



HAL
open science

Cardiopharyngeal Mesoderm specification into cardiac and skeletal muscle lineages in gastruloids

Laurent Argiro, Céline Chevalier, Caroline Choquet, Nitya Nandkishore, Adeline Ghata, Anaïs Baudot, Stéphane Zaffran, Fabienne Lescroart

► **To cite this version:**

Laurent Argiro, Céline Chevalier, Caroline Choquet, Nitya Nandkishore, Adeline Ghata, et al.. Cardiopharyngeal Mesoderm specification into cardiac and skeletal muscle lineages in gastruloids. 2023. <hal-04255863>

HAL Id: hal-04255863

<https://hal.science/hal-04255863v1>

Preprint submitted on 24 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

1 **Cardiopharyngeal Mesoderm specification into cardiac and skeletal muscle**
2 **lineages in gastruloids.**

3
4

5 **Running title: Cardiopharyngeal mesoderm specification in gastruloids.**

6

7 Laurent Argiro ^{1, §}, Céline Chevalier ^{1, §}, Caroline Choquet ^{1, §}, Nitya Nandkishore ¹, Adeline
8 Ghata ¹, Anaïs Baudot ¹, Stéphane Zaffran ^{1, £*}, Fabienne Lescroart ^{1, £*}

9

10 ¹ Aix-Marseille Université, INSERM, MMG U1251, 13005 Marseille, France.

11 [§] These authors contributed equally to this work.

12 [£] co-last authors.

13

14 * Corresponding authors: fabienne.lescroart@univ-amu.fr, stephane.zaffran@univ-amu.fr

15

16

17 **Key words:** Cardiopharyngeal mesoderm, Pharyngeal head muscles, Myogenesis, Skeletal
18 muscles, Cardiac muscles, Gastruloids, Lineage specification.

19 **Abstract**

20
21 Cardiopharyngeal mesoderm contributes to the formation of the heart and head muscles.
22 However, the mechanisms governing cardiopharyngeal mesoderm specification remain unclear.
23 Indeed, there is a lack of an *in vitro* model replicating the differentiation of both heart and head
24 muscles to study these mechanisms. Such models are required to allow live-imaging and high
25 throughput genetic and drug screening. Here, we show that the formation of self-organizing or
26 pseudo-embryos from mouse embryonic stem cells (mESCs), also called gastruloids, reproduces
27 cardiopharyngeal mesoderm specification towards cardiac and skeletal muscle lineages. By
28 conducting a comprehensive temporal analysis of cardiopharyngeal mesoderm establishment
29 and differentiation in gastruloids and comparing it to mouse embryos, we present the first
30 evidence for skeletal myogenesis in gastruloids. By inferring lineage trajectories from the
31 gastruloids single-cell transcriptomic data, we further suggest that heart and head muscles formed
32 in gastruloids derive from cardiopharyngeal mesoderm progenitors. We identify different
33 subpopulations of cardiomyocytes and skeletal muscles, which most likely correspond to different
34 states of myogenesis with “head-like” and “trunk-like” skeletal myoblasts. These findings unveil
35 the potential of mESC-derived gastruloids to undergo specification into both cardiac and skeletal
36 muscle lineages, allowing the investigation of the mechanisms of cardiopharyngeal mesoderm
37 differentiation in development and how this could be affected in congenital diseases.

38 **Introduction**

39

40 Deployment of progenitors and their proper allocation to correct cell lineages are
41 fundamental for the formation of organs. Defects in the specification and differentiation of
42 progenitors into particular cell lineages lead to congenital anomalies. For example, a recent
43 report showed improper cardiopharyngeal mesoderm progenitor development in a 22q11.2DS
44 mouse model, where head muscle and heart morphogenesis were impaired ¹.

45

46 **Cardiopharyngeal mesoderm includes bipotent progenitors for heart and head muscles**

47 The heart forms from two cell populations, namely the first and second heart fields. The
48 first heart field forms essentially the left ventricle with small contributions to the right ventricle
49 and atria ^{2,3}. Second heart field progenitors, located in cardiopharyngeal mesoderm, contribute
50 to cardiac muscles (myocardium) of the outflow tract, right ventricle and atria ²⁻⁷.
51 Cardiopharyngeal mesoderm populates the pharyngeal arches and also gives rise to head and
52 a subset of neck skeletal muscles, in addition to cardiac muscles ⁸⁻¹¹. Clonal analyses in the
53 mouse model have revealed the existence of bipotent progenitors in cardiopharyngeal
54 mesoderm that form both head and heart muscles in vertebrates with different contributions
55 along the antero-posterior axis of the pharyngeal region of the embryo ^{8,12,13}. Bipotent
56 progenitors were thus found for distinct subsets of myocardium and specific head and neck
57 skeletal muscles ^{12,13}. Lineage tracing of early nascent mesoderm expressing the bHLH
58 transcription factor *Mesp1* showed that bipotent head and heart muscle progenitors are in fact
59 present at the onset of gastrulation ¹⁴. Cardiopharyngeal mesoderm is evolutionary conserved
60 since similar multipotent cardiopharyngeal progenitors have been shown to give rise to heart
61 and pharyngeal muscles in tunicates ^{9,15}.

62

63 **Cardiopharyngeal mesoderm contributes to skeletal muscles of the head and neck** 64 **with a particular genetic program**

65 For its skeletal muscles derivatives, cardiopharyngeal mesoderm contributes to specific
66 head/neck muscles ^{9,10}. Skeletal muscles derived from cardiopharyngeal mesoderm have
67 distinct genetic regulatory programs compare to trunk and limb skeletal muscles ¹⁶. Trunk and
68 limb skeletal muscles derive from somitic paraxial mesoderm. Somitic myogenesis is
69 regulated by *Pax3* (a homeobox paired-domain transcription factor), which regulates the
70 myogenic regulatory transcription factors *MyoD* and *Myf5*. Strikingly, *Pax3/Myf5* mutant
71 embryos failed to develop trunk and limb muscles while head muscles were broadly not
72 affected, thus decoupling trunk (somatic) from head (cardiopharyngeal) myogenesis ¹⁷. In
73 cardiopharyngeal mesoderm, the myogenic factors *MyoD/Myf5* are regulated by a different
74 core of transcription factors that together mark the cardiopharyngeal mesoderm but are

75 individually not strictly restricted to this population. They include Tbx1, Lhx2, MyoR (also called
76 Msc), Tcf21 (also called Capsulin) and Pitx2^{18–22}. Myogenin (MyoG) then marks the start of
77 myoblast differentiation. The next steps of muscle differentiation are common between
78 cardiopharyngeal and somitic mesoderm derived progenitors¹⁶.

79

80 **Insights on how cardiopharyngeal mesoderm specification occurs mostly arise from** 81 **studies in tunicates**

82 Major insights into the specification of multipotent cardiopharyngeal mesoderm
83 progenitors and the mechanisms of cardiopharyngeal lineage segregation have come from
84 studies in Ciona. In this tunicate model, Tbx1/10 activates COE/Ebf for the specification of the
85 muscular lineage^{23,24}. It is however not completely clear how comparable are tunicates and
86 vertebrates. The heart and pharyngeal muscles derive from only 2 Mesp+ progenitors in
87 tunicates, while about 250 Mesp1+ progenitors contribute to heart development in mouse^{9,25}.
88 Despite the number of important studies using lineage tracing and clonal analysis, it is
89 challenging to analyze cardiopharyngeal mesoderm specification in vertebrates due to its
90 difficult accessibility and the absence of restricted and specific markers. There is a need of a
91 simple model that could faithfully recapitulate vertebrate, and more specifically mammalian,
92 cardiopharyngeal mesoderm development and could allow live-imaging and large throughput
93 analyses.

94

95 **Pluripotent stem cells-derived models to study Cardiopharyngeal mesoderm**

96 Pluripotent stem cells have emerged as an interesting tool in general to model how
97 transcription factors and signaling molecules interact to control cell fate decisions and lineage
98 specification^{26,27}. Using mouse embryonic stem cells (mESCs), Chan et al. have shown that
99 Mesp1+-derived progenitors have a dual heart and skeletal muscle differentiation potential,
100 when cultured in a pro-cardiogenic or a pro-skeletal myogenic culture medium²⁸. Mouse and
101 human ESCs have also been used to investigate the differences in molecular cues involved in
102 myogenic specification between cardiopharyngeal versus somitic mesoderm²⁹. Both studies
103 however, required changes in cell culture medium and growth factors to promote either cardiac
104 or skeletal muscle differentiation^{28,29}. Therefore, there is a lack of an *in vitro* model that allows
105 parallel differentiation of cardiac and skeletal muscle tissue and faithfully recapitulates *in vivo*
106 cardiopharyngeal mesoderm early development.

107

108 **Gastruloids model early cardiac development**

109 The establishment of gastruloids provides an alternative model for cardiopharyngeal
110 mesoderm development. The model of gastruloids was first described in 2014 with axial
111 elongation and symmetry breaking from the aggregation of a restricted number of mESCs

112 cultured with a pulse of Wnt activation (with Chiron treatment) ^{30,31}. This model faithfully
113 recapitulates early mouse development mimicking *in vivo* gene expression ³². Recently, the
114 model has been pushed further towards early organogenesis, with development of heart-like
115 structures with first and second heart field components ^{33,34} or somites ^{35,36}. Despite the
116 formation of somites, skeletal myogenesis, either cardiopharyngeal or somitic, has not yet
117 been demonstrated in gastruloids.

118 Here, we have adapted the self-organizing mESCs-based gastruloid protocol for longer
119 times of culture to investigate whether gastruloids can model cardiopharyngeal mesoderm
120 specification and differentiation. Multiplex fluorescent *in situ* hybridization shows that
121 cardiopharyngeal mesoderm specification occurs in gastruloids in a similar spatio-temporal
122 organization as observed in mouse embryos. Using single-cell RNAseq analysis along a time-
123 course from day 4 to late day 11 of culture, we demonstrate the presence of three
124 subpopulations of cardiomyocytes and two subpopulations of myoblasts. We find that
125 gastruloids can undergo myogenesis and that gastruloid-derived myoblasts can arise from
126 both cardiopharyngeal and somitic paraxial mesoderm. Therefore, gastruloids can be used to
127 reconstruct cardiopharyngeal mesoderm development.

128

129 **Results**

130

131 **Cardiopharyngeal mesoderm markers are expressed in gastruloids**

132 To generate gastruloids that form cardiopharyngeal mesoderm and characterize their
133 differentiation to cardiac and skeletal myogenic fates, we optimized previous protocol described
134 by Rossi et al. ³⁴ to culture gastruloids for an extended time, until at least day 11 (Fig. 1a) (see
135 Methods). In brief, mESCs were aggregated at day 0, following centrifugation, and treated with a
136 Wnt agonist (Chiron) for 24h from day 2. Cardiogenic factors (bFGF, VEGF and ascorbic acid)
137 were added to the culture media at 96 h (day 4) for 3 days. We noted elongation from day 4 and
138 beating regions appearing at day 7, on the anterior part of the gastruloid (Fig. 1b.).

139 To investigate whether key markers of cardiopharyngeal mesoderm and its derivatives
140 (Fig. 1c) were expressed during the time-course culture, we used quantitative RT-PCR. We
141 observed the transient expression of the bHLH transcription factor *Mesp1* and the expression of
142 the cardiopharyngeal mesoderm transcription factors *Islet1* (*Isl1*) and *Tbx1*, as previously
143 described ³⁴ (Fig. 1d). We also observed an increasing expression of transcripts encoding the
144 cardiac specific myosin (*MyI7*) and troponin (*Tnnt2*) at day 5 (Fig. 1e). Interestingly, the
145 cardiopharyngeal marker, *Tcf21* ^{20,24}, began to be expressed at day 4. The myogenic transcription

146 factors *Myf5* and *MyoD* were expressed at day 6 and day 8 respectively (Fig. 1f). Importantly,
147 similar kinetics of expression were also observed for *Mesp1*, *Isl1* and *Tcf21* with another wild-
148 type mESC line (see Methods for details on the lines and Supplementary Fig. S1). These data
149 demonstrate that cardiopharyngeal mesoderm and downstream myogenic transcriptional
150 programs are robustly activated in gastruloids under these culture conditions.

151

152 **Comparison with mouse embryo shows similar spatio-temporal gene expression**

153 To explore whether gastruloids faithfully mimic mouse cardiopharyngeal mesoderm
154 development, we compared gastruloids with mouse embryos at equivalent developmental
155 stages. Our goal was to investigate the expression of markers of the cardiopharyngeal
156 mesoderm, cardiomyocytes, and the robustness of the cardiopharyngeal mesoderm
157 specification. We used multiplex fluorescent *in situ* hybridization or RNAscope to explore gene
158 expression. Comparison of day 4 gastruloids with E7.75 cardiac crescent embryos showed
159 expression of *Mesp1* in non-overlapping domains from the expression of *Isl1* and *Tbx1* (Fig.
160 1g-h). In the embryo, *Mesp1* was observed in the posterior mesoderm (likely corresponding to
161 the somitic mesoderm), while *Tbx1* and *Isl1* expression overlapped in the anterior region of the
162 embryo (Fig. 1g). Similarly, *Tbx1* and *Isl1* expression overlapped in gastruloids (Fig. 1h). We
163 then investigated *Tbx1*, *Isl1*, and *Nkx2-5* expression at later time points in E8.5 embryos and
164 day 6 gastruloids (Fig. 1i-j). In the embryo, *Nkx2-5* expression labelled the heart tube while *Isl1*
165 and *Tbx1* were expressed in second heart field progenitors, located in pharyngeal mesoderm
166 behind the differentiated heart tube (Fig. 1i). *Nkx2-5* was also expressed in the anterior part of
167 the gastruloid, marking cardiomyocytes. In day 6 gastruloids, *Tbx1* and *Isl1* were expressed in
168 a domain adjacent to that of *Nkx2-5* (Fig. 1j). Similarly, we found that *Tcf21* and *Ebf3* were
169 expressed in pharyngeal mesoderm of E8.5 embryos behind the heart tube expressing *Tnnt2*
170 (Fig. 1k). In day 7 gastruloids, *Ebf3* and *Tcf21* were also found in a close domain of expression
171 adjacent to the cardiomyocytes (*Tnnt2*⁺) (Fig. 1l). Taken together these findings show that
172 cardiopharyngeal mesoderm markers are expressed in a similar spatio-temporal pattern in
173 both gastruloids and mouse embryos.

