Single cell multi-omic analysis identifies a \textit{Tbx1}-dependent multilineage primed population in murine cardiopharyngeal mesoderm

Hiroko Nomaru\textsuperscript{1}, Yang Liu\textsuperscript{1,10}, Christopher De Bono\textsuperscript{1,10}, Dario Righelli\textsuperscript{2,3}, Andrea Cirino\textsuperscript{4,5}, Wei Wang\textsuperscript{6}, Hansoo Song\textsuperscript{1}, Silvia E. Racedo\textsuperscript{1}, Anelisa G. Dantas\textsuperscript{1,7}, Lu Zhang\textsuperscript{8}, Chen-Leng Cai\textsuperscript{8}, Claudia Angelini\textsuperscript{2}, Lionel Christiaen\textsuperscript{6}, Robert G. Kelly\textsuperscript{9}, Antonio Baldini\textsuperscript{4,5}, Deyou Zheng\textsuperscript{1} & Bernice E. Morrow\textsuperscript{1}

The poles of the heart and branchiomerical muscles of the face and neck are formed from the cardiopharyngeal mesoderm within the pharyngeal apparatus. They are disrupted in patients with 22q11.2 deletion syndrome, due to haploinsufficiency of \textit{TBX1}, encoding a T-box transcription factor. Here, using single cell RNA-sequencing, we now identify a multilineage primed population within the cardiopharyngeal mesoderm, marked by \textit{Tbx1}, which has bipotent properties to form cardiac and branchiomerical muscle cells. The multilineage primed cells are localized within the nascent mesoderm of the caudal lateral pharyngeal apparatus and provide a continuous source of cardiopharyngeal mesoderm progenitors. \textit{Tbx1} regulates the maturation of multilineage primed progenitor cells to cardiopharyngeal mesoderm derivatives while restricting ectopic non-mesodermal gene expression. We further show that \textit{TBX1} confers this balance of gene expression by direct and indirect regulation of enriched genes in multilineage primed progenitors and downstream pathways, partly through altering chromatin accessibility, the perturbation of which can lead to congenital defects in individuals with 22q11.2 deletion syndrome.
The heart develops from two successive waves of mesodermal progenitor cells during early embryogenesis. The first heart field (FHF) constitutes the first wave of cardiac progenitors and results in the primitive beating heart, while the second heart field (SHF) forms the second wave that builds upon the two poles of the heart. The SHF can be anatomically partitioned to the anterior SHF (aSHF) and posterior SHF (pSHF), whose cells migrate to the heart via the dorsal pericardial wall to the outflow tract or inflow tract, respectively. Expression of Mesp1 at gastrulation marks the earliest mesodermal cells that will form the heart. Using single-cell RNA-sequencing (scRNA-seq) of the Mesp1 lineage, it was discovered that the FHF, aSHF, and pSHF are specified at gastrulation.

Retrospective clonal analysis and lineage tracing studies revealed that the branchiomesodermal skeletal muscles (BrM) of the craniofacial region and neck share a clonal relationship with the SHF. The bipotential nature of these cardiac and skeletal muscle progenitor cells is supported by studies in the ascidian, Ciona, an invertebrate chordate, in which single cells give rise to both cardiac and skeletal muscle cells. When taken together, a new term, cardiopharyngeal mesoderm (CPM), was introduced to clearly include both SHF cardiac and skeletal muscle progenitor populations. A cartoon of these populations in the mouse embryo is shown in Fig. 1a. The Tbx1 gene, encoding a T-box transcription factor, and gene haploinsufficient in 22q11.2 deletion syndrome (22q11.2DS), is expressed in the CPM and is required for cardiac outflow tract and BrM development, implicating its essential roles in the CPM.

A total of 60–75% of patients with 22q11.2DS have cardiac outflow tract defects, which often require life-saving surgery during the neonatal period. In addition, most individuals with this condition have speech, feeding, and swallowing difficulties in infancy, due in part to BrM hypotonia. Further, heterozygous mutations of Tbx1 in rare, non-deleted individuals, phenocopy symptoms of the deletion. Inactivation of both alleles of Tbx1 in the mouse results in a persistent truncus arteriosus and significant failure to form BrMs. Gene expression profiling of wild-type versus Tbx1 null mutant embryos identified genes that changed in expression but it was unclear whether the changes were autonomous in the CPM or in other cell populations, such as neural crest cells. Therefore, we do not yet understand the functions of Tbx1 on a single-cell level, which is needed to elucidate the true molecular pathogenesis of 22q11.2DS.

The CPM is distributed throughout the embryonic pharyngeal apparatus during early gestation. The pharyngeal apparatus consists of individual bulges of cells termed arches that form in a rostral to caudal manner from mouse embryonic days (E)8–10.5. The formation of individual arches is a highly dynamic but regulated process that requires sufficient cell populations to elongate the OPT and to form diverse BrMs. The cellular and molecular mechanisms of how CPM cells in the pharyngeal apparatus both are maintained in a progenitor state and are allocated to form the heart and BrMs in mammals are unknown.

To fill these gaps, we performed scRNA-seq of mesodermal cell lineages from the pharyngeal apparatus plus heart at multiple stages during embryogenesis. We discovered a multilineage primed progenitor (MLP) population within the CPM, which is maintained and gradually matures from E8-10.5 and has differentiation branches toward cardiac and skeletal muscle fates, serving as common lineage progenitors. MLP cells are localized to the nascent lateral mesoderm of the pharyngeal apparatus, deploying cells to the heart and BrMs. We found that the Tbx1 cell lineage marks MLPs and Tbx1 activity is critical for their function. Inactivation of Tbx1 disrupts a MLP gene expression program needed for differentiation and results in ectopic expression of non-mesoderm genes. We further identify the gene regulatory network downstream of Tbx1 in the MLPs providing insights into the molecular mechanism of mammalian CPM function, essential for understanding the etiology of 22q11.2DS.

### Results

**Identification of common progenitor cells in the CPM.** To identify the various populations that constitute the CPM (Fig. 1a), we performed droplet-based scRNA-seq on fluorescence-activated cell sorted GFP expressing cells from microdissected Mesp1Cre;ROSA26-GFPf/+ (f, flox) embryos at E8.0, E8.25, E9.5, and E10.5 (Table 1, Fig. 1b, Supplementary Figs. 1a, b, 2a, b, 3, and Supplementary Table 1). These stages were chosen because they are the critical periods when Tbx1 is expressed, reaching the highest expression at E9.5, and when the pharyngeal apparatus is dynamically elongating; this is coordinated with heart development and BrM specification. To better understand the developmental sequence of events, we integrated the four time point datasets (Fig. 1c) and identified 20 cell clusters (Fig. 1d). Utilizing knowledge of the expression of known marker genes in each cluster (Fig. 1e, Supplementary Fig. 2c–e, and Supplementary Data 1), we identified cell types of all the clusters, half of which include cardiovascular progenitor cell populations or their derivatives (Fig. 1d, bold font).

We identified CPM clusters (C1, C3, C9, C15, and C18) with marker gene expression for the CPM including Tbx1, Isl1, Wnt5a, and Tcf21, among other genes (Fig. 1f, g and Supplementary Fig. 2e). Cluster C9 contains BrM progenitor cells identified by expression of Tcf21, Lhx2, and Myf5 (Fig. 1f and Supplementary Fig. 2e). Clusters C1 and C18 contain pSHF populations as identified by expression of Hoxb13, Tbx5, Foxf1, and Wnt23,24 (Fig. 1f and Supplementary Fig. 2e). Many of the pSHF cells are located more medially and caudally in the embryo and contribute to posterior organ development, such as the formation of the lung. The cells in C3 express Nkk2.5 and Mef2c and cardiac structural protein genes (Tmnt2, Tmnt1, Myh4 (Fig. 1f and Supplementary Fig. 2e)). In addition, subdomains express either Fgf or aSHF genes (Tbx5 but not Isl1 or Tcf21) or aSHF genes (Fig. 1f and Supplementary Fig. 2e). We discovered that cluster C15 expresses genes shared by CPM clusters, including Tbx1, Isl1, Wnt5a, Mef2c, Tcf21, and Foxf1 (Fig. 1f and Supplementary Fig. 2e). Based on this, we suggest that C15 as a multilineage primed progenitor (MLP) population within the CPM.

