEF2C KO cells experience a substantial delay or termination of
242 their normal differentiation. Notably, topics that are enriched for CM genes (e.g. *Myl2*,
243 *Tnnt1*) and cardiac TF binding motifs (e.g. TBX, MEF2, NKX) only mark the end of the
244 differentiation trajectory containing WT cells (Fig. 4I). In the complex branching inflow
245 tract trajectory, we noted that the terminal WT cells are enriched for topics containing
246 expected atrial genes, such as *Angpt1* and *Nr2f1*, and binding motifs for TBX5 and SNAI,
247 among others (Fig. 4M, Fig. S4A-B). Whereas, the MEF2C KO inflow tract cells stream
248 into one of two terminal branches enriched for genes such as *Wnt6, Jarid2, Tbx2,* and
249 *Zeb1* and binding motifs for ERF, GATA, and TEAD factors (Fig. 4M, Fig. S4A-B).
250 Together, these data illustrate that lineage-specific changes in gene expression and
251 chromatin accessibility occur as heart tube development progresses and that these
252 lineage-specific programs are uniquely altered upon the loss of MEF2C.
253
254 **Leveraging multiomic data to identify novel enhancers via zebrafish transgenesis**
255 We next sought to determine whether the observed lineage-specific patterns of gene
256 regulation by MEF2C were driven by segment-specific enhancers that are active during
257 heart tube development. To identify candidate enhancers, we utilized our E8.5 ArchR
258 dataset containing integrated chromatin accessibility and gene expression data. We
259 began by identifying all peaks that lost accessibility in the MEF2C KO (3,224), then filtered
260 to the peaks that demonstrated altered accessibility in only one of the three heart tube
261 segments (2,549) and contained a MEF2C binding motif (1,059). Next, we again utilized
262 ArchR's Peak2Gene functionality and filtered down to peaks correlated with the
263 expression of genes that were mis-regulated in the MEF2C KOs (130) and were located
264 in promoter regions (119). Finally, from this list of 119 candidates, we selected 12 of the
265 highest priority to screen based on the Log2FC value of the change in peak accessibility
266 and linkage to genes of interest (Fig. 5A, Table S6). We enumerated these candidate
267 regions with the prefix "MVEB," after the initials of authors Muncie-Vasic and Brower. To

268 further validate our selection of candidates, we overlayed our chromatin accessibility data
269 at these regions with published ChIP-seq datasets[26,32] to visualize the occupancy of
270 MEF2C and H3K27ac, a histone post-translational modification that marks active
271 enhancers. Importantly, each of our candidates had clear and strong occupancy of both
272 MEF2C and H3K27ac (Fig. 5B). Of note, one of our ventricular-specific candidates,
273 MVEB8, is located within the *Nppa*/*Nppb* super enhancer cluster[33]. Furthermore, two of
274 the candidates we identified form differential accessibility in IFT-CMs overlapped with
275 regions found in the VISTA enhancer database[34], with demonstrated enhancer activity
276 in the hearts of E11.5 mouse embryos (Fig. 5C-D).

277 In order to determine whether our candidate regions are *bona fide* enhancers with
278 activity in early heart tube development, we employed the zebrafish Tol2 Transgenesis
279 system[35,36], allowing us to screen candidates with relatively high-throughput. We
280 reasoned that because *Mef2c* has the orthologs *mef2ca* and *mef2cb* in zebrafish, and
281 because there is a high degree of overlap between the key cardiac transcriptional
282 regulators in zebrafish and mouse[37], there should be a fairly high concordance between
283 the activity of our candidate regions identified in the mouse and their activity in the fish.
284 Moreover, it has been previously demonstrated that regulatory elements can exhibit
285 activity in the heart of both species[38–40]. To perform our screen, we cloned our candidate
286 enhancer regions into vector backbones just upstream of a minimal E1b promoter and
287 the coding sequence for eGFP, flanked by Tol2 recognition sites. As a negative control,
288 we also tested an empty reporter construct consisting of only the minimal E1b promoter
289 and eGFP. To control for injection efficacy, we utilized a vector consisting of the ß-actin
290 promoter upstream of the minimal E1b promoter and BFP, flanked by Tol2 sites. These
291 vectors were injected along with transposase mRNA into zebrafish embryos from
292 Tg(*cmlc2:mCherry*)^{s890} and WT parents, allowing us to screen F0 embryos for enhancer
293 activity (eGFP) within the heart (mCherry) (Fig. 5E). To determine whether a candidate
294 enhancer demonstrated activity, we quantified the percentage of BFP-positive embryos
295 that exhibited eGFP expression in the heart. We were unable to successfully clone one
296 of our twelve candidates (MVEB10) due to a highly repetitive “GT” sequence, so our final
297 screen consisted of eleven candidate enhancers.

298 Upon screening, we detected enhancer activity in the developing heart for seven
299 of the eleven candidates (Fig. 5F, Fig. S5A-B). Both candidates that overlapped with
300 regions found in the VISTA enhancer database (MVEB1, MVEB2) exhibited activity in the
301 heart by 72 hours post-fertilization (hpf) (Fig. 5F). Two additional candidates we identified
302 from differential accessibility in the IFT-CMs, MVEB3 and MVEB6, demonstrated robust
303 and specific activity in the zebrafish heart at 72 hpf (Fig. 5F, Fig. S5A-B). Finally, we also
304 observed activity from two candidates that we identified from differential accessibility in
305 the V-CMs, MVEB7 and MVEB8, and one candidate identified from differential
306 accessibility in the OFT-CMs, MVEB12 (Fig. 5F, Fig. S5A-B). To our knowledge, these
307 five candidates (MVEB3, MVEB6, MVEB7, MVEB8, MVEB12) have not been previously
308 reported and represent novel, MEF2C-depenedent enhancers with activity in the
309 developing heart. Interestingly, we also observed a range in the timing of enhancer activity
310 in the embryonic zebrafish heart. Candidates MVEB2 and MVEB6 were clearly detectable
311 as early as 24 hpf, while the others were not active until 72 hpf (Fig. 5G-H). This variation
312 in the onset of enhancer activity could result from differences in the timing of accessibility
313 of these regulatory elements, different levels of sensitivity to *Mef2c* expression, or
314 cooperation with other TFs or transcriptional regulators.
315

316 **An overactive "posteriorized" gene regulatory network driven by NR2F2 and**
317 **GATA4 underlies the expanded inflow tract phenotype in MEF2C KO**

318 The posteriorized gene signature (Fig. 2H-J) and chromatin accessibility landscape (Fig.
319 3F, 3H) that occurs in the heart tube of MEF2C KO embryos, together with the multi-
320 staged branching structure of the inflow tract lineage trajectory (Fig. 4L), reveals a
321 previously unappreciated complexity of the transcriptional network governing inflow tract
322 development that we sought to further dissect. Moreover, the expanded but aberrantly
323 formed inflow tract that forms in MEF2C KO embryos (Fig. 1C) suggests that loss of
324 MEF2C leads to a more complex heart tube phenotype than a simple failure of cardiac
325 differentiation and growth. To determine what transcriptional regulators may drive the
326 aberrant phenotype in the posterior of the heart tube, we performed TF binding motif
327 enrichment analysis on the regions of chromatin that gain accessibility in the MEF2C KO
328 IFT-CMs at E8.5 and found that Nuclear Receptor (NR) and GATA motifs are among the

329 most enriched (Fig. 6A). From this observation, we hypothesized that the loss of MEF2C
330 in the posterior heart tube causes a redistribution of NR and GATA factors to regions of
331 chromatin where they would normally not have activity. More specifically, we hypothesize
332 that MEF2C co-binds certain regulatory elements along with NR or GATA factors in the
333 WT condition, and that these NR and GATA factors become "free" to bind different
334 regulatory elements in the MEF2C KO. If this were true, one would expect that sites of
335 accessible chromatin that mark WT IFT-CMs and contain an NR or GATA motif would be
336 likely to also contain a MEF2C motif, whereas the sites that gain accessibility in MEF2C
337 KO and contain an NR or GATA motif would be more likely to not also contain a MEF2C
338 motif. Indeed, this is exactly what we observed (Fig. 6B).

339 These data are particularly interesting given the importance of GATA factors, and
340 in particular GATA4, in cardiac development[41] and the knowledge that nuclear receptor
341 subfamily 2 group F member 2 (NR2F2, also known as COUP-TFII) is a necessary driver
342 of atrial development[42,43]. Using the web tool Galaxy[44] and publicly available ChIP-seq
343 datasets[42,45], we found NR2F2 and GATA4 occupancy at a number of the regions of
344 chromatin that gained accessibility in the MEF2C KO IFT-CMs at E8.5 (Fig. S6A-B),
345 suggesting that these DARs could indeed be sites of potential NR2F2 and GATA4
346 regulation. To determine whether NR2F2 and GATA4 may be redistributed and driving the
347 "posteriorization" of the heart tube (Fig. 2H-I), we next asked whether NR2F2 and GATA4
348 target genes were upregulated in the inflow tract of MEF2C KO embryos. Using publicly
349 available datasets to identify the target genes of each TF[46,47], we calculated the odds
350 ratio (OR) of finding NR2F2 or GATA4 target genes compared to non-target genes
351 amongst our upregulated differentially expressed genes at E8.5. We found strong
352 evidence that NR2F2 (OR = 1.52, $p = 1.79e-4$) and GATA4 (OR = 4.23, $p < 2.2e-16$)
353 targets were enriched amongst the upregulated genes (Fig. 6C). Next, we took the list of
354 NR2F2 and GATA4 target genes that were upregulated (Table S7A-B) and calculated a
355 "module score" for the expression of these genes in our dataset[48]. In agreement with our
356 working model that these NR2F2 and GATA4 targets are driving the "posteriorized" heart
357 tube phenotype in MEF2C KO, we found the NR2F2 and GATA4 module scores were
358 increased in all segments of the heart tube at E8.5 (Fig. 6D).

359 To further examine the altered roles of NR2F2 and GATA4 in the transcriptional
360 networks governing heart tube development, we used CellOracle[49] to construct gene
361 regulatory networks for each segment of the heart tube at each timepoint in our dataset
362 (Fig. 6E). In brief, CellOracle uses integrated gene expression and chromatin accessibility
363 data to infer interactions between TFs and target genes based on the accessibility of a
364 given TF's binding motif near a potential target gene and the correlation of expression
365 between the TF and the potential target. To validate that these inferred CellOracle
366 networks effectively captured the true gene regulatory networks that govern inflow tract
367 development, we deleted *Mef2c in silico* in the E8.5 inflow tract network and compared
368 the predicted changes in gene expression to the measured changes in gene expression
369 in our E9 dataset (Fig. 6F). We found that the networks correctly predicted the gene
370 expression changes with 92% accuracy (Fig. 6G). Having validated the accuracy of the
371 CellOracle networks, we next visualized the core cardiac subnetworks of altered genes
372 in the WT and MEF2C KO inflow tracts (Fig. 6H). These subnetworks consist of key
373 cardiac transcription factors and their inferred interactions with the top 100 most
374 differentially regulated genes in the E8.5 MEF2C KO IFT-CMs. Next, we further narrowed
375 in on the direct interactions between NR2F2 and GATA4 and their target genes in the WT
376 and MEF2C KO inflow tract networks (Fig. 6I-J). From these plots we observe similar
377 behavior for both TFs – approximately half of the WT interactions are maintained in the
378 MEF2C KO, while nearly twice as many new direct interactions are established. This is
379 consistent with our working model that in response to MEF2C KO, aberrant NR2F2 and
380 GATA4 activity regulate additional genes that drive the inflow tract phenotype in the
381 posterior of the heart tube.
382 Given this hypothesis, we theorized that reducing the activity of NR2F2 may rescue
383 the inflow tract phenotype in the MEF2C KO heart tube. To test this, we generated
384 *Mef2c*$^{del/+}$;*Nr2f2*$^{del/+}$ male mice and crossed them to *Mef2c*$^{del/+}$ females. Although there is
385 no apparent heart tube phenotype at E9.5 for *Nr2f2*$^{del/+}$ (Fig. 6K), it was previously shown
386 that these animals are recovered at lower numbers than expected Mendelian ratios[43],
387 suggesting that *Nr2f2* is a dosage-sensitive gene and that *Nr2f2* heterozygosity can be
388 expected to alter transcriptional activity. Strikingly, we observed a notable partial rescue
389 in ventricle and inflow tract development at E9.5 (18-24 somites) in the