174 To further investigate whether myoblast differentiation occurs and is faithfully
175 recapitulated in gastruloids, we also compared the expression of myoblast markers in both
176 models. At E10.5, *Myf5* and *MyoD* positive cells were detected in the pharyngeal arches of
177 mouse embryos, close to the heart (Fig. 1m and o). Similarly, *Myf5* and *MyoD* were detected
178 close to cardiomyocytes (*Tnnt2*-positive cells) in gastruloids collected at day 15 (Fig. 1m-p).
179 We noted however that *Myf5* and *MyoD* did not show a compact expression as observed in
180 the embryo indicating the absence of arch morphogenesis. These data support the existence
181 of skeletal myogenesis in gastruloids cultured for more than 10 days. Together, this

182 comparison shows that gastruloids and mouse embryos display similar spatio-temporal gene
183 expression during cardiopharyngeal mesoderm development.

184 185 **Single-cell RNA-sequencing reveals cardiopharyngeal mesoderm subpopulations in** 186 **gastruloids**

187 To further assess the potential of gastruloids to form cardiopharyngeal mesoderm and
188 investigate in detail the different cell populations, we performed single-cell RNA-sequencing
189 (scRNA-seq) on gastruloids following a time course. Gastruloid cells were collected at day 4,
190 day 5, day 6 and day 11. For each time point, at least 8 gastruloids were dissociated into single-
191 cell suspensions and approximately 7,000 cells were sequenced. After quality controls, we
192 analyzed 6,646 cells at day 4, 6,704 cells at day 5, 5,284 cells at day 6 and 8,024 cells at day 11.
193 We performed Leiden clustering and differential gene expression analysis at each time point (Fig.
194 2 and Supplementary Table S1). Integration of the gastruloid single-cell data with the published
195 atlas of mouse embryonic cells ranging from embryonic day (E) 6.5 to E8.5³⁷ (see Methods 4-5)
196 showed that cells collected at day 4, day 5 and day 6 likely overlap with cells of E7.25-E7.75,
197 E8.0-E8.25 and E8.5 embryos respectively (Supplementary Fig. S2), in agreement with our
198 fluorescent *in situ* hybridization experiments (Fig. 1g-p). However, gastruloid cells collected at
199 day 11 do not overlap in the UMAP with cells of the atlas, maybe reflecting distinct transcriptional
200 profiles (Supplementary Fig. S2j-k). Gastruloids cells at day 11 might thus reflect more
201 differentiated states, usually found after E8.5 in mouse embryos. This observation indicates that
202 while we can confidently compare our gastruloid single-cell dataset at day 4, day 5 and day 6 with
203 the cells of the atlas, distinct annotation criteria had to be applied to day 11 clusters. Cell type
204 annotation transfer, from the atlas, was first applied to the primary clusters of day 4, 5 and 6 (See
205 Methods 4). Manual annotation was then performed based on this label transfer and on differential
206 gene expression (Supplementary Table S1). The clusters of day 11 were only annotated manually
207 based on differential gene expression analysis (Supplementary Table S1) and tissue database
208 (<https://tissues.jensenlab.org/About>).

209 To assess the cellular heterogeneity of the gastruloids, we analyzed the different clusters
210 obtained at each time point. At day 4, we observed a significant number of clusters of mesodermal
211 cells (Fig. 2a-b; Supplementary Fig. S3) with four clusters showing high *Mesp1* expression
212 (clusters 1, 3, 5 and 14). Analysis of the single-cell data also showed high expression of the
213 *Gata4* and *Gata6* transcription factors in most clusters (all clusters except clusters 8, 9 and 13)
214 (Fig. 2b; Supplementary Fig. S4). These transcription factors are known to be involved in
215 cardiopharyngeal mesoderm specification^{38,39}. *Hand1*, a marker of the extraembryonic mesoderm
216 and cardiac progenitor cells in the recently identified juxtacardiac field^{40,41} was also highly
217 expressed at day 4 (except in clusters 8, 9 and 10). Interestingly, with scRNA-seq, we detected
218 *Myf7* as early as day 4. Cluster 9 showed expression of the pluripotency marker *Pou5f1* indicative

219 of epiblast-like cells (Fig. 2b; Supplementary Fig. S3). *Mesp2*, the closest homolog of *Mesp1*, was
220 only expressed at very low levels in gastruloids (Supplementary Fig. S4). At day 5, the cell type
221 annotation transfer from the atlas was enriched in mesenchymal and pharyngeal mesodermal
222 cells while nascent mesoderm was barely observed (Supplementary Fig. S5). We identified 5
223 clusters annotated as pharyngeal mesoderm (clusters 1, 3, 6, 7, and 13) (Fig. 2c). The first fully
224 differentiated cells were found in cluster 8. This cluster corresponds to endothelial cells and is
225 marked by the expression of *Sox7* (Fig. 2c-d Supplementary Fig.S5). At this stage, *Mesp1* started
226 to be downregulated while *Gata6* was still highly expressed (Supplementary Fig. S6). Expression
227 of *Isl1* is also detected and *Tbx1* at a very low level (Fig. 2d, Supplementary Fig. S6). At day 6,
228 we uncovered 16 clusters (Fig. 2e, Supplementary Fig. S7). We detected two clusters of
229 differentiating or differentiated cardiomyocytes (clusters 6 and 2) expressing high level of *Myf7*
230 and other cardiac myosin genes. All these data are consistent with the previous report on
231 gastruloids³⁴. We also found 5 clusters of pharyngeal mesodermal cells (clusters 3, 5, 8, 10 and
232 11). Expression of the transcription factors involved in cardiopharyngeal mesoderm development,
233 including *Tbx1*, *Tcf21*, *Lhx2* and *Ebf3*, was also detected (Fig. 2f and Supplementary Fig.S8).
234 These data reveal the emergence of cardiopharyngeal mesoderm in differentiating gastruloids.

235 Finally, to investigate the potential differentiation of cardiopharyngeal mesoderm
236 derivatives in gastruloids, we analyzed scRNA-seq data at day 11. We identified 20 clusters (Fig.
237 2g). One cluster (cluster 20) likely corresponds to visceral endoderm tissue, with expression of
238 *Epcam*, *Sox17*. We also uncovered 4 different clusters (clusters 10, 12, 13 and 17) of ectodermal
239 cells with enrichment of genes linked with neural derivatives, including neural crest-like cells
240 (Supplementary Tables S1). We identified a cluster of endothelial cells, with expression of *Sox7*
241 (cluster 7) (Fig. 2h) and a cluster of blood/hematopoietic lineages (cluster 19). Clusters 1, 2, 3, 4,
242 5, 11 and 14 were annotated as mesoderm and included mesenchymal cells and undifferentiated
243 progenitors (Fig. 2g, Tables S1). The presence of high expression levels of *Myf7* in clusters 6, 8
244 and 9 indicates that these clusters contain cardiomyocytes. Strikingly, we found 2 clusters
245 (clusters 16 and 18) with expression of the myogenic regulatory genes *MyoD* and *Myogenin*
246 (*Myog*) (Fig. 2g-h). Together, our data thus reveal different cardiopharyngeal mesoderm
247 subpopulations in the gastruloid model and show that cardiomyocytes and skeletal myoblasts
248 markers are found in this model.

249

250 **Three subtypes of cardiomyocytes differentiate in gastruloids**

251 In order to investigate whether the cardiomyocyte clusters at day 11 represent different
252 cardiac subpopulations, we performed a detailed analysis of the clusters 6, 8 and 9, i.e., the three
253 clusters containing cardiomyocytes. Among the genes differentially expressed in cluster 6, we
254 found *Actg1*, *Cald1*, *Actb*, *Acta2*, *Vsnl1*, *Gucy1a1*, *Ptma*, *Map1b*, *Shox2* and *Ptn* (Fig. 3a-b and
255 Supplementary Table S2). This cluster thus showed the expression of *Cald1*, *Acta2*, expressed

256 in smooth-muscle cells and throughout the myocardium of the immature heart tube^{42,43} as well
257 as genes involved in sinus venosus/sinus atrial node development (*Vsn1*, *Shox2*, *Tbx3*)^{44–46} (Fig.
258 3a-b). In cluster 9, *Nppa*, *Itga6*, *Myl7*, *Myh6* were among the genes differentially expressed (Fig.
259 3c-d and Supplementary Table S1). These genes are usually enriched in atrial cardiomyocytes.
260 Cluster 8 on the other hand was enriched in genes such as *Myl2*, *Myl3*, *Myh7*, *Mpped2*, *Pln* (Fig.
261 3e-f and Supplementary Table S1). These genes were previously shown to be the signature of
262 ventricular cardiomyocytes in different independent studies^{47–49}. ScRNA-seq thus identified 3
263 cardiomyocyte clusters with distinct transcriptional signatures. We can further speculate that
264 gastruloids contain ventricular, atrial and conductive-like cardiomyocytes.

265 In order to validate these results, we performed lineage tracing with a cell line that labels
266 specific subpopulations of cardiomyocytes in gastruloids. For that purpose, we rederived mESCs
267 from *Mef2c-AHF-enhancer-Cre*; *Rosa^{tdTomato/+}* mouse blastocysts. *Mef2c-AHF-enhancer-Cre*
268 mouse line is known to specifically label progenitors that contribute to outflow tract, right ventricle
269 and a subpopulation of venous pole myocardium (dorsal mesenchymal protrusion) as well as non-
270 cardiac cardiopharyngeal mesoderm derivatives^{50–52} (Fig. 3g). *Mef2c-AHF-enhancer-Cre*;
271 *Rosa^{tdTomato/+}* (*Mef2c-Cre*; *Rosa-tdTomato*) derived mESCs were able to form beating gastruloids
272 with expression of tdTomato in their anterior domain as expected, but without notable spatial
273 organization (Fig. 3h). Flow cytometry at day 11 revealed that *Mef2c-Cre*; *Rosa-tdTomato*-derived
274 gastruloids contain a significant proportion of cardiomyocytes, as shown by the expression of
275 cardiac troponin T (cTnT) (Fig. 3i). Thus, tdTomato⁺ cardiomyocytes represented an average of
276 58.2% (ranging from 30 to 74.4%) of all cardiomyocytes in gastruloids (Fig. 3j). These data
277 indicate that gastruloids can undergo differentiation into two distinct cardiomyocyte fates: either a
278 tdTomato positive second heart field fate (outflow tract, right ventricle for example) or a first heart
279 field enriched tdTomato negative fate.

280 Flow cytometry in E10.5/E11.5 mouse embryos showed less variability with between 0.9
281 and 1.6% of tdTomato⁺ cells. In addition, staining of mouse embryos with cTnT antibody showed
282 that *Mef2c-AHF-enhancer-Cre* derived (tdTomato⁺) cardiomyocytes accounted for about 22.9%
283 of all cardiomyocytes in embryos (Fig. 3k). These data validate the existence of at least two
284 subpopulations of cardiomyocytes in gastruloids with proportions significantly different from
285 mouse embryos.

286 To further validate these data, we performed *in situ* hybridization with cardiomyocyte
287 specific probes followed by flow cytometry using the HCR approach (see Methods). We found a
288 significant shift of the cloud of cells in the flow cytometry chart indicative that a significant
289 proportion of cells *Myl7*-positive alone (corresponding to clusters 6, 8 and 9 of the sc-RNAseq
290 data) or in combination with *Myl2* (differentially expressed in cluster 8) in gastruloids at day 11
291 (Fig. 3l). This result suggests that *Myl7*⁺, *Myl2*⁺ ventricular-like cardiomyocytes and *Myl7*⁺*Myl2*⁻
292 non-ventricular cardiomyocytes are present in gastruloids. This result further validates the

293 scRNAseq analysis and confirms the existence of different subpopulations of cardiomyocytes in
294 gastruloids.

295

296 **Skeletal myogenesis in gastruloids**

297 To ascertain whether gastruloids could generate skeletal muscle cells, we analyzed the
298 two clusters (clusters 16 and 18) with expression of genes associated with myoblast differentiation
299 (Fig. 4a). Differential gene expression between clusters 16 and 18 was not conclusive
300 (Supplementary Table S3). We thus decided to investigate in detail the genes involved in
301 cardiopharyngeal and/or somitic mesoderm muscle differentiation and their expression in the two
302 myoblast clusters (Fig. 4b). We noticed an increased expression of genes associated with muscle
303 precursors such as *Pax7* and *Myf5* in cluster 16 compared to cluster 18, as shown by the violin
304 plots (Fig. 4c-d). While *Met* has been shown to be downregulated along differentiation in some
305 particular cardiopharyngeal or somites derived myoblasts^{53,54}, *Met* was expressed in more cells
306 and at higher levels in cluster 16 than in cluster 18 (Fig. 4e). In contrast, markers of more
307 committed myoblasts such as *MyoD*, *Myog* and *Myh3* were enriched in cluster 18 (Fig. 4f-h).
308 These data indicate that myogenesis occurs in the gastruloid model and that clusters 16 and 18
309 express markers of muscle precursors and committed myoblasts, respectively.