**Multilineage primed progenitors of the CPM differentiate into cardiac and skeletal muscles.** We next investigated the relationship between the MLPs and more differentiated CPM cells using partition-based graph abstraction (PAGA). Several clusters that are not part of the CPM and were already well separated in the above cluster analysis (C4, C5, C7, C13, and C17), were excluded from PAGA analysis (Supplementary Fig. 2d). The PAGA analysis partitioned the CPM cells into five branches (Fig. 2a and Supplementary Fig. 4a), connecting all mesodermal cell populations. The five branches include cardiomyocytes (CMs), pSHF with lung progenitor cells (pSHF, Lung PC), connective tissue (CT), branchiomesodermal muscle progenitor cells (BrM), and skeleton/limb (Sk/L). Convergent results from pseudotime analysis (Fig. 2b) and real-time point information (Fig. 2c and Supplementary Fig. 4b), infer that MLP cells (C15) in the center are in a progenitor state of the CPM while more differentiated cells are toward the outside in each branch (BrM, C9; CMs with aSHF, C3; part of the pSHF, C1 + C18; Fig. 2d and Supplementary Fig. 4c). The MLP cells exist not only in early time points (E8.0, E8.25) but also in later time points (E9.5, E10.5), but with fewer cells at E10.5 (Fig. 2e).
Based on the specific expression of genes (Supplementary Data 1) and distribution of gene expression patterns in the PAGA plot (Fig. 2f and Supplementary Fig. 4d–f), we focused on two marker genes for MLPs, **Aplnr** (Apelin receptor) and **Nrg1** (Neuregulin 1). **Aplnr** is expressed in the CPM27 but not known for the specific presence in MLPs, while **Nrg1** is not known to be a CPM gene, but it is required in the embryonic heart for the development of the chamber myocardium28. We examined the co-expression of genes in the cells in CPM populations. The heatmap in Fig. 2g (Supplementary Fig. 4e, f) shows the expression of genes enriched in MLPs, with the same genes also expressed in more differentiated CPM populations (BrM, CMs, and pSHF), but at different expression levels, indicating that they are multilineage primed. Taken together, these results further support the progenitor status of MLPs.
Fig. 1 Single-cell analysis of Mesp1+ lineages at E8.0 to E10.5 identifies CPM lineages. a Cartoon of an E9.5 embryo in a right lateral view (left, CPM, gray). The CPM includes the branchiomatic muscle progenitor cells (BrM; aqua), anterior SHF (aSHF; red), and part of the posterior SHF (pSHF; blue). Because cardiomyocyte progenitors (CMs; green) are from both the HHF and CPM, only CPM-derived CMs are included. b Whole-embryo images with GFP fluorescence of Mesp1Cre, ROSA26-GFPf/+ embryos at E8.0, E8.25, E9.5, and E10.5 used for scRNA-seq. The white dotted line represents the region that was dissected. Only GFP-positive cells were used for scRNA-seq. H head, T tail. Scale bar E8.0, E8.25 is 750 µm; E9.5 and E10.5 is 500 µm. c t-distributed stochastic neighbor embedding (tSNE) plot colored by developmental stage. d tSNE plot colored by clusters (C1-20). Cardiac relevant clusters are in bold font. C1: pSHFcl, posterior CPM includes cardiac and lung progenitor cells; C2: MCs, mesenchyme expressing epithelial-mesenchymal transition markers; C3: CMs cardiomyocyte progenitor cells; C4: Endo endocardium and endothelial cells; C5: Blood blood cells; C6: MCs mesenchyme expressing epithelial-mesenchymal transition markers; C7: Vendo vascular endothelial cells; C8: Sk/L skeleton/limb progenitor cells; C9: BrM branchiomeric muscle progenitor cells; C10: ST septum transversum progenitor cells; C11: PEO proepicardial organ; C12: CT connective tissue progenitor cells; C13: endotho endothelial cells; C14: Lung PC lung progenitor cells; C15: MLP multilineage progenitor cells; C16: CT connective tissue progenitor cells; C17: Vendo vascular endothelial cells; C18: pSHF posterior CPM of cardiac progenitor cells; C19: Endothelial cells; C20: NPG neural progenitor cells. e Heatmap of average gene expression of marker genes in representative clusters. Dot size indicates the fraction of cells expressing the genes in each cluster and color indicates scaled mean expression (bar below; dark red is the strongest expression). f, g tSNE plots showing expression of CPM marker genes (f) and Tbx1 (g). The color spectrum from gray to red indicates expression levels from low to high. h tSNE plots of CM progenitors showing the expression level of CM (Tbx5) and CPM (Isl1) marker genes.

| Time point | No. of embryos | Somites | Cell viability | Cell number (captured) | Mean reads/cell | Median genes/cell | Tissue |
|------------|----------------|---------|----------------|------------------------|----------------|------------------|--------|
| Mesp1Cre; ROSA26-GFPf/+ | | | | | | | |
| E8.0 | 7 | 3–4 | 90% | 7581 | 58,422 | 4380 | Half of the body |
| E8.25 | 5 | 6–7 | 95% | 2956 | 151,357 | 5363 | Half of the body |
| E9.5 | 6 | 19–22 | 80% | 4055 | 80,919 | 4036 | PA and heart |
| E9.5 | 2 | 21–22 | 81% | 7418 | 60,500 | 4106 | PA and heart |
| E10.5 | 4 | 30–31 | 80% | 8070 | 52,326 | 3673 | PA and heart |
| Mesp1Cre; ROSA26-GFPf/+; Tbx1f/+ | | | | | | | |
| E9.5 | 1 | 22 | 83% | 5157 | 65,997 | 5015 | PA and heart |
| E9.5 | 2 | 24 | 92% | 10,281 | 43,868 | 3617 | PA and heart |
| Tbx1Cre/−; ROSA26-GFPf/+ | | | | | | | |
| E8.5 | 7 | 8–9 | 88% | 1434 | 136,347 | 3683 | Half of the body |
| E9.5 | 10 | 19–21 | 87% | 5453 | 54,317 | 3994 | PA and heart with back |
| Tbx1Cre/−; ROSA26-GFPf/+ | | | | | | | |
| E9.5 | 4 | 9–10 | 92% | 1109 | 140,352 | 4374 | Half of the body |
| E9.5 | 3 | 19–22 | 97% | 6267 | 54,892 | 3561 | PA and heart with back |
| Tbx1+/− vs wild-type | | | | | | | |
| E9.5 KO WT | 4 | 22–23 | 89% | 12,050 | 37,218 | 4380 | PA and heart with back |
| E9.5 KO | 4 | 20–23 | 92% | 11,504 | 37,646 | 4138 | PA and heart with back |

The MLPs are bilaterally localized to the caudal pharyngeal apparatus. To elucidate whether MLPs are localized within a defined embryonic region in the pharyngeal apparatus, we performed RNAscope in situ hybridization analysis using probes for genes enriched in MLPs including Tbx1, Isl1, Aplnr, and Nrg1 with the Mesp1+ lineage marked by EGFP expression (Fig. 2h–m). The most caudal and lateral mesoderm is the least differentiated, while the most rostral and lateral mesoderm is the least (Fig. 2h). The pharyngeal arches form in a rostral to caudal manner where the most caudal and lateral mesoderm is the least differentiated, while the most rostral mesoderm has already migrated to the rostral mesoderm has already migrated to the cranial pharyngeal apparatus, while they deploy cells rostrally, medially, and dorsally thereby explaining in part the mechanism for the extension of the pharyngeal apparatus caudally (Fig. 2n). MLPs dynamically transition over time. An important question is whether MLPs as CPM progenitors, maintain the same state of gene expression over time. To address this, we examined differentially expressed genes in MLPs from E8-10.5. We identified core CPM genes that are expressed similarly at all time points, including Isl1, Mef2c, and Nkx2-5 (Fig. 3a). However, we also found that early expressing genes such as Aplnr, Nrg1l, Irx1-5, Fgf8/10, and Tbx1 (Fig. 3b) are reduced over time, with increasing expression of cardiac developmental genes such as Hand2, Gata3/5/6, Bmp4, and Sema3c (Fig. 3c). These gradient differences are also visualized in violin plots (Fig. 3d–f). Nkx2-5 and Sema3c are expressed in the caudal pharyngeal apparatus at E9.5, like that of Isl1, Tbx1, and Aplnr, defining MLPs (Fig. 3g). The MLP region is reduced in size in the caudal pharyngeal apparatus, pharyngeal arches 3–6, at E10.5 (Fig. 3g). Further, expression of the early MLP marker Aplnr is not observed at E10.5, while co-expression of Sema3c, Nkx2-5, and Isl1 occur strongly in the OFT (Fig. 3g). This is consistent with the model that the MLPs continuously allocate progenitor cells to BrMs and OFT-CMs, while showing a gradual maturation themselves by E10.5 (Fig. 3h).
The intersection of the *Tbx1* and *Mesp1* lineages helps to identify the CPM. *Tbx1* function in MLPs and roles in derivative CPM cells on a single-cell level are unknown. We therefore examined the *Mesp1* and *Tbx1* lineages in control embryos to understand how the CPM lineages compare in relation to *Tbx1*. The *Mesp1* lineage contributes more broadly to the embryonic mesoderm, while *Tbx1* is expressed in pharyngeal endoderm and distal pharyngeal ectoderm, in addition to the CPM. Although *Tbx1* is strongly expressed in the CPM, it is not expressed in the heart, neither in the FHF nor the caudal and medial pSHF at the timepoints analyzed. The intersection of these two datasets defines the CPM more precisely.