390 *Mef2c*$^{del/del}$;*Nr2f2*$^{del/+}$ embryos compared to *Mef2c*$^{del/del}$ (Fig. 6K-L), although the extent of
391 the rescue did vary (Fig. S6C-D). The MEF2C KO single-ventricle phenotype was
392 maintained in *Mef2c*$^{del/del}$;*Nr2f2*$^{del/+}$ embryos, but these ventricles were larger and more
393 expanded than Mef2c$^{del/del}$. Moreover, the posterior structures of the heart tube in
394 *Mef2c*$^{del/del}$;*Nr2f2*$^{del/+}$ more closely resembled WT inflow tracts with expanded prospective
395 atria, compared to the completely disrupted inflow tract in *Mef2c*$^{del/del}$ embryos (Fig. 6K-
396 L). This partial rescue of the ventricle and inflow tract in *Mef2c*$^{del/del}$;*Nr2f2*$^{del/+}$ embryos
397 provides clear evidence that the MEF2C KO phenotype in the posterior heart tube is
398 driven, at least in part, by an altered gene regulatory network with increased and aberrant
399 NR2F2 activity.
400
401 **DISCUSSION**
402 To date, MEF2C has been thought of as a broad activator of the myocyte transcriptional
403 program, important for driving the expression of sarcomeric proteins and related
404 contractile machinery[24,50]. Here, we demonstrate that in addition to this role, MEF2C is
405 involved in more complex, segment-specific networks that regulate gene expression and
406 chromatin accessibility in the developing linear heart tube. Overall, we found that loss of
407 MEF2C causes a “posteriorization” of the heart tube. Typical outflow tract markers were
408 downregulated in MEF2C KO mice, while inflow tract markers were upregulated and their
409 domain of expression expanded into the developing ventricle (Fig. 2H-J, Table S3). These
410 changes in gene expression were coupled to altered chromatin accessibility throughout
411 the developing heart tube (Fig. 3). In line with MEF2C functioning as both a myocyte
412 activator and a segment-specific regulator, we found sets of DARs that were linked to
413 altered gene expression of sarcomeric proteins throughout the entire heart tube, as well
414 as DARs that were linked to DEGs unique to specific heart tube segments (Fig. 3C-F).
415 Using a deep-learning based modeling approach[30] to construct pseudotime lineage
416 trajectories for WT and MEF2C KO CMs, we also found a previously unappreciated
417 complexity to the role of MEF2C in regulating development of the inflow tract (Fig. 4J-M).
418 Our analyses demonstrated that MEF2C is required for proper determination of inflow
419 tract fate at E8.5, as well as for proper differentiation to either atrial or atrioventricular
420 canal CM subtypes at E9. This was markedly different from what we observed in the

421 outflow tract and ventricular lineages, where loss of MEF2C resulted in either a single
422 branch point diverging from normal development (outflow tract) or a delay in differentiation
423 (ventricular).

424 Using our integrated snRNA-seq and snATAC-seq datasets, we identified seven
425 MEF2C-dependent enhancers with activity in the early zebrafish embryonic heart with a
426 hit-rate greater than 60% (Fig. 5). The accuracy of our screening approach resulted from
427 our ability to link accessibility directly to gene expression in a segment-specific manner
428 and prioritize our candidates using published MEF2C and H3K27ac ChIP-seq
429 datasets[26,32]. Interestingly, we observed a range in the timing of onset of enhancer
430 activity among the positive hits we identified. This variation is likely due to differences in
431 the underlying mechanisms of activity specific to each enhancer, which warrant further
432 study. Possible mechanisms include differences in the temporal dynamics of accessibility
433 of each enhancer region, differential sensitivity to *Mef2c* expression, and unique
434 cooperativity with other TFs and regulatory elements. Due to the mosaicism of zebrafish
435 transgenesis in F0 embryos, we were unable to determine whether our novel enhancers
436 demonstrated segment-specific activity in the zebrafish. We are currently in the process
437 of generating stable lines that will be better suited for examining the precise spatial
438 domains of enhancer activity. In the future, we also plan to perform transgenesis of our
439 novel enhancers in the mouse to more accurately determine their activity in the
440 mammalian heart.

441 By constructing lineage-specific gene regulatory networks, we identified a genetic
442 interaction between MEF2C and NR2F2 in the inflow tract of the developing linear heart
443 tube. Although *Nr2f2* is not differentially expressed in any CM subtype at E8.5 and is
444 modestly downregulated in A-CMs at E9 (Table S3), we identified an altered
445 transcriptional network occurs in MEF2C KO (Fig. 6H) that leads to increased gene
446 regulation by NR2F2 (Fig. 6I) at newly accessible regions of chromatin (Fig. 6A).
447 Interestingly, MEF2C and NR2F2 have been shown to physically interact to regulate gene
448 expression in Leydig cells in the male testes[51]. Our data suggest that MEF2C and NR2F2
449 also cooperate as co-factors in the linear heart tube. We envision a model in which the
450 loss of MEF2C leads to a redistribution of NR2F2 to other sites in the genome, altering
451 its transcriptional activity. In support of this model, we observed that newly accessible

452 regions of chromatin in MEF2C KO IFT-CMs containing an NR binding motif rarely also
453 contained a MEF2C motif (Fig. 6B). This suggests that in the WT condition, MEF2C may
454 be responsible for recruiting NR2F2 to the sites that are regulated by cooperativity
455 between MEF2C and NR2F2. Both MEF2C and NR2F2 contain well-characterized DNA
456 binding regions and regulate gene expression by binding DNA directly[52,53], so it remains
457 unclear why and how NR2F2 redistributes to new sites upon the loss of MEF2C. One
458 likely possibility is that interactions with new co-factors are involved in this process. It will
459 also be interesting to determine whether the normal interaction of MEF2C and NR2F2 in
460 the heart tube occurs directly or involves other key cardiac TFs or co-factors.

461 Our approach provides a generalizable framework for integrating snRNA-seq and
462 snATAC-seq data to build GRNs that provide specific and comprehensive models of gene
463 regulation. By constructing these GRNs using both WT and mutant data, we are able to
464 immediately test predictions made from the WT networks (Fig. 6F-G) and validate the
465 inferred regulatory interactions. These tools also allow specific networks to be built for
466 particular cell or tissue types, provided there are enough cells in the group of interest to
467 generate robust inferences. Moreover, the experimental mutant data reveals how the
468 overall networks are altered – both in ways that would and would not be predicted from
469 the WT networks. These unexpected differences help to illuminate indirect or
470 transactivating gene regulation. Insights from these networks lead to new hypotheses
471 about regulatory elements, TF interactions, DNA binding, and chromatin organization. As
472 these hypotheses are tested experimentally, the GRNs can be refined, leading to higher-
473 fidelity models that will have greater predictive power. These higher-fidelity networks will
474 yield important new insights into the transcriptional regulation of developmental
475 processes and will have the ability to reveal precisely how these processes are altered in
476 the context of disease-causing mutations.

MATERIALS AND METHODS

Mouse Models

Animal studies were performed in strict compliance with the UCSF Institutional Animal Care and Use Committee. Mice were housed in a standard 12-hour light/dark animal husbandry barrier facility at the Gladstone Institutes. The *Mef2c*$^{del/+}$ allele used was the same as the original generated by Lin et al.[19]. The *Smarcd3*-F6-eGFP reporter allele was generated in our lab and described previously[20]. The *COUP-TFII lacZ* knock-in allele (abbreviated here as Nr2f2$^{del/+}$) was obtained from the Mutant Mouse Resource & Research Centers and was originally generated by Takamoto et al.[54]. The *Mef2c*$^{del/+}$ and *Nr2f2*$^{del/+}$ lines were maintained in the C57BL6/J background (Jackson Laboratory #664). The *Smarcd3*-F6-eGFP allele was crossed into our *Mef2c*$^{del/+}$ line from a mixed background of C57BL6/J (Jackson Laboratory #664) and CD-1 (Charles River #022).

Timed Embryo Dissections and Genotyping

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

508 DNA detection. Primers used for genotyping were ordered from Integrated DNA
509 Technologies (IDT). For genotyping the *Mef2c* allele, a three-primer mix was used (P1:
510 ACTGGCTGAGAGTTGTACCCAC, P2: GCGTGATTTGCTAGATAGTGGTAGAC, P3:
511 ATGTGGAATGTGTGCGAGGC), resulting in a single 314 bp band for WT, a single 242
512 bp band for *Mef2c*$^{del/del}$, and both bands for Mef2c$^{del/+}$. For genotyping the *Nr2f2* allele, a
513 three-primer mix was used (P4: CATCCGGGATATGTTACTGTCCGG, P5:
514 TGGGGAAGCTAAGTGTTGATGTGATTCC, P6: GCCGTGGGTTTCAATATTGGCTTC),
515 resulting in a single 786 bp band for WT, a single ~1600 bp band for *Nr2f2*$^{del/del}$, and both
516 bands for *Nr2f2*$^{del/+}$.
517
518 **Whole-Mount Embryo Immunostaining and Imaging**
519 Embryos collected for immunostaining were fixed for 1 hour at room temperature in 4%
520 paraformaldehyde (PFA) freshly diluted from 16% weight/volume PFA aqueous solution
521 (Thermo Fisher Scientific, 043368-9M) in PBS (Life Technologies, 14190250). Embryos
522 were stored at 4 °C for up to 1 month prior to immunostaining. To permeabilize and block
523 non-specific antigens, embryos were incubated at 37 °C for 2 hours with gentle rotation
524 in a blocking buffer consisting of PBS (Life Technologies, 14190250), 5% normal donkey
525 serum (Sigma Aldrich, S30-M), and 0.8% Triton X-100 (Sigma Aldrich, X-100-5ML). This
526 was then removed and replaced with primary antibodies diluted in the same blocking
527 buffer and incubated overnight at 37 °C with gentle rotation. The next day, primary
528 antibody solution was removed and embryos were washed 3x 30 minutes in the same
529 blocking buffer at 37 °C with gentle rotation. Secondary antibodies and DAPI (Abcam,
530 ab228549, used at 1 µg/mL) were added in a blocking buffer consisting of PBS (Life
531 Technologies, 14190250), 5% normal donkey serum (Sigma Aldrich, S30-M), and 0.4%
532 Triton X-100 (Sigma Aldrich, X-100-5ML), and incubated for 3-4 hours at 37 °C with gentle
533 rotation. Secondary antibody solution was removed and embryos were washed 3x 30
534 minutes in PBS (Life Technologies, 14190250) with 0.1% Tween-20 (Thermo Fisher
535 Scientific, J20605.AP) at 37 °C with gentle rotation. Embryos were imaged in PBS using
536 an upright epifluorescence microscope (Leica MZFLIII, Leica DFC 3000G, Lumen
537 Dynamics XCite 120LED) and acquisition software LASX (Leica). Primary antibodies
538 used were sheep polyclonal MEF2C (R&D Systems, AF6786, used at 1:250) and rabbit

539 polyclonal cardiac Troponin T (Thermo Fisher Scientific, 15513-1-AP, used at 1:250).
540 Secondary antibodies used were Donkey anti-Sheep IgG (H+L) Cross-Adsorbed
541 Secondary Antibody, Alexa Fluor™ 647 (Thermo Fisher Scientific, A-21448, used at
542 1:1000) and Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody,
543 Alexa Fluor™ 555 (Thermo Fisher Scientific, A-31572, used at 1:1000).
544

545 **Whole-Mount Embryo Fluorescence *In Situ* Hybridization via Hybridization Chain**
546 **Reaction and Imaging**