310 In order to dissect the cardiopharyngeal versus somitic origin of these myoblasts, we
311 investigated the expression of transcription factors that were specific to either the
312 cardiopharyngeal or somitic muscle progenitors (Fig. 4b). Expression of the cardiopharyngeal
313 mesoderm genes, *Isl1*, *Pitx2*, *Tbx1* and *MyoR*, was found in both cluster 16 and 18 but only in a
314 small number of cells (Fig. 4i-l). *Tcf21* was not expressed in the myoblast clusters, likely having
315 already been downregulated. *Pax3* expression was also recorded in about 40 out of 243 cells of
316 cluster 16 and in 2 out of 67 cells of cluster 18 (Fig. 4m). Deeper analysis of cluster 16 showed
317 a mutually exclusive expression of *Tbx1* and *Pax3* in the population of muscle precursors (Fig.
318 4n), suggesting the activation of distinct myogenic programs. Our scRNA-seq analysis at day 11
319 thus indicates that both cardiopharyngeal and somitic progenitors undergo myogenesis in the
320 gastruloid model.

321 To validate this data and further support the existence of myogenesis in gastruloids, we
322 performed *in situ* hybridization with skeletal myocyte specific probes followed by flow cytometry
323 using the HCR approach (see Methods). Flow cytometry interestingly showed a shift in the cloud
324 of cells with two cell populations expressing *Myog* (Fig. 4o). As found in the single-cell data, we
325 identified a subpopulation of cells expressing *Myog* alone (cluster 16) and a subpopulation of cells
326 expressing *Myog* and *Myh3* (cluster 18). Overall, these observations demonstrate the existence
327 of convergent cardiopharyngeal and somitic myogenic programs in the gastruloid model.

328

329 **Muscular trajectories are found in gastruloids over time**

330 Time series scRNA-seq data allows the reconstruction of transcriptional trajectories
331 from a progenitor cell state toward differentiated cell states⁵⁵. We hypothesize that we could
332 similarly infer the transcriptional trajectories from cardiopharyngeal mesoderm progenitors
333 toward cardiomyocyte and myoblast cell states. We merged the single-cell data from the 4
334 different time points to create a transcriptional time series of the growing gastruloids (Fig. 5a).
335 We first used CellRank (see Methods)⁵⁶. We focused on a differentiated cell state (called
336 macrostate), that includes cells from the cardiomyocyte cluster 8 from the day 11 dataset, and
337 computed the fate probabilities towards this macrostate. Interestingly, we found high fate
338 probability with cells from clusters 6 and 9 from day 11, as well with clusters annotated as
339 cardiomyocytes at day 6 (Fig. 5b). We then focused on a second macrostate that includes cells
340 from the myoblast cluster 18 from the day 11 dataset. We found high fate probabilities with
341 cells from the myoblast cluster 16 also from day 11 (Fig. 5c). These data indicate the existence
342 of muscular trajectories toward the cardiac and skeletal muscle states.

343 To reconstruct fate decision trees for the various populations, we used URD⁵⁷ (see
344 Methods) as a computational reconstruction method, with the merge of all time points (days 4,
345 5, 6 and 11). URD analysis computed transcriptional trajectories from cells at day 4 (named
346 root) towards the different clusters at day 11 (named tips) (Fig. 5d). We observed 5 branches
347 emerging from day 4. The first branch (branch A) heads towards the ectodermal lineages
348 (clusters 10, 12, 13 and 17) while the last branch (E) heads towards the endoderm lineages
349 (cluster 20). The 3 other branches (branches B, C and D) are mesodermal, and include the
350 endothelial (cluster 7) and hematopoietic lineages (cluster 19). Notably, one of these
351 mesodermal branches, branch D, heads towards the cardiomyocytes (clusters 6, 8 and 9) and
352 myoblast clusters (clusters 16 and 18). Thus, the URD analysis shows that transcriptional
353 trajectories can be found toward both the cardiomyocytes and skeletal muscles.

354 To further dissect these trajectories, we specifically focused on branch D. This branch
355 also includes clusters 4 and 14. These two clusters were previously defined as mesodermal
356 derivatives (Fig. 2g). We investigated gene expression in branches M and H to characterize
357 further these clusters (Fig. 5d). We found that branches M and H showed *Gata6* expression
358 as well as *Hand1* and *Mab21l2*, which mark the juxta-cardiac field^{40,41} (Supplementary Fig.
359 S10). We also showed expression of *Tbx18* and *Wt1* (Supplementary Fig. S10), which mark
360 the proepicardial organ and their derivatives⁵⁸. These data suggest that these clusters might
361 contain epicardial cells. We then investigated how key markers of the cardiopharyngeal
362 mesoderm and their derivatives are expressed in the tree. We found that *Isl1*, *Tcf21*, but also
363 *Lhx2* and *Ebf3* are expressed in branch D, in the root as well as along the trajectories towards
364 the cardiomyocyte and myoblast clusters (Fig. 5f-g and Supplementary Fig. S10). *Myf7* is
365 specifically expressed in the three cardiomyocyte clusters (6, 8 and 9) and the cardiomyocyte
366 branching point I (Fig. 5h), while *Myf2* is expressed after the branching bifurcation only in

367 cluster 8 (Fig. 5i). As expected *Myog* is expressed in branch L that heads towards the myoblast
368 cluster 18 (Fig. 5j). *Tbx1*, *MyoR*, markers of the cardiopharyngeal muscle progenitors, *Pax3*,
369 markers of the somitic muscle progenitors, and the early myogenic transcription factor *Myf5*
370 are all expressed in branch F that heads towards the two myoblast clusters 16 and 18
371 (Supplementary Fig. S10). We also explored where cells of clusters from days 5 and 6 would
372 be found in this tree (Fig. 5k-m). Strikingly, cells of the cardiopharyngeal mesoderm 1 cluster,
373 identified at day 5, were found almost at the root of the tree (branch D), before the branching
374 bifurcation between the cardiomyocytes and myoblast lineages (Fig. 5k). Similarly, cells of the
375 cardiopharyngeal mesoderm 2 cluster, identified at day 6, were found before the branching
376 point (branch D) (Fig. 5l). However, we found that cells of the cardiopharyngeal mesoderm 3
377 cluster, identified at day 6, were restricted to the cardiomyocyte lineages (branch I). These
378 cells are located after the branching bifurcation between the cardiomyocytes and myoblast
379 lineages (branch D) (Fig. 5m). We then focused on the two myoblast clusters identified at
380 day 11. Cells of cluster 16 were also found in branch F, before the branching point that heads
381 toward the myoblast clusters 16 and 18 (Fig. 5n). Cells of cluster 18, on the other hands, were
382 found only after branch bifurcation, in branch L (Fig. 5o). These data further confirm that
383 myoblast cells from cluster 18 are in a more differentiated state than cells from cluster 16.
384 Together these trajectory inferences show that muscular transcriptional trajectories are found
385 in gastruloids. They also show that a specific branch of the tree includes the cardiopharyngeal
386 mesoderm and heads towards both cardiomyocytes and myoblasts clusters.

387

388 **Discussion**

389

390 Our results provide evidence that cardiopharyngeal mesoderm markers are expressed in
391 a similar spatio-temporal pattern in gastruloids and the mouse embryo. We also showed that
392 cardiopharyngeal mesoderm differentiates in the gastruloid model into different subpopulations of
393 cardiomyocytes as well as skeletal myoblasts. Gastruloids thus faithfully recapitulate the
394 developmental timing of cardiopharyngeal mesoderm specification and differentiation (Fig. 6).

395

396 To the best of our knowledge, here we provide the first *in vitro* model for parallel
397 differentiation of cardiomyocytes and skeletal muscles. Indeed, in addition to cardiomyocyte
398 differentiation previously reported by Rossi and colleagues³⁴, our study demonstrates that
399 skeletal myogenesis occurs in gastruloids. Using scRNA-seq analysis we found that the
400 mesodermal origin (cardiopharyngeal or somitic) was not reflected in the clustering of skeletal
401 muscle progenitor cells. In contrast, we identified two clusters that segregate different states of
402 myogenesis. This finding indicates that transcriptional similarity due to the state of myogenesis
403 might predominate over its developmental origin, raising challenges in addressing whether

404 myogenesis in gastruloids derives from cardiopharyngeal or somitic mesoderm. We first showed
405 the expression of key markers of cardiopharyngeal progenitors such as *Tcf21*, *Tbx1*, *Lhx2*, *MyoR*,
406 and *Pitx2*^{16,20,23,24}. Additionally, we identified transcriptional trajectories from the cardiopharyngeal
407 mesoderm towards both myoblasts and cardiomyocytes. These results together with the close
408 proximity between the cardiomyocytes and myoblasts in gastruloids indicate that skeletal
409 myogenesis, at least in part, occurs from the cardiopharyngeal mesoderm. Based on the exclusive
410 expression of *Tbx1* (cardiopharyngeal marker) and *Pax3* (somitic marker) in myoblast cluster 16
411 we propose that myoblasts also derive from somitic mesoderm and that gastruloids thus allow
412 skeletal myogenesis of both head and trunk programs. Furthermore, transcriptomic trajectories
413 computed with URD showed a branch with both cardiomyocytes and myoblasts. But expression
414 of both somitic (*Pax3*) and cardiopharyngeal mesoderm markers is found in that particular branch.
415 This highlights the known limitations with the use of single-cell transcriptomics approaches for
416 lineage inference. ScRNAseq alone provides information only on gene expression, and attempts
417 to reconstruct lineage trajectories from transcriptomic data can be misleading and do not allow a
418 complete lineage reconstruction⁵⁹. Only clonal analysis will confirm whether cardiopharyngeal
419 mesoderm bipotent progenitors are found in gastruloids but this requires specific tools and
420 analysis.

421
422 Interestingly, different subpopulations of cardiomyocytes were identified in gastruloids.
423 Gene expression analysis suggests that these subpopulations correspond to ventricular (*Myf2*+,
424 *Myf3*+, *Myf7*+), atrial (*Myf6*+, *Nppa*+) and immature/conductive cardiomyocytes (*Shox2*+,
425 *Vsnl1*+) ^{44-47,49}. However, atrial and ventricular myocardium share many genes in common during
426 development. They are distinguished later by acquisition of their mature electro-mechanical
427 properties (conduction, voltage tension, some structural proteins and ions channels). Thus, few
428 genes are differentially expressed *in vivo*: *Myf7* is only restricted to atrial myocardium in fetal
429 mouse embryo, while *Myf2* is restricted in ventricular myocardium. Finally, *Nppa* is detected in the
430 atrium and the ventricle during development, but its expression became atrial specific immediately
431 before birth⁶⁰. It is thus not clear how mature are the cardiomyocytes in gastruloids. Further work
432 will be required to characterize these different subpopulations, their spatial organization and their
433 origin.

434
435 Our results reveal robust myogenic differentiation to heart and head muscle in the
436 gastruloid system. This is supported by reproducibility with similar observations from different
437 replicates. We also showed redundancy with similar conclusions obtained when using scRNA-
438 seq, RNAscope, HCR, lineage tracing and immunofluorescence. The use of different mESC lines
439 showed the robustness of the gastruloids for the specification of the cardiopharyngeal mesoderm.
440 However, our gastruloid model also presents some limitations. In contrast to gastruloids from

441 human induced pluripotent stem cells, which have been shown to form a heart-tube like structure
442 ³³ or human cardioids which have been shown to form multi-chambers heart like structures ⁶¹, our
443 murine gastruloid model does not show robust heart morphogenesis. We also have not observed
444 cardiac crescent like structures as was previously described ³⁴. This indicates that further studies
445 and protocol adaptations are required to form these different chamber structures and to improve
446 cardiac morphogenesis. Interactions between pharyngeal endoderm and the heart are known to
447 be crucial for heart morphogenesis ^{62,63}. We hypothesize that with the robust development of
448 cardiopharyngeal mesoderm in our study, the spatial proximity between the cardiac derivatives
449 and the endoderm/gut structure might be perturbed. Affecting the proximity with the endoderm
450 might impact cardiac morphogenesis.

451
452 Morphogenesis and physical forces are also different in gastruloids, compared to embryos.
453 Our results indicate that while cardiopharyngeal mesoderm is formed in gastruloids, there is no
454 indication of pharyngeal arch morphogenesis. Further studies are required to determine how
455 these transient structures are formed in the embryo. The gastruloid model offers a possibility to
456 reconstruct such processes and investigate whether physical constraints could play a role as
457 shown for example with matrigel embedding for somitogenesis ^{35,36}. Additionally, we know that
458 pharyngeal arches are also colonized by neural crest cells, which are largely absent in the
459 gastruloid model ³². Adapting the protocol to support both neural crest cell and cardiopharyngeal
460 mesoderm development might improve pharyngeal arch morphogenesis. While cranial/cardiac
461 neural crest cells are known to be critical for outflow tract and valve development ⁶⁴⁻⁶⁶ as well as
462 for the patterning of head skeletal muscles ⁶⁷, our results support that neural crest cells are not
463 required for cardiac and cardiopharyngeal mesoderm specification.