The *Tbx1Cre* allele has a knockin of Cre that inactivates one copy of the *Tbx1* gene. The *Tbx1Cre* mice do not have cardiac or
aortic arch defects in the Swiss Webster background36. We then performed scRNA-seq of the Tbx1Cre lineage at E9.5 and integrated this with data from the Mesp1Cre lineage at the same stage to compare the characteristics of the two lineages (Fig. 4a–c and Supplementary Data 2). Data integration provides consistency in defining common cell types among different samples, in addition to removing batch effects37. We found that the CPM can be identified in both populations (Fig. 4a, bold font). As expected, the Mesp1 lineage includes the FHF, which is not included in the Tbx1 lineage, whereas the Tbx1 lineage includes the pharyngeal epithelia and otic vesicle, not included in the Mesp1 lineage (Fig. 4b, c). The relative proportions of CPM populations are shown in Fig. 4d. The pSHF in the Mesp1Cre lineage includes the caudal pSHF with lung progenitors that is not included in the Tbx1Cre lineage33. The MLPs are found in both lineages, marked by the expression of Isl1, Aplnr, Tbx1, and Nrg1 (Fig. 4e). Therefore, the data from scRNA-seq using Tbx1Cre help define the CPM better and serves as a replication for the data on the CPM from scRNA-seq using Mesp1Cre, which is relevant to Tbx1 as shown in Fig. 4f.

Tbx1 regulates MLP development by promoting gene expression needed for differentiation and restricting the expression of non-mesodermal genes. Mesp1Cre and Tbx1Cre mediated Tbx1 conditional null embryos have similar phenotypes, including hypoplasia of the caudal pharyngeal apparatus and a fully penetrant persistent truncus arteriosus16,32. This supports the utility of both Cre lines in this study and the importance of the mesoderm domain of Tbx1 expression in mediating its function. We inactivated Tbx1 in the Mesp1 and Tbx1 lineages (Fig. 5a, b) and performed scRNA-seq to ascertain how its loss affects the CPM, using both lineages as a comparison and replication for each other (Table 1). We generated embryos that were Mesp1Cre; Tbx1+/− (Mesp1Cre Ctrl) vs Mesp1Cre; Tbx1−/− (Mesp1Cre cKO) at E9.5 (Supplementary Fig. 1a, b, Fig. 5a, and Table 1) and generated two biological replicates. We also generated Tbx1Cre+ (Tbx1Cre Ctrl) vs Tbx1Cre− (Tbx1Cre cKO) embryos at E8.5 and E9.5 (Supplementary Fig. 1c, d, Fig. 5b, and Table 1). The E8.5 stage used (8–10 somites) is only very slightly different from the E8.25 stage (6–7 somites) used for the Mesp1Cre experiment (Table 1). As observed, the ROSA26-GFPβgal reporter was used to purify the lineages (Supplementary Fig. 1).

With the conditional knockout data, we performed two integrated and clustering analyses; one for Mesp1Cre Ctrl vs Mesp1Cre cKO (two replicates; Supplementary Fig. 6 and Supplementary Data 3) and another for Tbx1Cre Ctrl vs Tbx1Cre cKO embryos (Supplementary Fig. 7 and Supplementary Data 4–5), and then again focusing on the CPM. With increased resolution afforded by the single time point experiments, we found that the aSHF population is continuous with the somatic mesoderm (SoM; Fig. 5c–f). The SoM is adjacent to the aSHF and in embryos and gives rise to the ventral pericardial wall (expressing Msx1, Msx2, Epha3, but not Nkx2-5; Nkx2-5 and Msx2 are shown in Fig. 5g, h).

The ratio of the number of cells compared to the total number of cells in each replicate of Mesp1Cre Ctrl versus cKO shows a modest variation (first replicate of WT (WT1) is 0.07 (284/4046 cells), WT2 is 0.129 (912/7048) vs first replicate of KO (KO1) is 0.101 (474/4689) and KO2 is 0.108 (1056/9750), nonetheless, MLPs are present in both (Fig. 5c–f). To understand how Tbx1 affects gene expression within the MLPs and derivative cell types leading to the observed phenotypes at later stages, we analyzed differentially expressed genes (DEGs) with the scRNA-seq datasets. We identified DEGs in each cluster in the two replicates of Mesp1Cre Ctrl vs cKO embryos (Supplementary Data 4) and separately DEGs in Tbx1Cre Ctrl vs Tbx1Cre cKO embryos at E9.5 (Supplementary Data 6). The DEGs in the two Mesp1Cre Ctrl vs cKO replicates overlapped significantly, and the final DEGs were from the pooled replicates. Furthermore, to focus on the most reproducible alterations, we examined only DEGs shared in both Mesp1 Cre replicates and between Mesp1 and Tbx1 scRNA-seq datasets at E9.5, which change in the same direction, with adjusted P value <0.05 and have an absolute log2-fold change >0.25 (Fig. 6a). In the MLPs, we identified 651 DEGs; 468 genes were decreased and 183 were increased in both Mesp1 and Tbx1 cKO embryos at E9.5 (Fig. 6b, Supplementary Data 2). Gene ontology (GO) analysis was used to identify enriched biological functions of the downregulated genes shared in both Mesp1 and Tbx1 cKO embryos (Supplementary Data 8). Genes affected in MLPs are involved in cell migration, organ development, and muscle development (Fig. 6c; e.g., Sox9, Mef2c, Grem1, Hey1, Bmp7, Fgf10, Foxj1, Wnt5a), Me2c, Hey1, Bmp7, Fgf10, and Wnt5a are required for normal cardiac development. Enriched GO terms for the upregulated genes included axonogenesis such as Bdnf or Pax6, inner ear development (placodal formation), and cell fate determination (Fig. 6d). Further, genes not normally expressed in the mesoderm at detectable levels were expressed, such as Pax8 (Fig. 6d). The ratio of the number of Isl1 and Aplnr expressing MLPs in Mesp1Cre Ctrl embryos was reduced in Tbx1 cKO embryos (WT1, 0.246 (70/284 cells); WT2, 0.214 (195/912) vs cKO1, 0.011 (5/474); KO2, 0.004 (5/1,056)), while the ratio of the number of Pax8 expressing MLPs was dramatically increased (WT1, 0.011 (3/284 cells), WT2, 0.003 (3/912); KO1, 0.395 (187/474); KO2, 0.181 (191/1,056)), further indicating functional changes in MLPs.
the MLPs of Mesp1Cre and Tbx1Cre cKO embryos (Fig. 6e, f and Supplementary Fig. 8). In serial transverse sections of both sets of cKO and control embryos at E9.5, reduced levels of Aplnr expression in the MLPs in the posterior pharyngeal apparatus were observed, while Pax8 was increased in the same region containing EGFP + cells (Fig. 6g, h and Supplementary Fig. 9).

We next performed scRNA-seq of wild-type versus Tbx1−/− embryos at E9.5, without cell sorting, as an independent replication for these experiments, and we found similar gene expression changes as when Mesp1Cre or Tbx1Cre lines were used for gene inactivation (Supplementary Fig. 10). We note that in this experiment, the ability to identify individual cell types is reduced.
Fig. 3 MLPs transition as cells are allocated to more differentiated states over time. a Heatmap of expression of core genes in MLPs at E8, E8.25, E9.5, and E10.5. Row indicates the expression of each cell. b Heatmap of expression of the genes enriched in expression in earlier stage-MLPs (E8, E8.25) and shown in all four stages. Row indicates the expression of each cell. c Heatmap of expression of the genes enriched in expression in intermediate (E9.5) and later stage MLPs (E10.5) and shown in all four stages. Row indicates the expression in each cell. d Violin plots of the expression of core genes (Isl1, Mef2c, Nkx2-5, Mpped2) in MLPs over time. e Violin plots of expression of early-MLP genes (Aplnr, Nrg1, Irx1, Tbx1) in MLPs over time. f Violin plots of expression of late-MLP genes (Hand2, Gata6, Bmp4, Sema3c) in MLPs over time. g Whole embryo RNAscope with Isl1 (green), Nkx2-5 (red), Sema3c (purple), Aplnr (green), and Tbx1 (red) at E9.5 (19 and 22 somites) and E10.5 (30 somites) to identify MLPs in the distal pharyngeal apparatus (n = 3 for each stage). Images on the right of each row are a composite of all three probes. V ventricle, OFT cardiac outflow tract, MLP, PA1-4 are indicated. Scale bar, 200 µm. h Cartoon of MLP transitions and cell allocation over time. The MLPs in the nascent pharyngeal mesoderm migrate dorsally and will differentiate to BrMs or ventrally and medially to CMs. Blue arrow indicates migration to form BrMs and red arrow(s) indicates migration to the poles of the heart to form CMs. The color spectrum from blue to yellow indicates differentiation from MLPs to their derivative cell types.

Fig. 4 CPM clusters were found in both Mesp1Cre and Tbx1Cre lineages. a UMAP plot of integrated data of Mesp1Cre/+ and Tbx1Cre/+ scRNA-seq data at E9.5, colored by cluster. CPM clusters are shown in bold font. CPM clusters are shown in bold font. b UMAP plot colored by samples (Mesp1Cre, coral; Tbx1Cre, aqua). c UMAP plots, colored by cluster, separated by samples. d The ratio of cell populations of CPM lineages in Mesp1Cre/+ and Tbx1Cre/+ scRNA-seq data. A two proportion Z test was performed in each cluster with 95% confidence interval (MLP: P value = 2.27e-13; pSHF: P value = 2.2e-16; SoM/aSHF: P value = 2.72e-5; OFT + CMs: P value = 0.70). e UMAP plots for showing MLP marker genes, separated by samples. The color spectrum from gray to red indicates expression level from low to high. f Intersection of scRNA-seq data between the two Mesp1Cre and Tbx1Cre populations. Black triangle shows genes shared in both populations and this represents the CPM. Genes shared in both populations have been used for further study in this report. Genes expressed in one versus the other have not been further investigated.
due to the presence of other cell populations, nevertheless, we found similar changes in MLPs with this independent comparison. Overall, the three independent scRNA-seq datasets suggest that Tbx1 promotes lineage maturation but restricts ectopic expression of non-mesodermal genes in MLPs.

