FRONT MATTER

Title

Gastrulation-stage gene expression in *Nipbl*$^{+/-}$ mouse embryos foreshadows the development of syndromic birth defects

Authors

[1, 2, *]Stephenson Chea, [3]Jesse Kreger, [1, 2]Martha E. Lopez-Burks, [3]Adam L. MacLean, [1, 2]Arthur D. Lander, [4, 1, 2, *]Anne L. Calof

Affiliations

[1]Department of Developmental and Cell Biology, School of Biological Sciences, University of California Irvine, Irvine, CA 92697

[2]Center for Complex Biological Systems, University of California Irvine, Irvine, CA 92697

[3]Department of Quantitative and Computational Biology, Dornsife College of Letters, Arts, and Sciences, University of Southern California, Los Angeles, CA

[4]Department of Anatomy and Neurobiology, School of Medicine, University of California Irvine, Irvine, CA 92697

*Corresponding author. Email: schea2@uci.edu (S.C.); alcalof@uci.edu (A.L.C)

Abstract

In animal models, *Nipbl*-deficiency phenocopies gene expression changes and birth defects seen in Cornelia de Lange Syndrome (CdLS), the most common cause of which is *Nipbl*-haploinsufficiency. Previous studies in *Nipbl*$^{+/-}$ mice identified aberrant gene expression and heart defects as early as cardiac crescent (CC) stage. Here, we performed single-cell RNA-sequencing on wildtype (WT) and *Nipbl*$^{+/-}$ mouse embryos at CC- and earlier (gastrulation) stages. *Nipbl*$^{+/-}$ embryos had fewer mesoderm cells than WT and altered proportions of mesodermal cell subpopulations. These findings were associated with an underexpression of genes implicated in driving specific mesodermal lineages. *Nipbl*$^{+/-}$ embryos also misexpressed developmentally-

critical genes, including the transcription factor, *Nanog*, and genes governing left-right and anterior-posterior patterning. These events of cell misallocation and transcriptional dysregulation foreshadowed defects in tissue composition and patterning that arise later in *Nipbl*$^{+/-}$ mice, offering insights into early developmental contributions to birth defects in CdLS.

Teaser

Gene expression changes during gastrulation of *Nipbl*-deficient mice shed light on early origins of structural birth defects.

MAIN TEXT

Introduction

Each year, 1 of every 33 babies in the United States is born with a birth defect (*1*), the most prevalent of which are congenital heart defects (CHDs), neural tube defects, and cleft lip/palate (*2*). Given the significant impact of birth defects on both mortality and morbidity (*3, 4*), there is a need to elucidate their origins, but their diversity and sporadic nature pose challenges for identifying causal mechanisms. A promising approach is to study genetic syndromes that present with multiple, concurrent defects in various body parts, many of which mirror common isolated birth defects. Studying them might thus provide insights into the causes and development of isolated birth defects.

Cornelia de Lange Syndrome (CdLS) affects an estimated 1 in 10,000 to 1 in 30,000 live births (*5*), and is characterized by craniofacial anomalies, delays in growth and maturation, intellectual disability, neurological impairments, abnormalities of limbs, especially arms and hands, coupled with issues in the visual, auditory, gastrointestinal, genitourinary, and cardiopulmonary systems (*6-9*).

The majority of CdLS cases – more than 55% – are caused by heterozygous mutations in the gene *Nipped-B-like* (*NIPBL*), named for its homology to the Drosophila gene *Nipped-B* (*9-13*). These mutations often produce an absent or shortened protein, suggesting CdLS arises from

51 haploinsufficiency (*14*). Remarkably, even a subtle 15% reduction in NIPBL gene expression can
52 produce a mild yet recognizable CdLS phenotype (*15, 16*). Such observations highlight the
53 importance of precise NIPBL gene dosage in human development.

54 The *NIPBL* gene encodes a universally conserved protein that plays a role in loading
55 cohesin onto chromosomes (*17-20*). Cohesin, similarly conserved and ubiquitous, is a four-
56 subunit protein complex (Smc1, Smc3, Rad21, and either SA1 or SA2) essential for chromosome
57 organization and genome stability (*21*). Mutations in the cohesin subunits *SMC1* and *SMC3*
58 account for a small proportion of clinically mild CdLS (~5% and <1%, respectively) (*22-24*).
59 Additionally, mutations in *HDAC8*, which catalyzes cohesin release from chromatin during
60 mitosis, are found in a distinct subset of CdLS patients (*25*). Mutations in *RAD21* have also
61 surfaced in individuals exhibiting a CdLS-like phenotype but with substantially milder cognitive
62 impairment (*26*). Collectively, these findings reinforce the idea that impairment of cohesin
63 function contributes significantly to CdLS (*9, 27*).

64 Cohesin was initially identified for its role sister chromatid cohesion during mitosis (*21*).
65 However, pronounced defects in sister chromatid cohesion or irregularities in mitosis have not
66 been observed in either CdLS patients or *Nipbl*-haploinsufficient (*Nipbl*$^{+/-}$) mice (*28-31*),
67 suggesting that cohesin has additional functions. Studies in CdLS patient cells and *Nipbl*-
68 haploinsufficiency animal models indicate that cohesin is involved in transcriptional regulation
69 (Dorsett & Krantz, 2009). Specifically, *Nipbl*-haploinsufficiency leads to alterations in the
70 expression of many hundreds to thousands of genes (*16, 28, 32, 33*). Interestingly, many of the
71 affected genes are regulated by long-distance enhancers (*34*), aligning with the emerging concept
72 of NIPBL and cohesin as critical determinants in DNA-looping (*35-38*).

73 Interestingly, most gene expression changes in *Nipbl*-deficient animals are small, usually
74 less than 2-fold. Though likely to be inconsequential individually, such small changes can act
75 collectively to produce structural and functional defects. For example, in zebrafish, joint depletion

76 of two developmental genes downregulated by *nipbl*-deficiency produced a CdLS-like phenotype
77 (*32*), suggesting that altered gene expression is the ultimate cause of developmental and
78 physiological abnormalities. Thus, CdLS exemplifies a class of genetic disorders known as
79 "transcriptomopathies" (*39*), in which the cumulative or synergistic effects of minor disturbances
80 to gene expression lead to developmental abnormalities.

81 Attempts to associate such changes with the production of birth defects have featured
82 studies of CHDs in *Nipbl*-haploinsufficient mice. 30% of *Nipbl*$^{+/-}$ mice display CHDs, primarily
83 atrial septal defects. *Nipbl*$^{+/-}$ mice also exhibit heart abnormalities at earlier developmental stages:
84 At embryonic day 13.5 (E13.5), 70% of *Nipbl*$^{+/-}$ mice display delays in ventricular septal fusion,
85 whereas three days earlier, at E10.5, an overwhelming 100% exhibit right ventricle hypoplasia
86 (*40*). Even earlier, these mice display reduced expression of transcription factors crucial for early
87 heart progenitor cell growth and differentiation: specifically, *Nkx2-5* in the second heart field at
88 E7.75, and *Mesp1* in mesoderm progenitors at E7.5 (*40*). That significant abnormalities may
89 begin even earlier is suggested by results in in *nipbl*-morphant zebrafish, in which defects in the
90 initial migration of cardiogenic mesoderm can be detected as early as 18 hours post-fertilization
91 (*32*). These results suggest that, at least for CHDs, causal events may occur as early as
92 gastrulation, when the three primary germ layers form and the earliest progenitor cells of major
93 tissues and organs begin to differentiate (*41*).

94 In this study, we sought to identify developmental alterations during gastrulation that
95 might account for birth defects in *Nipbl*$^{+/-}$ mice. We used single-cell RNA sequencing
96 (scRNAseq) to compare the cellular compositions, lineage trajectories, and transcriptional
97 landscapes of *Nipbl*$^{+/-}$ mouse embryos to their WT counterparts at two stages spanning the end of
98 gastrulation (*42*): the late bud- (LB) stage (E7.5) and cardiac crescent- (CC) stage (E7.75). Our
99 findings reveal that *Nipbl*$^{+/-}$ embryos have the same cell populations as WT embryos but display
100 subtle misallocation of specific mesodermal cell populations. Our evidence suggests this occurs as

101 a result of alterations in specific cell fate decisions, including the choice by mesoderm cells to
102 progress toward a non-cardiac versus cardiac fate. Our observations strongly suggest these events
103 cannot be attributed to changes in apoptosis, cell proliferation, global developmental delay, or the
104 structure of cell lineages. As in earlier research, we observed that most gene expression changes
105 in *Nipbl*-deficient tissues at these stages were small (*16, 28, 32, 33*), but we also identified several
106 key developmental genes that are more markedly misexpressed in *Nipbl*$^{+/-}$ embryos, the most
107 notable of which was *Nanog*. *Nipbl*$^{+/-}$ embryos failed to downregulate *Nanog* at the end of
108 gastrulation, which normally occurs in all but germ cells. As a result, the misexpression of *Nanog*
109 target genes was observed in all germ layers. We also saw substantial underexpression of *Hox*
110 genes and overexpression of Nodal signaling genes, which play roles in anterior-posterior and
111 left-right patterning, respectively. As a result, we propose a model in which birth defects in CdLS
112 arise from the prolonged overexpression of *Nanog* and dysregulation of developmental pathways
113 governing axial patterning, resulting in the misdirection of cell fate decisions, and misallocation
114 of specific progenitor cell populations.

115 **Results**

116 **Gastrula-stage *Nipbl*$^{+/-}$ mice display the same cell populations as are found in WT mice**

117 To investigate early factors influencing birth defects in CdLS, we generated WT and
118 *Nipbl*$^{+/-}$ mouse embryos using a floxed *Nipbl* allele (*40*). To ensure an accurate accounting of cell
119 populations and proportions at well-defined stages of early embryonic development, we generated
120 a large excess of embryos and sorted them into groups of narrowly defined stage based on
121 morphological criteria (*43*). Using the 10X Genomics Chromium Single Cell Expression platform
122 (*44*), we performed scRNAseq on 5 WT and 6 *Nipbl*$^{+/-}$ samples of late bud stage (LB) embryos
123 (E7.5), and 8 WT and 8 *Nipbl*$^{+/-}$ samples of cardiac crescent stage (CC) embryos (E7.75) (Fig. 1A
124 and fig. S1). Given the low cell numbers in individual LB-stage embryos, and the 10X Genomics
125 Chromium Single Cell Expression system's capture efficiency, samples at the LB stage consisted

126 of pairs of embryos, whereas at the CC-stage they consisted of single embryos. Pairing LB-stage
127 embryos facilitated the capture of rare cell groups, such as primordial germ cells, which typically
128 comprise about 50 cells per embryo at this stage (*45*).

129 At LB-stage, we captured a median of 2,537 cells per sample, with a median of 17,301
130 RNA transcripts per cell (fig. S2 and table S1). At CC-stage, we captured a median of 4,116 cells
131 per sample, with a median of 14,997 RNA transcripts per cell (table S1). As expected, all *Nipbl*$^{+/-}$
132 embryos expressed both *Cre* and *Betageo*, which report an inactivated *Nipbl* allele (*40*), while
133 WT embryos did not (fig. S3, A and B). *Nipbl*$^{+/-}$ embryos across both stages expressed *Nipbl* at
134 levels that were approximately 50% lower than WT counterparts (fig. S3C).

135 We corrected for batch effects among embryos of identical stage and genotype using
136 Seurat's integration protocol (*46*) (Fig. 1B and fig. S4), which identifies anchors (cells of similar
137 gene expression across samples of the same biological condition) and uses them to align cells into
138 a space shared by all samples. Recognizing that all cells in *Nipbl*$^{+/-}$ tissues exhibit substantial gene
139 expression changes (*28*), and considering that clustering algorithms rely on differences in gene
140 expression among cells (*46*), we implemented measures to prevent our cell clustering from being
141 skewed by gene expression alterations attributable to *Nipbl*-haploinsufficiency. We did this by
142 first clustering cells from WT samples (Fig. 1B), using a robust iterative clustering method that
143 considered both intra-cluster stability and inter-cluster variation (fig. S5) to determine the optimal
144 number of clusters for cells at each developmental stage. WT cells were clustered into 18 cell
145 populations at LB-stage (Fig. 1C) and 26 cell populations at CC-stage (Fig. 1D). We then
146 projected *Nipbl*$^{+/-}$ cells onto these WT cell populations at corresponding stages (Fig. 1, B to D).
147 We observed that all *Nipbl*$^{+/-}$ cells could be projected to WT cell populations, with each WT
148 population receiving some *Nipbl*$^{+/-}$ cells (Fig. 1, C and D). This pattern held true across individual
149 *Nipbl*$^{+/-}$ samples, as well as the total *Nipbl*$^{+/-}$ cells in aggregate at each stage (Fig. 1, E and F).

150 These results suggest that *Nipbl*$^{+/-}$ cells contribute to all clusters, but as described later (Fig. 3),
151 they do so in proportions that are not always identical.

152 **scRNAseq captured cells from all germ layers and progenitors of major tissues**

153 To assign biological identities to clustered cell populations, we used WT cells as a
154 reference. For each WT cell population, we performed differential gene expression analysis
155 (DGEA) by comparing it to all other cells within embryos of identical stage (Mann-Whitney U
156 test). We detected markers indicative of germ layer identity in every cluster (data S1, data S2, and
157 table S2), enabling us to associate each cluster with a specific germ layer at each developmental
158 stage (Fig. 2A). Figure 2B displays all cells from all samples, categorizing them into germ layers
159 at both LB- and CC-stages of development. Figures 2C and 2D further depict the assignment of
160 cells based on the expression of germ layer markers, at LB- and CC-stages, respectively. These
161 markers include those for ectoderm, such as *Utf1, Sox2,* and *Pou3f1*, mesoderm, including *T,*
162 *Hand1, Vim, Twist1,* and *Prrx2*, and endoderm, *Ttr*.