547 The protocol for *in situ* hybridization via hybridization chain reaction (HCR™) was adapted
548 from the optimized manufacturer's protocol for whole-mount mouse embryos (Molecular
549 Instruments). Embryos were fixed overnight at 4 °C in 4% PFA freshly diluted from 16%
550 weight/volume PFA aqueous solution (Thermo Fisher Scientific, 043368-9M) in PBS (Life
551 Technologies, 14190250). Embryos were then washed 3x 5 minutes in PBS (Life
552 Technologies, 14190250) with 0.1% Tween-20 (Thermo Fisher Scientific, J20605.AP),
553 abbreviated PBST, and then processed through 10-minute serial incubations in a
554 dehydration series of 25%, 50%, 75%, and 100% methanol solutions in PBST on ice.
555 Embryos were stored in 100% methanol at −20 °C for up to 6 months prior to *in situ*
556 hybridization. Embryos were then rehydrated through 10-minute serial incubations in a
557 series of 75%, 50%, 25%, and 0% methanol solutions in PBST on ice. Embryos were
558 then brought to room temperature and permeabilized by 10-minute incubation with 10
559 µg/mL Proteinase K (Sigma Aldrich, 70663-5) in PBST, washed 2x 5 minutes in PBST at
560 room temperature, post-fixed in 4% PFA freshly diluted from 16% weight/volume PFA
561 aqueous solution (Thermo Fisher Scientific, 043368-9M) in PBS (Life Technologies,
562 14190250), and then washed 3x 5 minutes in PBST at room temperature. Embryos were
563 pre-hybridized in Probe Hybridization Buffer (Molecular Instruments) for 5 minutes at
564 room-temperature followed by an additional 30 minutes at 37 °C with fresh buffer. Single
565 or multiplexed hybridization probes were then diluted in pre-warmed Probe Hybridization
566 Buffer (Molecular Instruments) and embryos were hybridized for 48 hours at 37 °C.
567 Embryos were then washed 4x 15 minutes with pre-warmed Probe Wash Buffer
568 (Molecular Instruments) at 37 °C, followed by 2x 5 minutes with 5xSSCT (diluted in
569 nuclease-free water from 20xSSCT, Thermo Fisher Scientific, 15557044) at room

570 temperature. During washes, amplification hairpins were snap-cooled by heating to 95 °C
571 for 90 seconds and then allowed to cool to room temperature in the dark. Embryos were
572 pre-amplified in Amplification Buffer (Molecular Instruments) for 5 minutes at room
573 temperature. Snap-cooled hairpins and DAPI (Abcam, ab228549, used at 1 µg/mL) were
574 diluted in Amplification Buffer (Molecular Instruments) and embryos were incubated in this
575 solution overnight in the dark at room temperature. The following day, embryos were
576 washed through a series of 5xSSCT washes (diluted in nuclease-free water from
577 20xSSCT, Thermo Fisher Scientific, 15557044) in the dark at room temperature (2x 5
578 minutes, 2x 30 minutes, 1x 5 minutes). Embryos were imaged in PBS using an upright
579 epifluorescence microscope (Leica MZFLIII, Leica DFC 3000G, Lumen Dynamics XCite
580 120LED) and acquisition software LASX (Leica). To maximize the utility of collected
581 embryos, probes were then stripped by overnight incubation in 80% formamide in water
582 (Thermo Fisher Scientific, BP228100) at room temperature, followed by 3x 5 minutes
583 washes in PBST at room temperature. Embryos could then be probed for additional
584 targets, beginning the protocol again at the pre-hybridization step. Probes used in this
585 study were ordered from the Molecular Instruments catalogue, when possible, or custom-
586 designed and manufactured by Molecular Instruments. Hybridization probes used were:
587 *Tnnt2*-B3 (used at 4 nM), *Ttn*-B2 (used at 4 nM), *Nppa*-B2 (used at 20 nM), *Nkx2.5*-B1
588 (used at 20 nM), *Tbx5*-B1 (used at 20 nM), *Gata4*-B1 (used at 20 nM), *Wnt2*-B2 (used at
589 20 nM), *Tdgf1*-B2 (used at 20 nM). Amplification probes used were all used at 60 nM: B1-
590 Alexa Fluor™ 647, B2-Alexa Fluor™ 546, B2-Alexa Fluor™ 647, B3-Alexa Fluor™ 488.
591

592 **Embryo Preparation for 10x Multiome**

593 For E7.75 embryos, whole embryos were dissected and harvested for single-nucleus
594 library generation. For E8.5 and E9 embryos, the head folds and posterior trunk were
595 removed by microdissection prior to harvesting for library generation to enrich the relative
596 capture of cardiac cell types. Two embryos were collected per genotype per embryonic
597 stage. Following dissection and rapid PCR genotyping, embryo samples selected for the
598 multiome experiment were incubated in 200 µl TrypLE Select (Thermo Fisher Scientific,
599 12563-011) for 5 minutes at 37 °C, triturated gently by pipetting up and down, and then
600 incubated an additional 3 minutes at 37 °C. The dissociated cell suspension was

601 quenched with 600 µl of PBS (Life Technologies, 14190250) with 1% FBS (Thermo Fisher
602 Scientific, 10439016), singularized by passage through the 35 µm mesh of a 5 mL
603 Falcon™ Round-Bottom Polystyrene Test Tube with Cell Strainer Snap Cap (Thermo
604 Fisher Scientific, 08-771-23), pelleted by centrifugation at 300 g for 5 minutes at 4 °C,
605 and resuspended in 50 µl of PBS (Life Technologies, 14190250) with 0.04% bovine serum
606 albumin (BSA, Sigma Aldrich, A1595). At this stage, the manufacturer's protocol for Nuclei
607 Isolation for Single Cell Multiome ATAC + Gene Expression Sequencing, Appendix, Low
608 Cell Input Nuclei Isolation (10x Genomics, CG000365 Rev B) was followed exactly to
609 prepare nuclei for library generation.
610
611 **Generation of snRNA-seq and snATAC-seq Libraries and Sequencing**
612 Libraries for snRNA-seq and snATAC-seq were prepared according to the manufacturer's
613 protocol for Chromium Next GEM Single Cell Multiome ATAC + Gene Expression (10x
614 Genomics, CG000338 Rev B), using the 10x Genomics Chromium controller, Chromium
615 Next GEM Single Cell Multiome ATAC + Gene Expression Reagent Bundle (10x
616 Genomics, PN-1000283), Chromium Next GEM Chip J Single Cell Kit (10x Genomics,
617 PN-1000234), Single Index Kit N Set A (10x Genomics, PN-1000212), and Dual Index Kit
618 TT Set A (10x Genomics, PN-1000215). Nuclei isolated from multiome samples were
619 processed directly into the snATAC-seq transposition reaction and then captured in 10x
620 Genomics GEMs via the Chromium Controller. GEMs were stored at -80 °C for up to 4
621 weeks, allowing for collection of replicate embryos of each genotype at each development
622 stage. All subsequent library preparation steps were performed together for all embryos
623 of a given developmental stage to reduce the likelihood of technical batch artifacts. A
624 targeted maximum recovery of 10,000 cells per sample were loaded onto the 10x
625 Genomics Chromium controller instrument and each sample was indexed with a unique
626 sample identifier (Single Index Kit N Set A for snATAC-seq libraries, Dual Index Kit TT Set
627 A for snRNA-seq libraries). Final libraries were quality-controlled using an Agilent
628 Bioanlyzer instrument with Agilent High Sensitivity DNA Kit (Agilent, 5067-4626). The
629 DNA concentration of each library was measured using KAPA Library Quantification
630 qPCR Kit (Thermo Fisher Scientific, 50-196-5234), and then libraries were pooled and
631 sequenced on NovaSeq6000 S4 lanes (Illumina). All snRNA-seq libraries were

632 sequenced to a depth of at least 20,000 mean raw aligned reads per cell and snATAC-
633 seq libraries were sequenced to a depth of at least 35,000 mean raw aligned read pairs
634 per cell (most libraries >60,000).
635
636 **Processing of Raw Sequencing Data**
637 Raw sequencing reads were processed using the 10x Genomics Cell Ranger Arc v2.0.0
638 pipeline. Reads were demultiplexed using cellranger-arc mkfastq and aligned with
639 cellranger-arc count to the mm10 reference genome containing an additional sequence
640 for eGFP to map reads of the *Smarcd3*-F6-eGFP reporter transcript.
641
642 **Seurat Analysis of snRNA-seq Data**
643 Gene expression data were analyzed using Seurat v4.3.0 in R[21]. Datasets for each
644 developmental stage (E7.75, E8.5, E9) were analyzed as separate Seurat objects.
645 Counts matrices from each sample that were output by the Cell Ranger Arc pipeline were
646 used as inputs to the Read10X() and CreateSeuratObject() functions. Data were filtered
647 for quality control based on number of detected genes, unique molecular identifier (UMI)
648 counts, percent ribosomal transcripts, and percent mitochondrial transcripts.
649 SCTransform v2[22] was used to normalize and scale data, with regressions performed
650 with respect to percentage of ribosomal transcripts and percentage of mitochondrial
651 transcripts. Principal component analysis (PCA) was performed, followed by batch
652 correction with Harmony[55], and cells were clustered and visualized as UMAP
653 embeddings based on the top 30 principal components using FindNeighbors(),
654 FindClusters(), RunUMAP(), and DimPlot(). Cell types were annotated using the
655 FindAllMarkers() function with Wilcoxon rank-sum test (min.pct = 0.25, logfc threshold =
656 0.25, p_val_adj < 0.05) to identify cluster-specific marker genes (Table S1A-C, Fig. S2D-
657 F). Cardiac progenitors, cardiomyocytes, and related mesoderm cell types were subset
658 into new Seurat objects. PCA analysis, batch correction, visualization, and clustering were
659 then performed on these new objects as described above. Cell types in the subset objects
660 were annotated using the FindAllMarkers() function with Wilcoxon rank-sum test (min.pct
661 = 0.25, logfc threshold = 0.5, p_val_adj < 0.05) to identify cluster-specific marker genes
662 that enabled more specific cell type labeling than in the full datasets (Table S2A-C, Fig.

S2G-I). Differential gene expression was performed between MEF2C KO and WT cell types of interest using the FindMarkers() function with Wilcoxon rank-sum test (min.pct = 0.1, logfc threshold = 0.25, p_val_adj < 0.05). The DotPlot() and RidgePlot() functions was used to visualize gene expression levels of differentially expressed genes across cell types of interest. The AddModuleScore() function[48] was used to add module scores for up-regulated (MEF2C KO-vs-WT) NR2F2 and GATA4 target genes and the RidgePlot() function was used to visualize these module scores in cell types of interest.

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

Initial Dataset Processing

Chromatin accessibility and integrated multiome data was analyzed using the ArchR[28] software package v1.0.2 in R. The fragments files output by the Cell Ranger Arc pipeline were used as inupts to the createArrowFiles() function to create Arrow Files for each sample. These Arrow Files were used to create individual ArchR projects for developmental stage (E7.75, E8.5, E9). The gene expression data was added using the import10xFeatureMatrix() and addGeneExpressionMatrix() functions, using the matching cell barcodes to combine the snRNA-seq and snATAC-seq data for each cell. Data were filtered for quality control based on the number of unique fragments, TSS Enrichment score, expression UMI counts, and number of detected genes. Doublets were filtered using the addDoubletScores() and filterDoublets() functions. Dimensionality reduction was performed using ArchR's iterative latent semantic index (LSI) implementation individually for the snATAC-seq data (varFeatures = 25,000, dimsToUse = 1:30) and snRNA-seq data (varFeatures = 5,000, dimsToUse = 1:30). Then, the reduced dimensions were combined using addCombinedDims() and batch correction was performed on the combined dimensions using Harmony[55]. Cells were then embedded in UMAP space with addUMAP(), clustered using addClusters(), and visualized using plotEmbedding(). Cell types were annotated by transferring metadata labels from the previously analyzed Seurat objects. The getMarkerFeatures() function with Wilcoxon rank-sum test (FDR <= 0.01, Log2FC >= 0.5) was used to identify cluster-specific marker genes to confirm cell type labels. Cardiac progenitors, cardiomyocytes, and related mesoderm cell types were subset into new ArchR projects. Dimensionality reduction,