In addition to MLPs, we examined DEGs in other CPM populations (BrM, SoM+aSHF, pSHF, and CMs; Supplementary Fig. 11 and Supplementary Data 9). Of interest, Tbx1 is expressed strongly in the BrM populations. We found that there are 667 genes decreased and 163 increased in expression in the BrM population at E9.5 (Supplementary Fig. 11a).

Lhx2 and Myf5 are representative genes that are specific for and downregulated only in the BrM populations (Supplementary Data 9). This suggests the potential of direct transcriptional regulation. We also identified DEGs in other CPM-derivative populations, where Tbx1 is not strongly expressed (Supplementary Fig. 11), and examined GO
pathways affected (Supplementary Fig. 12 and Supplementary Data 10). We compared DEGs specific for MLP versus derivative CPM populations (Supplementary Fig. 13a). Representative DEGs are decreased in MLPs and derivative cells include *Bmp7*, *Fgf10*, *Fgf1*, *Fn1*, *Wnt5a*, *Efnb1*, *Mef2c*, *Kdr*, *Cdh40*, *Anxa6*, *Ets1*, *Cdh13*, *Gbx2*, *Enho2*, *Grem1*, *Fgf8*, *Lox2*, *Lama5*, *Prkd1*, *Hdac5*, *Apoe*, *Rhoj*, *Semad6*, *Scl*1, *Prkd2*, *Sema6d*, *Angptl6*, *Pk3c3*, *Arsb*, *Bmp7*, *Fgf10*, *Foxf1*, *Usbn1c*, *Shox2*, *Hoxd3*, *Edn1*. Other DEGs are reduced only in derivative CPM populations, even where *Tbx1* is not strongly expressed (Supplementary Fig. 13a), suggesting indirect regulation and changes in cell fate acquisition.

**TBX1 defines a gene regulatory network in the MLPs for cardiac and BrM formation.** To better understand how TBX1...
regulates the expression of genes in the CPM at the chromatin level, we used two biological replicates of ATAC-seq experiments of Tbx1^Cre/+ (Tbx1 Ctrl) versus Tbx1^Cre/− (Tbx1 cKO) mutant embryos (Supplementary Fig. 1c, d, Fig. 7a, Supplementary Fig. 14a–c, and Table 2). We chose stage E9.5 because this is when Tbx1 expression is the highest and when the OFT is elongating. The ATAC-seq peaks were separated into commonly accessible regions (CARs) or differentially accessible regions (DARs; FDR < 0.05, Fig. 7b and Supplementary Fig. 14d) between Tbx1 Ctrl and Tbx1 cKO mutant samples. Among 5872 DARs, 5859 decreased and 13 increased in chromatin accessibility in Tbx1 cKO embryos (Supplementary Fig. 14d). To focus on the CPM, we next performed ATAC-seq on Mesp1^Cre/+ cells in two biological replicates and used the peaks to exclude CARs and DARs that were found only in Tbx1 Ctrl and Tbx1 cKO samples, to eliminate changes in non-mesoderm lineages (Fig. 7c, d, gray regions). The remaining CARs and DARs are thus reported as CARs-Mesp1 and DARs-Mesp1 (Fig. 7c, d). In DARs-Mesp1, several transcription factors binding motifs related to the heart or BrM differentiation including T-box motifs were enriched (Fig. 7e and Supplementary Data 11). More peaks in DARs-Mesp1 were found in distal intergenic regions than in CARs-Mesp1 (Supplementary Fig. 13e, f), suggesting that Tbx1 inactivation has a large effect on the regulation of genes through putative enhancer regions.

Next, we annotated DARs-Mesp1 (2185 peaks) to predicted target genes (2652 genes; Fig. 7f and Supplementary Data 12). A total of 160 of the 468 DEGs that were downregulated (down in cKO) in MLPs (Fig. 6b) were associated with DARs-Mesp1 (Fig. 7f, P < 1e-16, chi-square test). GO analysis of the 160 genes indicated that they are involved in cell migration and muscle development, important for the function of Tbx1 (Fig. 7g). On the other hand, those without DARs-Mesp1 that were reduced in expression in mutant embryos were associated with the MAPK pathway and other functions (Fig. 7h). Of interest, Tbx1 might indirectly affect MAPK signaling in the MLPs that were not detected by ATAC-seq analysis and this is potentially mediated by well-known Tbx1-dependent FGF signaling.47,48

To determine which genes with DARs could be direct target genes of Tbx1, we performed ChIP-seq with our new Avi-tagged Tbx1 mouse line and created double homozygous mice harboring the biotin ligase (BiRa) gene (Tbx1^Avi-BiRa, Supplementary Fig. 15). We used input as background (sequencing depth replicate 1 is 26,318,633; replicate 2 is 29,821,823; replicate 3 is 33,965,309) to compare with peaks found by ChIP (sequencing depth replicate 1 is 28,070,687; replicate 2 is 32,561,521; replicate 3 is 32,907,512). We considered peaks found in at least two of the three replicates, each containing 20 embryos of 19–22 somites, with aligned peaks (replicate 1, 294; replicate 2, 128; replicate 3, 536 peaks) as high confidence Tbx1-binding sites (Supplementary Fig. 16a–c). Out of 255 peaks, 176 peaks had a T-box motif (Fig. 8a, Supplementary Fig. 16d, e, and Supplementary Data 13), supporting TBX1 occupancy; 104 peaks (41%) in the ChIP-seq data overlapped with DARs (Fig. 8a, b, P < 0.001, permutation test). Comparing Tbx1 Ctrl and Tbx1 cKO ATAC-seq data in 255 ChIP-seq peak regions, we found that these regions were mostly open and accessible in Ctrl embryos but closed in Tbx1 cKO embryos (Fig. 8c and Supplementary Fig. 16f). Of the 255 ChIP-seq peaks, 151 (59%) did not show significant chromatin accessibility changes (i.e., overlapping with DARs), which include some TBX1-binding sites in closed regions that did not change in accessibility in the mutant data (Fig. 8c), suggesting a diverse role of TBX1 in promoting chromatin remodeling.

ChIP-seq peaks were annotated to predicted direct transcriptional target genes (443 genes, 470 peaks, Supplementary Data 14). We then intersected the DEGs reduced in MLPs, the annotated genes from Mesp1-DARs and TBX1 ChIP-seq targets (Fig. 8b). We found 21 known genes (Fig. 8b and Supplementary Data 14) common in all three datasets (P < 0.001; permutation test). Among them, eight had a DAR that overlapped with a Tbx1 ChIP-seq peak (Slit1-intron, Crebg3-intron, Nrg1-upstream, Trps1-downstream, Sox9-upstream, Trmr1b-downstream, Fln1-upstream, and Cric2-promoter region). Data is consistent with TBX1 binding to accessible chromatin in control embryos, but the interval is not accessible when Tbx1 is inactivated. The rest had a DAR that did not overlap with a Tbx1 ChIP-seq binding site (Ifgl, Apln, Tshz3, Rcsd1, B3galnt12, Ppml1, Spool1, Mpped2, Tbx18, Daum1, Parvh); perhaps regulation is by long-range chromatin interaction with TBX1 binding. We show two representative examples, Aplnr and Nrg1, which are MLP enriched genes that are TBX1 direct transcriptional targets (Fig. 8d). In the Nrg1 locus, the TBX1-binding region was closed in Tbx1 cKO embryos. We examined the ENCODE ChIP-seq tracks in the UCSC genome browser tracks. For Nrg1, the co-localized peak is within an ENCODE cis-regulatory element (cCRE) that is a poised enhancer in mouse embryonic heart. For the Aplnr gene region, the TBX1-binding site that is just downstream of the 3’UTR (Fig. 8d), is in an ENCODE cCRE that is a poised enhancer in mouse embryonic heart, but it was not in a DAR found in our data. The DAR that was identified in the Aplnr locus, is in an ENCODE cCRE (E0701748/enhD) and is an ATAC-seq peak region in the embryonic heart. Overall, the regions found appear to be regulatory regions, but TBX1 might not always affect chromatin accessibility, indicating that multiple mechanisms of regulation occur.

Taking the results from the three types of functional genomic data in this report, we can generate a putative gene regulatory network for Tbx1 function in the MLPs as summarized in Fig. 8e. Here, we distinguish four categories of genes potentially regulated by Tbx1: (1) Direct target genes with or (2) without chromatin changes, and
indirect target genes, (3) with chromatin changes, that contain transcription factor binding sites, and (4) without chromatin changes. Overall, we suggest that TBX1 with Isl1, Fox, Six, Pitx, and E-box proteins such as Tcf21, (Fig. 8c), act together to regulate the progression of MLPs to more differentiated states in the CPM.