163 To identify cell types within germ layers, we re-applied DGEA after clustering individual
164 germ layers. For the endoderm, this enabled us to identify definitive endoderm progenitors (DEP)
165 (*Sox4* and *Foxa2*), as well as extraembryonic endoderm (XEnd) (*Hnf4a* and *Tbx3)* at LB-stage
166 (Fig. 2E, data S3, and table S3). In CC-stage embryos, we captured identical cell types in the
167 endoderm, and additionally, visceral endoderm (VE) (*Hnf4a* and *Esx1*) (Fig. 2F, data S4, and
168 table S3). DEPs and XEnd give rise to the gut and placenta, respectively.

169 Within the ectoderm of LB-stage embryos, we identified early progenitor populations
170 including intraembryonic ectoderm progenitors (IEPs) (*Pou3f1* and *Otx2*), neural ectoderm
171 progenitors (NEPs) (*Cdx1* and *Nkx1-2)*, and surface ectoderm (SE) (*Id1, Msx2,* and *Msx1*) (Fig. 2,
172 E and G, data S5, and table S4). Additionally, we detected a small portion of extraembryonic
173 ectoderm (XEct) (*Efl5* and *Tead4*), which was expected, as a small quantity of XEct cells was
174 included in our dissection. By CC-stage, IEPs, NEPs, and XEct were no longer detectable, a

175 phenomenon previously noted (*47*). Rather, we continued to detect surface ectoderm (SE). We
176 also detected six distinct neural populations (Fig. 2, F and H, data S6, and table S4). These neural
177 populations included: prosencephalic neural plate (*Six3* and *Hesx1*), mesencephalic neural plate
178 (*En1* and *Pax2*), rhombencephalic neural plate (*Hoxa1* and *Hes3*), midline neural plate (*Foxa2*
179 and *Nkx2-9*), spinal neural plate (*Sox2* and *Ezr*), and neural crest (*Pax3* and *Foxd3*) (data S6).
180 Collectively, these populations constitute the nascent progenitors of the brain, spinal cord, and
181 peripheral nervous system.

182 Within the mesoderm of LB-stage embryos, we identified six distinct groups of cells,
183 detailed in Fig. 2, E and G, data S7, and table S5. These consisted of: 1) neuromesodermal
184 progenitors (NMPs): expressing markers *T,* and *Nkx1-2*; 2) derivatives of NMPs: including
185 *Mesp1*-expressing mesoderm progenitors (MMPs) (Mesp1), two types of first heart field cells
186 (FHF1 and FHF2) marked by *Hand1*; cells of the second heart field (SHF) with *Id2* and *Mef2c*;
187 and three kinds of paraxial mesoderm cells, further categorized into paraxial mesoderm
188 progenitors (PMPs) (*Tbx6* and *Meis2)*, trunk paraxial mesoderm (TPM) (*Foxc2* and *Tcf15*), and
189 head paraxial mesoderm (HPM) (*Alx1* and *Tcf15*); 3) hematopoietic cells: including
190 hemangioblasts (HMBs) (*Tal1* and *Etv2*) and primitive erythrocytes (PEs) (*Gata1* and *Tal1*); and
191 4) two rare populations: comprising cells of the primitive node (PN) (*Noto* and *Foxj1*), and
192 primordial germ cells (PGCs) (*Tfap2c* and *Msx1*). At CC-stage, we detected all mesodermal
193 populations found earlier, along with several that emerge later in development. These comprised
194 lateral plate mesoderm progenitors (LPMP) (*Tlx2,* and *T*), additional populations of *Mesp1*-
195 expressing mesoderm progenitors (MMP1 and MMP2) (*Mesp1*), cardiomyocytes (CDMs) (*Mef2c*
196 and *Nkx2-5*), and presomitic mesoderm (PSM) (*Tcf15* and *Meox1*) (Fig. 2, F and H, data S8, and
197 table S5). Collectively, these progenitors are responsible for the formation of structures including
198 skeletal muscle, bone, blood, and heart.

199 ***Nipbl*$^{+/-}$ mice have fewer mesoderm cells, primitive erythrocytes, and first heart field cells,**
200 **and more paraxial mesoderm cells**

201 We analyzed the cellular composition of embryos at both LB- and CC-stage, quantifying
202 the proportion of cells within each germ layer that contributed to the total cell count. LB-stage
203 mutants exhibited 13% fewer mesoderm cells (Fig. 3, A and B, data S9), while concurrently
204 displaying a greater proportion of endoderm and ectoderm cells. We sought to ascertain if this
205 decrease in mesoderm was generalized or restricted to specific subpopulations. Focusing on LB-
206 stage embryos, we merged mesodermal cell subpopulations of similar biological identity together
207 (FHF1 + FHF2 into FHF, and PSM + CPM into PM) and calculated the percentage of cells across
208 these subpopulations, relative to the overall mesoderm (Fig. 3C). The findings (Fig. 3D) revealed
209 a 77% reduction in PEs *Nipbl*$^{+/-}$ embryos (data S10). Additionally, differences were observed in
210 two mesodermal derivatives of the neuromesoderm: 1) Mutants had 24% fewer first heart field
211 cells (FHF) (Fig. 3D); and 2) *Nipbl*$^{+/-}$ embryos showed a contrasting pattern in paraxial mesoderm
212 (PM), a 33% increase in PM cells compared to WT (Fig. 3D).

213 ***Nipbl*$^{+/-}$ embryos do not show changes in developmental timing**

214 Cellular composition changes very quickly during and after gastrulation; one possible
215 explanation for differences in proportions of cell types would be a change in the overall pace of
216 development, such that embryos of one genotype were slightly delayed or accelerated relative to
217 the other. To investigate whether this was the case for *Nipbl*$^{+/-}$ embryos, we employed
218 pseudotemporal ordering. To ensure that pseudotime accurately mirrored real developmental
219 time, we augmented our analysis with scRNAseq data from an additional two pairs of WT
220 embryos at the early bud (EB) stage and two pairs of WT embryos at the early head fold (EHF)
221 stage (Fig. 3E and fig. S6). EB-stage embryos are two hours younger and EHF-stage embryos are
222 two hours older than LB-stage embryos, so these supplementary samples provided the temporal
223 resolution necessary to discern even minor changes (less than 2 hours) in developmental timing

224 (*43*). We eliminated low-quality cells and doublets (Fig. 3F, fig. S7 and table S6) before
225 integrating cells of the same stage (fig. S8). We then clustered EB-stage cells into 8 distinct
226 populations and EHF-stage cells into 13 (fig. S9, A and B). Each population was subsequently
227 annotated as ectoderm (Ect), mesoderm (Mes), or endoderm (End) (Fig. 3G, and fig. S9, C and D)
228 based on the expression of specific germ layer markers (data S11 and S12, and table S7).

229 We ordered cells from each embryo at every developmental stage (EB, LB, and EHF)
230 using URD (*48*), which constructs a diffusion map of transition probabilities and, starting with an
231 assigned group of root cells, performs a probabilistic breadth-first graph search using the
232 transition probabilities. When we visualized the arrangement of cells from each stage based on
233 their density in pseudotime, we found that cells from WT embryos ordered in accordance to their
234 developmental stage: EB came first, followed by LB, and finally EHF (Fig. 3H, data S13), while
235 showing partial overlap between cells of different stages. This confirmed that the pseudotime
236 values we acquired were an accurate representation of actual developmental timing. When the
237 pseudotime ordering of WT and *Nipbl*$^{+/-}$ LB-stage embryos were compared with each other, we
238 found no statistically significant deviations (Kolmogorov-Smirnov test) (Fig. 3I). These findings
239 argue that the observed reduction in mesoderm cells in *Nipbl*$^{+/-}$ mice at LB-stage cannot be
240 attributed to overall developmental delay or acceleration. The fact that *Nipbl*-haploinsufficiency
241 does not result in global changes in developmental timing strongly suggests that birth defects in
242 CdLS result from specific cellular or molecular irregularities within individual developmental
243 pathways, and not a globally altered developmental timeline.

244 **_Nipbl_$^{+/-}$ embryos do not show changes in cell lineage trajectory**

245 To discern whether alterations in the sizes of cell population in *Nipbl*$^{+/-}$ embryos were
246 attributable to changes in the structures of cell lineages, we used scVelo (*49*) to compute the RNA
247 velocity for all mesoderm cells from LB-stage embryos. RNA velocity provides a predictive
248 metric for a cell's future transcriptional state, gauging the equilibrium between the synthesis of

249 spliced mRNA from unspliced mRNA and mRNA degradation. scVelo was used to build cell
250 lineage trajectories separately for WT and *Nipbl*$^{+/-}$ embryos at LB-stage (Fig. 4A). In WT
251 mesoderm, five distinct lineages were discernible (Fig. 4B, data S14), with neuromesodermal
252 progenitors (NMPs) giving rise to three terminal fates via four specific pathways. These pathways
253 encompassed the differentiation of NMPs into FHF, SHF, and PM. It's intriguing to note that
254 *Mesp1*-expressing cells have the versatility to differentiate into both cardiac and paraxial
255 mesoderm. While *Mesp1* is typically recognized as a cardinal factor for cardiac specification (*50*),
256 our findings suggest its influence may extend beyond this role into paraxial specification as well.
257 As a result, the paraxial mesoderm (PM) has dual progenitors, PMPs and MMPs. We also
258 observed a fifth lineage, independent of NMP lineages, where HMBs differentiate into PEs.
259 Notably, PGCs and PN cells did not form part of any identified lineages. Examination of the
260 mesoderm in *Nipbl*$^{+/-}$ embryos revealed lineage trajectories identical to those in WT (Fig. 4A,
261 data S15). These observations suggest that variations in the overall structures of cell lineage
262 pathways do not explain differences in the number of FHF and PM cells in *Nipbl*$^{+/-}$ embryos (Fig.
263 3, B and D).

264 **_Nipbl_$^{+/-}$ embryos misdirect mesoderm cells into paraxial mesoderm at the expense of the first**
265 **heart field**

266 To explore whether variations in cell fate decisions might be responsible for the observed
267 differences in the numbers of FHF and PM cells in *Nipbl*$^{+/-}$ embryos, we employed CellRank (*51*)
268 to compute the likelihood of cells from the mesoderm (excluding HMBs, PEs, PGCs, and PN
269 cells) transitioning terminally into cells of SHF, FHF, or PM fate (data S16 and data S17).
270 HMBs, PEs, PGCs, and PN were omitted from the calculation because our trajectories derived
271 from RNA velocity indicated that these populations do not ultimately transition into SHF, FHF, or
272 PM. In CellRank, the absorption probability defines the likelihood that a specific cell type will
273 make a transition to a terminal state, derived using RNA velocity-directed, random walks from

274 initial to terminal cell states (*51*). We represented these probabilities in UMAP space, with shades
275 of blue to signify low probabilities and red for high probabilities. Fig. 4, C and D illustrates that
276 the probabilities of NMPs, MMPs, and PMPs making a terminal transition solely into SHF do not
277 vary significantly between *Nipbl*$^{+/-}$ and WT embryos, as quantified in the adjacent violin plots.
278 However, the transition probabilities into FHF are 27-48% lower in *Nipbl*$^{+/-}$ embryos (Fig. 4, E
279 and F), while the probabilities for transitions into PM are 9-11% higher (Fig. 4, G and H). Taken
280 together, these results support the hypothesis that disparities in cell numbers in *Nipbl*$^{+/-}$ embryos
281 stem from a misdirection of cells from FHF pathway towards a PM fate (Fig. 4I).

282 **_Nipbl_$^{+/-}$ embryos do not show evidence of altered apoptotic activity or cell proliferation**

283 Altered cell numbers in different mesodermal subpopulations of *Nipbl*$^{+/-}$ embryos could
284 conceivably also stem from modifications in rates of cell death or proliferation. Focusing on
285 apoptotic activity, we carried out differential gene expression analysis (DGEA) on WT and
286 *Nipbl*$^{+/-}$ embryos at LB-stage, using genes associated with apoptosis from two well-known gene
287 sets: Reactome Apoptosis (Hardwick et al., 2004) and Hallmark Apoptosis (*52*). We ranked each
288 gene by minimum Q-value, arranging them from lowest to highest, and visualized their
289 expression across all germ layers in *Nipbl*$^{+/-}$ embryos as heatmaps of the fold change relative to
290 their WT counterparts (fig. S10, A and B, and data S18). For both gene sets, we saw no visible
291 pattern of differences in the expression of genes with even the lowest Q-values between any of
292 the germ layers in both WT and *Nipbl*$^{+/-}$ embryos. Although genes from these gene sets tend to
293 mark cells with the capacity for apoptosis, their expression levels are not necessarily indicative of
294 cells actively undergoing apoptosis. We therefore also turned to a gene set identified in a study
295 that compared 180 apoptosis-associated genes in hematopoietic cells that were either healthy or
296 undergoing apoptosis (*53*), and identified 93 apoptosis-associated genes that were differentially
297 expressed. When we examined these genes in the germ layers of *Nipbl*$^{+/-}$ embryos, we also saw no

298 visible difference with WT embryos (fig. S10C and data S18). These data suggest that changes in
299 apoptotic activity are unlikely to be a major contributor to differences in cell type proportions.
300 To examine cell proliferation, we again employed DGEA across all germ layers of both
301 WT and *Nipbl*$^{+/-}$ embryos at LB-stage, focusing on the meta-PCNA gene set, an ensemble of
302 genes that exhibit the strongest positive correlation with PCNA expression, a recognized
303 biomarker of cell proliferation (*54*). When we visualized the expression of these genes in the
304 same manner as the gene sets above, we found no discernible variations in the expression levels
305 of the meta-PCNA genes between the germ layers of WT and *Nipbl*$^{+/-}$ embryos (fig. S11A and
306 data S19). As another measure of cell proliferation, we used Seurat to assign cells from LB-stage
307 WT and *Nipbl*$^{+/-}$ embryos into phases (G1, S, and G2/M) of the cell cycle (*55*). Seurat does this by
308 calculating a cell cycle phase score based on the expression of canonical S phase and G2/M phase
309 markers. Seurat considers these marker sets to be anticorrelated in their expression, so when cells
310 express neither, they are considered to be in G1 phase. As expected, in all germ layers, cells from
311 LB-stage WT and *Nipbl*$^{+/-}$ embryos assigned into all phases of the cell cycle (fig. S11B and data
312 S20). In all germ layers, there was no statistically significant difference in the proportions of cells
313 in each cell cycle between WT and *Nipbl*$^{+/-}$ embryos (fig. S11C). We therefore concluded that
314 extensive alterations in cell proliferation are not likely responsible for disparities in cell
315 subpopulations.