694 batch correction, clustering, and visualization were performed as described above. Cell
695 types in the subset projects were annotated by transferring metadata labels from the
696 previously analyzed subset Seurat objects. The getMarkerFeatures() function with
697 Wilcoxon rank-sum test (FDR <= 0.01, Log2FC >= 0.5) was used to identify cluster-
698 specific marker genes to confirm subset cell type labels. Due to the small number of
699 MEF2C KO OFT-CMs in the E8.5 dataset, an ArchR project with combined E8.5 and E9
700 samples was created and processed as described above for downstream
701 analyses/comparisons of OFT-CMs.
702
703 *Peak Calling, Differential Peak Testing, and Motif Enrichment*
704 Cells were pseudobulked into replicates based on their cell type and genotype labels (e.g.
705 WT V-CMs), with two pseudobulked replicates created corresponding to the two embryos
706 collected for each genotype at each developmental stage. Replicates were created using
707 the addGroupCoverages() function (minCells = 20, maxCells = 500, sampleRatio = 0.8).
708 Peaks were then called using addReproduciblePeakSet() with MACS2[29] (reproducibility
709 = 2, peaksPerCell = 500, minCells = 20). Differential peak accessibility testing was
710 performed using getMarkerFeatures() with Wilcoxon rank-sum test (FDR <= 0.15,
711 Log2FC >= 0.5 or Log2FC <= -0.5) for each cell type of interest with MEF2C KO cells
712 used as the group of interest and WT cells used as the background. Transcription factor
713 binding motifs were annotated using addMotifAnnotations() with the Vierstra motif
714 archetypes[56] and motif enrichments for regions of gained or lost accessibility were
715 calculated using peakAnnoEnrichment() with cutoffs of FDR <= 0.15 and Log2FC <= -0.5
716 or Log2FC >= 0.5.
717
718 *Peak2Gene Linkages and Browser Track Visualization*
719 Peak2Gene Linkages were identified using the addPeak2GeneLinks() and
720 getPeak2GeneLinks() functions. Code written in this study was used to filter Peak2Gene
721 links based on presence of a MEF2C motif and linkage to specific genes (i.e. differentially
722 expressed genes). Peak2Gene links were visualized using the plotBrowserTrack()
723 function. The getGroupBW() function was used to export the accessibility coverage data

724 in bigwig format for visualization along with published ChIP-seq data using the IGV
725 software package[57].
726
727 **MIRA Analysis of Integrated snRNA-seq and snATAC-seq Data**
728 *Preparing Merged Datasets*
729 We used the Multimodal models for Integrated Regulatory Analysis[30] (MIRA) python
730 package to examine dynamic changes in gene expression and chromatin accessibility in
731 WT and MEF2C KO embryos across the time course of developmental stages that we
732 collected (E7.75, E8.5, E9). We first created a merged Seurat object and ArchR project
733 containing only our cardiac cell types of interest. We then used sceasy[58] to convert our
734 Seurat object and the peak matrix from the ArchR project to AnnData formats, which were
735 used as inputs for the MIRA analyses.
736
737 *Topic Modeling*
738 The first step of the MIRA pipeline is to train and tune topic models for both the gene
739 expression and chromatin accessibility data. Model training and tuning was performed
740 with GPU support on nodes of the Wynton HPC Co-Op cluster at UCSF. For our gene
741 expression topic model, we used the 4,088 most variable genes (min_disp = 0.1) as the
742 "exogenous" gene set and top 1,959 variable genes (dispersion threshold > 0.7) as the
743 "endogenous" gene set for model training. We then set the minimum and maximum
744 learning rate bounds, tuned the model (iters = 64, max_topics = 55), and MIRA's tuner
745 selected the best model that was used for all subsequent analyses, which contained 33
746 gene expression topics. To aid in the interpretation of results, we exported the top 200
747 genes for each of the 33 gene expression topics. We then repeated these steps to train
748 and tune the accessibility topic model, using all features (i.e. there are no "exogenous" or
749 "endogenous" peak sets), which resulted in a model containing 17 accessibility topics. To
750 aid in the interpretation of these topics, we performed TF binding motif enrichment
751 analysis on each topic and exported the TF motif enrichment scores (top_quantile = 0.2)
752 for each of the 17 accessibility topics.
753
754 *Pseudotime Analysis and Lineage Trajectories*

755 For the heart tube segment lineages we analyzed (OFT, V, IFT), we created separate
756 subsetted AnnData objects containing only the cell types that contribute to each lineage.
757 We generated UMAP embeddings for each lineage based on joint representations of gene
758 expression and accessibility topics. We then performed pseudotime analysis by
759 calculating the diffusion map, selecting the appropriate number of diffusion map
760 components, and generated a K nearest neighbor graph based on these diffusion map
761 components. In order to create pseudotime trajectories, we randomly selected a WT cell
762 in the earliest progenitor cell type of each lineage to serve as the initial cell, and then
763 manually selected terminal cells from the appropriate cell types based on the structure of
764 each trajectory (for OFT lineage: E9 WT OFT-CM and E9 KO OFT-CM; for V lineage: E9
765 WT V-CM; for IFT lineage: E8.5 KO IFT-CM, E9 WT A-CM, E9 WT AVC-CM, E9 KO A-
766 CM, E9 KO AVC-CM). Given these initial and terminal cells, along with the KNN
767 pseudotime graph, the probability of reaching each terminal state was calculated for every
768 cell and these probabilities we parsed to create the bifurcating tree lineage trajectory plots
769 (stream plots). Gene expression and accessibility topics of interest were then plotted
770 along these stream plots to interpret the dynamic changes in gene expression and
771 chromatin accessibility that occur in WT and MEF2C KO cells as they differentiate through
772 each lineage.
773
774 **Cloning of Reporter Constructions for Zebrafish Transgenesis**
775 To generate the Tg(*MVEB:egfp*)^{sfc26} zebrafish lines, 650-900 nucleotide fragments
776 containing the candidate enhancers were cloned by PCR from mouse genomic DNA.
777 Candidate nucleotide sequences and the primers used to amplify them are listed in Table
778 S6. The resulting fragments were cloned into the Tol2 transgenic vector E1b-eGFP-
779 Tol2[35] (RRID: Addgene_37845) using the In-Fusion Cloning system (Takara Bio). All
780 plasmid sequences were confirmed by long-read sequencing before zebrafish injection
781 (Plasmidsaurus).
782
783 **Zebrafish Transgenesis Injections, Screening, and Imaging**
784 Zebrafish were raised under standard laboratory conditions at 28 °C. The outbred
785 wildtype zebrafish EKW (*Danio rerio-Ekwill*) and Tg(*cmlc2:mCherry*)^{s890} lines have been

786 described previously[55]. One-cell zebrafish embryos obtained from
787 Tg(*cmlc2:mCherry*)^{s890} and wildtype EKW (Ekwill strain) parents were injected with 1nl of
788 injection mix containing 15pg of candidate-enhancer-eGFP and 15pg of ß-actin-BFP
789 reporter plasmids along with 30pg of *Tol2* mRNA. The ß -actin promoter drives expression
790 of BFP strongly in somite muscles and was used as a positive injection control. As a
791 negative control, we also injected and screened embryos with an empty reporter construct
792 containing only the minimal E1b promoter and eGFP. Injected embryos were screened at
793 24 and 72 hpf for eGFP and BFP expression. Representative embryos were imaged for
794 eGFP expression as a readout of candidate enhancer activity and for mCherry expression
795 as a cardiac marker. Enhancer activity was scored as a percentage of embryos
796 expressing eGFP in the heart compared to the total number of BFP-positive (BFP+)
797 embryos. A minimum of at least 50 BFP+ embryos were screened per construct, except
798 for MVEB1, for which only 17 BFP+ embryos were obtained. The exact numbers screened
799 for each construct are displayed in Fig. S5B.
800
801 **Comparative Analyses with Published Datasets**
802 NR2F2 and GATA4 target genes were downloaded from the Harmonizome[46] and
803 TFLink[47] databases, respectively. Fisher's exact test was used to determine the Odd's
804 Ratios and p-values for identifying NR2F2 and GATA4 targets among genes upregulated
805 in MEF2C KO IFT-CMs at E8.5. Publicly available ChIP-seq datasets were downloaded
806 from the Gene Expression Omnibus (GEO) in wig or bigwig format for MEF2C
807 (GSE124008)[26], H3K27ac (GSE52386)[32], GATA4 (GSE52123)[45], and NR2F2
808 (GSE46498)[42]. Datasets provided in wig format were converted to bigwig format using
809 the Galaxy[44] web tool wigToBigWig. CrossMap[60] was used together with chain files from
810 the UCSC Genome Browser[61] to lift genome annotations from mm9 to mm10 reference
811 genome assemblies, and vice versa. For the browser tracks in Fig. 5B-C, the MEF2C and
812 H3K27ac ChIP-seq bigwigs were lifted from mm9 to mm10, the Galaxy[44] web tool
813 bigwigAverage was used to generate a single average MEF2C and H3K27ac occupancy
814 track from the individual replicate tracks, and the results were displayed along with our
815 pseudobulked snATAC-seq data using the IGV software package[57]. For the heat maps
816 displayed in Fig. S6A-B, we first created bed format files for the gained DARs in MEF2C

KO IFT-CMs containing NR or GATA motifs. We then lifted these bed regions from mm10 to mm9 and used the Galaxy[44] web tools computeMatrix and plotHeatmap with the downloaded NR2F2 and GATA4 ChIP-seq data to visualize NR2F2 and GATA4 occupancy at the DARs. At the computeMatrix step, missing data was converted to zeroes. At the plotHeatmap step, kmeans clustering was used to generate two clusters of DARs (effectively “occupied” and “not occupied”) and these were sorted by decreasing mean value prior to being plotted.

Gene Regulatory Network Inference and Analysis

We used CellOracle[49] to infer GRNs for WT and MEF2C KO IFT-CMs at E8.5 and E9. The GRN inference process included the following steps:

Base GRN Construction

We used OFT-CM, V-CM, and IFT-CM snATAC-seq data to construct a base GRN that defines the accessible TF binding motifs near genes for each of the following conditions: E8.5-WT, E8.5-KO, E9-WT, and E9-KO. Pooling the OFT-CM, V-CM, and IFT-CMs mitigated artifacts caused by data sparsity. The binarized PeakMatrix output from ArchR was used as an input to Cicero’s[62] run_cicero() function (with default parameters) to compute co-accessibility scores among peak pairs within 500kb. 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 500bp 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[63] library to produce a set of TF motifs present in accessible regions within 500kb of each gene. The TFs that bind to these motifs have the potential to regulate the given gene.

848

849 *GRN Inference*

850 Using CellOracle, we constructed GRNs in which edge strengths are quantified based on
851 the TF-to-gene correlation, constrained by the accessible regions defined in the base
852 GRN. We inferred GRNs for E8.5 WT, E8.5 KO, E9 WT, and E9 KO IFT-CMs, using the
853 computed base GRNs to determine permissible TF-to-gene links. GRN inference used
854 the top 5,000 differentially expressed genes in the snRNA-seq dataset. We used the
855 CellOracle oracle.get_links() function (with default parameters), fitting linear models with
856 L2 weight regularization to derive coefficients relating TF expression to target gene
857 expression. The Bagging Ridge model was used to fit coefficients, providing posterior
858 distributions for each TF-to-gene relationship. Positive coefficients indicated activating,
859 while negative coefficients indicated inhibiting relationships. We then filtered the inferred
860 network to include 4,000 edges with the largest magnitude coefficients.

861

862 *Simulation of MEF2C Knockout Effects*

863 To predict the effect of MEF2C KO on gene expression, we identified genes directly
864 regulated by MEF2C in the E8.5 WT IFT-CM network. We predicted that reduced
865 expression of *Mef2c* in the network would downregulate genes directly activated by
866 MEF2C and upregulate genes directly inhibited by MEF2C in the E8.5 WT IFT-CM
867 network. The simulation results were compared against differentially expressed genes in
868 the E9 snRNA-seq IFT-CM experimental data (WT-vs-KO). Differentially expressed genes
869 were identified by selecting for genes with logfold change magnitude >1.5.

870

871 *Network Visualization*

872 Whole-network visualizations were generated using Gephi[(64)], applying the Yifan Hu
873 layout, with node sizes scaled by betweenness centrality. PyVis[(65)] was used to visualize
874 subnetworks that included manually selected TFs (*Mef2a, Mef2c, Nr2f2, Tbx5, Gata4,*
875 *Meis2, Plagl1, Isl1, Pitx2, Tbx20, Hand2, Nkx2-5*) and the the top 100 differentially
876 expressed genes (ranked in ascending order by p value) identified by comparing the E8.5
877 WT and MEF2C KO IFT-CM snRNA-seq experimental data. Visualizations for direct
878 interactions between TFs (*Nr2f2* and *Gata4*) and targets in the E8.5 WT and KO IFT-CMs

879 were developed using Python's NetworkX package[66], with a circular layout and edge
880 widths scaled by coefficient values.