**Discussion**

In this report, we discovered a progenitor cell population within the CPM that we term MLP. We discovered that MLPs are localized bilaterally within the posterior nascent mesoderm of the pharyngeal apparatus. The MLPs function to both maintain a
progenitor state and promote differentiation to derivative cell types as shown in the model in Fig. 8f. Tbx1 is required for the MLPs to express critical genes for cardiac and BrM development, and to prevent ectopic expression of non-mesodermal genes, needed for maturation of the CPM-derivative cells (Fig. 8f). This is mediated by direct and indirect transcriptional regulation and chromatin accessibility allowing for cell state progression. Previous work showed that the aSHF, pSHF, and BrM cells comprising the CPM, derive from a relatively small number of Mesp1 expressing progenitor cells at gastrulation8. Based on the work presented in this report, it is most likely that not all cells have committed to final CPM fates at gastrulation and MLPs provide a source of progenitors as the pharyngeal arches form. Retrospective clonal analysis has shown that there is a direct clonal relationship between progenitor cells that form the muscles of mastication and right ventricle, which are derived from the first pharyngeal arch, with distinct clones that form both the OFT and facial expression muscles, from the second arch, while other clones contribute to different arches. It is possible that MLPs are destined to express myogenic transcription factor genes as they become restricted to form BrM skeletal muscle cells. Besides the MLPs, Tbx1 is also expressed in BrM progenitor cells, and therefore, some of the gene expression changes we observed in these cells might be due to Tbx1 expression in the BrM progenitor cells themselves.

Although we focus on the CPM as it relates to cardiac and skeletal muscle development, it was shown that the CPM also contributes to mesenchyme of connective tissue, including cartilage in the neck58. Thus, it is possible that the MLPs could promote connective tissue fates in the craniofacial region that are dependent on Tbx1. Further work will be needed to assess their lineage relationships. Given that most patients with 22q11.2DS have craniofacial malformations in the face and neck, it is important to understand the developmental trajectories of Tbx1-dependent connective tissue progenitor cells.

A strength of this study is that we used two different Cre lines (Mesp1Cre or Tbx1Cre) and analyzed genes altered in both lines. This is because neither Mesp1Cre nor Tbx1Cre is sufficient to define the CPM, and further, both label non-relevant cell lineages. Since the cell number isolated from these embryos was relatively small, it is also possible that rare populations would be missed, so that using this strategy reduces the concern of under sampling. Further, there are limitations of using the Tbx1Cre allele to compare with gene expression changes in Tbx1Cre conditional null mutant embryos. In Tbx1Cre/+ control embryos, one copy of Tbx1 is inactivated. However, we did not observe cardiac defects in Tbx1Cre heterozygous embryos or Tbx1+/−/embryos in which a different region of Tbx1 was inactivated, as maintained in the SwissWebster background. This is in comparison to Tbx1 heterozygous mice as maintained in C57Bl/6, in which heterozygous mice have pharyngeal arch artery defects at reduced penetrance16–18. This suggests that the mild phenotype observed is not due to localized genetic modifiers in the Tbx1 locus itself but rather the genetic background.

The effect of genetic background upon the pharyngeal arch artery phenotype of Tbx1 heterozygous mice has been known for many years. Early work noted that the original 129 strain, in which the Tbx1Cre, Tbx1+/+, and Tbx1+/−/mutations were generated, has a suppressive effect upon the penetrance of such defects (similarly to SwissWebster), while the C57Bl/6 background has an enhancing effect, and research excluded that this modifying effect was due to modifications in the localized genomic region59. Further, we have not observed the presence of novel or unusual transcripts in the Tbx1 locus in different alleles in our scRNA-seq data. Finally, this work uses the null phenotype to draw conclusions about the role of Tbx1 in development, not the

---

**Table 2 Summary of ATAC-seq samples.**

| No. of embryos | Somites | Cell number | GFP(+) | Aligned peaks |
|---------------|---------|-------------|--------|---------------|
| Mesp1Cre; GFP/+ | 3       | 20-21       | 10,000 | 25,485        |
| E9.5-1        | 2       | 19-21       | 10,000 | 30,535        |
| E9.5-3        | 5       | 21-22       | 10,000 | 21,161        |
| Tbx1Cre; GFP/+ | 5       | 19-22       | 10,000 | 49,490        |
| E9.5-1        | 5       | 19-21       | 10,000 | 43,298        |
| E9.5-3        | 5       | 19-21       | 10,000 | 47,013        |
| Tbx1Cre; F/P+/+ | 4       | 19-22       | 10,000 | 29,251        |
| E9.5-1        | 2       | 21-22       | 10,000 | 24,042        |
| E9.5-3        | 2       | 21          | 10,000 | 19,772        |

recently34, indicate that rather than changes in expression, there are instead cell population changes in Tbx1 mutant embryos. Therefore, this scRNA-seq study discerns better between expression versus population changes depending on Tbx1, which is often a key challenge in interpreting developmental phenotypes. More work needs to be done in the future to better understand how these borders are formed and maintained.

In addition to deploying cells to the DPW and then to the heart, the MLPs express genes required for BrM formation in each arch. The BrMs form segmentally in a rostral to the caudal manner, in which the muscles of mastication form first from the first arch and the other muscles of the face and neck form thereafter from more caudal arches. The BrMs express myogenic regulatory transcription factors, including Tcf21, Msc, Myf5, and later Myod155. In addition, transcription factor genes such as Isl111, Pitx256, Tbx119, Lhx225, and Ebf genes (Ciona27) are expressed in the CPM and are required for BrM formation. A subset of MLPs expressing Tbx1 will later express BrM genes as they migrate to the core of the pharyngeal arches. These cells progressively express myogenic transcription factor genes as they become restricted to form BrM skeletal muscle cells. Besides the MLPs, Tbx1 is also expressed in BrM progenitor cells, and therefore, some of the gene expression changes we observed in these cells might be due to Tbx1 expression in the BrM progenitor cells themselves.

---

| E9.5-1 | 3 | 20-21 | 10,000 | 25,485 |
| E9.5-2 | 2 | 19-21 | 10,000 | 30,535 |
| E9.5-3 | 5 | 21-22 | 10,000 | 21,161 |
| Tbx1-Cre; GFP/+ | 5 | 19-22 | 10,000 | 49,490 |
| E9.5-1 | 5 | 19-21 | 10,000 | 43,298 |
| E9.5-3 | 5 | 19-21 | 10,000 | 47,013 |
| Tbx1-Cre; GFP/+ | 4 | 19-22 | 10,000 | 29,251 |
| E9.5-1 | 2 | 21-22 | 10,000 | 24,042 |
| E9.5-3 | 2 | 21 | 10,000 | 19,772 |

---

| E9.5-1 | 3 | 20-21 | 10,000 | 25,485 |
| E9.5-2 | 2 | 19-21 | 10,000 | 30,535 |
| E9.5-3 | 5 | 21-22 | 10,000 | 21,161 |
| Tbx1-Cre; GFP/+ | 5 | 19-22 | 10,000 | 49,490 |
| E9.5-1 | 5 | 19-21 | 10,000 | 43,298 |
| E9.5-3 | 5 | 19-21 | 10,000 | 47,013 |
| Tbx1-Cre; GFP/+ | 4 | 19-22 | 10,000 | 29,251 |
| E9.5-1 | 2 | 21-22 | 10,000 | 24,042 |
| E9.5-3 | 2 | 21 | 10,000 | 19,772 |
heterozygous phenotype. The null phenotype, as two decades of publications indicate, is remarkably similar across multiple genetic backgrounds. Therefore, the genetic background-dependence of the heterozygous phenotype would not affect the major conclusions of the manuscript.

Although there are likely some gene expression changes in \( \text{Tbx1}^{\text{Cre}} / + \) embryos versus wild-type embryos, we obtained similar findings when we directly compared scRNA-seq results from \( \text{Mesp1}^{\text{Cre}} \) versus \( \text{Tbx1}^{\text{Cre}} \) experiments. When taken together, using both \( \text{Mesp1}^{\text{Cre}} \) and \( \text{Tbx1}^{\text{Cre}} \) alleles and investigating changes that occurred only in both, allowed for a more complete analysis of the CPM with respect to \( \text{Tbx1} \). We also performed a scRNA-seq experiment using wild-type versus \( \text{Tbx1}^{-/-} \) embryos at E9.5 and obtained comparable results.
To understand the molecular mechanisms of *Tbx1* function we performed ATAC-seq and TBX1 ChIP-seq. We generated a gene regulatory pathway downstream of *Tbx1* in MLPs that is required for the maturation of cells to CPM derivatives. We obtained robust data from the ATAC-seq experiments and found alteration of chromatin accessibility when *Tbx1* was inactivated, from which some harbored putative T-box-binding sites. The TBX1 ChIP-seq provided hundreds of direct target genes, of which some are also reduced in expression in mutant embryos and show a change in chromatin accessibility. One note is that we identified hundreds rather than thousands of direct transcriptional targets that were expected based upon studies of other transcription factors. This could be because we used whole embryos for the ChIP-seq with lower tissue yield than microdissection. Nonetheless, the ChIP-seq data supports the ATAC-seq findings. Further, this work shows that both direct and indirect regulation occurs downstream from TBX1 because not all differentially accessible sites have T-box motifs, and not all differentially expressed genes have either differentially accessible sites or ChIP-seq peaks. Additionally, TBX1 protein can regulate the protein level of serum response factor, SRF without changing the expression level of RNA, thus there are possible other functions of TBX1 that do not involve binding to DNA. We identified MAPK pathway genes are altered indirectly downstream of *Tbx1*, suggesting that this is a result of the alteration of well-known *Tbx1*-dependent FGF signaling. A subset of genes found with differentially accessible regions in mutant embryos was expressed in MLPs as well as derivative cell types. This is particularly true for the BrM progenitor cells because this is where *Tbx1* is actively expressed.