316 **_Nipbl_$^{+/-}$ mice underexpress genes predicted to drive the transition of mesoderm cells into**
317 **first heart field**

318 Next, we sought to investigate whether the misdirection of mesoderm cells in *Nipbl*$^{+/-}$
319 embryos at the LB-stage could be attributed to the misexpression of genes that drive the transition
320 of mesoderm cells into either FHF or PM. To answer this question, we also employed CellRank to
321 identify potential driver genes of the FHF and PM fates, using those mesoderm cells that
322 specifically contribute to these developmental pathways (Fig. 5, A and D). CellRank achieves this

323 by calculating a correlation coefficient between the likelihood (absorption probabilities) of cells
324 progressing towards a particular lineage fate and the expression level of individual genes (*51*). In
325 this context, genes with positive correlation coefficients are considered drivers, as their
326 expression elevates alongside an increase in absorption probabilities. Conversely, genes with
327 negative correlation coefficients are deemed anti-drivers, as their expression diminishes with
328 increasing absorption probabilities. To minimize the risk of false discoveries, we set a threshold,
329 accepting genes with correlation coefficients greater than 0.25 as drivers and those less than -0.25
330 as anti-drivers.

331 CellRank identified 383 genes as drivers of FHF fate (Fig. 5B and data S21). Intriguingly,
332 11% of them were transcription factors. Among these were *Hand1* and *Tbx20*, transcription
333 factors previously recognized for their roles in promoting FHF development (*56, 57*). Several
334 transcription factors that have not been previously linked to FHF development were also predicted
335 to be FHF drivers, including *Msx1, Msx2, Foxf1, Hoxb6,* and *Pitx1*. Of these, *Hoxb6* and *Pitx1*
336 play significant roles in segmentation and patterning. Gene set over-representation analysis using
337 MSigDB's Hallmark gene sets (*52*), which characterize well-defined biological states and
338 processes, revealed that FHF drivers were significantly enriched in genes associated with TGF-
339 beta signaling (Fig. 5C), which has been previously described in cardiomyocyte proliferation (*58,*
340 *59*) Conversely, CellRank predicted 225 genes to be anti-drivers of FHF fate, with 16% being
341 transcription factors. Among these were genes known to be involved in pluripotency (*Pou5f1*),
342 mesoderm specification/expansion (*T, Cdx1, Sp5,* and *Zic3*), anterior-posterior patterning (*Hoxb1*
343 and *Hoxa1*), and paraxial mesoderm development (*Hes7, Meis2, Tbx6, Gbx2,* and *Foxb1*).

344 CellRank predicted 66 genes to be drivers and 115 genes to be anti-drivers of PM fate
345 (Fig. 5E and data S22). A substantial proportion (22%) of PM drivers were transcription factors.
346 These included *Tcf15* and *Foxc2*, transcription factors that have been previously recognized for
347 their role in regulating PM development (Fig. 5E) (*60, 61*). Potentially novel PM drivers included

348 *Irx* transcription factors *Irx5, Irx1,* and *Irx3*. Historically, the Irx family has been associated with
349 defining patterns and establishing boundaries in developing body parts and segments (*62*). PM
350 drivers did not exhibit significant over-representation of any Hallmark gene sets (Fig. 5F).
351 Of all PM anti-drivers, 20% were identified as transcription factors, including *Hand1, Cdx2,*
352 *Msx2, Msx1,* and *Cdx4*. Intriguingly, *Hand1, Msx2, and Msx1* had also been predicted as drivers
353 for FHF. This overlap prompted us to conduct a comparative analysis between PM anti-drivers
354 and FHF drivers. Our examination revealed that 40% of all PM anti-drivers were concurrently
355 FHF drivers (Fig. 5, G and H). The confluence of expressed genes between FHF drivers and PM
356 anti-drivers seems to indicate that during normal development, increased expression of these
357 shared genes by mesoderm cells predisposes them to transition into FHF, while decreased
358 expression of these same genes steers them towards PM (Fig. 5I). This set of shared genes may
359 thus be considered to be a "lineage branching signature".

360 How does *Nipbl*-haploinsufficiency impact the expression of driver and anti-driver genes?
361 Examination of *Nipbl*$^{+/-}$ embryos revealed significant misexpression in NMP, MMP, and PMP
362 cells of numerous driver and anti-driver genes associated with both FHF (first heart field) and PM
363 (paraxial mesoderm) (Fig. 5J, Fig. 5K, data S23, and data S24). Interestingly, most affected genes
364 were downregulated. To ascertain whether this was more than coincidental, we employed a Chi-
365 square analysis, which revealed a striking pattern: *Nipbl*$^{+/-}$ embryos underexpressed FHF drivers
366 at a rate far higher than they overexpressed (166 out of 174 DEGs). Similarly, they commonly
367 underexpressed PM anti-drivers (40 out of 51), aligning with expectations since many PM anti-
368 drivers serve dual roles as FHF drivers. Among the underexpressed FHF drivers in *Nipbl*$^{+/-}$
369 embryos, 6% were transcription factors (Fig. 5J), including well-known regulators of FHF
370 development like *Hand1* as well as other factors such as *Foxf1, Prdm6, Msx1, Hoxb6,* and *Pitx1*.
371 These results suggest that, in the context of *Nipbl*-haploinsufficiency, skewed differentiation of

372 mesoderm cells towards the PM lineage at the expense of FHF lineage may be attributable to the
373 underexpression of genes driving FHF fate.

374 **_Nipbl_$^{+/-}$ mice show large changes in the expression of major developmental regulators in all**
375 **germ layers**

376 Using DGEA comparing the germ layers of LB-stage WT and *Nipbl*$^{+/-}$ embryos, we found
377 that *Nipbl*$^{+/-}$ embryos misexpressed hundreds of genes across all germ layers, with
378 underexpression being more common than overexpression in all cases (Fig. 6A). The majority of
379 gene expression changes were subtle, i.e., less than 2-fold (Fig. 6A). This pattern agrees with
380 earlier studies (*16, 28, 32, 33*), supporting a broad role of *Nipbl* in enhancing gene expression.
381 Gene set enrichment analysis (GSEA) for MSigDB's Hallmark gene sets found that *Nipbl*$^{+/-}$
382 embryos showed enrichment, and no de-enrichment, for 4 out of 40 gene sets: oxidative
383 phosphorylation in the mesoderm, Mtorc1 signaling in the ectoderm, and G2M checkpoint, Myc
384 targets, and E2F targets in the endoderm (Fig. 6B). In these cases, however, enrichment was
385 driven by a relatively small number of the genes in these sets; the vast majority were either not
386 differentially expressed in *Nipbl*$^{+/-}$ embryos or showed changes much smaller than 1.5-fold (Fig.
387 6C).

388 Genes exhibiting substantial changes in expression – greater than 2-fold upregulated or
389 downregulated – were nevertheless observed across every germ layer of *Nipbl*$^{+/-}$ embryos (Fig.
390 6A). Individual curation of these genes unveiled associations with developmental processes (Fig.
391 6D, data S25, data S26, and data S27). Within the mesoderm, two- to six-fold upregulation of
392 genes involved in Nodal signaling, including *Tdgf1, Eomes, Lefty2, Lhx1*, was observed (*63-66*) .
393 At the same time, three clusters of genes were robustly downregulated: those associated with
394 primitive erythropoiesis, including *Gata1* and *Klf1* (2-3 fold) (*67, 68*), *Hox* genes such as *Hoxb8,*
395 *Hoxb5,* and *Hoxa5* (2-fold), and genes associated with Notch signaling, including *Hey2* and *Nr2f2*
396 (2-fold) (*69, 70*) (Fig. 6D).

397 In the ectoderm, five transcription factor genes tied to the hedgehog signaling pathway
398 were downregulated more than 2-fold, including *Pax2, Sox9, Nkx6-1, Sox1,* and *Irx2* (*71-75*) (Fig.
399 6D). In the endoderm, there was a distinct 2-3 fold upregulation of several adult hemoglobin
400 genes such as *Hbb-bt, Hba-a2, Hba-a1,* and *Hbb-bs* (Fig. 6D). Intriguingly, consistent patterns
401 were evident in both the ectoderm and endoderm, mirroring the mesoderm's trends. Genes related
402 to Nodal signaling were upregulated, and *Hox* genes were noticeably downregulated across all
403 germ layers. The magnitude of changes in Nodal signaling and *Hox* genes, being greater than 2-
404 fold in both directions, and their critical role in various germ layers strongly suggest that
405 disruptions in these pathways could be major contributors to phenotypes observed in *Nipbl*$^{+/-}$
406 mice.

407 **_Nipbl_$^{+/-}$ mice overexpress _Nanog_ during and after gastrulation**

408 To uncover developmental processes that might be the most broadly impacted in *Nipbl*$^{+/-}$
409 embryos, we used the large, significant, changes in gene expression (greater than 2-fold up or
410 down) as input data for STRING (*76*), a database and algorithm that constructs a network of
411 potential gene interactions (data S28), in this case limiting predictions to those informed by
412 experiments demonstrating co-expression or protein-protein interactions. The outcome of this
413 analysis (Figs. 7, A-B) revealed that many genes with substantial expression changes in *Nipbl*$^{+/-}$
414 embryos are predicted to interact either directly or indirectly with *Nanog*. Indeed, *Nanog* itself
415 was overexpressed across all three germ layers in *Nipbl*$^{+/-}$ embryos, with a more pronounced
416 overexpression, exceeding 2-fold, in the mesoderm and endoderm (Fig. 7B).

417 *Nanog*, which encodes a transcription factor, is notably expressed at two junctures during
418 mouse embryonic development. Initially, it appears at the blastocyst stage, where it plays an
419 essential role in sustaining the pluripotency of the inner cell mass cells (*77*). Subsequently, during
420 gastrulation, *Nanog* expression temporarily surges, only to be silenced as cells transition out of
421 the primitive streak (*78*). This expression pattern could be observed directly in our data: In Fig.

422 7C, we ordered mesodermal cell populations from LB-stage embryos by their stage in
423 gastrulation, as inferred from their RNA velocity positions (Fig. 4A). In WT NMPs, *Nanog* is
424 expressed at a relatively high level. As cells move through successive stages of gastrulation
425 within WT embryos, *Nanog's* expression exhibits a consistent decline (Fig. 7C and data S29),
426 reaching near zero in cells of the first and second heart field (FHF and SHF). An exception to this
427 trend occurs in primordial germ cells, which are known to maintain elevated *Nanog* expression
428 (*79*). In contrast, *Nipbl*$^{+/-}$ embryos (Fig. 7C) *Nanog* expression remained elevated across all cell
429 populations. Although expression still falls after the NMP stage, in *Nipbl*$^{+/-}$ cells, *Nanog* never
430 falls to WT levels. This is evident even in primordial germ cells. Moreover, the overexpression of
431 *Nanog* in *Nipbl*$^{+/-}$ embryos is sustained up to the CC-stage (Fig. 7D and table S7), where it
432 reaches a 10-fold increase over WT levels. These observations indicate that *Nipbl*-
433 haploinsufficiency leads to a marked failure of *Nanog* downregulation.

434 Another transcription factor that plays a role in maintaining pluripotency is *Pou5f1* (Oct4).
435 From studies in embryonic stem cells, it is known that *Pou5f1* promotes the expression of *Nanog*,
436 and conversely, *Nanog* promotes the expression of *Pou5f1* (*80*). Given this, we were curious
437 about whether *Nipbl*$^{+/-}$ embryos might also overexpress of *Pou5f1*. Despite the fact that *Pou5f1* is
438 more highly expressed than *Nanog*, we did not observe any statistically significant difference in
439 *Pou5f1* expression across mesodermal cell types or throughout LB-stage *Nipbl*$^{+/-}$ embryos (Fig.
440 7E, Fig. 7F, data S30, and table S8). However, a more than 2-fold elevation in *Pou5f1* expression
441 was detected in *Nipbl*$^{+/-}$ embryos at CC-stage (Fig. 7F).

442 We sought to determine whether elevated *Nanog* expression in *Nipbl*$^{+/-}$ embryos during
443 gastrulation might reflect some sort of non-specific overactivity of the *Nanog* gene. To investigate
444 this hypothesis, we looked at *Nanog* expression in WT and *Nipbl*$^{+/-}$ embryonic stem cells (ESCs),
445 as ESCs are known to express *Nanog* (as do to the blastocyst inner cell mass cells from which
446 ESCs are derived). We generated WT and *Nipbl*$^{+/-}$ (ESCs) by treating *Nipbl*$^{Flex/+}$ (*Nipbl*$^{+/-}$) ESCs

447 with Flp recombinase (which inverts a gene trap in *Nipbl*-Flex ESCs in such a way that reverses
448 gene trapping) to produce *Nipbl*$^{Flrt/+}$ (WT) ESCs (Fig. 7G) (*40*). We assessed *Nanog* expression in
449 nine independent clones from each genotype, using quantitative reverse transcription polymerase
450 chain reaction (RT-qPCR),. We observed no discernible difference in *Nanog* expression between
451 *Nipbl*$^{+/-}$ and WT ESCs, a pattern that was also observed for *Pou5f1* (Fig. 7H and data S31). As a
452 control, we showed that *Nipbl*$^{+/-}$ ESCs exhibit the expected reduced expression of *Nipbl*, as also
453 seen in *Nipbl*$^{+/-}$ embryos. Collectively, these findings support the conclusion that the
454 overexpression of *Nanog* in *Nipbl*$^{+/-}$ LB-stage and CC-stage embryos is result of a specific failure
455 to appropriately suppress *Nanog* following gastrulation, rather than unusually elevated *Nanog*
456 expression persisting from early embryogenesis.