ACKNOWLEDGEMENTS

Sequencing was performed at the UCSF CAT, supported by UCSF PBBR, RRP IMIA, and NIH 1S10OD028511-01 grants. 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. J.M.M-V. is grateful for assistance from Dr. Kavitha Rao in learning Seurat, as well as from Dr.'s 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 project, and help with experiments.

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.B. performed mouse colony maintenance and genotyping. J.M.M-V. and E.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; R01-HL160665 and R01-HL162925 to J.J.S.; R01-HL114948 and R01-

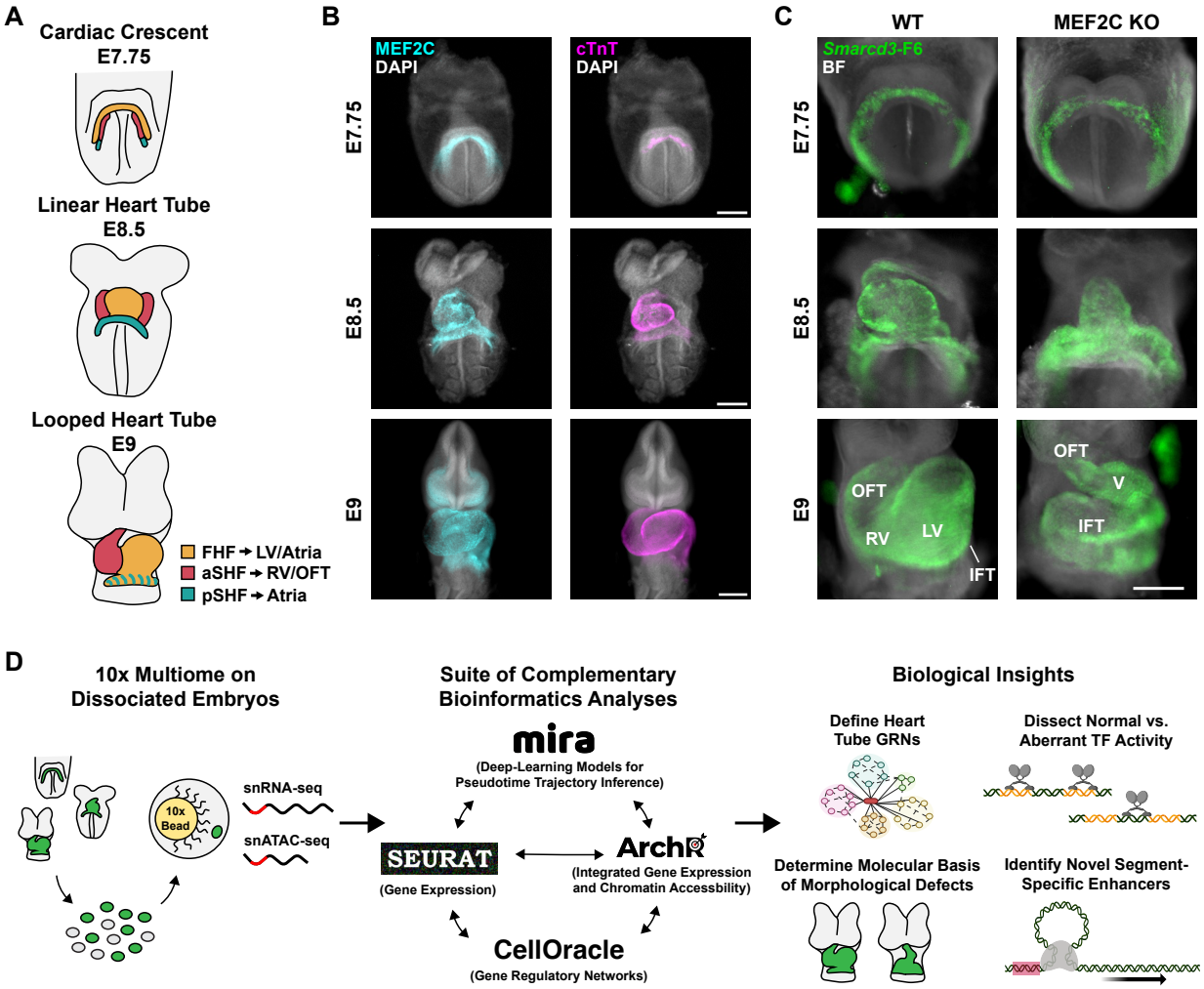
HL155906 to B.G.B.), the Roddenberry Foundation (B.G.B.), Additional Ventures (B.G.B.), and the Younger Family Fund (B.G.B.). J.M.M-V. was supported by fellowships from the NHLBI (F32-HL162450) and the American Heart Association together with The Children's Heart Foundation (24POST1191660). A.P.C. was supported by the NHLBI (T32-HL007284).

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 and plot generation is available in the GitHub repository associated with this manuscript: https://github.com/jmmuncie/Muncie-Vasic_bioRxiv_2024.

924 **FIGURES AND CAPTIONS**

925 **Figure 1**



926

927

928 **Figure 1: Although MEF2C is expressed throughout the developing heart tube, loss**
929 **of MEF2C causes segment-specific defects.**

930 A) Schematic of cardiac progenitors and their contributions to linear heart tube
931 development from cardiac crescent (E7.75) to looped heart tube (E9) stage. B)
932 Immunofluorescent staining of MEF2C (cyan) and cardiac Troponin T (cTnT, magenta) in
933 E7.75, E8.5, and E9 WT embryos. C) Representative images of WT and MEF2C KO
934 embryos at E7.75, E8.5 and E9. Cardiac progenitors are marked by the *Smarcd3*-F6-
935 eGFP reporter transgene (green). BF, brightfield. D) Schematic of the methodology and
936 biological insights presented in the current study. Elements of this panel were created in

BioRender. B, B. (2024) https://BioRender.com/i72e213. Scale bars = 200 μm. FHF, first heart field; aSHF, anterior second heart field; pSHF, posterior second heart field; LV, left ventricle; RV, right ventricle; V, ventricle; IFT, inflow tract; OFT, outflow tract.

Figure 2

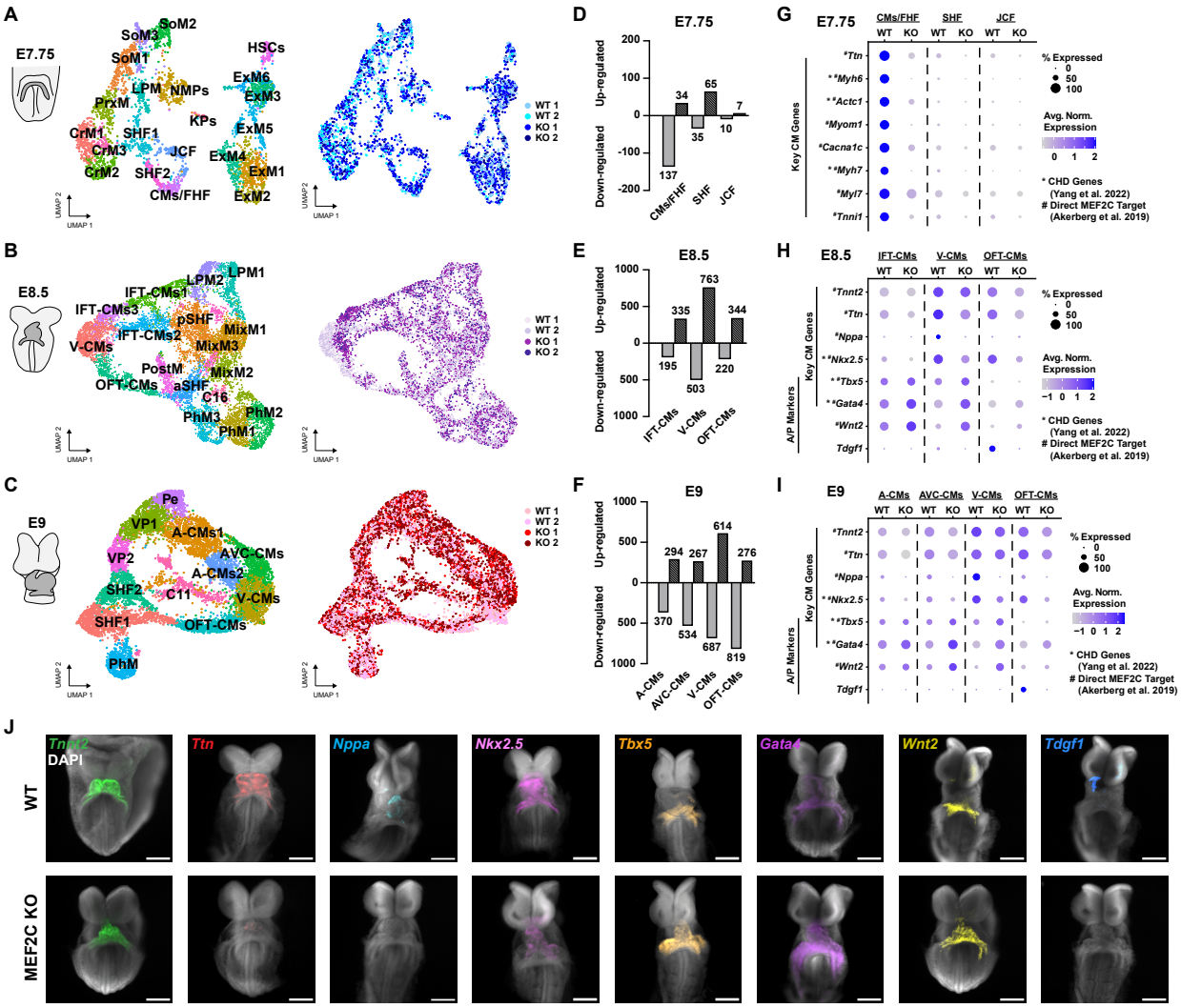
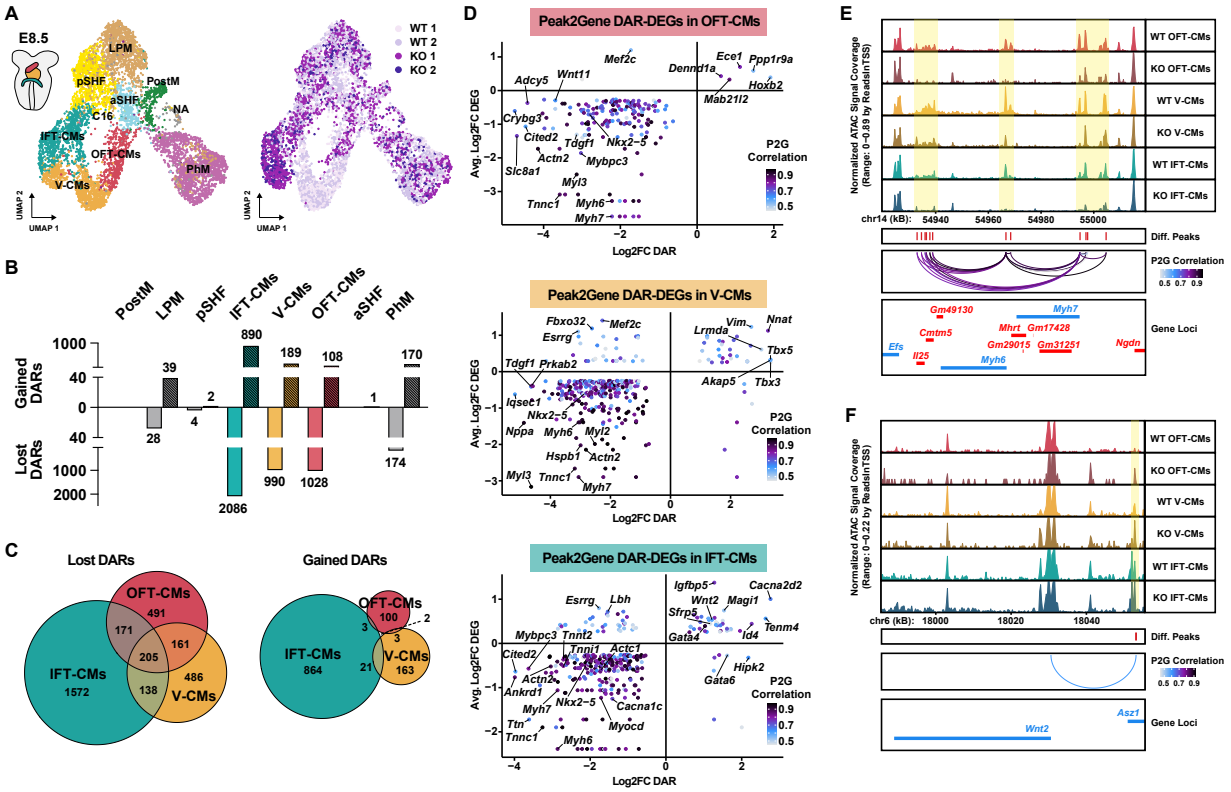


Figure 2: Loss of MEF2C leads to broad changes in gene expression throughout the heart tube, including mis-regulation of key cardiomyocyte genes and anterior/posterior markers.