By performing these multi-omic studies, we identified *Aphn* and *Nrg1* among the genes enriched in expression in MLPs. *Aphn* encodes the APJ/Apelin G-protein coupled receptor that binds Apelin or Elabela/Toddler peptide ligands that have many embryonic and adult functions. In zebrafish, knockdown of *Aphn* disrupts normal migration of cells during gastrulation including that of cardiac progenitors resulting in severe defects. Unexpectedly, their role in early embryogenesis is not recapitated fully in mouse models, implicating perhaps functional redundancy with other G-protein-coupled receptors or ligands. *Aphn* is expressed in the CPM, and from stem cell studies, has a role in cardiomyocyte development. *Nrg1* is also of particular interest. In contrast to *Aphn*, *Nrg1* is not expressed in the DPW. *Nrg1* encodes an EGF family ligand that binds to ErBb receptor tyrosine kinases and has multiple roles in cardiac development and function. Interestingly, both *Aphn* and *Nrg1* are direct target genes of TBX1 based on our ATAC-seq and ChIP-seq results, however, more work will need to be done to know whether these genes have functional importance in MLPs or in relation to *Tbx1*.

Among the genes that were differentially expressed and differentially accessible in *Tbx1* conditional null embryos are *Isl1*, *Foxc2*, *Six1/2*, *Pitx2*, *Tcf21*, *Is1*, *Foxc2*, and *Six2* may be direct transcriptional target genes of TBX1 based upon ChIP-seq analysis. Some known downstream target genes of *Tbx1* were not identified in the multi-omic data, such as *Wnt5a*, *Fgf10*, and *Nkx2-5*, possibly due to low transcript abundance, incomplete set of TBX1 target genes from ChIP-seq, or non-autonomous functions in neighboring CPM cells.

The distal pharyngeal apparatus is hypoplastic when *Tbx1* is inactivated in the mesoderm. This is in part because the loss of *Tbx1* severely affects pharyngeal endoderm-mediated segmentation affecting neighboring neural crest cell populations. These functions are non-autonomous between the CPM and neural crest cells, given that *Tbx1* is not expressed in neural crest cells that contribute to OFT septation. It is possible that altered signaling from affected pharyngeal endoderm cells or lack of neural crest cells could influence MLP or CPM differentiation, besides cell or tissue autonomous effects.

Global inactivation of *Tbx1* results in a persistent truncus arteriosus and hypoplastic muscles of mastication and failure to form the facial/neck muscles. Here, we report the existence of a mesodermal cell population termed MLPs. We show that inactivation of *Tbx1* results in dysregulation of gene expression in MLPs affecting their cellular state. Once this occurs, differentiation is affected as well as signaling to adjacent cells thereby altering migration and/or survival of derivative CPM cells leading to the observed phenotypes. The MLPs are thus a continuous but evolving source of CPM cells that is maintained during the development of the pharyngeal apparatus and plays a critical role in craniofacial and cardiac development as well as pathology.
mice (Supplementary Fig. 1c). All the mice are maintained on the SwissWebster genetic background. The PCR strategies for mouse genotyping have been described in the Genetic Resources section. All the PCR conditions for amplification is one cycle 5° 72°C 30” (seconds) 98° 12 cycles, 98° 30”, 63°C 1, 72°C. PCR products were purified using the MinElute PCR purification kit (Qiagen, Cat# 28004).

ATAC-seq. The ATAC-seq method has been previously described28; however, we provide details in this section. The embryos at E9.5 were isolated from a euthanized mother. GFP-positive embryos were selected under a StereO Discovery.V12 microscope (Carl Zeiss, Jena, Germany) in ice-cold DPBS with Ca2+ and Mg2+ ( Gibco, Cat# 14040-133). The rostral half of the embryos were collected at E8.0, E8.5, and E8.5. The pharyngeal apparatus with heart was collected at E9.5 and E10.5, as shown in Fig. 1b. The microdissected tissues were kept on ice and pooled in DEMEM (4°C, Gibco, Cat# 11885-084) until all the dissections were completed. Following centrifugation (4°C, 100 × g, 5 min), the pellet was resuspended in PBS w/o Ca2+ and Mg2+. (Corning, Cat# 21-031-cv) with 10% FBS (heat-inactivated, ATCC, Cat# 30-2021) at RT. Then FBS (heat-inactivated, ATTC, Cat# 30-2021) was added to stop the reaction. After centrifugation (4°C, 300 x g, 5”), the cells were resuspended in PBS w/o Ca2+ and Mg2+ (Corning, Cat# 21-031-cv) with 10% FBS and passed through the 100-µm cell strainer. DAPI (Thermo Fisher Scientific, Cat# D13004) was added before cell sorting. The GFP + DAPI cells were sorted. The图书馆 Aria II system (Becton, Dickinson Biosciences, Franklin Lakes, NJ) was used to collect the cells sorted. When rehydrated in 10x Chromatin, then incubated 4°C overnight on a rotator. Strepavidin magnetic beads were washed with washing buffer containing 10 mM Tris-HCl pH 8.0, 0.1% SDS, 0.5% IGEPAL CA-630, 150 mM NaCl, 0.1% SDS, 1% Triton-X100, 2 mM EDTA, high salt wash buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% SDS, 1% Triton-X100, 2 mM EDTA), LiCl wash buffer containing 10 mM Tris-HCl pH 8.0, 0.25 M LiCl (MilliporeSigma, Cat# L7026), 1% IPEAL™ CA-630, 0.5% Sodium deoxycholate (MilliporeSigma, Cat# D6760), 1 mM EDTA. After DNA was purified using the MinElute PCR purification kit. After measuring the DNA concentration, the sections were incubated with Opal 570 Multiplex Fluoroscent v2 reagents (Advanced Cell Diagnostics, Cat# 323100), according to the manufacturer’s instructions (Advanced Cell Diagnostics). The concentrations of the libraries were measured with Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA) using Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Cat# Q32851). The details of each sample are provided in Table 1.

RNAseq. The RNA libraries were sequenced using an Illumina HiSeq2500 system (at Eugene Epigenomics Core Facility), Illumina HiSeq4000 system (at Genewiz, South Plainfield, NJ) or NovaSeq6000 system (at Novageno, Sacramento, CA), with paired-end, 100 bp read length.

ChIP-seq. Whole embryos from Tbx1–Avi;BirA double homozygous embryos or Bira homozygous embryos (Ctrl) at E9.5 were collected and microdissected in ice-cold PBS. Following centrifugation (4°C, 200 × g, 5”), tissues were cross-linked with 1% formaldehyde (Thermo Fisher Scientific, Cat# 28996), 30” at RT. A total of 2.5 M glycine (MilliporeSigma, Cat# M9760) was added to a final concentration of 0.25% glycine chemically. From this point on, all the dissections were performed in ice-cold conditions. Following 30 minutes of ice-cold PBS wash, and centrifuged (4°C, 200 × g, 5”), the pellets were frozen in dry ice and stored at −80°C. We used 2 embryos for each sample and performed three biological replicates. The frozen tissues were homogenized by bead-mix Microtube Homogenizer (Benchmark Scientific, Edinson, NJ) in lysis buffer, which contained 50 mM HEPES pH 7.5 (MilliporeSigma, Cat# E7889), 140 mM NaCl, 1 mM EDTA (MilliporeSigma, Cat# E7889), 0.5% IPEAL™ CA-630 (MilliporeSigma, Cat# R8896), 0.25% Trition-X100 (MilliporeSigma, Cat# X100), and Protease Inhibitor cocktail (MilliporeSigma, Cat# P8340). Samples were incubated on ice for 10”. Following centrifugation (4°C, 2000 × g, 5”), extracted nuclei were washed in Wash buffer containing 10 mM Tris-Cl pH 8.0 (MilliporeSigma, Cat# T2694), 200 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0 (MilliporeSigma, Cat# E889) and Protease Inhibitor Cocktail. The nuclei were resuspended in Shearing buffer containing 10 mM Tris-Cl pH 8.0, 1 mM EDTA pH 8.0, 0.1% SDS (BIO-RAD), Cat# 1610418, and Protease Inhibitor Cocktail. Resuspended nuclei were incubated with a 2:1 ratio of the nuclei and Protein G for immunoprecipitation (Thermo Fisher Scientific, Cat# 28122) at 4°C for 1 h. Following washing with high salt wash buffer, blocked streptavidin magnetic beads were added to the sheared chromatin, then incubated 4°C overnight on a rotator. Streptavidin magnetic beads were washed with TE buffer (10 mM Tris-HCl pH 8.0, 0.1% SDS, 1% Triton-X100, 2 mM EDTA), high salt wash buffer (20 mM Tris-Cl pH 8.0, 150 mM NaCl, 0.1% SDS, 1% Triton-X100, 2 mM EDTA), LiCl wash buffer containing 10 mM Tris-Cl pH 8.0, 0.25 M LiCl (MilliporeSigma, Cat# L7026), 1% IPEAL™ CA-630, 0.5% Sodium deoxycholate (MilliporeSigma, Cat# D6760), 1 mM EDTA. After washing with TE buffer, DNA was eluted in Elution buffer (10 mM Tris-Cl pH 8.0, 1 mM EDTA, 1% SDS) with Proteinase K (Promega, Cat# V3021) at 65°C, 5 h. Eluted DNA was purified with the MinElute PCR purification kit. After measuring the DNA concentration with the Qubit® 2.0 Fluorometer using Qubit dsDNA HS Assay Kit, the libraries were prepared using Accel-NCS ZS Plus DNA Library Kit (Swift Bioscience, Cat# 21024) and 25 Set A Indexing Kit (Swift Bioscience, Cat# 26148); then the DNA concentration was measured with Qubit® 2.0 Fluorometer using Qubit dsDNA HS assay kit.