457 ***Nanog* overexpression accounts for many of the gene expression changes in *Nipbl*$^{+/-}$ mice**

458 As a critical regulator of pluripotency in early embryonic development, *Nanog* directly
459 influences the expression of a multitude of other genes. This relationship prompted us to ask how
460 many of the gene expression differences in *Nipbl*$^{+/-}$ embryos might be attributable to the
461 overexpression of *Nanog*. To address this question, we analyzed data from a recent study (*81*) in
462 which mice were engineered to express *Nanog* under doxycycline-inducible (Dox) control. In that
463 study, bulk RNA sequencing was performed on untreated (*Nanog* Dox-) and Dox-treated (*Nanog*
464 Dox+) embryos at stages E7.5 and E9.5, with E7.5 corresponding closely to the LB-stage
465 embryos analyzed here. The E7.5 *Nanog* Dox+ embryos that were sequenced had been treated
466 with doxycycline from E4.5 to E7.5, while E9.5 *Nanog* Dox+ embryos were treated with
467 doxycycline from E6.5 to E9.5. Differential gene expression analysis, performed by the authors,
468 between *Nanog* Dox- and *Nanog* Dox+ embryos found 726 and 635 differentially expressed genes
469 (DEGs) at E7.5 and E9.5, respectively.

470 We analyzed the DEGs in E7.5 and E9.5 *Nanog* Dox+ embryos in three ways. First, we
471 plotted the fold change in expression of genes in *Nipbl*$^{+/-}$ embryos, as a whole, from that of WT

472 embryos, at both LB- and CC-stages (Fig. A & C, data S32, and data S33). We colored Genes
473 significantly overexpressed in *Nipbl*$^{+/-}$ embryos pink and those significantly underexpressed, light
474 blue. We colored DEGs in *Nipbl*$^{+/-}$ embryos that are also significantly overexpressed or
475 underexpressed in *Nanog* Dox+ embryos, red and blue, respectively. Remarkably, at the LB-
476 stage, *Nipbl*$^{+/-}$ embryos overexpressed 84.4% and underexpressed 76.5% of the same genes
477 identified in E7.5 *Nanog* Dox+ embryos (Fig. 8B). At CC-stage, *Nipbl*$^{+/-}$ embryos overexpressed
478 76.8% and underexpressed 59.8% of the same genes detected in E9.5 *Nanog* Dox+ embryos (Fig.
479 8D). Statistical evaluation using both the Chi-Square and Binomial tests affirmed that the
480 probability of such overlaps occurring by chance were exceedingly low (Fig. 8, B and D). At CC-
481 stage, we discovered that *Nipbl*$^{+/-}$ embryos also bore substantial similarities to E9.5 *Nanog* Dox+
482 embryos in their gene expression profiles (data S34).

483 Second, we examined how closely *Nipbl*$^{+/-}$ embryos mirrored the gene expression
484 alterations of *Nanog* Dox+ embryos, not only in terms of the genes affected but also the
485 magnitude of their fold changes. To do this, we categorized genes into two groups at each stage:
486 those overexpressed in *Nanog* Dox+ embryos (ranked from top to middle) and those
487 underexpressed (ranked from bottom to middle), ordering them by ascending Q-value as in (*81*).
488 Next, we constructed a heatmap that visually captures the fold changes in gene expression within
489 *Nanog* Dox+ whole embryos, contrasting them with the corresponding fold changes in the
490 individual germ layers of *Nipbl*$^{+/-}$ embryos (Fig. 8, E and F). E7.5 *Nanog* Dox+ embryos
491 overexpressed genes linked with the Nodal signaling pathway (*Tdgf1, Eomes, Fgf5, Lhx1, Foxd3,*
492 *Lefty2,* and *Nodal*) (*63-66, 82-84*) and those integral to maintaining pluripotency (*Nanog* and
493 *Pou5f1*) (*80*). Concurrently, they underexpressed genes associated with primitive erythropoiesis
494 (*Hbb-bh1, Tal1,* and *Lmo2*) (*85, 86*) and specific *Hox* genes (*Hoxa1, Hoxb3os,* and *Hoxb1*).
495 Strikingly, this pattern of fold changes was mirrored, to a remarkable degree, in the germ layers of
496 LB-stage *Nipbl*$^{+/-}$ embryos (Fig. 8E). E9.5 *Nanog* Dox+ embryos overexpressed not only *Nanog*

497 but also several genes connected to the Wnt signaling pathway, such as *Ror2, Ptk7,* and *Fzd2* (*87-*
498 *89*). In parallel, a pronounced underexpression was observed in numerous hemoglobin genes
499 (*Hbb-bh1, Hbb-y, Hba-x,* and *Hba-a1*) as well as genes encoding definitive erythrocyte proteins,
500 including *Gypa, Spta1, Ank1,* and *Sptb* (*90-93*). This pattern of underexpressed genes was
501 strongly mirrored in the germ layers of CC-stage *Nipbl*$^{+/-}$ embryos. In the CC-stage *Nipbl*$^{+/-}$
502 embryos, only a select few overexpressed genes were replicated, the most prominent of which
503 was *Nanog* itself.

504 Third, we again plotted the fold change in expression of genes in *Nipbl*$^{+/-}$ embryos from
505 that of WT embryos, but this time across all germ layers, rather than across whole embryos (Fig.
506 8, G and H). Genes colored red and blue are DEGs in *Nipbl*$^{+/-}$ embryos that are also significantly
507 overexpressed or underexpressed in *Nanog* Dox+ embryos, respectively. When we manually
508 curated the red and blue genes that were expressed with more than 2-fold change, we found that at
509 LB-stage, the mesoderm of *Nipbl*$^{+/-}$ embryos underexpressed genes associated with primitive
510 erythropoiesis (*Hbb-bh1*, *Klf1*, and *Gata1*) (*67, 68*) and hedgehog signaling (*Pax2* and *Sox9*) (*71,*
511 *72*) (Fig. 8G). All germ layers showed upregulation of genes involved with Nodal signaling
512 (*Tdgf1, Lefty2, Eomes, Lhx1,* and *Ffg5*) (*63-66, 82*) and underexpression of *Hox* genes (*Hoxb8*
513 *and Hoxb6*). At CC-stage, all germ layers of *Nipbl*$^{+/-}$ embryos showed downregulation of genes
514 associated with definitive erythropoiesis, encompassing adult hemoglobins including *Hbb-y, Hba-*
515 *a1,* and *Hba-x* (Fig. 8H, data S35, data S36, and data S37). These findings lead us to infer that
516 many of the largest changes in gene expression that occur in *Nipbl*$^{+/-}$ embryos, as well as the
517 developmental pathways they regulate, are likely attributable to the overexpression of *Nanog*.

518 **_Nipbl_$^{+/-}$ mice exhibit delayed expression of anterior _Hox_ genes**

519 We noticed in LB-stage *Nipbl*$^{+/-}$ embryos exhibited a significant underexpression,
520 exceeding 2-fold, in *Hox* genes across all germ layers, especially genes within the *Hoxb* cluster
521 (Fig. 8G). As *Hox* genes are integral to spatial patterning in development, we compared the

522 expression of all *Hox* genes in *Nipbl*$^{+/-}$ embryos with their WT counterparts. The results revealed
523 extensive misexpression at both LB- and CC-stage. At LB-stage, any *Hox* gene that was
524 significantly differentially expressed was underexpressed in *Nipbl*$^{+/-}$ embryos in every germ layer
525 where expression was detectable (Fig. 9A). At CC-stage, we observed a trend of underexpression
526 across all germ layers for the majority of misexpressed *Hox* genes (Fig. 9B), but exceptions were
527 found in the mesoderm and ectoderm, where a select few *Hox* genes were overexpressed. These
528 included *Hoxb1* and *Hoxb2* in the mesoderm and *Hoxa1* in the ectoderm, all of which are
529 characterized as anterior *Hox* genes.

530 *Hox* genes are organized into chromosomal clusters that turn on expression in a wave-like
531 temporal pattern, producing highly structured and sequential activation during development,.
532 Specifically, genes at the 5' end of the cluster, known as anterior *Hox* genes, are activated early in
533 development, and progressively silenced as genes further forward the 3' end (posterior *Hox* genes)
534 are turned on (*94*). To investigate whether our data reflected this known pattern, we examined the
535 *Hoxb* genes, as they were the most prominently expressed *Hox* genes at both LB- and CC-stage.
536 After ordering them according to their chromosomal positions, we plotted their expression within
537 the mesoderm and ectoderm of WT embryos at the corresponding developmental stages. In the
538 mesoderm of LB-stage embryos anterior *Hoxb* genes were actively expressed, while posterior
539 *Hoxb* genes were only faintly detectable, if at all (Fig. 9C). At CC-stage, the expression of
540 anterior *Hoxb* genes decreased from their LB-stage levels, accompanied by the emergence of
541 posterior *Hoxb* gene expression. This dynamic shift in expression aligns closely with the
542 established understanding of *Hox* gene regulation during development. In contrast, the ectoderm
543 exhibited a less pronounced pattern, with both anterior and posterior *Hoxb* genes expressing at
544 minimal levels. Furthermore, no discernible difference in the expression of anterior *Hox*b genes
545 was detected in the ectoderm between LB- and CC-stage embryos (Fig. 9D).

546 Next we looked at the expression of *Hoxb* genes in the mesoderm of *Nipbl*$^{+/-}$ embryos. As
547 in WT embryos, anterior *Hoxb* genes were actively expressed at LB-stage, while posterior *Hoxb*
548 genes were silent (Fig. 9E). However, *Hoxb1* and *Hoxb2* were expressed at levels that were
549 conspicuously lower than in WT (Fig. 9E). At CC-stage, *Nipbl*$^{+/-}$ embryos reflected the general
550 trend found in WT embryos, with anterior *Hoxb* genes waning in expression as the posterior *Hoxb*
551 genes began their ascent (Fig. 9F, note axis scale). Interestingly, the two most anterior Hox genes,
552 *Hoxb1* and *Hoxb2* did not decline as much, proportionally, as they did in WT embryos, with the
553 outcome being that, by CC-stage, expression of *Hoxb1* and *Hoxb2* in *Nipbl*$^{+/-}$ embryos was
554 actually higher than in WT (Fig. 9F). These results collectively provide evidence for a temporal
555 disruption in the *Hox* gene expression program in *Nipbl*$^{+/-}$ embryos. Specifically, Hox genes
556 appear to be delayed in initiating expression at the LB-stage and similarly tardy in becoming
557 suppressed at the CC-stage.

558 ***Nipbl*$^{+/-}$ mice show anteriorization of thoracic vertebrae, with left-right asymmetry**

559 A previous study (*95*) showed that a knockout of *Hoxb1-Hoxb9* in mice led to the
560 anteriorization of the axial skeleton, characterized by an increased number of thoracic vertebrae
561 with ribs. Since *Nipbl*$^{+/-}$ mice underexpress *Hox* genes and show delayed regulation of anterior
562 *Hox* genes, we were curious if similar anteriorizations could be detected in *Nipbl*$^{+/-}$ mice (Fig. 9).
563 To explore this, we analyzed the vertebrae and ribs of 15 WT and 12 littermate *Nipbl*$^{+/-}$ embryos
564 at E18.5, using alizarin red and alcian blue stains. In WT embryos, we confirmed the presence of
565 the normal count of 13 ribs (Fig. 10A). In contrast, many *Nipbl*$^{+/-}$ embryos displayed 14 thoracic
566 vertebrae, the 14th of which bore a range of rib growths, often displaying left-right asymmetry
567 (Fig. 10A). We classified the growths of these 14th vertebrae based on severity, ranking them
568 from low (S1) to high (S5), and documented left-right differences (Fig. 7B). Three out of 7
569 *Nipbl*$^{+/-}$ embryos displayed partial growth of a 14th rib on the left side only (S1-S2). These
570 growths were either cartilage only or disconnected formation of rib nub with cartilage, but never

571 any whole ribs. Four out of 7 *Nipbl*$^{+/-}$ embryos showed bilateral growth (S3-S5). Partial growth
572 occurred on both the left and right sides, but complete formation a whole rib only ever occurred,
573 in one instance, on the right side (S5). In total, 58% of E18.5 *Nipbl*$^{+/-}$ embryos demonstrated some
574 form of rib growth from a 14th thoracic vertebra (Fig. 10C). These findings suggest that
575 anteriorization of the axial vertebrae does occur in *Nipbl*$^{+/-}$ embryos, possibly as a result of
576 delayed activation (and inactivation) of anterior Hox gene expression.

577 **Discussion**

578 The *Nipbl*$^{+/-}$ mouse, a model of the multisystem birth defects syndrome known as Cornelia
579 de Lange Syndrome (CdLS) (*96*), offers a distinct lens through which to explore the genetic
580 origins of birth defects. *Nipbl*$^{+/-}$ mice display many of the same birth defects observed in human
581 CdLS and a key feature of *Nipbl*-deficiency across various organisms is quantitative alterations in
582 gene expression, including both upregulation and downregulation, affecting at least hundreds of
583 genes in every tissue (*28, 40, 97*). Previous research into animal models of *Nipbl*-deficiency
584 suggested that the root causes of birth defects in CdLS likely manifest during a period when
585 progenitor cell populations are being formed for all major tissue and organ lineages (*32, 40*).
586 Therefore, to elucidate how early, cell-type-specific changes in gene expression might contribute
587 to the onset of birth defects, we used scRNAseq to conduct a comparative analysis between
588 *Nipbl*$^{+/-}$ embryos and their WT littermates on key developmental stages, focusing on the
589 conclusion of gastrulation, LB-stage (approximately E7.5) to early CC-stage (approximately
590 E7.75) (*41*).