464
465 As previously reported, we have also noted variability among the lines of mESC used.
466 Notably, mESC derived from *Mef2c-AHF-Cre/tdTomato* blastocysts seem to have less potential
467 for skeletal myogenesis. We speculate that it could be due to a different genetic background.
468 Alternatively, established and rederived mESC lines from mouse embryos might not strictly have
469 the same pluripotency status. In particular, we found that different mESC clones exhibited a truly
470 different beat and myogenic potential. Due to the lack of specific markers, it was also difficult to
471 address whether other cardiopharyngeal mesoderm derivatives such as connective tissue or
472 smooth muscle ^{51,53,68}, could develop in gastruloids.

473
474 Several reports have already indicated that gastruloids, somitoids, segmentoids and
475 embryonic trunk-like structures, from mouse or human pluripotent stem cells, could faithfully
476 recapitulate paraxial mesoderm early specification with the formation of somites ^{35,36,69-71}. None
477 of these studies however showed the initiation of myogenesis. However, most of their protocols

478 stopped after 5 to 7 days of culture whereas we have extended the culture of gastruloids up to 15
479 days. Our kinetics of expression showed that markers of myogenesis are initiated only after day
480 8. This suggests that somitoids, segmentoids and embryonic trunk-like structures might also have
481 the potential to undergo myogenesis if cultured for a longer period of time.

482
483 Together our results show the relevance of the gastruloid model to investigate
484 cardiopharyngeal mesoderm specification and differentiation, and demonstrate that this *in vitro*
485 model could be used as an alternative model to complement or reduce the use of animal models.
486 The protocol is relatively simple and allows the culture of a large number of gastruloids in a single
487 experiment. It thus offers a powerful *in vitro* platform to model congenital heart disease and
488 muscle dystrophies (22q11.2DS for example), and to deconstruct cardiopharyngeal mesoderm
489 specification as well as myogenesis. Cardiopharyngeal mesoderm development is still not fully
490 understood *in vivo*. The field lacks reporter lines and face difficulties to carefully track the
491 progenitors and their derivatives along a course of about 3 to 5 days, from the emergence of the
492 progenitors at gastrulation to their differentiation into cardiomyocytes or skeletal myoblasts.
493 Cardiopharyngeal mesoderm-containing gastruloids additionally offer the possibility for high
494 resolution live imaging and cell tracking. Furthermore, the model will likely be useful for large scale
495 genetic and drug screens to better understand and potentially identify therapeutic strategies for
496 congenital diseases affecting head and heart muscles. In this context, we expect gastruloids to
497 become indispensable tools in the field of cardiac and skeletal muscle specification to address
498 the mechanisms of lineage specification as early as gastrulation.

499 **Methods**

500

501 **Mice colonies**

502 *Mef2c-AHF-Cre* and *Rosa-tdTomato* (*Gt(ROSA)26Sor^{tm9(CAG-tdTomato)Hze}*) mice were previously
503 described^{50,72}. We used CD1/Swiss mice (Charles River) as wildtype animals. Mouse colonies
504 were maintained in certified animal facilities (agreement #C 13 013 08) in accordance with
505 European guidelines. The experiments were approved by the local ethical committee (CEBEA)
506 and the study is compliant with all relevant ethical regulations regarding animal research
507 (Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche;
508 Authorization N 32-08102012).

509

510 **Cell culture**

511 R1/E #1036 mouse embryonic ES cells were purchased from ECACC (cat. Number
512 07072001). Zx1 mouse ESCs were obtained from M. Kyba⁷³. *Mef2c-cre Rosa tdTomato* cells
513 were derived in the lab from E3.5 blastula collected from mouse⁵⁰. All cell lines were maintained
514 on 0.1% gelatin coated labware in GMEM-BHK21 medium (Gibco ref 21710-025) supplemented
515 with 8% Fetal Bovine Serum (Pansera, ref P30-2602), 1% non-essential amino acids (Gibco
516 11140-035), 1% Sodium Pyruvate (Gibco ref 11360-039), 1X Penicillin/Streptomycin (Gibco ref
517 15070-063), 0.1 mM beta-mercaptoethanol (Sigma ref M3148-100ML). This basal medium was
518 completed with 5ng/ml recombinant mouse LIF (Sigma ref LIF2050), 3µM CHIR99021 (Sigma
519 ref SML1046-5MG) and 1µM PD0325901 (Sigma ref PZ0162-5MG). Cells were passaged at least
520 twice but no more than 10 times before differentiation.

521

522 **Gastruloid production**

523 Gastruloids were produced according to previously described protocols^{32,34} with slight
524 modifications (centrifugation for aggregation and culture in N2B27 medium after day7). Briefly,
525 mES reaching a 60% confluency were dissociated with 0,5% Trypsin-EDTA (Gibco ref 25300-
526 054). After enzyme neutralization, cells were carefully washed 2 times with D-PBS (Gibco ref
527 14190-094) and resuspended in N2B27 medium (Neurobasal Medium (Gibco ref 21103049),
528 DMEM/F12 (Gibco ref 11320074), 0,5% B27 supplement (Gibco ref 17504-044), 0,5% N2
529 supplement (Gibco ref 17502-048), 1X Glutamax (Gibco ref 35050-038), 0.1 mM beta-
530 mercaptoethanol (Sigma ref M3148-100ML). 300 to 600 individual cells were distributed into 40µl
531 of N2B27 in each well of a Ultra-Low-Attachment 96 well round bottom plate (Corning ref 0707).
532 Plates were then centrifuged at 100g for 5 min and placed in a 5% CO2 incubator during 48 hours.
533 Then, 150 µl of N2B27 with 3µM CHIR99021 were added in each well and plates were incubated
534 for 24h. After this period, medium was replaced by N2B27 without CHIR99021 for another 24h.
535 Then from day 4 to day 7, N2B27 was supplemented with 30 ng/ml bFGF (R&D ref 233-FB), 5

536 ng/ml VEGF (Gibco ref PHC9394) and 0,5mM Ascorbic Acid (Sigma ref 255564-5G) and culture
537 plates were placed on an orbital shaker at 100 rpm. After day 8, N2B27 was used and changed
538 every other day.

539

540 **Gene expression analysis by QPCR**

541 Day 0 undifferentiated mES cells and different time points gastruloids were collected during the
542 experiments. Samples were then lysed in Buffer RA1 and RNA was extracted with NucleoSpin
543 RNA kit (Macherey-Nagel ref 740902.250). Total RNAs were then retrotranscribed to cDNA using
544 AffinityScript Multiple Temperature cDNA Synthesis Kit (Agilent ref 200436) and QPCR were run
545 on QuantStudio 5 according to PowerUp™ SYBR® Green instructions (Thermo ref A25742). List
546 of QPCR primers can be found in supplementary information (Table S4). Data were all
547 standardized using *Tbp* as a housekeeping reference gene and on day 0 results and gene
548 expression fold changes were calculated with the $\Delta\Delta\text{Ct}$ method.

549

550 **Single cell RNA-sequencing analysis**

551 Gastruloids collected on day 4, 5, 6 and 11 were dissociated by incubation 10 min at 37°C in 0.5%
552 Trypsin-EDTA. Cells were washed 2 times with D-MEM + 10% FBS and filtrated through a 40µm
553 cell strainer. Cell count and viability were determined on a Countess II (Thermo ref AMQAX100).
554 Single cell-RNA seq libraries were made from 10,000 individual cells using 10X Genomics
555 Chromium Next GEM Single Cell 3' kit v3.1 (PN-1000268) and the Chromium Controller. Libraries
556 were then sequenced on an Illumina NextSeq 500 at the MMG Sequencing platform.

557

558 Single-cell analysis was performed using 1) Seurat, the R toolkit for single-cell genomics⁷⁴, 2)
559 CellRank, to discover the cellular dynamics⁵⁶, and 3) URD to obtain transcriptional trajectories⁵⁷.
560 The analysis was implemented in R v4.1.3 and Python v3.9.10.

561

562 **Preparation of the Pijuan-Sala et al. atlas**

563 The Pijuan-Sala et al. atlas has been constructed from mouse embryos to describe gastrulation
564 and early organogenesis³⁷. It contains single-cell analysis at multiple developmental stages
565 between 6.5 to 8.5 days post-fertilization³⁷. The atlas was subsetted to select the cells with more
566 than 200 feature gene counts and the features detected in at least 4 cells. Then, normalization
567 was performed on the feature counts with the “LogNormalize” method. An atlas must be
568 comparable to the dataset under study. Here, and because we work on gastruloids mimicking
569 gastrulation, the stage “mixed_gastrulation” and the cells annotated “endoderm” or
570 “extraembryonic endoderm” were removed from the atlas. Finally, the identification of the highly
571 variable genes, the data scaling, and the PC analysis and UMAP dimensional reduction were
572 performed as detailed in step 2 of the next section.

573

574 **Processing of the four individual single-cell datasets**

575 Four single-cell RNA sequencing datasets were generated from gastruloids' cells collected at days
576 4, 5, 6, and 11. Each dataset was analyzed with the R package Seurat v4.1.1 using the following
577 six steps ⁷⁴.

578

579 **1. Quality controls**

580 Low quality cells were filtered out with different cutoffs on gene expression. All these cutoffs were
581 set manually and independently on each dataset. The proportion of transcripts mapping to
582 mitochondrial and ribosomal genes were evaluated. We set cutoffs on high mitochondrial and low
583 ribosomal gene expression. In addition, we set two cutoffs on the UMI counts of each cell. Finally,
584 the data were subsetted to keep cells with more than 200 feature gene counts and features
585 detected in at least 4 cells.

586

587 **2. Preprocessing workflow**

588 Feature counts were normalized using the "LogNormalize" method. Highly variable genes were
589 identified with the "vst" selection mode, and selected the 2000 most variable genes. Then, the
590 data were scaled on all the features and Principal Component (PC) analysis dimension reduction
591 was performed. The 30 first PCs were kept for the rest of the analysis. To visualize the data,
592 Uniform Manifold Approximation and Projection (UMAP) was applied ⁷⁵. Finally, the normalized
593 data were structured with a shared nearest neighbor graph. The number of nearest neighbors
594 was set to 20 (default). We used the Leiden algorithm with the resolution set to 1 to cluster the
595 graph ⁷⁶.

596

597 **3. Doublets detection**

598 Doublets detection was carried out with the R package DoubletFinder v2.0.3 ⁷⁷. A cell type
599 annotation transfer was first applied. Indeed, according to DoubletFinder recommendations, a cell
600 type annotation transfer allows a better estimate of the number of detectable doublets ⁷⁷. The
601 Seurat tutorial "Mapping and annotating query datasets" was applied to transfer the annotation of
602 cell types, using the atlas of Pijuan-Sala et al. as a reference ^{37,74}. We estimated the number of
603 heterotypic doublets from the percentage of doublets of the used reagent kit of 10x Genomics
604 Chromium 3' chemistry v3.1. In our case, the rate of doublets amounts to 8%. For each dataset
605 independently, the pK parameter was optimized with the function paramSweep_v3. The first 30
606 PCs were considered. As a result of the process, every cell was labeled as either singlet or
607 doublet. We selected the cells labeled as singlets, and reapplied the Step 2 preprocessing
608 workflow to the singlet subset.

609

610 **4. Cell type annotation transfer from the atlas**

611 Like for doublets detection, cell type annotations were transferred using the Seurat functions with
612 default parameters ⁷⁴.

613

614 **5. Global structure-preserving UMAP plots**

615 The global structure-preserving UMAP dimension reduction was set up using the R package uwot
616 v0.1.14 ⁷⁵. The aim was to keep a similar visualization structure among the different datasets. The
617 reference global structure was created using the Pijuan-Sala et al. atlas ³⁷. We first integrated our
618 dataset with the Pijuan-Sala et al. atlas using the Seurat framework for datasets integration with
619 default parameters.

620 Besides, UMAP dimension reduction was carried out on a subset of the atlas containing 200 cells
621 per cell type. From the resulting UMAP coordinates, the centroid of each cell type was calculated.
622 We randomly assigned positions to all the cells of the integrated object (dataset and atlas) around
623 the centroid of their corresponding cell type. The UMAP algorithm was run on the integrated
624 object, with the random positions around cell type centroids as the starting points. In a following
625 step, the UMAP algorithm was run again on the dataset only, with the dataset's UMAP coordinates
626 in the integrated object as the starting points.

627

628 **6. Differential expression analysis**

629 Differentially expressed genes (DEGs) were determined using a Wilcoxon Rank Sum test with the
630 Seurat function FindAllMarkers. The function compares each cluster (group 1) to the rest of the
631 data (group 2). Default values were used for min.pct, logfc.threshold, and return.thresh
632 parameters: min.pct allows to speed up the function by selecting features that are expressed over
633 a given proportion of cells within each group, return.thresh selects p-values lower than a given
634 threshold, and logfc.threshold selects features showing at least a 2-fold change between the two
635 groups of cells. We selected the positive markers only.

636 For each cluster, we also retrieved the most represented cell type and its corresponding
637 percentage of cells in the cluster (see Table S1).

638

639 **Processing of the merged single-cell datasets**

640 The four single-cell datasets were also merged. The datasets were reduced to the previously
641 identified high-quality cells (see steps 1 and 3 of the processing pipeline). The identification of the
642 highly variable genes, the data scaling, and the PCA and UMAP dimensional reductions were
643 performed as described in step 2 from the previous section. Steps 4 to 6 were then applied to the
644 merged dataset.