Sequencing. The DNA libraries were sequenced using an Illumina HiSeq2500 system (at Eugene Epigenomics Core Facility), Illumina HiSeq4000 system (at Genewiz, South Plainfield, NJ) or NovaSeq6000 system (at Novageno, Sacramento, CA), with paired-end, 100 bp read length.
75%, 50%, 25% methanol, and PBS + 0.1% Tween (Sigma, Cat# P7949). The embryos were permeabilized using Protease III (Advanced Cell Diagnostics, Cat# 322100) for 20 minutes or a wash with PBS + 0.01% Tween at RT. Embryos were incubated in AMP1 for 30’, AMP2 for 30 min, and AMP3 for 15 min at 40°C with washes in between. Tyramide Signal Amplification (TSA) was prepared at 1:2000 for TSA-CY3 (Aksya Biosciences, Cat# NEL744001KT), 1:1000 for TSA-CY5 (Aksya Biosciences, Cat# NEL744001KT), 1:500 for TSA-Fluor (Aksya Biosciences, Cat# NEL741001KT). Embryos were incubated in HRP-C1 for 15 at 40°C followed by 30’in their chosen TSA for C1. Amplification was blocked in HRP-Blocker for 15’. The previous two steps were then repeated for HRP-C2 for C2 probes and HRP-C3 for C3 in their chosen TSA for C1. According to their protocols for each step.

ChIP-seq data analysis. We used adapters using cutadapt with a –output and set of adapters detected with FASTQC. Sequences were then clipped using TrimGalore with option –length 0. Then sequences were aligned to the mouse genome (mm9) using Bowtie2 2.3.4. with default parameters. Reads were removed with mates mapping to different chromosomes, or with discordant pairs oriented in opposite mate-pair distance > 2 kb, or PCR duplicates (defined as when both mates are aligned to the same genomic coordinate). Only uniquely mappable reads were retained. ChIP-seq peaks in each sample were identified using MACS2 2.1.2 with default parameters. Then, a consensus list of enriched regions was obtained using the intersectfunction from the BEDTools 2.29 with the default maximum overlap and retaining only the peak regions common to at least two out of the three replicates. Peaks were filtered by removing those overlapping with blacklist regions (Encode mm9 black regions Version 2) using findOverlapOfPeaks of ChiPpeakAnno. The transcription factor binding motifs were obtained using the findMotifs Genome program with -size parameter of the HOMER suite. For peak annotation as cis-regulatory regions, GREAT was used with default settings of mm9. The comparison of the gene list from DARs and DEGs was performed using standard R scripts. The clusterProfiler v3.10.1 was used for Gene Ontology pathway analysis. Coverage heatmaps and average enrichment profiles (TSS +/- 10 Kb) in each experimental condition were obtained using ngsplot or deepTools. We applied statistical analysis in the processes of all the software and packages with default settings.

scRNA-seq data analysis. We utilized Cell Ranger (v 3.1.0, from 10x Genomics) to align reads of scRNA-seq data to the mouse reference genome (assembly, mm10). All the samples passed quality control measures for Cell Ranger (Table 1), and the filtered gene-barcode matrices were used for the following analyses. For MespT4 four time point dataset analyses, Scan v1.10.2 was used to normalize the individual datasets with the computeSumFactors function with the normalization method for scaling normalization. The cells were clustered by densityClust v0.3 with the density peak clustering algorithm. We found batch effects existed in the scRNA-seq datasets from different time points (MespT4 data) or experimental perturbations (MespT4 or Tbx1Cre Ctrl vs CKO). Therefore, we performed batch correction before we comprehensively analyzed gene expression values across these scRNA-seq datasets. We employed the MNN (mutual nearest neighbors) method to identify shared cell types across datasets, and corrected batches, according to the shared cell types, by the MNN batch correction method. More specifically, we removed batch effects by fastMNN function of the Scran package, with the default settings for the corrected counts from individual datasets. The MespT4 Ctrl data from different stages are comprised of homogeneous cell types (transcripts of each cell type are concordant across datasets), but the MespT4 Ctrl with CKO and then separately, the Tbx1Cre Ctrl versus CKO, are comprised of heterogeneous cell types (the same cell types from different conditions, Ctrl vs CKO, with some degree of expression). Therefore, we performed gene expression analysis for Ctrl vs CKO data using the RPCI (reference principal component integration) method in RISC, which utilizes the global gene reference to calibrate the gene expression changes of heterogeneous cell types. In detail, we combined individual datasets by the scMultiIntegrate function of the RISC package and outperformed the corrected gene expression values after the clustering by Seurat v3.1.5. Then, the corrected datasets were processed on Scapy v4.3.9 for cell trajectory analysis, with the PAGA approach. The RISC software was also used to identify differentially expressed genes of Tbx1 Ctrl vs CKO embryos in each cluster by a Negative Binomial generalized linear model (scDEG function of RISC package), with adaptive P value correction (log fold change) >0.25 or < -0.25. The clusterProfiler v3.10.1 was used for the Ontology pathway analysis, with adaptive P values < 0.05.

ATAC-seq data analysis. ATAC-seq analysis pipeline has been described previously, however, we describe the methods here as well. We removed Nextera Transpose Sequences primer pairs in the range 33–47 bp with cutadapt with the following option –a CTGTCCTCTATACATACATCGCCGACCAGCAGCA -A CTGTCTTCTATACATACATCGCCGACCAGCAGCA. Sequences were then aligned to the mouse genome (mm9) using Bowtie2 2.3.4.99 with default parameters. Only uniquely mappable reads were retained. We removed reads with mates mapping to different chromosomes, or with discordant pairs orientation, or with a mate-pair distance >2 kb, or PCR duplicates (defined as when both mates are aligned to the same genomic coordinate). Removed reads mapping to the same genomic region were also removed. ATAC peaks in each sample were identified using MACS2 2.1.2.100 with the option --nomodel --shift100 --extsize 200. The differentially enriched regions (DARs) of Tbx1Ctrl Ctrl vs CKO were obtained using DiffBind 2.14.0 by loading all the MACS2 peaks, default parameters were used except for the common peak-mapping to DISCOVRY. We used the DISCOVRY regulatory regions, GREAT was used with the default settings of mm9. The comparison of the gene lists from DARs and DEGs was performed using standard R scripts. The clusterProfiler v3.10.1 was used for Gene Ontology pathway analysis. Coverage heatmaps and average enrichment profiles (TSS +/- 10 Kb) in each experimental condition were obtained using ngsplot or deepTools. We applied statistical analysis in the processes of all the software and packages with default settings.

Quantification and statistical analysis. Besides the methods for scRNA-seq analysis, specific statistical tests were described in “Results” and Figure legends. Briefly, to determine the relative proportions of the CPM populations, a two proportion Z test was performed in each cluster with 95% confidence interval. For ATAC-seq and TBX1 ChIP-seq, default statistical analysis was used using DiffBind, Homer, and GREAT. The heatmap of ATAC-seq read densities in TBX1 ChIP-seq peaks on ATAC-seq signals were determined using enrichPeakOverlap present in ChiPseeker. For Venn diagrams, Pearson’s Chi-squared test was performed.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. The datasets generated during this study are available at the GEO repository under the accession numbers: scRNA-seq “GSE167493”, scRNA-seq “GSE167491”, ATAC-seq “GSE173700”, and ChiP-seq “GSE173521”. The scRNA-seq data can be viewed at https://scviewer.shinyapps.io/heartMLP. All other relevant data supporting the key findings of this study are available within the article and its Supplementary Information files or from the corresponding author upon reasonable request.