591 **Changes in the sizes of different mesodermal cell populations in *Nipbl*$^{+/-}$ embryos**
592 **foreshadow pathological changes in tissue composition and patterning in CdLS**

593 As shown in Fig. 1, we showed that *Nipbl*$^{+/-}$ embryos do not lack any cell populations
594 found in WT embryos. However, LB-stage *Nipbl*$^{+/-}$ embryos possess fewer total mesoderm cells,
595 and within the mesoderm, they had fewer PEs and FHF cells, and more PM cells (Fig. 2). These

596 changes, while not always large, may be physiologically significant. For example, the observation
597 that PEs are reduced in number is consistent with defects in blood formation and circulation
598 observed in *Nipbl*-morphant zebrafish (*32*), and thrombocytopenia in CdLS (*98*).

599 In addition, the reduction in FHF cell number could contribute to the presence of CHDs in
600 *Nipbl*$^{+/-}$ mice and individuals with CdLS (*40, 99*). In a prior study in *Nipbl*$^{+/-}$ mice, the majority of
601 CHDs observed were in heart regions (right ventricle, atrial/ventricular septa) considered to be
602 derivatives of the second heart field (SHF) (*40, 100*). Although FHF cells are not thought to give
603 rise directly to "SHF structures", they likely exert indirect effects on the development of SHF
604 derivatives. This echoes the conclusions of (*40*), who found that complex interactions between
605 different cell types (even between non-cardiogenic and cardiogenic cell types), influence the
606 incidence of CHDs in *Nipbl*$^{+/-}$ mice.

607 The increased numbers of PM cells at LB-stage are intriguing, as it correlates with defects
608 in PM derivatives, such as the axial skeleton (Fig. 10) and limbs in *Nipbl*$^{+/-}$ mice (*101*), both of
609 which are notable features of CdLS (*102, 103*). How an overabundance of PM cells, however,
610 could contribute to limb defects is not obvious. Nevertheless, these findings suggest that birth
611 defects may arise not only from the absence of specific progenitor cell populations, but also from
612 progenitor cell misallocation events that alter relative proportions of cell types.

613 **Progenitor cell misallocation in *Nipbl*$^{+/-}$ correlates with changes in expression of cell fate**
614 **driver genes**

615 *Nipbl*$^{+/-}$ embryos at LB-stage did not exhibit global changes in apoptotic activity, cell
616 proliferation, developmental timing, or overall lineage trajectories (Fig. 3 and Fig. 4, A and B).
617 However, *Nipbl*$^{+/-}$ nascent mesoderm cells appeared to differentiate more frequently than WT
618 cells into PM than into FHF (Fig. 4, C to I). Therefore, we conclude that misallocation of
619 mesoderm cells in *Nipbl*$^{+/-}$ embryos is driven by alterations in early cell fate choices, resulting in
620 disparities in cell numbers, occurring in such a way that *Nipbl*-haploinsufficiency leaves the main

621 pathways of lineage trajectories intact, while subtly influencing cell fate decisions along these
622 trajectories. The reduced differentiation into FHF may occur as a result of underexpression of
623 genes predicted to drive FHF differentiation (Fig. 4).

624 **Is *Nanog* overexpression responsible for gene expression changes in *Nipbl*$^{+/-}$ embryos?**

625 Intriguingly, we observed some large shifts in gene expression (exceeding two-fold)
626 across all germ layers of *Nipbl*$^{+/-}$ embryos at both LB- and CC-stages (Figs. 6D and 8H), a finding
627 distinct from prior studies in which gene expression differences of this magnitude were rarely
628 detected (*16, 28, 32, 33*). One of the largest changes in gene expression we observed was
629 overexpression of *Nanog,* which occurred in LB-stage *Nipbl*$^{+/-}$ embryos and persisted to CC-stage
630 (Figs. 7 and 8). Using data from (*81*), we demonstrated that many of the gene expression changes
631 in *Nipbl*$^{+/-}$ embryos mirror those seen in embryos with induced *Nanog* overexpression at
632 corresponding stages: 77% of total DEGs at LB-stage, and 72% of total DEGs at CC-stage (Fig.
633 8, A to F). We referred to such genes as "*Nanog*-overexpression-linked" genes.

634 A question that arises from these observations is whether all gene expression alterations
635 observed in *Nipbl*$^{+/-}$ embryos might be due to *Nanog* overexpression. For a number of reasons, we
636 do not think that this is the case. First, for the largest gene expression shifts (greater than two-fold
637 over- or underexpressed), only 39% at LB-stage and 18% at CC-stage were misexpressed in the
638 same direction as DEGs in embryos in which *Nanog* was overexpressed (Fig. 8). Second,
639 although *Nanog* continues to be overexpressed in *Nipbl*$^{+/-}$ embryos from LB- to CC-stage, the
640 magnitude of its expression decreases substantially by CC-stage (Fig. 7). Third, *Nipbl*-
641 haploinsufficiency itself causes significant large expression changes in genes other than the
642 "Nanog-overexpression-linked" genes. For instance, we observed the reduced expression of genes
643 linked to the Notch signaling pathway in the mesoderm and an increased expression of adult
644 hemoglobin genes in the endoderm (Fig. 6). Therefore, we conclude that the extensive gene
645 expression changes in *Nipbl*$^{+/-}$ embryos are unlikely to be due to *Nanog* overexpression alone.

646 **Delayed anterior *Hox* gene expression and its relationship to skeletal anomalies in *Nipbl*$^{+/-}$**
647 **embryos**

648 In Figs. 4D and 8G, we showed that large decreases in *Hox* gene expression occur across
649 all germ layers of LB-stage *Nipbl*$^{+/-}$ embryos. While only a subset of *Hox* genes is underexpressed
650 by more than two-fold, many *Hox* genes across all germ layers show small (less than two-fold)
651 but significant decreases in expression (Fig. 9A). We also provide evidence for a temporal
652 disturbance in anterior *Hox* gene expression in *Nipbl*$^{+/-}$ embryos, whereby *Nipbl*-
653 haploinsufficiency postpones the onset of anterior *Hox* gene expression at LB-stage and delays
654 their subsequent deactivation at CC-stage (Fig. 9, C to F). These shifts in *Hox* gene expression
655 dynamics mirror those in a previous study of *nipbl*-morphant zebrafish, in which genome-
656 location-specific misexpression of *hox* genes was also observed (*104*). In Fig. 10, we showed that
657 anteriorization of thoracic vertebrae occurs in E18.5 *Nipbl*$^{+/-}$ embryos (Fig. 9). Such alterations in
658 axial skeletal development parallel skeletal anomalies that have been identified in individuals
659 with CdLS, including fused, absent, or misshapen ribs (*105*). Since *Nipbl*-deficiency is the
660 dominant form of CdLS, it is interesting to speculate that these skeletal anomalies might originate
661 from underexpression of and delayed anterior *Hox* gene activity during gastrulation.

662 **Differential Nodal signaling in *Nipbl*$^{+/-}$ embryos: potential contributors to left-right**
663 **patterning defects in CdLS**

664 In Fig. 4D, we showed that *Nipbl*$^{+/-}$ embryos at LB-stage show large overexpression genes
665 associated with the Nodal signaling pathway across all germ layers. Nodal signaling, pivotal for
666 left-right patterning (*106*), ensures the appropriate positioning and morphogenesis of the
667 musculoskeletal system and internal organs (*107*). In Fig. 10, we showed that growths of a 14th
668 vertebrae in E18.5 *Nipbl*$^{+/-}$ embryos displayed left-right asymmetry, with the growth of a whole
669 14th rib only occurring on the right side of the embryo. Such patterning alterations are consistent
670 with heightened incidence of right-sided skeletal anomalies in CdLS (*102, 108, 109*). Changes in

671 left-right patterning are also consistently associated with intestinal malrotation and some types of
672 cardiac anomalies in *Nipbl*$^{+/-}$ mice, *nipbl*-morphant zebrafish, and CdLS (*28, 40, 110*). These data
673 suggest that left-right patterning defects in CdLS might be due to overexpression of Nodal
674 signaling genes as early as gastrulation.

675 ***Nipbl*, global regulation of gene expression, and chromatin architecture**

676 This study, along with prior research involving mouse and zebrafish models, reinforces
677 the notion that the chromosomal architecture surrounding genes significantly influences their
678 sensitivity to variations in *Nipbl* levels. For instance, in E13.5 and E17.5 *Nipbl*$^{+/-}$ mouse brains,
679 we observed that most expression alterations in the 22-gene protocadherin beta (*Pcdhb*) cluster
680 were localized to genes situated at the 3' end of the cluster (*28*). Similarly, in developing pectoral
681 fins of nipbl-morphant zebrafish at 32-36 hours post-fertilization, Hox genes located near the 3'
682 end of three separate Hox clusters (*Hoxa*, *Hoxc*, and *Hoxd*) showed a unique pattern of
683 overexpression (*104*). In the present study, the data in Fig. 9 demonstrate that anterior *Hoxb* genes
684 – located at the 3' end of the *Hoxb* cluster – are overexpressed in the mesoderm of *Nipbl*$^{+/-}$
685 embryos at gastrulation. Thus, we consistently observe a strong influence of chromosomal
686 location on gene sensitivity to *Nipbl* levels, particularly within gene clusters, and this influence is
687 already apparent at gastrulation.

688 In this study, we showed widespread changes in gene expression, in all cell types, evident
689 at gastrulation. This apparently leads to changes in patterning, cell fate decisions, and
690 misallocation within early progenitor cell lineages. We show data relating these changes to
691 structural birth defects in *Nipbl*$^{+/-}$ mice. We conclude that the structural changes observed in
692 *Nipbl*$^{+/-}$ animals and, by extension, CdLS, likely result from the interplay of these processes
693 during the course of development. What is needed next is to profile chromatin topology and how
694 it relates to changes in gene expression in *Nipbl*$^{+/-}$ embryos during gastrulation. For example, in
695 the present study, we find that only a few genes are highly over- or underexpressed in early

696 *Nipbl*$^{+/-}$ embryos. What makes these genes, such as *Nanog*, uniquely susceptible to *Nipbl*-
697 deficiency at this developmental time? Tracking temporal changes in *Nipbl's* role in regulating
698 accessibility of these genes and their enhancers/silencers could provide insight into this question.

699 **Materials and Methods**

700 **Generation of WT and *Nipbl*$^{+/-}$ Mice**

701 WT and *Nipbl*$^{+/-}$ mouse littermates were generated by mating *Nanog*$^{Cre/+}$ mice with
702 *Nipbl*$^{Flox/Flox}$ mice. *Nipbl*$^{Flox/Flox}$ mice were generated as described in (*40*). They have an inverted
703 gene trap cassette encoding *Bgeo* that is flanked by target sites for *Cre* recombinase in intron 1 of
704 both alleles of the *Nipbl* gene. In this inverted orientation, which we call Flox, there is no trapping
705 of the *Nipbl* gene and *Nipbl* is expressed normally. However, when this cassette is exposed to Cre
706 recombinase, the gene trap cassette gets inverted into a non-inverted orientation which we call
707 FIN. In this non-inverted orientation, trapping of the *Nipbl* gene occurs and *Bgeo* is expressed as a
708 reporter of successful gene trapping. *Nanog*$^{Cre/+}$ mice were generated as described in (*111*). They
709 have a transgene encoding a *Cre* recombinase downstream of a promoter of the *Nanog* gene and
710 recombines in the first cells of the developing embryo. Thus, mating *Nanog*$^{Cre/+}$ mice with
711 *Nipbl*$^{Flox/Flox}$ mice results in littermates that are either *Nipbl*$^{Flox/+}$ or *Nipbl*$^{FIN/+}$, entirely. *Nipbl*$^{Flox/+}$
712 littermates express *Nipbl* at wildtype levels and phenocopy WT mice, and as such, are essentially
713 WT. *Nipbl*$^{FIN/+}$ mice express *Nipbl* at levels lower than WT and phenocopy *Nipbl*$^{+/-}$ mice, and as
714 such, are essentially *Nipbl*$^{+/-}$.

715 **Timing of Mouse Pregnancies**

716 To generate female mice that were pregnant on the same day, male mice were singly
717 housed and female mice were group-housed in groups of five for a minimum of one week to
718 synchronize their estrous cycles, thus taking advantage of the Lee-Boot effect (*112*). At the
719 beginning of the night cycle, the bedding from female cages were discarded and the bedding from
720 at least two male cages were transferred into each of them. At the beginning of the third night

721 cycle after which the females were exposed to male bedding, all females were transferred into
722 male cages, resulting in two females per male, thus taking advantage of the Whitten effect (*113*).
723 At the end of the third night cycle, the females were inspected for vaginal plugs. Those that had
724 vaginal plugs were considered potentially pregnant. To time the dissection of embryos from
725 potentially pregnant females, we considered the end of the 12-hour night cycle after which the
726 vaginal plug was discovered as E0 and dissected the embryos at the following times post E0: E7.5
727 (7 days + 12 hours) for late bud stage, E7.75 (7 days + 18 hours) for cardiac crescent stage, E7.41
728 (7 days + 10 hours) for early bud stage, E7.58 (7 days + 14 hours) for early head fold stage
729 embryos, and at E18.5 (18 days + 12 hours).