645

646 **Day11 differential expression analysis for cardiomyocytes and myoblasts characterization**

647 Additional differential expression (DE) analyses were performed on two subsets of the day11
648 dataset. First, to better characterize the cardiomyocytes, the clusters identified as cardiomyocytes
649 (ie. clusters 6, 8, and 9) were compared. In this case, each cluster was compared to the two
650 others. The cluster 18 of day 11 was also identified as a cluster of cardiomyocytes (see Table
651 S1). However, the day 11 is poorly annotated because the reference atlas does not reach this
652 time points. We manually assigned the “myoblast” label to this cluster. DE analysis was performed
653 between the clusters 16 and 18, two clusters of myoblast cells. In both analyses, the parameters
654 used for the DE analyses were the same as detailed in step 6 of the previous section.

655

656 **Data preparation for file format conversion**

657 CellRank and URD necessitate different file format. We started from raw data and removed the
658 cells and features that were filtered out during the preprocessing analysis of the four datasets
659 (step 1 and 3). The obtained datasets were then merged and the feature counts were normalized
660 using the “LogNormalize” method. The object was saved as a .rds file (R).

661

662 **Cell Rank time series framework**

663 In order to use CellRank⁵⁶ v1.5.1, the Seurat⁷⁴ object (R) was converted to an AnnData⁷⁸ v0.7.8
664 object (Python). The prepared dataset was converted into .h5ad (Python) (see Data preparation
665 for file format conversion section).

666 The .h5ad object was preprocessed with loading, normalization, logarithmization, highly variable
667 genes identification, PC analysis, and KNN graph computing, using the Scanpy v1.9.1 tool⁷⁹. We
668 added the metadata of UMAP coordinates, cell type assignation, and Leiden clustering to each
669 cell. CellRank was used with the Waddington Optimal Transport method (WOT), the time series
670 flavor of CellRank⁸⁰. The WOT kernel was then computed. Due to insufficient proliferation genes
671 in highly variable genes, we were not able to compute the initial growth rates. Hence, according
672 to CellRank specifications, the transition matrix was computed with the parameter `growth_iters =`
673 `3`⁵⁶. A symmetric transition matrix was also computed based on transcriptomic similarity, called
674 connectivity kernel. WOT kernel and connectivity kernel both contributed to the estimator with the
675 respective proportions of 0.8 and 0.2. The estimator was initialized with the Generalized Perron
676 Cluster Cluster Analysis (GPCCA) method⁸¹. Based on Schur decomposition, 6 macrostates
677 were computed⁸². Then, fate probabilities were computed towards the macrostates with the
678 function `compute_absorption_probabilities`, with default parameters. Putative driver genes were
679 identified with the function `compute_lineage_drivers`.

680

681 **Trajectory inference**

682 The R package URD v1.1.1 was used for trajectory inference⁵⁷. The prepared dataset was
683 converted into an urd object (see Data preparation for file format conversion section). The variable

684 features within each dataset were identified with the findVariableGenes function. A KNN graph
685 was constructed with distances between cells' gene expression, considering the 60 nearest
686 neighbors of each cell. We removed the outliers, identified visually according to the distance
687 between the first and the 60th neighbours of each cell. 92 outlier cells were removed from the
688 analysis. Then, a diffusion map was computed with the parameters sigma=3 and knn=60. The
689 pseudotime was simulated 5 times starting from any cell of the day 4 and the results were
690 averaged. All the clusters of day11 were defined as tips. The function
691 pseudotimeDetermineLogistic was used with the parameters optimal.cells.forward=40 and
692 optimal.cells.backward=80 to bias the input transition matrix of the random walk simulations. 10000
693 random walk simulations were carried out. Finally, the tree was constructed by merging the tips
694 as soon they share a common path among all the random walk simulations. To build the tree, the
695 parameters of the function buildTree were set as follows: bins.per.pseudotime.window=8,
696 cells.per.pseudotime.bin=28, divergence.method='preference', p.thresh=0.001.

697

698 **RNAscope experiments on sections**

699 Gastruloids collected at different time points were washed once in PBS + 0,01% Tween 20
700 (PBST) and then incubated in 4% PFA overnight at 4°C. After 3 washes in PBST, approximately
701 8 Gastruloids were embedded in heat liquified Histogel (Eprelia HG-4000-012). Solidified blocks
702 were then post-fixed with 4% PFA during 1h at room temperature and dehydrated with successive
703 incubations of 30 min in Ethanol / PBST: 70%, 96%, twice 100%. Xylene was then added twice
704 for 15 min at room temperature. Histogel blocs were finally embedded in heat liquid Paraplast
705 plus (Sigma ref P3683-1KG). Tissue sections cut at 7 µm were then processed according to the
706 protocol of the RNAscope Multiplex Fluorescent v2 kit (ACD-Bio cat. no.323110). The following
707 probes were used: *mm-Ebf3-C3* (576871-C3), *mm-Is1-C2* (cat no. 451931-C2), *mm-Mesp1-C2*
708 (cat no. 436281-C2), *mm-Mesp1-C3* (cat no. 436281-C3), *mm-Myf5-C1* (cat no. 492911), *mm-*
709 *Myod1-C2* (cat no. 316081-C2), *mm-Nkx2-5-C2* (cat no. 428241-C2), *mm-Tbx1-C1* (cat no.
710 481919), *mm-Tcf21-C2* (cat no. 508668-C2). Sections were imaged using an LSM800 confocal
711 microscope (Zeiss).

712

713 **Flow cytometry**

714 Gastruloids collected on day 11 were dissociated by incubation 10 min at 37°C in 0,5% Trypsin-
715 EDTA. Cells were washed 2 times with D-MEM + 10% FBS and filtrated through a 40µ cell
716 strainer (Falcon 352340). After a PBS wash, 1 million cells were incubated for 20 min at room
717 temperature with Zombie viability dye (Biolegend ref 77143). Then cells were incubated in
718 PBS + 10% FBS to block Fc receptors, followed by a permeabilization/fixation with
719 Cytofix/Cytoperm BD kit (ref 554714). After washing, anti-cTnt antibody (Thermo ref MA5-
720 12960) was incubated at a 1/400 dilution in BD wash buffer during 20 min at room temperature.

721 Then cells were again washed and followed by an incubation with anti-mouse IgG coupled to
722 PE-Cy7 (Biolegend ref 406613) at a 1/200 dilution in BD wash buffer during 20 min at room
723 temperature. After three washes, cells were resuspended in PBS + 1% BSA (Sigma ref A8412-
724 100ML), filtered on a 40 μ cell strainer and analyzed on a BD FACS Aria II cytometer.

725 Gastruloids made with Mef2c-cre Rosa-tdTomato were prepared by the same procedure up to
726 Zombie viability dye labelling and directly analyzed on FACS Aria II to allow sorting of the
727 Tomato positive and negative cells. RNA from sorted cells were extracted with NucleoSpin
728 RNA extraction kit.

729

730 **HCR RNA flow analysis**

731 Gastruloids collected on day 11 were dissociated by incubation 10 min at 37°C in 0,5% Trypsin-
732 EDTA. Cells were washed 2 times with PBS + 10% FBS and filtrated through a 40 μ cell. After
733 PBS wash, 1 million cells were incubated for 20 min at room temperature with Zombie viability
734 dye. Cells were then fixed by incubation with 4% PFA (EMS15714) during 1h at room
735 temperature. HCR multiplexed RNA detection was realized according to Molecular Instruments
736 protocols. Cells were washed in PBST: PBS + 0,1% Tween 20, resuspended in 70% Ethanol
737 and kept overnight at 4°C. The day after, cells were washed twice in PBST and incubated at
738 37°C in hybridization buffer. Probes specific for the different genes were then added at 16nM
739 concentration and incubated overnight at 37°C. Cells were then washed 5 times with wash
740 buffer and 2 times with 5x SSC (Gibco ref 15557-044) + 0,1% Tween 20. Then incubation of
741 the cells with amplification buffer was done during 30 min at room temperature. Amplification
742 hairpin h1 and h2 were prepared separately by heating at 95°C for 90 sec and cooling to room
743 temperature for 30 min. Then hairpins were pooled at 60 nM and incubated with cells overnight
744 at room temperature. After 6 washes with 5x SSC + 0,1% Tween 20. Samples were filtered on
745 Flowmi 70 μ m (Bel-Art H136800070) and analyzed on a Beckman Coulter Cytoflex LX
746 (AMUTICYT platform). The following HCR probes were used: *Myl2-B2*, *Myl7-B3*, *Myh3-B2* and
747 *Myog-B1* with the following amplifier: B1-Alexa647, B2-Alexa546 and B3-Alexa488.

748 **Data availability**

749 Single cell data analysis can be found at:
750 https://github.com/BAUDOTlab/gastruloid_timeserie_scrna-seq.

751

752 **Acknowledgments**

753 We want to thank M. Kyba for providing the Zx1 mES cell line. Single-cell isolation, sequencing
754 and alignment were performed with Veracyte or in the MMG GBIM facility. We thank the cells and
755 animal imaging platform and the animal phenotyping core platform (MMG). Flow Cytometry was
756 performed on the Amu-Flow Cytometry platform (Manon Richaud) and AMUTICYT platform
757 (Stéphane Robert). We thank Lionel Spinelli and the CIML bioinformatics platform for expert
758 advice on bioinformatics analyses. We thank Robert Kelly and Anabela Bensimon-Brito for careful
759 reading of the manuscript.

760

761 **Competing interests**

762 The authors declare no competing or financial interests.

763

764 **Author contributions**

765 L.A., F.L., S.Z designed the experiments. L.A. and C. Choquet performed most of the biological
766 experiments. C. Chevalier performed bioinformatic analysis for all the single-cell RNA-seq data.
767 L.A. performed HCR and FACS experiments. C. Choquet performed RNAscope on gastruloids
768 and mouse embryos. N.N. performed the immunofluorescence on Mef2c-Cre mouse embryos.
769 L.A. and N.N rederived Mef2c-Cre/tdTomato mouse ESCs. A.G. provided technical support. F.L.,
770 S.Z. and A.B. wrote the manuscript. All authors read and approved the final manuscript.

771

772 **Funding**

773 F.L.'s laboratory was supported by the INSERM ATIP-Avenir program. C. Choquet has been
774 supported by the Fondation Lefoulon Delalande and N. Nandkishore by the Fondation pour la
775 Recherche Médicale (FRM). A.B. and C. Chevalier received funding from the "Association
776 Française contre les Myopathies" [MoThARD-Project]. Computations were run on the Core
777 Cluster of the Institut Français de Bioinformatique (IFB) (ANR-11-INBS-0013). S.Z.'s laboratory
778 was supported by the Agence Nationale pour la Recherche (ANR-Heartbox) and the "Association
779 Française contre les Myopathies" [MoThARD-Project].

780

781 **References**

782

- 783 1. Nomaru, H. *et al.* Single cell multi-omic analysis identifies a Tbx1-dependent
784 multilineage primed population in murine cardiopharyngeal mesoderm. *Nat Commun* **12**, 6645
785 (2021).
- 786 2. Buckingham, M., Meilhac, S. & Zaffran, S. Building the mammalian heart from two
787 sources of myocardial cells. *Nat Rev Genet* **6**, 826–835 (2005).
- 788 3. Meilhac, S. M., Esner, M., Kelly, R. G., Nicolas, J.-F. & Buckingham, M. E. The Clonal
789 Origin of Myocardial Cells in Different Regions of the Embryonic Mouse Heart. *Developmental*
790 *Cell* **6**, 685–698 (2004).
- 791 4. Kelly, R. G., Brown, N. A. & Buckingham, M. E. The Arterial Pole of the Mouse Heart
792 Forms from Fgf10-Expressing Cells in Pharyngeal Mesoderm. *Developmental Cell* **1**, 435–440
793 (2001).
- 794 5. Zaffran, S., Kelly, R. G., Meilhac, S. M., Buckingham, M. E. & Brown, N. A. Right
795 Ventricular Myocardium Derives From the Anterior Heart Field. *Circulation Research* **95**, 261–
796 268 (2004).
- 797 6. Waldo, K. L. *et al.* Conotruncal myocardium arises from a secondary heart field.
798 *Development* **128**, 3179–3188 (2001).
- 799 7. Mjaatvedt, C. H. *et al.* The Outflow Tract of the Heart Is Recruited from a Novel Heart-
800 Forming Field. *Developmental Biology* **238**, 97–109 (2001).
- 801 8. Lescroart, F., Dumas, C. E., Adachi, N. & Kelly, R. G. Emergence of heart and
802 branchiomic muscles in cardiopharyngeal mesoderm. *Experimental Cell Research* **410**,
803 112931 (2022).
- 804 9. Diogo, R. *et al.* A new heart for a new head in vertebrate cardiopharyngeal evolution.
805 *Nature* **520**, 466–473 (2015).
- 806 10. Tzahor, E. Heart and craniofacial muscle development: A new developmental theme of
807 distinct myogenic fields. *Developmental Biology* **327**, 273–279 (2009).
- 808 11. Vyas, B., Nandkishore, N. & Sambasivan, R. Vertebrate cranial mesoderm:
809 developmental trajectory and evolutionary origin. *Cell. Mol. Life Sci.* **77**, 1933–1945 (2020).
- 810 12. Lescroart, F. *et al.* Clonal analysis reveals common lineage relationships between head
811 muscles and second heart field derivatives in the mouse embryo. *Development* **137**, 3269–
812 3279 (2010).
- 813 13. Lescroart, F. *et al.* Clonal analysis reveals a common origin between nonsomite-
814 derived neck muscles and heart myocardium. *Proc Natl Acad Sci USA* **112**, 1446–1451 (2015).
- 815 14. Lescroart, F. *et al.* Early lineage restriction in temporally distinct populations of Mesp1
816 progenitors during mammalian heart development. *Nat Cell Biol* **16**, 829–840 (2014).
- 817 15. Stolfi, A. *et al.* Early Chordate Origins of the Vertebrate Second Heart Field. *Science*
818 **329**, 565–568 (2010).
- 819 16. Grimaldi, A. & Tajbakhsh, S. Diversity in cranial muscles: Origins and developmental
820 programs. *Current Opinion in Cell Biology* **73**, 110–116 (2021).
- 821 17. Tajbakhsh, S., Rocancourt, D., Cossu, G. & Buckingham, M. Redefining the Genetic
822 Hierarchies Controlling Skeletal Myogenesis: Pax-3 and Myf-5 Act Upstream of MyoD. *Cell* **89**,
823 127–138 (1997).
- 824 18. Theis, S. *et al.* The occipital lateral plate mesoderm is a novel source for vertebrate
825 neck musculature. *Development* **137**, 2961–2971 (2010).
- 826 19. Harel, I. *et al.* Distinct Origins and Genetic Programs of Head Muscle Satellite Cells.
827 *Developmental Cell* **16**, 822–832 (2009).
- 828 20. Harel, I. *et al.* Pharyngeal mesoderm regulatory network controls cardiac and head
829 muscle morphogenesis. *Proceedings of the National Academy of Sciences* **109**, 18839–18844
830 (2012).
- 831 21. Sambasivan, R. *et al.* Distinct Regulatory Cascades Govern Extraocular and
832 Pharyngeal Arch Muscle Progenitor Cell Fates. *Developmental Cell* **16**, 810–821 (2009).
- 833 22. Kelly, R. G., Jerome-Majewska, L. A. & Papaioannou, V. E. The del22q11.2 candidate
834 gene Tbx1 regulates branchiomic myogenesis. *Human Molecular Genetics* **13**, 2829–2840