Received: 8 January 2021; Accepted: 18 October 2021; Published online: 17 November 2021

References
1. Cai, C. L. et al. Isl1 identifies a cardiac progenitor population that proliferates primarily during differentiation. Isl1 contributes a majority of cells to the heart. Dev. Cell 5, 877–889 (2003).
2. Diogo, R. et al. A new heart for a new head in vertebrate cardiopharyngeal evolution. Nature 520, 466–473 (2015).
3. Kelly, R. G., Brown, N. A. & Buckingham, M. E. The arterial pole of the mouse heart forms from Fgf10-expressing cells in pharyngeal mesoderm. Dev. Cell 1, 435–440 (2001).
4. Mjaatvedt, C. H. et al. The outflow tract of the heart is recruited from a novel heart-forming field. Dev. Biol. 238, 97–109 (2001).
5. Stefanovic, S. et al. Hox-dedicated coordination of mouse cardiac progenitor cell patterning and differentiation. eLife 9, https://doi.org/10.7554/eLife.55124 (2020).
6. Tian, Y. et al. Characterization and in vivo pharmacological rescue of a Wnt2-Gata6 pathway required for cardiac inflow tract development. Dev. Cell 18, 275–287 (2010).
7. Saga, Y. et al. MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. Development 126, 3437–3447 (1999).
8. Lescroart, F. et al. Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. Science 359, 1177–1181 (2018).
9. Lescroart, F. et al. Clonal analysis reveals common lineage relationships between head muscles and second heart field derivatives in the mouse embryo. Development 137, 3269–3279 (2010).
10. Lescroart, F. et al. Clonal analysis reveals a common origin between nonomite-derived neck muscles and heart myocardium. Proc. Natl Acad. Sci. USA 108, 14461–14466 (2011).
11. Nathan, E. et al. The contribution of Islet1-expressing splanchnic mesoderm cells to distinct branchiomiocic mesoderm signals significant heterogeneity in head muscle development. Development 135, 647–657 (2008).
12. Stoll, A. et al. Early chordate origins of the vertebrate second heart field. Science 329, 565–568 (2010).
13. McDonald-McGinn, D. M. et al. 22q11.2 deletion syndrome. Nat. Rev. Prim. 1, 15071 (2015).
14. Kollara, L. et al. Velopharyngeal structural and muscle variations in children with 22q11.2 deletion syndrome: an unseated MRI study. Cleft Palate Craniofac J. 56, 1139–1148 (2019).
15. Yagi, H. et al. Role of TBX1 in human del22q11.2 syndrome. Lancet 362, 1366–1373 (2003).
16. Jerome, L. A. & Papaioannou, V. E. DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1. Nat. Genet. 27, 286–291 (2001).
17. Merscher, S. et al. TBX1 is responsible for cardiovascular defects in velo-cardio-facial/DiGeorge syndrome. Cell 104, 619–629 (2001).
18. Lai, D. et al. Mouse Tbx1 flank haploinsufficiency in the DiGeorge syndrome region causes aortic arch defects in mice. Nature 410, 97–101 (2001).
19. Kelly, R. G., Jerome-Majewska, L. A. & Papaioannou, V. E. The del22q11.2 candidate gene Tbx1 regulates branchiomycic myogenesis. Hum. Mol. Genet. 13, 2829–2840 (2004).
20. Liao, J. et al. Identification of downstream genetic pathways of Tbx1 in the second heart field. Circ. Res. 116, 524–537 (2015).
21. Kong, P. et al. Tbx1 is required autonomously for cell survival and fate in the pharyngeal core mesoderm to form the muscles of mastication. Hum. Mol. Genet. 23, 4215–4231 (2014).
22. Harel, I. et al. Pharyngeal mesoderm regulatory network controls cardiac and head muscle morphogenesis. Proc. Natl Acad. Sci. USA 109, 18839–18844 (2012).
23. Hoffmann, A. D. et al. Foxf genes integrate tbx5 and hedgehog pathways in the second heart field for cardiac septation. PLoS Genet. 10, e1004604 (2014).
24. Steinle, J. D. et al. Evolutionarily conserved Tbx5-Wnt2/2b pathway orchestrates cardiopleural development. Proc. Natl Acad. Sci. USA 115, E10615–E10624 (2018).
25. Srivastava, D. Making or breaking the heart: from lineage determination to differentiation of multipotent heart progenitors. Circ. Res. 124, 1420–1432 (2014).
26. Wolf, F. A. et al. PAGA: graph abstraction reconciles clustering with trajectory segmentation by single-cell RNA-seq. Cell Rep. 28, 3402 (2019).
27. Furrer, L. et al. Apelin and its receptor control heart field cardiomyocyte deployment and patterning of downstream genetic pathways of Tbx1 in the heart field. Proc. Natl Acad. Sci. USA 112, 14461–14466 (2015).
28. Lai, D. et al. Ptx2 is required for cardiac outflow tract morphogenesis. Dev. Biol. 433, 4891–4899 (2006).
29. Zeng, X. X., Wilm, T. P., Sepich, D. S. & Solnica-Krezel, L. Apelin and its receptor control heart field cardiomyocyte deployment and patterning of downstream genetic pathways of Tbx1 in the heart field. Proc. Natl Acad. Sci. USA 112, 1451–1455 (2015).
30. Soh, B. S. et al. Reduced dosage of beta-catenin provides significant rescue of cardiac outflow tract anomalies in a Tbx1 conditional null mouse model of 22q11.2 deletion syndrome. PLoS Genet. 13, e1006687 (2017).
31. Freyer, L. et al. Collier/OLF/EBF-dependent transcriptional dynamics control pharyngeal myocardium specification from primed cardiopharyngeal progenitors. Dev. Cell 29, 263–276 (2014).
32. Comai, G. et al. A distinct cardiopharyngeal mesoderm genetic hierarchy establishes antero-posterior patterning of esophagus striated muscle. eLife 8, https://doi.org/10.7554/eLife.47460 (2019).
33. Lai, D. et al. Neuregulin 1 sustains the gene regulatory network in both cardiac and pharyngeal outflow tract derivatives in the mouse embryo. Biochim. Biophys. Acta 1833, 795–798 (2013).
34. Soh, B. S. et al. N-cadherin prevents the premature differentiation of anterior heart field progenitors in the pharyngeal mesodermal microenvironment. Cell Res. 24, 1420–1432 (2014).
35. Scott, I. C. et al. The g protein-coupled receptor agtl1b regulates early trabecular and nontrabecular myocardium development. Circ. Res. 115, 4320–4331 (2014).
36. Guo, C. et al. Tbx1-Six1/Eya1-Fgf8 genetic pathway controls mammalian cardiovascular and craniofacial morphogenesis. J. Clin. Invest. 121, 1585–1595 (2011).
37. Gorkin, D. U. et al. Author Correction: An atlas of dynamic chromatin landscapes in mouse fetal development. Nature 589, E4 (2021).
38. Francou, A. et al. Second heart field myocardial progenitors in the early mouse embryo. Biochim. Biophys. Acta 1833, 795–798 (2013).
39. Lescroart, F. et al. Clonal analysis reveals common lineage relationships between head muscles and second heart field derivatives in the mouse embryo. Cleft Palate Craniofac J. 56, 1139–1148 (2019).
40. Lescroart, F. et al. Clonal analysis reveals common lineage relationships between head muscles and second heart field derivatives in the mouse embryo. Dev. Biol. 338, 432–437 (2007).
41. Taddei, I., Morishima, M., Huyhn, T. & Lindsay, E. A. Genetic factors are major determinants of phenotypic variability in a mouse model of the DiGeorge/ 22q11.2 syndromes. Proc. Natl Acad. Sci. USA 98, 11428–11431 (2001).
42. Comai, G. et al. A distinct cardiopharyngeal mesoderm genetic hierarchy establishes antero-posterior patterning of esophagus striated muscle. eLife 8, https://doi.org/10.7554/eLife.47460 (2019).
43. Seo, S. & Kume, T. Forkhead transcription factors, Foxc1 and Foxc2, are required for the morphogenesis of the cardiac outflow tract. Dev. Biol. 296, 421–426 (2006).
77. Sousa, V. H., Miyoshi, G., Hjerling-Leffler, J., Huynh, T., Chen, L., Terrell, P. & Baldini, A. A fate map of Tbx1 expressing cells identifies a subset of progenitor cells in the myocardium. *Dev. Biol.* 389, 320–331 (2014).

84. Cirino, A. et al. Chromatin and transcriptional response to loss of TBX1 in the heart. *Dev. Dyn.* 244, 1492–1496 (2014).

91. Ross-Innes, C. S. et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 481, 389–393 (2012).

98. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* 38, 576–589 (2010).

105. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* 28, 495–501 (2010).

112. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842 (2010).

119. Zhu, J. J. et al. ChippeakAnn: A Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* 11, 237 (2010).

126. Shen, L., Shao, N., Liu, X. & Nestler, E. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *Nucleic Acids Res.* 44, W160–W165 (2016).

133. Andrews, S. FastQC: a quality control tool for high throughput sequence data. *http://www.bioinformatics.babraham.ac.uk/projects/fastqc* (2010).