730 **Dissection of Mouse Embryos**

731 Pregnant female mice were euthanized by CO_2 inhalation followed by cervical dislocation.
732 The uterine horns were dissected out of their abdomens with dissection forceps and scissors and
733 placed in a petri dish with 1X DEPC PBS on ice. Individual deciduae were separated from one
734 another and transferred into their own petri dishes with 1X DEPC PBS on ice. Embryos were
735 dissected out of each deciduae under a dissection microscope as described in (*114*). The
736 Reichert's membrane was removed from each embryo. The ectoplacental cone was separated
737 from embryos collected for scRNAseq, and transferred by forcep into microcentrifuge tubes,
738 where they were kept on ice or stored at 20°C as tissue for polymerase chain reaction (PCR)
739 genotyping. The exocoelom was separated from embryos collected for scRNAseq and transferred
740 into a welled plate by wide bore pipette tip with fixative on ice as tissue for genotyping by X-gal
741 stain. Embryos for scRNAseq were transferred by wide bore pipette tip with 1X DEPC PBS into
742 microcentrifuge tubes and kept on ice until dissociation. Tails were separated from E18.5
743 embryos collected for Alcian Blue – Alizarin Red staining. Embryos for Alcian Blue – Alizarin
744 Red staining were transferred by forcep into 10% neutral buffered formalin in scintillation vials
745 and kept at 4°C for 24 hours.

746 **Genotyping of Mouse Embryos and ES Cells**

747 ***X-gal Stain***

748 Since *Nipbl*$^{FIN/+}$ (*Nipbl*$^{+/-}$) mouse embryos express Bgeo (beta-galactosidase), their tissues
749 will turn blue when they are treated with X-gal, a substrate that releases a blue chromophore when
750 enzymatically acted on by beta-galactosidase. The exocoeloms from embryos were transferred
751 into welled plates containing fixative (0.2% Glutaraldehyde, 5mM EGTA, and 2mM MgCl2 in
752 0.1M phosphate buffer, pH 7.5) and kept on ice for a minimum of 15 min. After fixation, the
753 exocoeloms were rinsed with a detergent rinse (0.02% Igepal, 0.01% Sodium Deoxycholate, and
754 2mM MgCl2 in 0.1M phosphate buffer, pH 7.5) before they were treated with X-gal stain (0.02%
755 Igepal, 0.01% Sodium Deoxycholate, 5mM Potassium Ferricyanide, 5mM Potassium
756 Ferrocyanide, and 2mM MgCl2 in 0.1M phosphate buffer, pH 7.5). The exocoeloms were
757 incubated in X-gal stain at 37°C in the dark for a minimum of 1 hour after which they were
758 inspected for coloration under a dissection microscope. Those that turned blue were considered
759 *Nipbl*$^{+/-}$ (*Nipbl*$^{FIN/+}$) and those that did not were considered WT.

760 ***Polymerase Chain Reaction***

761 Ectoplacental cones were treated with 50 ul 60 ug/ml proteinase K in PBND (50 mM KCl,
762 10 mM Tris-HCl, 2.5 mM MgCl2, 0.1 mg/ml gelatin, 0.45% Nonident P40, and 0.45% Tween
763 20), for 1 hour at 55°C to extract DNA for PCR. To deactivate proteinase K so that it does not
764 interfere with DNA amplification, they were then incubated at 95°C for 10 min. To genotype
765 mouse embryos as either *Nipbl*$^{Flox/+}$ (WT) or *Nipbl*$^{FIN/+}$ (*Nipbl*$^{+/-}$) and ES cells as either *Nipbl*$^{Flrt/+}$
766 (WT) or *Nipbl*$^{FLEX/+}$ (*Nipbl*$^{+/-}$), standard Taq PCR was performed on the extracted DNA using the
767 primers and thermocyling protocol previously described in (*40*). Primer 1: 5'-CTCCGC
768 CTCCTCTTCCTCCATC-3'. Primer 2: 5'-CCTCCCCCGTGCCTTCCTTGAC-3'. Primer 3: 5'-
769 TTTGAGGGGACGACGACAGTCT-3'. Thermocycling conditions: 1 cycle a 95°C for 30 sec,
770 30 cycles of 95°C for 1 min, 59°C for 30 sec, and 68°C for 1 min, 1 cycle at 68°C for 5 min, and

771 hold at 4°C. PCR products were treated with loading dye and electrophoresed in agarose gel
772 stained with SYBER Safe DNA Gel Stain (Invitrogen: S33102) without cleanup and visualized
773 under UV light. Flox conformation: 782 base pairs (bp). FIN conformation 518 bp. Flrt
774 conformation: 735 bp. FLEX conformation: 652 bp.

775 **Single Cell RNA Sequencing**

776 Embryos were transferred by wide bore pipette tip in 20 ul 1X DEPC PBS into
777 microcentrifuge tubes. 200 ul of 1X TrypLE Express Enzyme with phenol red (Gibco: 12605010)
778 prewarmed to 37°C was added to each embryo. Embryos were triturated 4x with a wide bore
779 pipette tip. Embryos were incubated at 37°C and triturated 4x with a wide bore pipette tip every
780 minute (4-8 min) until no tissue aggregates were visible under a dissection microscope. TrypLE
781 Expres Enzyme activity was inactivated with the addition of pre-ice-chilled 200 ul 0.04% (m/v)
782 non-acetylated BSA in DPBS (Sigma-Aldrich: B6917-100MG, Gibco: 14190144). The resulting
783 cell suspension was underlayed with 200 ul of 1% non-acetylated BSA in HBSS (Sigma: 55021C-
784 1000ML) using a gel loading pipette tip. The suspension was centrifuged in a swing bucket
785 centrifuge at 300 rcf for 5 min at 4°C. 600 ul of supernatant was removed without disturbing the
786 cell pellet. 600 ul of HBSS was added to the cell pellet and the cell pellet was resuspended in it
787 using a wide bore pipette tip. The resulting cell suspension was centrifuged in a swing bucket
788 centrifuge at 300 rcf for 5 min at 4°C. 600 ul of supernatant was removed without disturbing the
789 cell pellet. The cell pellet was resuspending in the remaining 20 ul of supernatant using a wide
790 bore pipette tip.

791 20 ul single cell suspensions were submitted to the Genomics High Throughput Facility
792 (GHTF) at the University of California, Irvine (*115*) for scRNAseq using 10X Genomics'
793 Chromium Next GEM Single Cell 3' Kit v3.1 (10x Genomics: 1000268), Chromium Next GEM
794 Chip G Single Cell Kit (10x Genomics: 1000120), and Chromium Controller. Passage through the
795 Chromium Controller resulted in sample, cell, and transcript barcoded cDNA, which GHTF

796 amplified by PCR. GHTF assessed the quality and quantity of the amplified cDNA by
797 electrophoresis using Agilent's Agilent High Sensitivity DNA Kit (Agilent: 5067-4626) and
798 Bioanalyzer before library construction. Constructed libraries, representing embryonic samples,
799 were multiplexed and sequenced by GHTF on the Illumina HiSeq 4000 to a minimum depth of 20
800 million read pairs per cell. GHTF demultiplexed Illumina's raw base call (BCL) files and returned
801 FASTQ files as deliverables.

802 **Read Mapping and Cell Calling**

803 Cell Ranger v3.0 was used to map reads onto the GRCm38/mm10 C57BL/6J Mus
804 musculus genome/transcriptome assembly and call cells. Cell Ranger does this by using a read
805 mapper called STAR (*116*), which performs splicing-aware mapping of reads to the genome.
806 Reads are considered confidentially mapped to the genome with a MAPQ of 255. Exonic reads
807 are further mapped to annotated transcripts. A read that is compatible with the exons of an
808 annotated transcript, and aligned to the same strand, is considered mapped to the transcriptome.
809 CellRanger called cells using the EmptyDrops method described in (*117*).

810 **Normalization of Library Depth**

811 Cell Ranger v3.0 was used to normalize the read depth between libraries of the same
812 stage. Cell Ranger does this by subsampling reads from higher-depth libraries until all libraries of
813 the same stage had an equal number of reads.

814 **Removal of Low-Quality Cells and Doublets**

815 Cells exceeding three median absolute deviations in any one of the following criteria
816 among cells of the same stage were considered either low-quality cells or doublets (*118*) and
817 removed: 1) percentage of mitochondrial genes expressed, 2) number of genes expressed, and/or
818 3) number of transcripts detected (fig. S2 and fig. S8).

819 **Normalization of Cell Depth**

820 Seurat v3.0 was used to normalize the read depth between cells of the same stage using the
821 SCtransform method, which is described in (*119*). SCtransform does this by modeling the read
822 counts in a regularized negative binomial model to determine the variation due to read depth and
823 then adjusting that variance according to genes of similar abundances. SCtransform was also
824 used to normalize the read depth between cells when cells were subset from the whole embryo
825 into germ layers and clusters.

826 **Batch Effect Correction**

827 Seurat v3.0 was used to correct for batch effects among libraries of the same stage and
828 genotype. Seurat does this by identifying a set of shared variable genes among the libraries being
829 considered, and using these genes, identifies pairs of cells between any two libraries whose
830 expression of these genes are similar to other. These pairs of cells act as anchors between libraries
831 for batch effect correction and integration (*46*).

832 **Clustering of WT Cells**

833 Seurat v3.0 was used to cluster WT cells. Seurat does this by first performing principal
834 component analysis on the shared variable genes identified during batch effect correction and
835 integration. Principal components whose explained variances exceeded 2 median absolute
836 deviations are used to calculate k-nearest neighbors and construct a shared nearest-neighbor
837 graph. Clusters are determined by optimization of the modularity function using the Louvain
838 algorithm, as described in (Waltman & Jan van Eck, 2013). The number of clusters was
839 controlled by modulating the resolution function.

840 The optimal number of clusters was determined by clustering cells at increasing
841 consecutive numbers of clusters and generating a clustering tree (fig. S5) visualizing how cell
842 cluster identities change as the number of clusters consecutively increase. Clusters are stable
843 when a large proportion of cells are derived from a single preceding cluster rather than multiple
844 preceding clusters. We adopted Shannon Entropy as a measure of these proportions as a measure

845 of intra-cluster stability. A low Shannon Entropy represents high intra-cluster stability. We
846 visually inspected the clustering tree to determine which number of clusters maximized the
847 number of clusters while at the same time minimizing the total Shannon Entropy across all
848 clusters at that number of clusters.

849 We further performed differential gene expression analysis between clusters and
850 visualized the expression of the top differentially expressed genes in a heatmap. We visually
851 inspected the heatmap to confirm that the number of clusters that was selected for intra-cluster
852 stability also displayed inter-cluster differences in gene expression.

853 **Differential Gene Expression Analysis Between Clusters**

854 Seurat v3.0 was used to perform differential gene expression analysis between clusters.
855 For each cluster, Seurat performs the Mann-Whitney U Test (a non-parametric test) between cells
856 in that cluster and all other cells using normalized read counts. P-values were corrected for false
857 discovery using the Bonferroni correction method. Genes with Q-values less than 0.05 were
858 considered statistically significant and differentially expressed.

859 **Projection of *Nipbl*$^{+/-}$ Cells onto WT Clusters**

860 Seurat v3.0 was used to project *Nipbl*$^{+/-}$ cells onto WT clusters. Seurat does this by
861 identifying a set of shared variable genes among WT samples, and using these genes, identifies
862 pairs of cells between WT cells and *Nipbl*$^{+/-}$ cells whose expression of these genes are similar to
863 other. These pairs of cells act as anchors between WT cells and *Nipbl*$^{+/-}$ cells for projecting and
864 sorting *Nipbl*$^{+/-}$ cells into WT clusters (*46*).

865 **Differential Gene Expression Analysis Between Genotypes**

866 Seurat v3.0 was used to perform differential gene expression analysis between WT and
867 *Nipbl*$^{+/-}$ cells. For each cluster, Seurat performs the Mann-Whitney U Test between WT and
868 *Nipbl*$^{+/-}$ cells using normalized read counts. P-values were corrected for false discovery using the

869 Bonferroni correction method. Genes with Q-values less than 0.05 were considered statistically
870 significant and differentially expressed.

871 **Pseudotime**

872 URD was used to calculate the pseudotime of cells from EB-, LB-, and EHF-stage
873 embryos (*48*). URD does this by constructing a diffusion map of transition probabilities and
874 starting with an assigned group of root cells, performs a probabilistic breadth-first graph search
875 using the transition probabilities. This moves step-wise outward from the root cells, until the
876 entire graph is visited. Several simulations are run, and then pseudotime is calculated as the
877 average iteration that visited each cell.

878 **Construction of Lineage Trajectories**

879 Velocyto (*120*) was used to count the numbers of spliced and unspliced transcripts per
880 gene, using default parameters. Reads aligning to exonic regions were counted as spliced. Reads
881 aligning to intronic regions were counted as unspliced. Reads aligning to exon-intron boundaries
882 were considered ambiguous and excluded from downstream analyses. scVelo (*49*) was used
883 estimate RNA velocities. Counts were normalized using the pp.filter_and_normalize() function,
884 moments of unspliced vs spliced abundances were computed using the pp.moments() function,
885 and velocities were computed using the tl.velocity() function, all using default parameters.
886 Lineage trajectories were visually inferred from stream plots of computed RNA velocities.

887 **Calculating Fate Probabilities**

888 CellRank was used to calculate the fate probabilities of mesoderm cells (*51*). CellRank
889 does this by performing RNA velocity-directed random walks from initial cell states to terminal
890 cell states. Fate probabilities correspond to the fraction of walks in which a cell was a part of that
891 that terminated in a particular terminal cell state. Terminal states were set with the
892 *set_terminal_states()* function, absorption probabilities were computed with the

893 *compute_absorption_probabilities()* function, and driver genes were computed with the
894 *compute_lineage_drivers()* function, all with default parameters.

895 **Cell Cycle Phase Assignment**

896 Seurat was used to assign cells into G1, S, or G2/M phase based on the expression of
897 markers of S phase and G2/M phase provided by Seurat. Using the CellCycleScoring() function,
898 Seurat calculated scores for the expression of S phase and G2/M phase markers, while
899 considering the expression of these marker genes to be anticorrelated to one another. When cells
900 express neither, they are considered to be in G1 phase.