- 835 (2004).
- 836 23. Razy-Krajka, F. *et al.* Collier/OLF/EBF-Dependent Transcriptional Dynamics Control
837 Pharyngeal Muscle Specification from Primed Cardiopharyngeal Progenitors. *Developmental*
838 *Cell* **29**, 263–276 (2014).
- 839 24. Wang, W. *et al.* A single-cell transcriptional roadmap for cardiopharyngeal fate
840 diversification. *Nat Cell Biol* **21**, 674–686 (2019).
- 841 25. Chabab, S. *et al.* Uncovering the Number and Clonal Dynamics of Mesp1 Progenitors
842 during Heart Morphogenesis. *Cell Reports* **14**, 1–10 (2016).
- 843 26. Murry, C. E. & Keller, G. Differentiation of Embryonic Stem Cells to Clinically Relevant
844 Populations: Lessons from Embryonic Development. *Cell* **132**, 661–680 (2008).
- 845 27. Kattman, S. J., Adler, E. D. & Keller, G. M. Specification of Multipotential Cardiovascular
846 Progenitor Cells During Embryonic Stem Cell Differentiation and Embryonic Development.
847 *Trends in Cardiovascular Medicine* **17**, 240–246 (2007).
- 848 28. Chan, S. S.-K. *et al.* Development of Bipotent Cardiac/Skeletal Myogenic Progenitors
849 from MESP1+ Mesoderm. *Stem Cell Reports* **6**, 26–34 (2016).
- 850 29. Nandkishore, N., Vyas, B., Javali, A., Ghosh, S. & Sambasivan, R. Divergent early
851 mesoderm specification underlies distinct head and trunk muscle programmes in vertebrates.
852 *Development* **145**, dev160945 (2018).
- 853 30. Turner, D. A. *et al.* Anteroposterior polarity and elongation in the absence of
854 extraembryonic tissues and spatially localised signalling in *Gastruloids*, mammalian
855 embryonic organoids. *Development* dev.150391 (2017) doi:10.1242/dev.150391.
- 856 31. van den Brink, S. C. *et al.* Symmetry breaking, germ layer specification and axial
857 organisation in aggregates of mouse embryonic stem cells. *Development* **141**, 4231–4242
858 (2014).
- 859 32. Beccari, L. *et al.* Multi-axial self-organization properties of mouse embryonic stem cells
860 into gastruloids. *Nature* **562**, 272–276 (2018).
- 861 33. Olmsted, Z. T. & Paluh, J. L. A combined human gastruloid model of cardiogenesis and
862 neurogenesis. *iScience* **25**, 104486 (2022).
- 863 34. Rossi, G. *et al.* Capturing Cardiogenesis in Gastruloids. *Cell Stem Cell* **28**, 230-240.e6
864 (2021).
- 865 35. van den Brink, S. C. *et al.* Single-cell and spatial transcriptomics reveal somitogenesis
866 in gastruloids. *Nature* **582**, 405–409 (2020).
- 867 36. Veenvliet, J. V. *et al.* Mouse embryonic stem cells self-organize into trunk-like
868 structures with neural tube and somites. *Science* (2020) doi:10.1126/science.aba4937.
- 869 37. Pijuan-Sala, B. *et al.* A single-cell molecular map of mouse gastrulation and early
870 organogenesis. *Nature* **566**, 490–495 (2019).
- 871 38. Song, M. *et al.* *GATA4/5/6 family transcription factors are conserved determinants of*
872 *cardiac versus pharyngeal mesoderm fate.* 2020.12.01.406140
873 <https://www.biorxiv.org/content/10.1101/2020.12.01.406140v1>
874 doi:10.1101/2020.12.01.406140. (2020)
- 875 39. Swedlund, B. & Lescroart, F. Cardiopharyngeal Progenitor Specification: Multiple
876 Roads to the Heart and Head Muscles. *Cold Spring Harb Perspect Biol* **12**, a036731 (2020).
- 877 40. Tyser, R. C. V. *et al.* Characterization of a common progenitor pool of the epicardium
878 and myocardium. *Science* **371**, eabb2986 (2021).
- 879 41. Zhang, Q. *et al.* Unveiling Complexity and Multipotentiality of Early Heart Fields. *Circ*
880 *Res* **129**, 474–487 (2021).
- 881 42. Dupuis, L. E. *et al.* Altered versican cleavage in ADAMTS5 deficient mice; a novel
882 etiology of myxomatous valve disease. *Dev Biol* **357**, 152–164 (2011).
- 883 43. Gurdziel, K., Vogt, K. R., Walton, K. D., Schneider, G. K. & Gumucio, D. L.
884 Transcriptome of the inner circular smooth muscle of the developing mouse intestine: Evidence
885 for regulation of visceral smooth muscle genes by the hedgehog target gene, cJun.
886 *Developmental Dynamics* **245**, 614–626 (2016).
- 887 44. Liang, D. *et al.* Cellular and molecular landscape of mammalian sinoatrial node
888 revealed by single-cell RNA sequencing. *Nat Commun* **12**, 287 (2021).
- 889 45. Ola, R., Lefebvre, S., Braunewell, K.-H., Sainio, K. & Sariola, H. The expression of

- 890 Visinin-like 1 during mouse embryonic development. *Gene Expression Patterns* **12**, 53–62
891 (2012).
- 892 46. van Eif, V. *et al.* Transcriptome analysis of mouse and human sinoatrial node cells
893 reveals a conserved genetic program. *Development* dev.173161 (2019)
894 doi:10.1242/dev.173161.
- 895 47. DeLaughter, D. M. *et al.* Single-Cell Resolution of Temporal Gene Expression during
896 Heart Development. *Developmental Cell* **39**, 480–490 (2016).
- 897 48. Gonzalez, D. M. *et al.* Dissecting mechanisms of chamber-specific cardiac
898 differentiation and its perturbation following retinoic acid exposure. *Development* dev.200557
899 (2022) doi:10.1242/dev.200557.
- 900 49. Li, G. *et al.* Transcriptomic Profiling Maps Anatomically Patterned Subpopulations
901 among Single Embryonic Cardiac Cells. *Developmental Cell* **39**, 491–507 (2016).
- 902 50. Verzi, M. P., McCulley, D. J., De Val, S., Dodou, E. & Black, B. L. The right ventricle,
903 outflow tract, and ventricular septum comprise a restricted expression domain within the
904 secondary/anterior heart field. *Developmental Biology* **287**, 134–145 (2005).
- 905 51. Adachi, N., Bilio, M., Baldini, A. & Kelly, R. G. Cardiopharyngeal mesoderm origins of
906 musculoskeletal and connective tissues in the mammalian pharynx. *Development* **147**,
907 dev185256 (2020).
- 908 52. De Bono, C. *et al.* T-box genes and retinoic acid signaling regulate the segregation of
909 arterial and venous pole progenitor cells in the murine second heart field. *Human Molecular*
910 *Genetics* **27**, 3747–3760 (2018).
- 911 53. Comai, G. *et al.* A distinct cardiopharyngeal mesoderm genetic hierarchy establishes
912 antero-posterior patterning of esophagus striated muscle. *eLife* **8**, e47460 (2019).
- 913 54. Prunotto, C. *et al.* Analysis of Mlc-lacZ Met mutants highlights the essential function of
914 Met for migratory precursors of hypaxial muscles and reveals a role for Met in the development
915 of hyoid arch-derived facial muscles. *Developmental Dynamics* **231**, 582–591 (2004).
- 916 55. Ding, J., Sharon, N. & Bar-Joseph, Z. Temporal modelling using single-cell
917 transcriptomics. *Nat Rev Genet* **23**, 355–368 (2022).
- 918 56. Lange, M. *et al.* CellRank for directed single-cell fate mapping. *Nat Methods* **19**, 159–
919 170 (2022).
- 920 57. Farrell, J. A. *et al.* Single-cell reconstruction of developmental trajectories during
921 zebrafish embryogenesis. *Science* **360**, eaar3131 (2018).
- 922 58. Carmona, R. *et al.* The embryonic epicardium: an essential element of cardiac
923 development. *Journal of Cellular and Molecular Medicine* **14**, 2066–2072 (2010).
- 924 59. Wagner, D. E. & Klein, A. M. Lineage tracing meets single-cell omics: opportunities and
925 challenges. *Nat Rev Genet* **21**, 410–427 (2020).
- 926 60. Ng, S. Y., Wong, C. K. & Tsang, S. Y. Differential gene expressions in atrial and
927 ventricular myocytes: insights into the road of applying embryonic stem cell-derived
928 cardiomyocytes for future therapies. *American Journal of Physiology-Cell Physiology* **299**,
929 C1234–C1249 (2010).
- 930 61. Schmidt, C. *et al.* Multi-chamber cardioids unravel human heart development and
931 cardiac defects. 2022.07.14.499699 Preprint at <https://doi.org/10.1101/2022.07.14.499699>
932 (2022).
- 933 62. Kidokoro, H., Yonei-Tamura, S., Tamura, K., Schoenwolf, G. C. & Saijoh, Y. The heart
934 tube forms and elongates through dynamic cell rearrangement coordinated with foregut
935 extension. *Development* dev.152488 (2018) doi:10.1242/dev.152488.
- 936 63. Nascone, N. & Mercola, M. An inductive role for the endoderm in *Xenopus*
937 cardiogenesis. *Development* **121**, 515–523 (1995).
- 938 64. Etchevers, H. C., Dupin, E. & Le Douarin, N. M. The diverse neural crest: from
939 embryology to human pathology. *Development* **146**, dev169821 (2019).
- 940 65. Keyte, A. & Hutson, M. R. The neural crest in cardiac congenital anomalies.
941 *Differentiation* **84**, 25–40 (2012).
- 942 66. Le Douarin, N. M., Creuzet, S., Couly, G. & Dupin, E. Neural crest cell plasticity and its
943 limits. *Development* **131**, 4637–4650 (2004).
- 944 67. Rinon, A. *et al.* Cranial neural crest cells regulate head muscle patterning and

- 945 differentiation during vertebrate embryogenesis. *Development* **134**, 3065–3075 (2007).
- 946 68. Gopalakrishnan, S. *et al.* A Cranial Mesoderm Origin for Esophagus Striated Muscles.
- 947 *Developmental Cell* **34**, 694–704 (2015).
- 948 69. Sanaki-Matsumiya, M. *et al.* Periodic formation of epithelial somites from human
- 949 pluripotent stem cells. *Nat Commun* **13**, 2325 (2022).
- 950 70. Budjan, C. *et al.* Paraxial mesoderm organoids model development of human somites.
- 951 *eLife* **11**, e68925 (2022).
- 952 71. Miao, Y. *et al.* Reconstruction and deconstruction of human somitogenesis in vitro.
- 953 *Nature* **614**, 500–508 (2023).
- 954 72. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization
- 955 system for the whole mouse brain. *Nat Neurosci* **13**, 133–140 (2010).
- 956 73. Iacovino, M., Roth, M. E. & Kyba, M. Rapid Genetic Modification of Mouse Embryonic
- 957 Stem Cells by Inducible Cassette Exchange Recombination. in *Gene Function Analysis* (ed.
- 958 Ochs, M. F.) 339–351 (Humana Press, 2014). doi:10.1007/978-1-62703-721-1_16.
- 959 74. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-
- 960 3587.e29 (2021).
- 961 75. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and
- 962 Projection for Dimension Reduction. Preprint at <https://doi.org/10.48550/arXiv.1802.03426>
- 963 (2020).
- 964 76. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-
- 965 connected communities. *Sci Rep* **9**, 5233 (2019).
- 966 77. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in
- 967 Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**, 329-337.e4
- 968 (2019).
- 969 78. Virshup, I., Rybakov, S., Theis, F. J., Angerer, P. & Wolf, F. A. anndata: Annotated
- 970 data. 2021.12.16.473007 Preprint at <https://doi.org/10.1101/2021.12.16.473007> (2021).
- 971 79. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression
- 972 data analysis. *Genome Biol* **19**, 15 (2018).
- 973 80. Schiebinger, G. *et al.* Optimal-Transport Analysis of Single-Cell Gene Expression
- 974 Identifies Developmental Trajectories in Reprogramming. *Cell* **176**, 928-943.e22 (2019).
- 975 81. Reuter, B., Weber, M., Fackeldey, K., Röblitz, S. & Garcia, M. E. Generalized Markov
- 976 State Modeling Method for Nonequilibrium Biomolecular Dynamics: Exemplified on Amyloid β
- 977 Conformational Dynamics Driven by an Oscillating Electric Field. *J Chem Theory Comput* **14**,
- 978 3579–3594 (2018).
- 979 82. Brandts, J. Matlab code for sorting real Schur forms. *Numerical Linear Algebra with*
- 980 *Applications* **9**, 249–261 (2002).
- 981
- 982
- 983
- 984