A-C) UMAPs of snRNA-seq data for cardiac progenitors, cardiomyocytes, and related mesoderm subtypes from E7.75 (A), E8.5 (B), and E9 (C) embryos labeled by cell type (left) and genotype/sample ID (right). D-F) Bar plots displaying the number of up-regulated and down-regulated genes (MEF2C KO-vs-WT) for cell types of interest at E7.75 (D), E8.5 (E), and E9 (F). G-I) Dot plots displaying gene expression of key CM genes and anterior/posterior (A/P) markers at E7.75 (G), E8.5 (H), and E9 (I). J)

952 Fluorescence *in situ* hybridization of key CM genes and A/P markers in E8-E8.5 WT and
953 MEF2C KO embryos. Scale bars = 200 μm. CMs, cardiomyocytes; FHF, first heart field;
954 SHF, second heart field; JCF, juxtacardiac field; CrM, cranial mesoderm; PrxM, paraxial
955 mesoderm; LPM, lateral plate mesoderm; SoM, somitic mesoderm; NMPs,
956 neuromesodermal progenitors; KPs, kidney progenitors; ExM, extraembryonic
957 mesoderm; HSCs, hematopoietic stem cells; V-CMs, ventricular cardiomyocytes; IFT-
958 CMs, inflow tract cardiomyocytes; OFT-CMs, outflow tract cardiomyocytes; aSHF, anterior
959 second heart field; pSHF, posterior second heart field; PostM, posterior mesoderm; PhM,
960 pharyngeal mesoderm; MixM, mixed mesoderm; A-CMs, atrial cardiomyocytes; AVC-
961 CMs, atrioventricular canal cardiomyocytes; Pe, proepicardium; VP, venous pole; *,
962 Genes known to be associated with CHDs[25]; #, Direct targets of MEF2C based on
963 MEF2C ChIP-seq data[26].

964 **Figure 3**



965

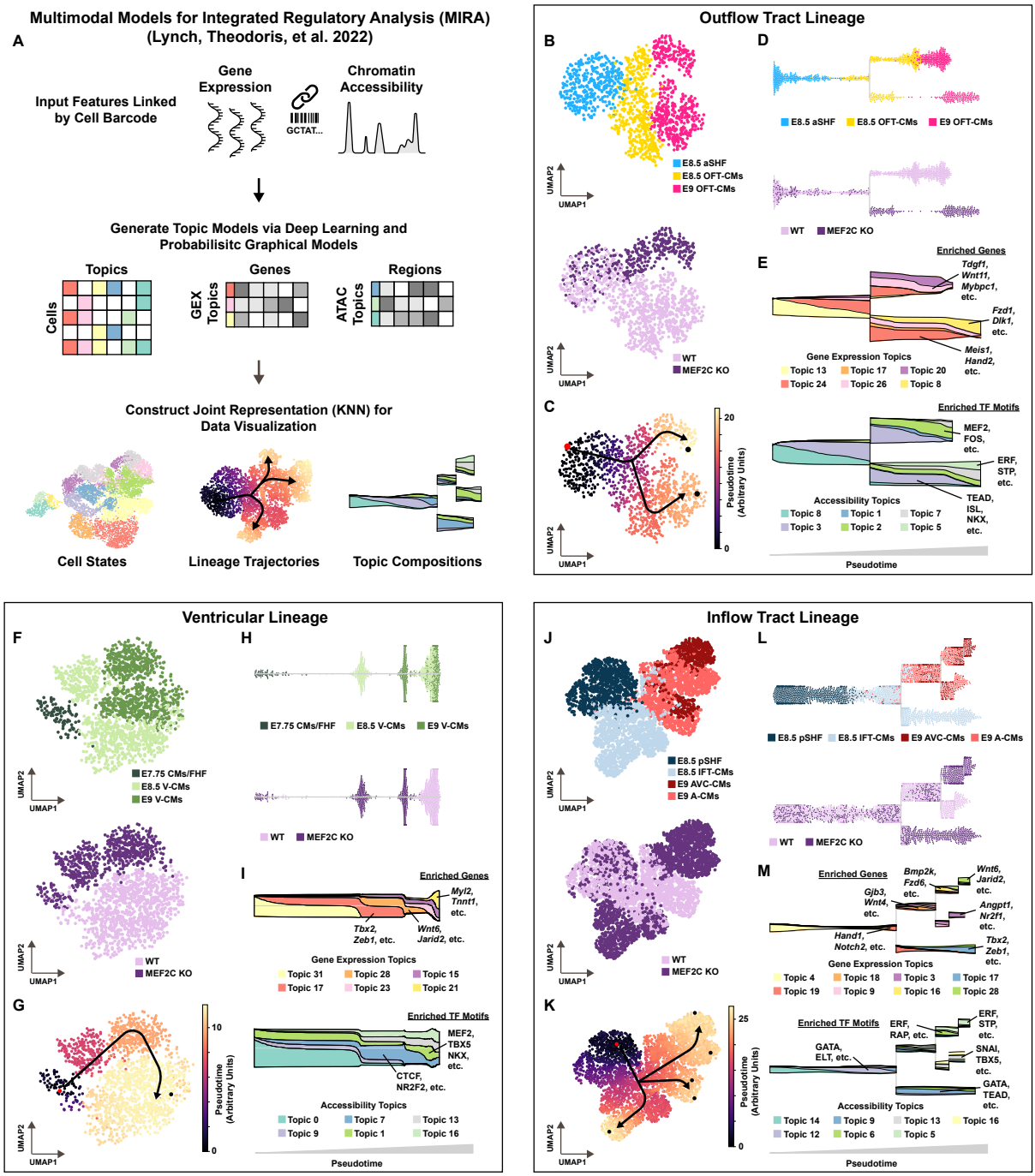
966

967 **Figure 3: MEF2C regulates chromatin accessibility broadly throughout the heart**
968 **tube and in a segment-specific manner.**

969 A) UMAP of integrated snRNA-seq and snATAC-seq data for cardiac progenitors,
970 cardiomyocytes, and related mesoderm subtypes at E8.5 labeled by cell types
971 determined from snRNA-seq clustering (left) and genotype/sample ID (right). B) Bar plots
972 displaying the number of gained and lost DARs (MEF2C KO-vs-WT) for cell types of
973 interest at E8.5. C) Venn diagrams displaying unique and overlapping DARs in the three
974 heart tube segments. D) Scatter plots displaying the relationship between the Log2 fold
975 change (Log2FC) values of the DEG and DAR analyses for all identified Peak2Gene
976 (P2G) links in the three heart tube segments. Dots are colored by the P2G correlation
977 score. E-F) Genome browser tracks displaying snATAC-seq accessibility profiles at the
978 *Myh6*/*Myh7* (E) and *Wnt2* (F) loci for the indicated pseudobulked cell types. DARs are
979 highlighted and the peaks are indicated by red bars (MEF2C KO-vs-WT IFT-CMs). Loops
980 indicate P2G links, colored by the P2G correlation score. V-CMs, ventricular

981 cardiomyocytes; IFT-CMs, inflow tract cardiomyocytes; OFT-CMs, outflow tract
982 cardiomyocytes; aSHF, anterior second heart field; pSHF, posterior second heart field;
983 LPM, lateral plate mesoderm; PostM, posterior mesoderm; PhM, pharyngeal mesoderm;
984 NA, cells not available in the snRNA-seq dataset.

985 **Figure 4**



986

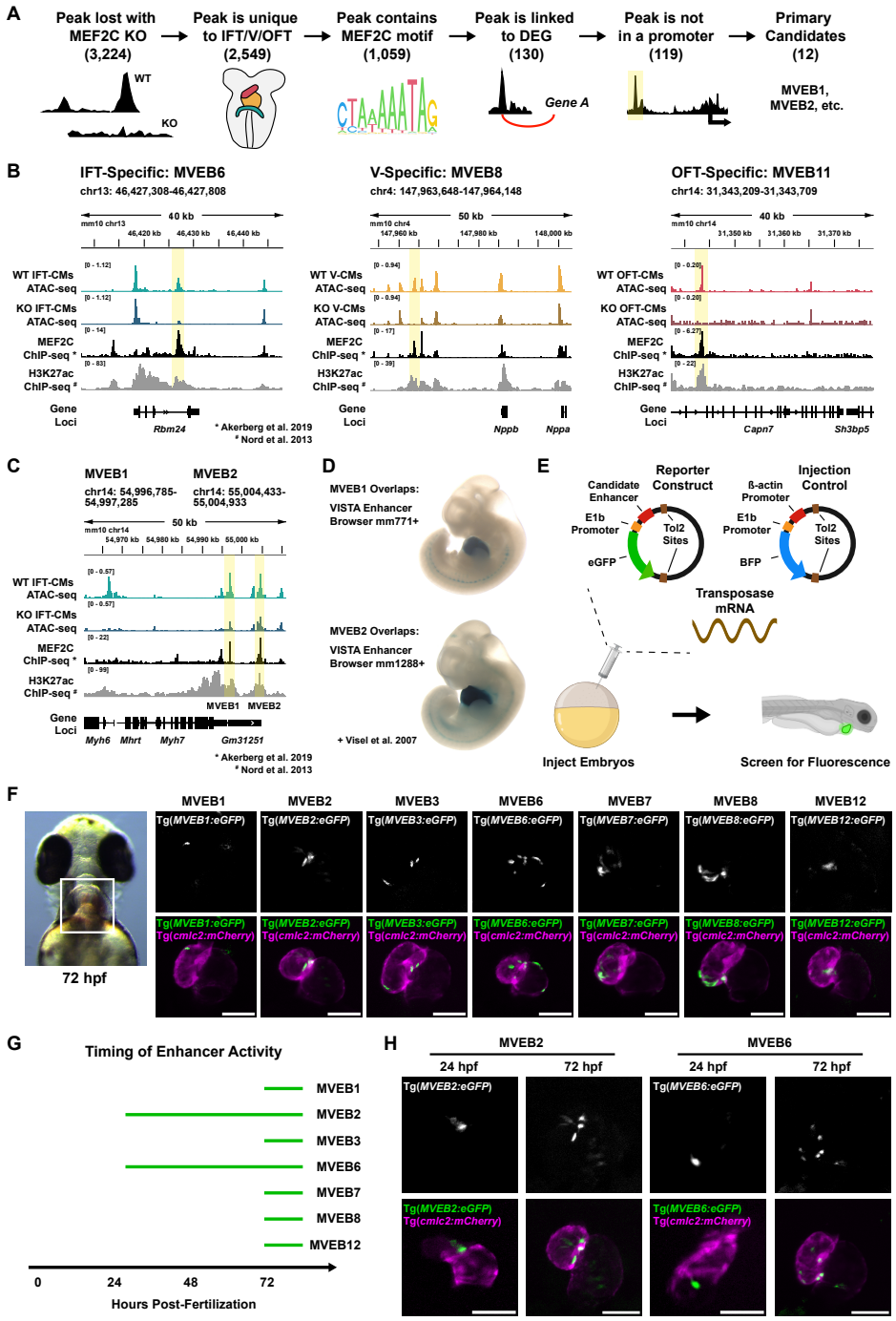
987

988 **Figure 4: Pseudotime analysis with MIRA reveals distinct MEF2C-depedent**
989 **developmental trajectories for each heart tube segment.**

990 A) Schematic of multimodal models for integrated regulatory analysis (MIRA) pipeline. B)
991 UMAP of outflow tract lineage cells plotted by MIRA topic models labeled by cell type (top)

992 and genotype (bottom). C) Pseudotime plot of outflow tract lineage cells. D) Lineage
993 trajectory stream plots for outflow tract lineage cells labeled by cell type (top) and
994 genotype (bottom). E) Stream plots displaying the flow of gene expression (top) and
995 chromatin accessibility (bottom) topics in the outflow tract lineage cells. Examples of the
996 top genes and TF binding motifs for dynamic topics are labeled. F) UMAP of vericular
997 lineage cells plotted by MIRA topic models labeled by cell type (top) and genotype
998 (bottom). G) Pseudotime plot of ventricular lineage cells. H) Lineage trajectory stream
999 plots for ventricular lineage cells labeled by cell type (top) and genotype (bottom). E)
1000 Stream plots displaying the flow of gene expression (top) and chromatin accessibility
1001 (bottom) topics in the ventricular lineage cells. Examples of the top genes and TF binding
1002 motifs for dynamic topics are labeled. J) UMAP of inflow tract lineage cells plotted by
1003 MIRA topic models labeled by cell type (top) and genotype (bottom). K) Pseudotime plot
1004 of inflow tract lineage cells. L) Lineage trajectory stream plots for inflow tract lineage cells
1005 labeled by cell type (top) and genotype (bottom). M) Stream plots displaying the flow of
1006 gene expression (top) and chromatin accessibility (bottom) topics in the inflow tract
1007 lineage cells. Examples of the top genes and TF binding motifs for dynamic topics are
1008 labeled.