901 **Identification of Drivers and Anti-drivers**

902 CellRank was used to identify the drivers and anti-drivers of FHF and PM fates. CellRank
903 does this by calculating a correlation coefficient between the fate probabilities of cells towards
904 their lineage fate and the expression of their genes. Those genes with positive correlation
905 coefficients are considered drivers, since their expression increases as absorption probabilities
906 increase and those with negative correlation coefficients are considered anti-drivers, since their
907 expression decreases as absorption probabilities decrease. To reduce the likelihood of false
908 discovery, we considered those genes with correlation coefficients greater than 0.25 as drivers and
909 those less than -0.25 as anti-drivers.

910 **Gene Set Over-Representation Analysis**

911 clusterProfiler (*121*) was used to perform gene set over-representation analysis. It does
912 this by performing a Fisher's Exact Test on a contingency table of the genes in a gene set that
913 match or do not match the genes of interest. P-values were corrected for false discovery using the
914 Bonferroni correction method. Gene sets with Q-values less than 0.05 were considered
915 statistically significant and over represented by the genes of interest.

916 **Gene Set Enrichment Analysis**

917 FGSEA (*122*) was used to perform gene set enrichment analysis.

918 **Generation of *Nipbl*$^{FLEX/+}$ and *Nipbl*$^{Flrt/+}$ ESCs**

919 ES cells were grown in Glasgow's MEM, 15% heat-inactivated FBS (ES cell qualified,
920 Hyclone SH30071.03E), 1X Glutamine, 1X Pen-Strep, 5mM mercaptoethanol, and 1000 U/ml
921 leukemia inhibitory factor (LIF: ESGRO Millipore). Following limited dilution single cell
922 cloning, a clone of EUC313f02 *Nipbl*$^{FLEX/+}$ ES cells (European Conditional Mouse Mutagenesis
923 Program) was transfected with pCAG-Cre:GFP (Addgene #13776) using Lipofectamine 2000
924 (Invitrogen) to convert the *Nipbl*^{FLEX} allele to the *Nipbl*^{Flrt} conformation in vitro. 48 hours post-
925 transfection, cells were plated at 1 cell/well into several 96 well plates. Clonal colonies were
926 isolated, and the clones were stained for β-galactosidase (lacZ) activity using X-gal. For LacZ
927 staining, the ES cells were fixed for 5 minutes in 2 mM MgCl2, 0.5% glutaraldehyde in 1X PBS,
928 followed by 3 washes with 1xPBS at room temperature. X-gal staining (5 mM K3Fe(CN)6, 5 mM
929 K4Fe(CN)6, 2 mM MgCl2, 1 mg/mL X-gal in 1X PBS) was performed at 37°C until the blue
930 precipitate was detected. Colonies positive for X-gal staining, *Nipbl*$^{FLEX/+}$ ES cells, and negative
931 for X-gal staining, *Nipbl*$^{Flrt/+}$, were verified by PCR genotyping, described above.

932 **Reverse Transcription-Quantitative Polymerase Chain Reaction**

933 Clones for both *Nipbl*$^{FLEX/+}$ and *Nipbl*$^{Flrt/+}$ were expanded, 9 each, and RNA was extracted
934 using the Monarch Total RNA Miniprep Kit (New England Biolabs). cDNA was made using
935 iSCRIPT reverse transcriptase (BioRad). qRT-PCR for *Nipbl*, *Nanog*, and *Pouf5f1* was performed
936 using iTaq SYBR green (BioRad) as per manufacturer's instructions, *Rpl4* was used as the
937 housekeeping gene. *Rpl4* primer 1: 5'-ATCTGGACGGAGAGTGCTTT-3'. *Rpl4* primer 2: 5'-
938 GGTCGGTGTTCATCATCTTG-3'. *Nipbl* primer 1: 5'-AGTCCATATGCCCCACAGAG-3'.
939 *Nipbl* primer 2: 5'-ACCGGCAACAATAGGACTTG-3'. *Nanog* primer 1: 5'-
940 AAATCCCTTCCCTCGCCATC-3'. *Nanog* primer 2: 5'-GCCCTGACTTTAAGCCCAGA-3'.
941 *Pou5f1* primer 1: 5'-CACCCTGGGCGTTCTCTTT-3'. *Pou5f1* primer 2: 5'-
942 GTCTCCGATTTGCATATCTCCTG-3'.

943 **Alcian Blue – Alizarin Red Staining**

944 E18.5 embryos were fixed in 10% neutral buffered formalin in scintillation vials at 4°C
945 overnight. Embryos were washed with H_2O twice over 2 days at room temperature. Embryos
946 were immersed in 95% EtOH for 1 week. Skin was removed from embryos with forceps under
947 dissection microscope. Embryos were stained with 0.02% Alcian Blue (10 mg Alcian Blue + 80
948 ml EtOH + 20 ml glacial acetic acid) over three days. Embryos were washed with 70% EtOH
949 twice in 1 day, 40% EtOH overnight, 15% EtOH in 1 day, and H_2O overnight. Embryos were
950 washed with 1% KOH twice over 3 days until they became translucent. Embryos were stained
951 with 0.015% Alizarin Red (15 mg Alizarin Red + 100 ml 1% KOH) over 3 days. Embryos were
952 washed with 1% KOH three times in 1 day. Fat pads and internal organs were removed from
953 embryos with forceps under dissection microscope. Embryos were washed with 20% glycerol +
954 1% KOH overnight, 50% glycerol + 1% KOH overnight, and 80% glycerol + 1% KOH overnight.
955 Embryos were stored in 100% glycerol at room temperature.

956 **References**

957 1. C. Centers for Disease, Prevention, Update on overall prevalence of major birth defects--
958 Atlanta, Georgia, 1978-2005. *MMWR Morb Mortal Wkly Rep* **57**, 1-5 (2008).

959 2. C. T. Mai *et al.*, National population-based estimates for major birth defects, 2010-2014.
960 *Birth Defects Res* **111**, 1420-1435 (2019).

961 3. R. Sun, M. Liu, L. Lu, Y. Zheng, P. Zhang, Congenital Heart Disease: Causes, Diagnosis,
962 Symptoms, and Treatments. *Cell Biochem Biophys* **72**, 857-860 (2015).

963 4. D. M. Ely, A. K. Driscoll, Infant Mortality in the United States, 2020: Data From the
964 Period Linked Birth/Infant Death File. *Natl Vital Stat Rep* **71**, 1-18 (2022).

965 5. L. Mannini, F. Cucco, V. Quarantotti, I. D. Krantz, A. Musio, Mutation spectrum and
966 genotype-phenotype correlation in Cornelia de Lange syndrome. *Hum Mutat* **34**, 1589-
967 1596 (2013).

968 6. J. M. Opitz, The Brachmann-de Lange syndrome. *Am J Med Genet* **22**, 89-102 (1985).

969 7. M. Ireland, D. Donnai, J. Burn, Brachmann-de Lange syndrome. Delineation of the
970 clinical phenotype. *Am J Med Genet* **47**, 959-964 (1993).

971 8. L. Jackson, A. D. Kline, M. A. Barr, S. Koch, de Lange syndrome: a clinical review of
972 310 individuals. *Am J Med Genet* **47**, 940-946 (1993).

973 9. J. Liu, I. D. Krantz, Cornelia de Lange syndrome, cohesin, and beyond. *Clin Genet* **76**,
974 303-314 (2009).

975 10. I. D. Krantz *et al.*, Cornelia de Lange syndrome is caused by mutations in NIPBL, the
976 human homolog of Drosophila melanogaster Nipped-B. *Nat Genet* **36**, 631-635 (2004).

977 11. E. T. Tonkin, T. J. Wang, S. Lisgo, M. J. Bamshad, T. Strachan, NIPBL, encoding a
978 homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is
979 mutated in Cornelia de Lange syndrome. *Nat Genet* **36**, 636-641 (2004).

980 12. D. Dorsett, I. D. Krantz, On the molecular etiology of Cornelia de Lange syndrome. *Ann N*
981 *Y Acad Sci* **1151**, 22-37 (2009).

982 13. A. Selicorni *et al.*, Clinical score of 62 Italian patients with Cornelia de Lange syndrome
983 and correlations with the presence and type of NIPBL mutation. *Clin Genet* **72**, 98-108
984 (2007).

985 14. L. A. Gillis *et al.*, NIPBL mutational analysis in 120 individuals with Cornelia de Lange
986 syndrome and evaluation of genotype-phenotype correlations. *Am J Hum Genet* **75**, 610-
987 623 (2004).

988 15. G. Borck *et al.*, Father-to-daughter transmission of Cornelia de Lange syndrome caused by
989 a mutation in the 5' untranslated region of the NIPBL Gene. *Hum Mutat* **27**, 731-735
990 (2006).

991 16. J. Liu *et al.*, Transcriptional dysregulation in NIPBL and cohesin mutant human cells.
992 *PLoS Biol* **7**, e1000119 (2009).

993 17. R. Ciosk *et al.*, Cohesin's binding to chromosomes depends on a separate complex
994 consisting of Scc2 and Scc4 proteins. *Mol Cell* **5**, 243-254 (2000).

995 18. P. J. Gillespie, T. Hirano, Scc2 couples replication licensing to sister chromatid cohesion
996 in Xenopus egg extracts. *Curr Biol* **14**, 1598-1603 (2004).

997 19. T. S. Takahashi, P. Yiu, M. F. Chou, S. Gygi, J. C. Walter, Recruitment of Xenopus Scc2
998 and cohesin to chromatin requires the pre-replication complex. *Nat Cell Biol* **6**, 991-996
999 (2004).

1000 20. E. Watrin *et al.*, Human Scc4 is required for cohesin binding to chromatin, sister-
1001 chromatid cohesion, and mitotic progression. *Curr Biol* **16**, 863-874 (2006).

1002 21. K. Nasmyth, C. H. Haering, Cohesin: its roles and mechanisms. *Annu Rev Genet* **43**, 525-
1003 558 (2009).

1004 22. A. Musio *et al.*, X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. *Nat*
1005 *Genet* **38**, 528-530 (2006).

1006 23. M. A. Deardorff *et al.*, Mutations in cohesin complex members SMC3 and SMC1A cause
1007 a mild variant of cornelia de Lange syndrome with predominant mental retardation. *Am J*
1008 *Hum Genet* **80**, 485-494 (2007).

1009 24. L. Mannini *et al.*, SMC1A codon 496 mutations affect the cellular response to genotoxic
1010 treatments. *Am J Med Genet A* **158A**, 224-228 (2012).

1011 25. M. A. Deardorff *et al.*, HDAC8 mutations in Cornelia de Lange syndrome affect the
1012 cohesin acetylation cycle. *Nature* **489**, 313-317 (2012).

1013 26. M. A. Deardorff *et al.*, RAD21 mutations cause a human cohesinopathy. *Am J Hum Genet*
1014 **90**, 1014-1027 (2012).

1015 27. P. Castronovo *et al.*, Somatic mosaicism in Cornelia de Lange syndrome: a further
1016 contributor to the wide clinical expressivity? *Clin Genet* **78**, 560-564 (2010).

1017 28. S. Kawauchi *et al.*, Multiple organ system defects and transcriptional dysregulation in the
1018 Nipbl(+/-) mouse, a model of Cornelia de Lange Syndrome. *PLoS Genet* **5**, e1000650
1019 (2009).

1020 29. M. Kaur *et al.*, Precocious sister chromatid separation (PSCS) in Cornelia de Lange
1021 syndrome. *Am J Med Genet A* **138**, 27-31 (2005).

1022 30. P. Castronovo *et al.*, Premature chromatid separation is not a useful diagnostic marker for
1023 Cornelia de Lange syndrome. *Chromosome Res* **17**, 763-771 (2009).

1024 31. M. G. Vrouwe *et al.*, Increased DNA damage sensitivity of Cornelia de Lange syndrome
1025 cells: evidence for impaired recombinational repair. *Hum Mol Genet* **16**, 1478-1487
1026 (2007).

1027 32. A. Muto, A. L. Calof, A. D. Lander, T. F. Schilling, Multifactorial origins of heart and gut
1028 defects in nipbl-deficient zebrafish, a model of Cornelia de Lange Syndrome. *PLoS Biol* **9**,
1029 e1001181 (2011).

1030 33. Y. Wu *et al.*, Drosophila Nipped-B Mutants Model Cornelia de Lange Syndrome in
1031 Growth and Behavior. *PLoS Genet* **11**, e1005655 (2015).

1032 34. R. Chien *et al.*, Cohesin mediates chromatin interactions that regulate mammalian beta-
1033 globin expression. *J Biol Chem* **286**, 17870-17878 (2011).

1034 35. M. H. Kagey *et al.*, Mediator and cohesin connect gene expression and chromatin
1035 architecture. *Nature* **467**, 430-435 (2010).

1036 36. D. Dorsett, Cohesin: genomic insights into controlling gene transcription and
1037 development. *Curr Opin Genet Dev* **21**, 199-206 (2011).

1038 37. Y. Guo *et al.*, CTCF/cohesin-mediated DNA looping is required for protocadherin alpha
1039 promoter choice. *Proc Natl Acad Sci U S A* **109**, 21081-21086 (2012).

1040 38. S. Remeseiro, A. Cuadrado, A. Losada, Cohesin in development and disease.
1041 *Development* **140**, 3715-3718 (2013).

1042 39. B. Yuan *et al.*, Global transcriptional disturbances underlie Cornelia de Lange syndrome
1043 and related phenotypes. *J Clin Invest* **125**, 636-651 (2015).

1044 40. R. Santos *et al.*, Conditional Creation and Rescue of Nipbl-Deficiency in Mice Reveals
1045 Multiple Determinants of Risk for Congenital Heart Defects. *PLoS Biol* **14**, e2000197
1046 (2016).

1047 41. M. A. Mole, A. Weberling, M. Zernicka-Goetz, Comparative analysis of human and
1048 mouse development: From zygote to pre-gastrulation. *Curr Top Dev Biol* **136**, 113-138
1049 (2020).