985 **Figure 1: Gastruloids faithfully recapitulates cardiopharyngeal mesoderm specification**
986 **and differentiation toward its myogenic fates. a.** Experimental scheme of the generation of
987 gastruloids from mouse embryonic stem cells (mESCs). Gastruloids were cultured in N2B27 with
988 the addition of small molecules as indicated below the time line. **b.** Representative pictures of
989 gastruloids at day 4, day 5, day 6 and day 11. A, anterior, P, posterior. **c.** Scheme of
990 cardiopharyngeal mesoderm specification in mouse toward the myogenic fates. Markers of each
991 state are indicated in italics. **d-f.** Expression profiles of *Mesp1*, *Isl1* and *Tbx1* (**d**), *Tnnt2* and *Myl7*
992 (**e**) and *Tcf21*, *MyoD* and *Myf5* (**f**) throughout the culture of gastruloids as measured by
993 quantitative RT-PCR. Results are normalized on the expression of *Tbp*. Fold changes are
994 represented relative to expression on day 0. (mean with standard error of mean (SEM)). **g-h.**
995 Representative confocal images of sections across an E7.75 mouse embryo (**g**) and a gastruloid
996 at day 4 (**h**) after RNAscope experiment with *Mesp1* (green), *Isl1* (red) and *Tbx1* (white) probes.
997 **i-j.** Representative confocal images of sections across an E8.5 mouse embryo (**i**) and a gastruloid
998 at day 6 (**j**) after RNAscope experiment with *Isl1* (green), *Tbx1* (red) and *Nkx2-5* (white) probes.
999 **k-l.** Representative confocal images of sections across an E8.5 mouse embryo (**k**) and a
1000 gastruloid at day 7 (**l**) after RNAscope experiment with *Ebf3* (green), *Tcf21* (red) and *Tnnt2* (white)
1001 probes. **m-p.** Representative confocal images of sections across E10.5 mouse embryos (**m** and
1002 **o**) and gastruloids at day 15 (**n** and **p**) after RNAscope experiment with *Tnnt2* (green) and *MyoD*
1003 (**m-n**) or *Myf5* (**o-p**). r, right; l, left. v, ventral; d, dorsal; beat., beating; non beat., non-beating.,
1004 ant., anterior, post., posterior, phA, pharyngeal arch. NT, neural tube. Scale bars: 100 μ m.
1005

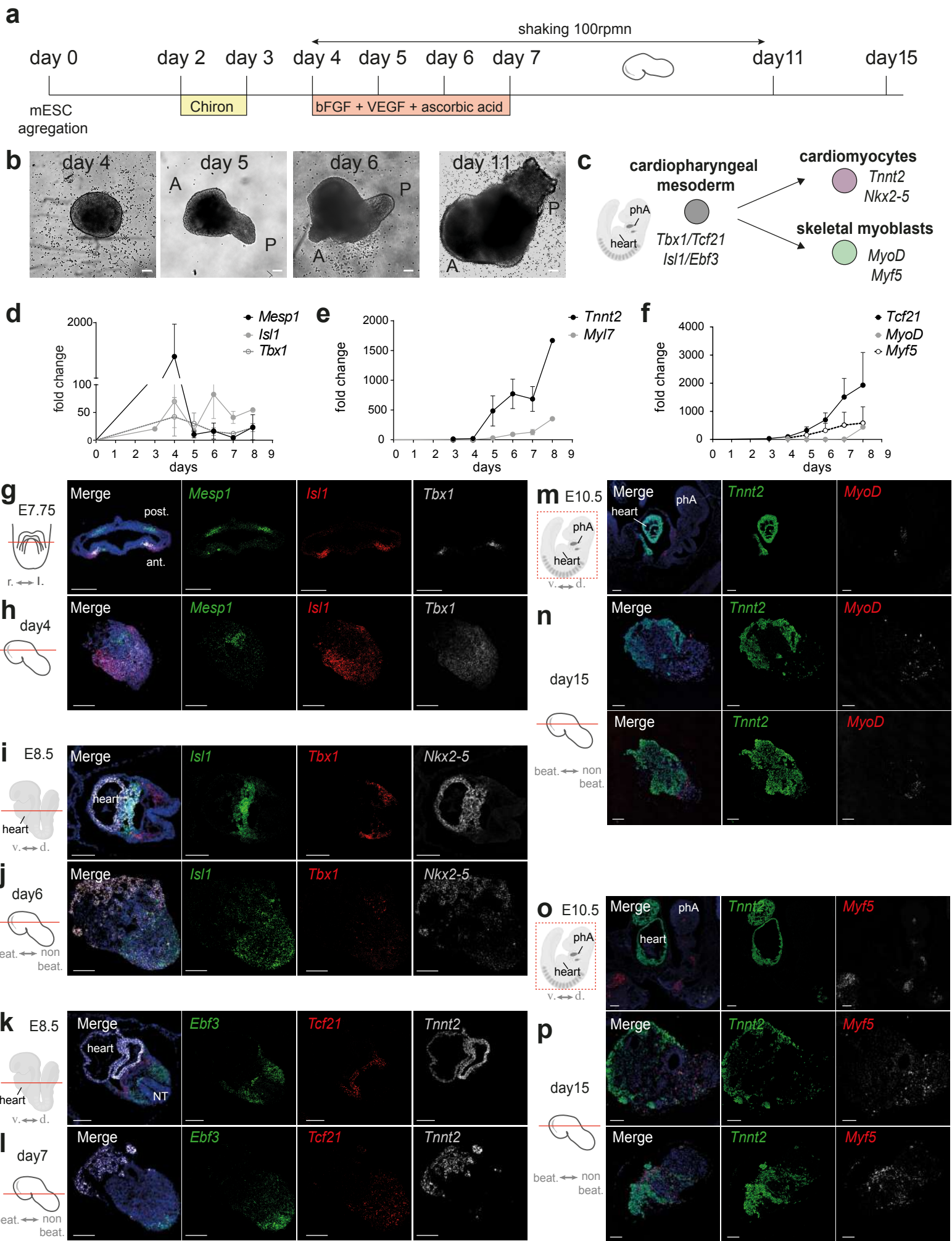


Figure 1

1008 **Figure 2: Transcriptomic analysis reveals cardiopharyngeal mesoderm subpopulations in**
1009 **gastruloids.** UMAP representation of Leiden clustering of gastruloids at day 4 (**a**), day 5 (**c**),
1010 day 6 (**e**) and day 11 (**g**). FeaturePlots of key markers in gastruloids at day 4 (**b**), day 5 (**d**), day 6
1011 (**f**) and day 11 (**h**). Scale bars represent expression levels. Mes., mesoderm; CardioPharyng.,
1012 cardiopharyngeal; Intermed., intermediate; Visc., visceral; Som. Somitic; Diff., differentiating; PS,
1013 primitive streak.
1014
1015

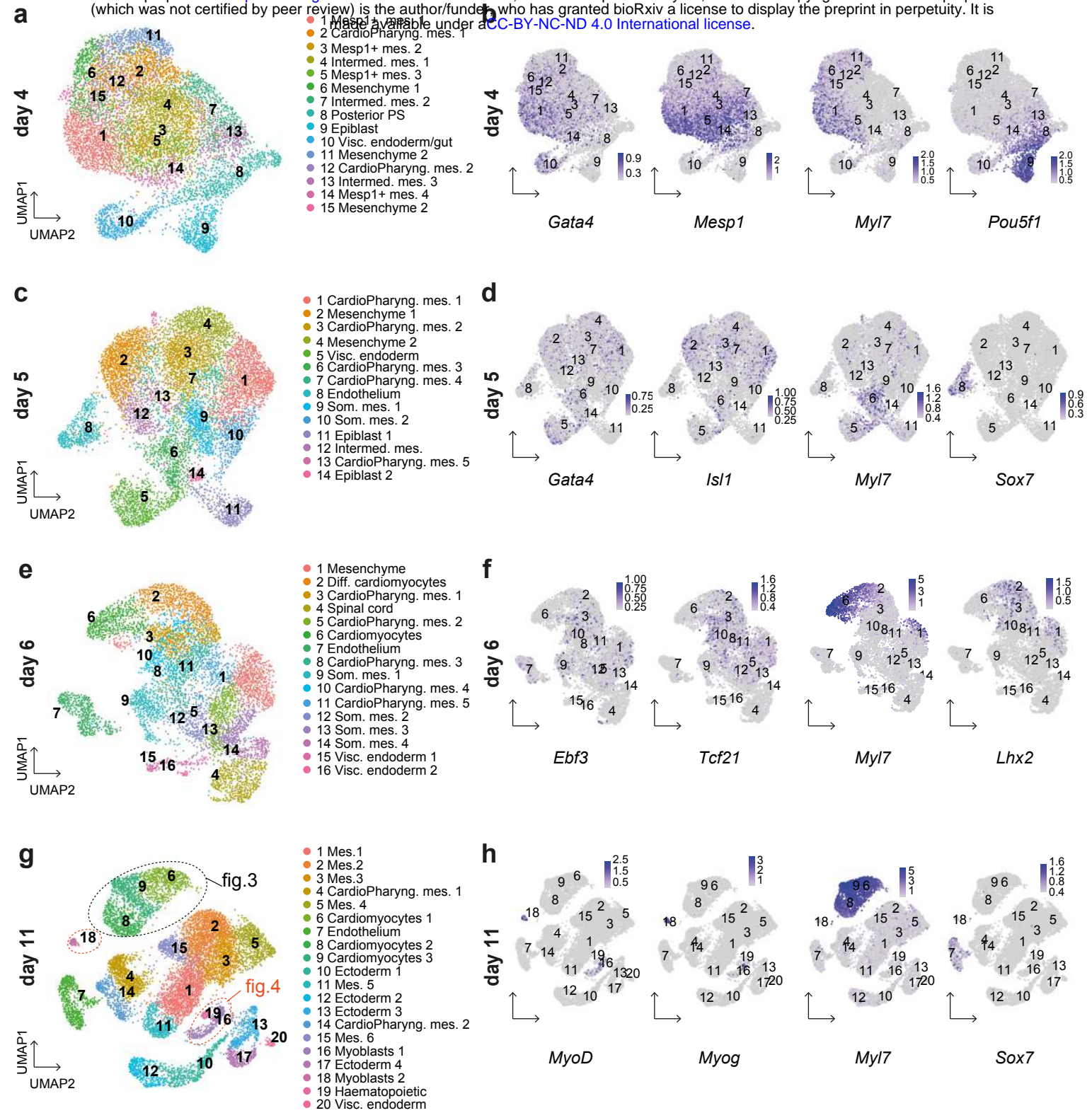


Figure 2

1016 **Figure 3: Three subtypes of cardiomyocytes differentiate in gastruloids. a, c, e.** UMAP
1017 representation of day 11 cardiomyocytes. Red dots represent cells of cluster 6 (a), cluster 9 (c)
1018 or cluster 8 (e). The most differentially expressed genes of the cluster are shown in red. **b, d, f.**
1019 Feature and violin plots of key differentially expressed genes in the cardiomyocyte cluster 6 (b),
1020 cluster 9 (d) and cluster 8 (f) at day 11. **g.** Scheme of the experimental design with the use of the
1021 *Mef2c-AHF-enhancer-Cre; Rosa-tdTomato (Mef2c-Cre; Rosa-tdTomato)* line that label the right
1022 ventricle (RV) and outflow tract. La, left atrium; RA, right atrium; LV, left ventricle; ao, aorta; pt,
1023 pulmonary trunk. **h.** Confocal image (maximum intensity projection) of a *Mef2c-Cre; Rosa-*
1024 *tdTomato* gastruloid at day 10 after immunostaining with a marker of cardiomyocytes (cardiac
1025 troponin T – cTnT in green). tdTomato positive cells and nuclei are depicted in red and white,
1026 respectively. Arrowheads indicate cTnT+ tdTomato+ cells. Scale bar: 100 μ m. **i.** FACS analysis of
1027 the combined expression of tdTomato and cTnT expression in all of the living gastruloids' cells at
1028 day 11. **j-k.** Graphs showing the proportion of tdTomato+ (Tom+) cells in cardiomyocytes (cTnT+
1029 cells) derived from day 11 gastruloids (j) and embryonic cells at E10.5/E11.5 (k) (in at least 4
1030 independent experiment (mean with standard error of mean (SEM))). **l.** FACS analysis of the
1031 combined expression of *MyI7* and *MyI2* expression in all of the living gastruloids' cells at day 11
1032 after HCR experiment. Negative controls with no probe are shown in left.
1033
1034