1009 **Figure 5**



1010

1011

1012 **Figure 5: Leveraging multiomic data to identify novel enhancers via zebrafish**

1013 **transgenesis.**

1014 A) Schematic of selection process to identify candidate MEFC-dependent enhancers from

1015 integrated snRNA-seq and snATAC-seq data. B) Genome browser tracks displaying

1016 snATAC-seq accessibility profiles, MEF2C ChIP-seq occupancy profiles[26], and H3K27ac
1017 ChIP-seq occupancy profiles[32] at example loci containing IFT-specific (left), V-specific
1018 (middle), and OFT-specific (right) candidate enhancers (yellow highlights). C) Genome
1019 browser tracks displaying snATAC-seq accessibility profiles, MEF2C ChIP-seq
1020 occupancy profiles[26], and H3K27ac ChIP-seq occupancy profiles[32] at the *Myh6*/*Myh7*
1021 locus, which contains two candidate enhancers (yellow highlights, MVEB1 and MVEB2)
1022 with IFT-specific altered accessibility that overlap with regions found in the VISTA
1023 Enhancer Browser database[34]. D) Images of E11.5 mouse embryos from the Vista
1024 Enhancer Browser database[34] demonstrating positive enhancer activity of regions that
1025 overlap with candidates MVEB1 and MVEB2. E) Schematic of the Tol2 Transgenesis
1026 assay used to screen candidate enhancers in zebrafish. Elements of this panel were
1027 created in BioRender. B, B. (2024) https://BioRender.com/h83r503. F) Representative
1028 ventral view images of zebrafish embryos at 72 hpf injected with candidate enhancers
1029 that demonstrated positive activity in the heart. Boxed area in the representative
1030 brightfield image (left) indicates the anatomical region of interest captured in the
1031 fluorescent images. G) Schematic representation of the observed onset of enhancer
1032 activity for candidate enhancers that demonstrated positive activity in the heart. H)
1033 Representative ventral view images of zebrafish embryos at 24 and 72 hpf injected with
1034 MVEB2:eGFP or MVEB6:eGFP reporter constructs. Scale bars = 100 μm.

Figure 6

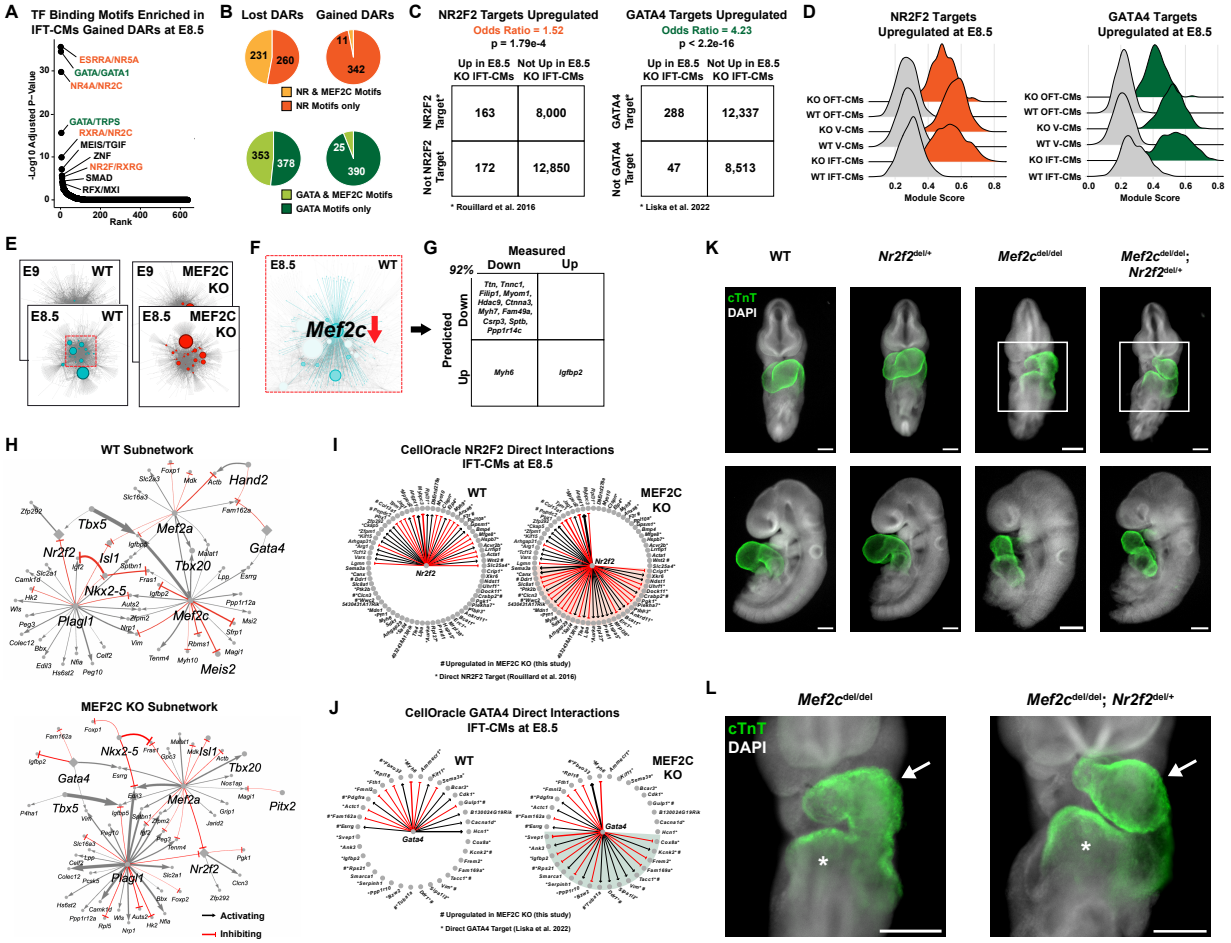


Figure 6: An overactive "posteriorized" GRN driven by NR2F2 and GATA4 underlies the expanded inflow tract phenotype in MEF2C KO.

A) TF binding motif enrichment analysis for gained DARs (MEF2C KO-vs-WT) in IFT-CMs at E8.5. B) Pie charts showing the proportion of gained or lost DARs (MEF2C KO-vs-WT) containing NR and MEF2C motifs or only NR motifs (top), and those containing GATA and MEF2C motifs or only GATA motifs (bottom). C) Odds ratio analysis for NR2F2 target genes[46] (left) and GATA4 target genes[47] (right) amongst DEGs up-regulated in MEF2C KO IFT-CMs. p-values were calculated using Fisher's exact test. D) Ridge plots displaying module scores for up-regulated NR2F2 targets (left) and GATA4 targets (right) in MEF2C KO and WT heart tube segments. E) Inferred GRNs constructed for MEF2C KO and WT IFT-CMs at E8.5 and E9. F) Schematic of GRN validation by predicting gene expression

with *Mef2c* knockdown in the E8.5 WT network and comparing to measured gene expression changes at E9. G) Results of GRN validation displaying predicted and measured gene expression changes. H) Visualizations of cardiac subnetworks within the WT and MEF2C KO E8.5 IFT-CM GRNs. I-J) Visualizations of direct NR2F2 (I) and GATA4 (J) interactions in the WT and MEF2C KO E8.5 IFT-CM GRNs. Direct interactions that occur upon MEF2C KO are highlighted. #, up-regulated DEG in MEF2C KO IFT-CMs at E8.5; *, Direct target of NR2F2 or GATA4[46,47]. K-L) Immunofluorescent staining of cardiac Troponin T (cTnT, green) in representative E9.5 embryos (18-24 somites) collected from *Mef2c*$^{del/+}$;*Nr2f2*$^{del/+}$ to *Mef2c*$^{del/+}$ crosses. Boxed regions in (K) are shown at higher magnification in (L). Arrows indicate more expanded ventricle phenotype in *Mef2c*$^{del/del}$;*Nr2f2*$^{del/+}$ embryos. Asterisks indicate better developed and looping atria in *Mef2c*$^{del/del}$;*Nr2f2*$^{del/+}$ embryos. n=5 *Mef2c*$^{del/del}$ and n=4 *Mef2c*$^{del/del}$;*Nr2f2*$^{del/+}$ embryos from 5 independent litters. Scale bars = 200 μm.

1062 **REFERENCES**

1063 1. Bruneau BG. Signaling and Transcriptional Networks in Heart Development and
1064 Regeneration. Cold Spring Harb Perspect Biol. 2013;5(3):a008292.

1065 2. Cohen M, Briscoe J, Blassberg R. Morphogen interpretation: the transcriptional logic
1066 of neural tube patterning. Curr Opin Genet Dev. 2013;23(4):423–8.

1067 3. McQueen C, Towers M. Establishing the pattern of the vertebrate limb. Development.
1068 2020;147(17):dev177956.

1069 4. Hafner A, Boettiger A. The spatial organization of transcriptional control. Nat Rev
1070 Genet. 2023;24(1):53–68.

1071 5. Wagh K, Stavreva DA, Upadhyaya A, Hager GL. Transcription Factor Dynamics: One
1072 Molecule at a Time. Annu Rev Cell Dev Biol. 2023;39(1):277–305.

1073 6. Spitz F, Furlong EEM. Transcription factors: from enhancer binding to developmental
1074 control. Nat Rev Genet. 2012;13(9):613–26.

1075 7. Lee TI, Young RA. Transcriptional Regulation and Its Misregulation in Disease. Cell.
1076 2013;152(6):1237–51.

1077 8. Mercatelli D, Scalambra L, Triboli L, Ray F, Giorgi FM. Gene regulatory network
1078 inference resources: A practical overview. Biochim Biophys Acta (BBA) - Gene Regul
1079 Mech. 2020;1863(6):194430.

1080 9. Olson EN. Gene Regulatory Networks in the Evolution and Development of the Heart.
1081 Science. 2006;313(5795):1922–7.

1082 10. Kelly RG, Buckingham ME, Moorman AF. Heart Fields and Cardiac Morphogenesis.
1083 Cold Spring Harb Perspect Med. 2014;4(10):a015750.

1084 11. Evans SM, Yelon D, Conlon FL, Kirby ML. Myocardial Lineage Development. Circ
1085 Res. 2010;107(12):1428–44.

1086 12. Parameswaran M, Tam PPL. Regionalisation of cell fate and morphogenetic
1087 movement of the mesoderm during mouse gastrulation. Dev Genet. 1995;17(1):16–28.

1088 13. Lopez-Sanchez C, Garcia-Martinez V, Schoenwolf GC. Localization of Cells of the
1089 Prospective Neural Plate, Heart and Somites within the Primitive Streak and Epiblast of
1090 Avian Embryos at Intermediate Primitive-Streak Stages. Cells Tissues Organs.
1091 2001;169(4):334–46.