1035 **Figure 4: Skeletal myogenesis takes place in gastruloids from the cardiopharyngeal and**
1036 **somitic mesoderm. a.** UMAP representation of day 11 myoblast clusters highlighted in purple
1037 (cluster 16) and in pink (cluster 18). **b.** Scheme of the genetic regulation of skeletal myogenesis
1038 in mouse from the cardiopharyngeal (red) or somitic mesoderm (green). Skeletal myogenesis
1039 starts from muscle precursors expressing *Pax7* and *Myf5* that differentiate toward committed
1040 myoblasts expressing *MyoD*, *Myog* and *Myh3*. **c-h.** Violin plots showing the level of expression of
1041 *Pax7* (**c**), *Myf5* (**d**), *Met* (**e**), *MyoD* (**f**), *Myog* (**g**) and *Myh3* (**h**) in the different clusters of gastruloids
1042 at day 11. **i-m.** Violin plots showing the level of expression of *Isl1* (**i**), *Pitx2* (**j**), *Tbx1* (**k**), *MyoR* (**l**)
1043 and *Pax3* (**m**) specifically in the two myoblast clusters (clusters 16 and 18) at day 11. **n.** UMAP
1044 representation of day 11 cluster 16. Blue and red dots represent *Tbx1* and *Pax3* expressing cells,
1045 respectively. **o.** FACS analysis of the combined expression of *Myh3* and *Myog* expression in
1046 gastruloid cells at day 11 after HCR experiment. Negative controls with no probe are shown on
1047 the left.
1048
1049

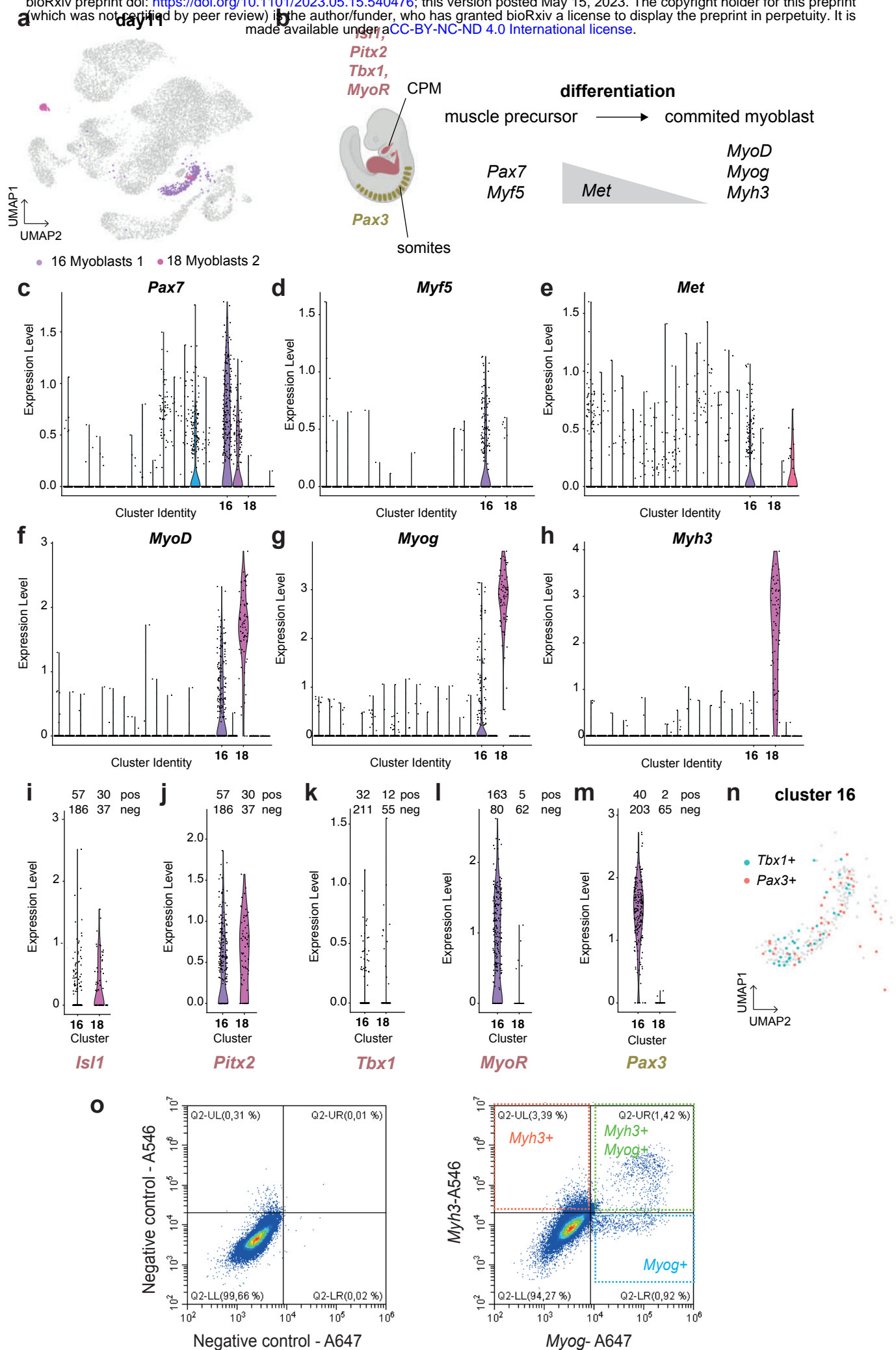
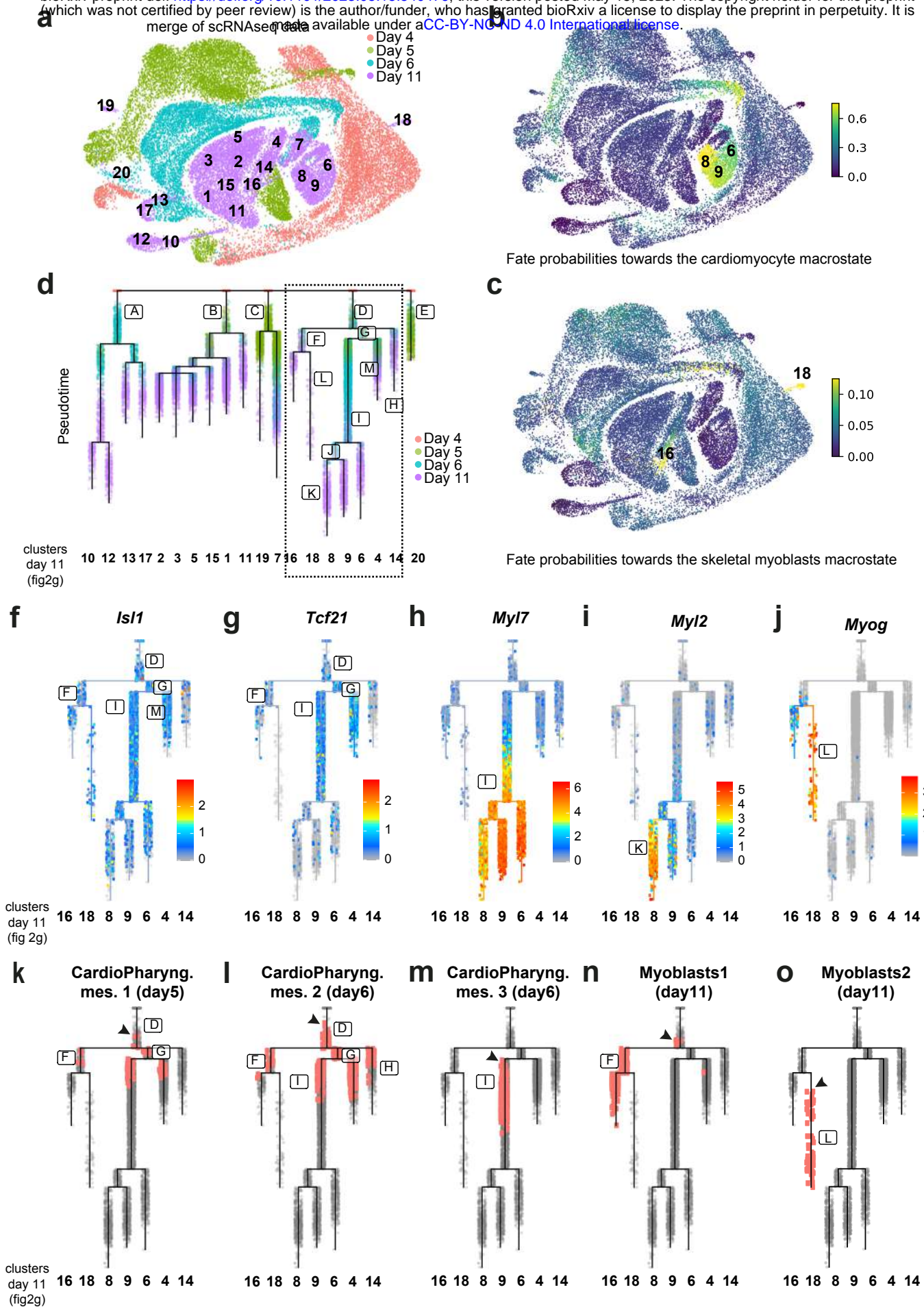


Figure 4

1049 **Figure 5: Myogenic trajectories are found in gastruloids over time. a.** UMAP representation
1050 of all gastruloid cells merged from day 4 (red), day 5 (green), day 6 (blue) and day 11 (purple).
1051 Numbers represent identified clusters of day 11 (as defined in Fig. 2g). **b-c.** Fate probabilities
1052 towards the cardiomyocyte macrostate (**b**) or towards the skeletal myoblast macrostate (**c**). Scale
1053 bars represent absorption probabilities. Numbers represent key cardiomyocytes and myoblast
1054 clusters of day 11 **d.** URD trajectories showing transition from cells at day 4 (root) towards cells
1055 at day 11 (tips). Framed letters represent branches along the tree. The dotted box represents the
1056 lineage branch detailed in f-o. **f-j.** Expression of *Isl1* (**f**), *Tcf21* (**g**), *Myl7* (**h**), *Myl2* (**i**) and *Myog* (**j**)
1057 across URD trajectory. Scale bars represent expression levels. **k-o.** Representation of clusters'
1058 cells across URD trajectory. Red dots represent cells of the cardiopharyngeal mesoderm
1059 (CardioPharyng. mes.) cluster 1 of day 5 (in **k**, Fig. 2c, cluster 1), of the cardiopharyngeal
1060 mesoderm (CardioPharyng. mes.) cluster 2 of day 6 (in **l**, Fig. 2e, cluster 5), of the
1061 cardiopharyngeal mesoderm (CardioPharyng. mes.) cluster 3 of day 6 (in **m**, Fig. 2e, cluster 8),
1062 of the myoblast cluster 1 of day 11 (in **n**, Fig. 2g, cluster 16) and of the myoblast cluster 2 of
1063 day 11 (in **o**, Fig. 2g, cluster 18).
1064
1065



1066 **Figure 6: Schematic representation of the developmental timing of cardiopharyngeal**
1067 **mesoderm specification and differentiation in gastruloids in comparison with the mouse**
1068 **model.**
1069

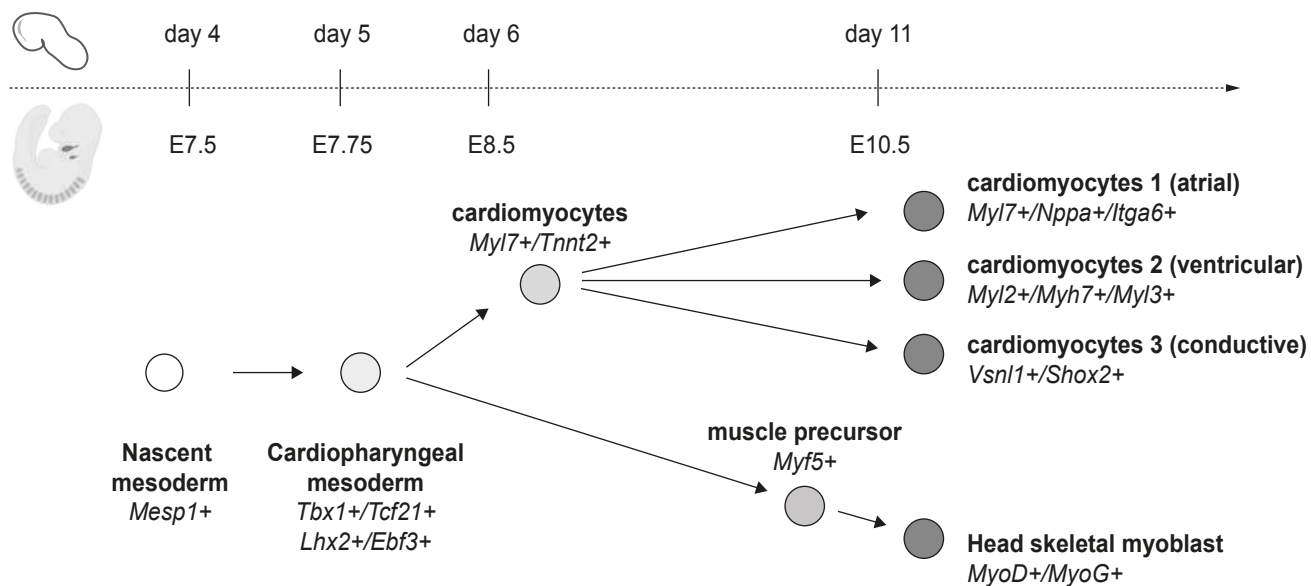


Figure 6