1092 14. Stalsberg H, DeHaan RL. The precardiac areas and formation of the tubular heart in
1093 the chick embryo. Dev Biol. 1969;19(2):128–59.

1094 15. DeRuiter MC, Poelmann RE, Vries IV de, Mentink MMT, Groot ACG de. The
1095 development of the myocardium and endocardium in mouse embryos. Anat Embryol.
1096 1992;185(5):461–73.

1097 16. Lescroart F, Chabab S, Lin X, Rulands S, Paulissen C, Rodolosse A, et al. Early
1098 lineage restriction in temporally distinct populations of Mesp1 progenitors during
1099 mammalian heart development. Nat Cell Biol. 2014;16(9):829–40.

1100 17. Devine WP, Wythe JD, George M, Koshiba-Takeuchi K, Bruneau BG. Early patterning
1101 and specification of cardiac progenitors in gastrulating mesoderm. eLife. 2014;3:e03848.

1102 18. Meilhac SM, Kelly RG, Rocancourt D, Eloy-Trinquet S, Nicolas JF, Buckingham ME.
1103 A retrospective clonal analysis of the myocardium reveals two phases of clonal growth in
1104 the developing mouse heart. Development. 2003;130(16):3877–89.

1105 19. Lin Q, Schwarz J, Bucana C, Olson EN. Control of Mouse Cardiac Morphogenesis
1106 and Myogenesis by Transcription Factor MEF2C. Science. 1997;276(5317):1404–7.

1107 20. Krup AL, Winchester SAB, Ranade SS, Agrawal A, Devine WP, Sinha T, et al. A
1108 Mesp1-dependent developmental breakpoint in transcriptional and epigenomic
1109 specification of early cardiac precursors. Development. 2023;150(9):dev201229.

1110 21. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated
1111 analysis of multimodal single-cell data. Cell. 2021;184(13):3573-3587.e29.

1112 22. Choudhary S, Satija R. Comparison and evaluation of statistical error models for
1113 scRNA-seq. Genome Biol. 2022;23(1):27.

1114 23. Tyser RCV, Ibarra-Soria X, McDole K, Jayaram SA, Godwin J, Brand TAH van den,
1115 et al. Characterization of a common progenitor pool of the epicardium and myocardium.
1116 Science. 2021;371(6533).

1117 24. Black BL, Olson EN. TRANSCRIPTIONAL CONTROL OF MUSCLE DEVELOPMENT
1118 BY MYOCYTE ENHANCER FACTOR-2 (MEF2) PROTEINS. Annu Rev Cell Dev Biol.
1119 1998;14(1):167–96.

1120 25. Yang A, Alankarage D, Cuny H, Ip EKK, Almog M, Lu J, et al. CHDgene: A Curated
1121 Database for Congenital Heart Disease Genes. Circ: Genom Precis Med.
1122 2022;15(3):e003539.

1123 26. Akerberg BN, Gu F, VanDusen NJ, Zhang X, Dong R, Li K, et al. A reference map of
1124 murine cardiac transcription factor chromatin occupancy identifies dynamic and
1125 conserved enhancers. Nat Commun. 2019;10(1):4907.

1126 27. Barnes RM, Harris IS, Jaehnig EJ, Sauls K, Sinha T, Rojas A, et al. MEF2C regulates
1127 outflow tract alignment and transcriptional control of Tdgf1. Development.
1128 2016;143(5):774–9.

1129 28. Granja JM, Corces MR, Pierce SE, Bagdatli ST, Choudhry H, Chang HY, et al. ArchR
1130 is a scalable software package for integrative single-cell chromatin accessibility analysis.
1131 Nat Genet. 2021;53(3):403–11.

1132 29. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-
1133 based Analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.

1134 30. Lynch AW, Theodoris CV, Long HW, Brown M, Liu XS, Meyer CA. MIRA: joint
1135 regulatory modeling of multimodal expression and chromatin accessibility in single cells.
1136 Nat Methods. 2022;19(9):1097–108.

1137 31. Zhou W, Lin L, Majumdar A, Li X, Zhang X, Liu W, et al. Modulation of morphogenesis
1138 by noncanonical Wnt signaling requires ATF/CREB family–mediated transcriptional
1139 activation of TGFβ2. Nat Genet. 2007;39(10):1225–34.

1140 32. Nord AS, Blow MJ, Attanasio C, Akiyama JA, Holt A, Hosseini R, et al. Rapid and
1141 Pervasive Changes in Genome-wide Enhancer Usage during Mammalian Development.
1142 Cell. 2013;155(7):1521–31.

1143 33. Man JCK, Duijvenboden K van, Krijger PHL, Hooijkaas IB, Made I van der, Vries C
1144 de G de, et al. Genetic Dissection of a Super Enhancer Controlling the Nppa-Nppb Cluster
1145 in the Heart. Circ Res. 2021;128(1):115–29.

1146 34. Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA Enhancer Browser—a
1147 database of tissue-specific human enhancers. Nucleic Acids Res.
1148 2007;35(suppl_1):D88–92.

1149 35. Birnbaum RY, Clowney EJ, Agamy O, Kim MJ, Zhao J, Yamanaka T, et al. Coding
1150 exons function as tissue-specific enhancers of nearby genes. Genome Res.
1151 2012;22(6):1059–68.

1152 36. Fisher S, Grice EA, Vinton RM, Bessling SL, Urasaki A, Kawakami K, et al. Evaluating
1153 the biological relevance of putative enhancers using Tol2 transposon-mediated
1154 transgenesis in zebrafish. Nat Protoc. 2006;1(3):1297–305.

1155 37. Hinits Y, Pan L, Walker C, Dowd J, Moens CB, Hughes SM. Zebrafish Mef2ca and
1156 Mef2cb are essential for both first and second heart field cardiomyocyte differentiation.
1157 Dev Biol. 2012;369(2):199–210.

1158 38. Zlatanova I, Sun F, Wu RS, Chen X, Lau BH, Colombier P, et al. An injury-responsive
1159 mmp14b enhancer is required for heart regeneration. Sci Adv. 2023;9(48):eadh5313.

1160 39. Yan R, Cigliola V, Oonk KA, Petrover Z, DeLuca S, Wolfson DW, et al. An enhancer-
1161 based gene-therapy strategy for spatiotemporal control of cargoes during tissue repair.
1162 Cell Stem Cell. 2023;30(1):96-111.e6.

1163 40. Kang J, Hu J, Karra R, Dickson AL, Tornini VA, Nachtrab G, et al. Modulation of tissue
1164 repair by regeneration enhancer elements. Nature. 2016;532(7598):201–6.

1165 41. Tremblay M, Sanchez-Ferras O, Bouchard M. GATA transcription factors in
1166 development and disease. Development. 2018;145(20):dev164384.

1167 42. Wu S pin, Cheng CM, Lanz RB, Wang T, Respress JL, Ather S, et al. Atrial Identity Is
1168 Determined by a COUP-TFII Regulatory Network. Dev Cell. 2013;25(4):417–26.

1169 43. Pereira FA, Qiu Y, Zhou G, Tsai MJ, Tsai SY. The orphan nuclear receptor COUP-
1170 TFII is required for angiogenesis and heart development. Genes Dev. 1999;13(8):1037–
1171 49.

1172 44. Community TG, Abueg LAL, Afgan E, Allart O, Awan AH, Bacon WA, et al. The Galaxy
1173 platform for accessible, reproducible, and collaborative data analyses: 2024 update.
1174 Nucleic Acids Res. 2024;52(W1):W83–94.

1175 45. He A, Gu F, Hu Y, Ma Q, Ye LY, Akiyama JA, et al. Dynamic GATA4 enhancers shape
1176 the chromatin landscape central to heart development and disease. Nat Commun.
1177 2014;5(1):4907.

1178 46. Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG,
1179 et al. The harmonizome: a collection of processed datasets gathered to serve and mine
1180 knowledge about genes and proteins. Database. 2016;2016(2016):baw100.

1181 47. Liska O, Bohár B, Hidas A, Korcsmáros T, Papp B, Fazekas D, et al. TFLink: an
1182 integrated gateway to access transcription factor–target gene interactions for multiple
1183 species. Database. 2022;2022:baac083.

1184 48. Tirosh I, Izar B, Prakadan SM, II MHW, Treacy D, Trombetta JJ, et al. Dissecting the
1185 multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science.
1186 2016;352(6282):189–96.

1187 49. Kamimoto K, Stringa B, Hoffmann CM, Jindal K, Solnica-Krezel L, Morris SA.
1188 Dissecting cell identity via network inference and in silico gene perturbation. Nature.
1189 2023;614(7949):742–51.

1190 50. Molkentin JD, Black BL, Martin JF, Olson EN. Cooperative activation of muscle gene
1191 expression by MEF2 and myogenic bHLH proteins. Cell. 1995;83(7):1125–36.

1192 51. Di-Luoffo M, Brousseau C, Tremblay JJ. MEF2 and NR2F2 cooperate to regulate
1193 Akr1c14 gene expression in mouse MA-10 Leydig cells. Andrology. 2016;4(2):335–44.

1194 52. Molkentin JD, Black BL, Martin JF, Olson EN. Mutational Analysis of the DNA Binding,
1195 Dimerization, and Transcriptional Activation Domains of MEF2C. Mol Cell Biol.
1196 1996;16(6):2627–36.

1197 53. Pereira FA, Tsai MJ, Tsai* SY. COUP-TF orphan nuclear receptors in development
1198 and differentiation. Cell Mol Life Sci CMLS. 2000;57(10):1388–98.

1199 54. Takamoto N, You LR, Moses K, Chiang C, Zimmer WE, Schwartz RJ, et al. COUP-
1200 TFII is essential for radial and anteroposterior patterning of the stomach. Development.
1201 2005;132(9):2179–89.

1202 55. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive
1203 and accurate integration of single-cell data with Harmony. Nat Methods.
1204 2019;16(12):1289–96.

1205 56. Vierstra J, Lazar J, Sandstrom R, Halow J, Lee K, Bates D, et al. Global reference
1206 mapping of human transcription factor footprints. Nature. 2020;583(7818):729–36.

1207 57. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al.
1208 Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24–6.

1209 58. Cakir B, Prete M, Huang N, Dongen S van, Pir P, Kiselev VY. Comparison of
1210 visualization tools for single-cell RNAseq data. NAR Genom Bioinform.
1211 2020;2(3):lqaa052.

1212 59. Zhang R, Han P, Yang H, Ouyang K, Lee D, Lin YF, et al. In vivo cardiac
1213 reprogramming contributes to zebrafish heart regeneration. Nature.
1214 2013;498(7455):497–501.

1215 60. Zhao H, Sun Z, Wang J, Huang H, Kocher JP, Wang L. CrossMap: a versatile tool for
1216 coordinate conversion between genome assemblies. Bioinformatics. 2014;30(7):1006–7.

1217 61. Raney BJ, Barber GP, Benet-Pagès A, Casper J, Clawson H, Cline MS, et al. The
1218 UCSC Genome Browser database: 2024 update. Nucleic Acids Res.
1219 2023;52(D1):D1082–8.

1220 62. Pliner HA, Packer JS, McFaline-Figueroa JL, Cusanovich DA, Daza RM, Aghamirzaie
1221 D, et al. Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin
1222 Accessibility Data. Mol Cell. 2018;71(5):858-871.e8.

1223 63. Bruse N, Heeringen SJ van. GimmeMotifs: an analysis framework for transcription
1224 factor motif analysis. bioRxiv. 2018;474403.

1225 64. Bastian M, Heymann S, Jacomy M. Gephi: An Open Source Software for Exploring
1226 and Manipulating Networks. ICWSM. 2009;3:361–2.

1227 65. Perrone G, Unpingco J, Lu H minn. Network visualizations with Pyvis and VisJS.
1228 arXiv. 2020;

1229 66. Hagberg AA, Schult DA, Swart PJ. Exploring Network Structure, Dynamics, and
1230 Function using NetworkX. In: Varoquaux G, Vaught T, Millman J, editors. Proceedings of
1231 the 7th Python in Science Conference (SciPy 2008). 2008. p. 11–6.