1050 42. E. S. Bardot, A. K. Hadjantonakis, Mouse gastrulation: Coordination of tissue patterning,
1051 specification and diversification of cell fate. *Mech Dev* **163**, 103617 (2020).

1052 43. K. M. Downs, T. Davies, Staging of gastrulating mouse embryos by morphological
1053 landmarks in the dissecting microscope. *Development* **118**, 1255-1266 (1993).

1054 44. G. X. Zheng *et al.*, Massively parallel digital transcriptional profiling of single cells. *Nat*
1055 *Commun* **8**, 14049 (2017).

1056 45. M. Ginsburg, M. H. Snow, A. McLaren, Primordial germ cells in the mouse embryo
1057 during gastrulation. *Development* **110**, 521-528 (1990).

1058 46. T. Stuart *et al.*, Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902
1059 e1821 (2019).

1060 47. B. Pijuan-Sala *et al.*, A single-cell molecular map of mouse gastrulation and early
1061 organogenesis. *Nature* **566**, 490-495 (2019).

1062 48. J. A. Farrell *et al.*, Single-cell reconstruction of developmental trajectories during
1063 zebrafish embryogenesis. *Science* **360**, (2018).

1064 49. V. Bergen, M. Lange, S. Peidli, F. A. Wolf, F. J. Theis, Generalizing RNA velocity to
1065 transient cell states through dynamical modeling. *Nat Biotechnol* **38**, 1408-1414 (2020).

1066 50. Y. Saga *et al.*, MesP1 is expressed in the heart precursor cells and required for the
1067 formation of a single heart tube. *Development* **126**, 3437-3447 (1999).

1068 51. M. Lange *et al.*, CellRank for directed single-cell fate mapping. *Nat Methods* **19**, 159-170
1069 (2022).

1070 52. A. Liberzon *et al.*, The Molecular Signatures Database (MSigDB) hallmark gene set
1071 collection. *Cell Syst* **1**, 417-425 (2015).

1072 53. S. M. Langemeijer *et al.*, Apoptosis-Related Gene Expression Profiling in Hematopoietic
1073 Cell Fractions of MDS Patients. *PLoS One* **11**, e0165582 (2016).

1074 54. D. Venet, J. E. Dumont, V. Detours, Most random gene expression signatures are
1075 significantly associated with breast cancer outcome. *PLoS Comput Biol* **7**, e1002240
1076 (2011).

1077 55. S. Nestorowa *et al.*, A single-cell resolution map of mouse hematopoietic stem and
1078 progenitor cell differentiation. *Blood* **128**, e20-31 (2016).

1079 56. R. M. Barnes, B. A. Firulli, S. J. Conway, J. W. Vincentz, A. B. Firulli, Analysis of the
1080 Hand1 cell lineage reveals novel contributions to cardiovascular, neural crest, extra-
1081 embryonic, and lateral mesoderm derivatives. *Dev Dyn* **239**, 3086-3097 (2010).

1082 57. Y. Chen *et al.*, The Role of Tbx20 in Cardiovascular Development and Function. *Front*
1083 *Cell Dev Biol* **9**, 638542 (2021).

1084 58. D. W. Sorensen, J. H. van Berlo, The Role of TGF-beta Signaling in Cardiomyocyte
1085 Proliferation. *Curr Heart Fail Rep* **17**, 225-233 (2020).

1086 59. S. Saadat *et al.*, Pivotal Role of TGF-beta/Smad Signaling in Cardiac Fibrosis: Non-
1087 coding RNAs as Effectual Players. *Front Cardiovasc Med* **7**, 588347 (2020).

1088 60. C. Guibentif *et al.*, Diverse Routes toward Early Somites in the Mouse Embryo. *Dev Cell*
1089 **56**, 141-153 e146 (2021).

1090 61. B. Wilm, R. G. James, T. M. Schultheiss, B. L. Hogan, The forkhead genes, Foxc1 and
1091 Foxc2, regulate paraxial versus intermediate mesoderm cell fate. *Dev Biol* **271**, 176-189
1092 (2004).

1093 62. D. T. Farmer, P. Patel, R. Choi, C. Y. Liu, J. G. Crump, A comprehensive series of Irx
1094 cluster mutants reveals diverse roles in facial cartilage development. *Development* **148**,
1095 (2021).

1096 63. H. Preiss *et al.*, Regulation of Nodal signaling propagation by receptor interactions and
1097 positive feedback. *Elife* **11**, (2022).

1098 64. S. Nowotschin *et al.*, The T-box transcription factor Eomesodermin is essential for AVE
1099 induction in the mouse embryo. *Genes Dev* **27**, 997-1002 (2013).

1100 65. C. Meno *et al.*, Diffusion of nodal signaling activity in the absence of the feedback
1101 inhibitor Lefty2. *Dev Cell* **1**, 127-138 (2001).

1102 66. I. Costello *et al.*, Lhx1 functions together with Otx2, Foxa2, and Ldb1 to govern anterior
1103 mesendoderm, node, and midline development. *Genes Dev* **29**, 2108-2122 (2015).

1104 67. K. Briegel *et al.*, Regulation and function of transcription factor GATA-1 during red blood
1105 cell differentiation. *Development* **122**, 3839-3850 (1996).

1106 68. M. Siatecka, J. J. Bieker, The multifunctional role of EKLF/KLF1 during erythropoiesis.
1107 *Blood* **118**, 2044-2054 (2011).

1108 69. W. K. Shen *et al.*, AnimalTFDB 4.0: a comprehensive animal transcription factor database
1109 updated with variation and expression annotations. *Nucleic Acids Res* **51**, D39-D45
1110 (2023).

1111 70. S. J. Bray, Notch signalling: a simple pathway becomes complex. *Nat Rev Mol Cell Biol*
1112 **7**, 678-689 (2006).

1113 71. R. Macdonald *et al.*, Midline signalling is required for Pax gene regulation and patterning
1114 of the eyes. *Development* **121**, 3267-3278 (1995).

1115 72. J. Park *et al.*, Regulation of Sox9 by Sonic Hedgehog (Shh) is essential for patterning and
1116 formation of tracheal cartilage. *Dev Dyn* **239**, 514-526 (2010).

1117 73. P. H. Su *et al.*, NKX6-1 mediates cancer stem-like properties and regulates sonic
1118 hedgehog signaling in leiomyosarcoma. *J Biomed Sci* **28**, 32 (2021).

1119 74. C. Cai, J. Thorne, L. Grabel, Hedgehog serves as a mitogen and survival factor during
1120 embryonic stem cell neurogenesis. *Stem Cells* **26**, 1097-1108 (2008).

1121 75. S. W. Choy *et al.*, A cascade of irx1a and irx2a controls shh expression during
1122 retinogenesis. *Dev Dyn* **239**, 3204-3214 (2010).

1123 76. D. Szklarczyk *et al.*, The STRING database in 2023: protein-protein association networks
1124 and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids*
1125 *Res* **51**, D638-D646 (2023).

1126 77. K. Mitsui *et al.*, The homeoprotein Nanog is required for maintenance of pluripotency in
1127 mouse epiblast and ES cells. *Cell* **113**, 631-642 (2003).

1128 78. A. H. Hart, L. Hartley, M. Ibrahim, L. Robb, Identification, cloning and expression
1129 analysis of the pluripotency promoting Nanog genes in mouse and human. *Dev Dyn* **230**,
1130 187-198 (2004).

1131 79. M. Zhang *et al.*, Esrrb Complementation Rescues Development of Nanog-Null Germ
1132 Cells. *Cell Rep* **22**, 332-339 (2018).

1133 80. G. Shi, Y. Jin, Role of Oct4 in maintaining and regaining stem cell pluripotency. *Stem*
1134 *Cell Res Ther* **1**, 39 (2010).

1135 81. M. Tiana *et al.*, Pluripotency factors regulate the onset of Hox cluster activation in the
1136 early embryo. *Sci Adv* **8**, eabo3583 (2022).

1137 82. M. Mossahebi-Mohammadi, M. Quan, J. S. Zhang, X. Li, FGF Signaling Pathway: A Key
1138 Regulator of Stem Cell Pluripotency. *Front Cell Dev Biol* **8**, 79 (2020).

1139 83. L. L. Chang, D. S. Kessler, Foxd3 is an essential Nodal-dependent regulator of zebrafish
1140 dorsal mesoderm development. *Dev Biol* **342**, 39-50 (2010).

1141 84. M. M. Shen, Nodal signaling: developmental roles and regulation. *Development* **134**,
1142 1023-1034 (2007).

1143 85. C. Porcher, H. Chagraoui, M. S. Kristiansen, SCL/TAL1: a multifaceted regulator from
1144 blood development to disease. *Blood* **129**, 2051-2060 (2017).

1145 86. V. S. Stanulovic, P. Cauchy, S. A. Assi, M. Hoogenkamp, LMO2 is required for TAL1
1146 DNA binding activity and initiation of definitive haematopoiesis at the haemangioblast
1147 stage. *Nucleic Acids Res* **45**, 9874-9888 (2017).

1148 87. R. Guo, Q. S. Xing, Roles of Wnt Signaling Pathway and ROR2 Receptor in Embryonic
1149 Development: An Update Review Article. *Epigenet Insights* **15**, 25168657211064232
1150 (2022).

1151 88. H. Berger, A. Wodarz, A. Borchers, PTK7 Faces the Wnt in Development and Disease.
1152 *Front Cell Dev Biol* **5**, 31 (2017).

1153 89. T. S. Gujral *et al.*, A noncanonical Frizzled2 pathway regulates epithelial-mesenchymal
1154 transition and metastasis. *Cell* **159**, 844-856 (2014).

1155 90. J. Baum, R. H. Ward, D. J. Conway, Natural selection on the erythrocyte surface. *Mol Biol*
1156 *Evol* **19**, 223-229 (2002).

1157 91. S. Chonat *et al.*, The Spectrum of SPTA1-Associated Hereditary Spherocytosis. *Front*
1158 *Physiol* **10**, 815 (2019).

1159 92. T. L. Huang *et al.*, A de novo ANK1 mutation associated to hereditary spherocytosis: a
1160 case report. *BMC Pediatr* **19**, 62 (2019).

1161 93. Y. Liu *et al.*, A novel SPTB gene mutation in neonatal hereditary spherocytosis: A case
1162 report. *Exp Ther Med* **20**, 3253-3259 (2020).

1163 94. K. A. Hubert, D. M. Wellik, Hox genes in development and beyond. *Development* **150**,
1164 (2023).

1165 95. O. Medina-Martinez, A. Bradley, R. Ramirez-Solis, A large targeted deletion of Hoxb1-
1166 Hoxb9 produces a series of single-segment anterior homeotic transformations. *Dev Biol*
1167 **222**, 71-83 (2000).

1168 96. M. I. Boyle, C. Jespersgaard, K. Brondum-Nielsen, A. M. Bisgaard, Z. Tumer, Cornelia de
1169 Lange syndrome. *Clin Genet* **88**, 1-12 (2015).

1170 97. S. Kawauchi *et al.*, Using mouse and zebrafish models to understand the etiology of
1171 developmental defects in Cornelia de Lange Syndrome. *Am J Med Genet C Semin Med*
1172 *Genet* **172**, 138-145 (2016).

1173 98. M. P. Lambert *et al.*, The incidence of thrombocytopenia in children with Cornelia de
1174 Lange syndrome. *Am J Med Genet A* **155A**, 33-37 (2011).

1175 99. A. Ayerza Casas *et al.*, Cornelia de Lange syndrome: Congenital heart disease in 149
1176 patients. *Med Clin (Barc)* **149**, 300-302 (2017).

1177 100. K. Zhao, Z. Yang, The second heart field: the first 20 years. *Mamm Genome* **34**, 216-228
1178 (2023).

1179 101. M. E. Lopez-Burks, R. Santos, S. Kawauchi, A. L. Calof, A. D. Lander, Genetic
1180 enhancement of limb defects in a mouse model of Cornelia de Lange syndrome. *Am J Med*
1181 *Genet C Semin Med Genet* **172**, 146-154 (2016).

1182 102. D. Mehta *et al.*, Characterization of limb differences in children with Cornelia de Lange
1183 Syndrome. *Am J Med Genet C Semin Med Genet* **172**, 155-162 (2016).

1184 103. A. D. Kline *et al.*, Diagnosis and management of Cornelia de Lange syndrome: first
1185 international consensus statement. *Nat Rev Genet* **19**, 649-666 (2018).

1186 104. A. Muto *et al.*, Nipbl and mediator cooperatively regulate gene expression to control limb
1187 development. *PLoS Genet* **10**, e1004671 (2014).

1188 105. F. A. Lee, F. M. Kenny, Skeletal changes in the Cornelia de Lange syndrome. *Am J*
1189 *Roentgenol Radium Ther Nucl Med* **100**, 27-39 (1967).

1190 106. J. Brennan, D. P. Norris, E. J. Robertson, Nodal activity in the node governs left-right
1191 asymmetry. *Genes Dev* **16**, 2339-2344 (2002).

1192 107. B. Mohapatra *et al.*, Identification and functional characterization of NODAL rare variants
1193 in heterotaxy and isolated cardiovascular malformations. *Hum Mol Genet* **18**, 861-871
1194 (2009).

1195 108. M. Mariani *et al.*, Adolescents and adults affected by Cornelia de Lange syndrome: A
1196 report of 73 Italian patients. *Am J Med Genet C Semin Med Genet* **172**, 206-213 (2016).

1197 109. A. D. Kline *et al.*, Natural history of aging in Cornelia de Lange syndrome. *Am J Med*
1198 *Genet C Semin Med Genet* **145C**, 248-260 (2007).

1199 110. K. C. Chatfield *et al.*, Congenital heart disease in Cornelia de Lange syndrome: phenotype
1200 and genotype analysis. *Am J Med Genet A* **158A**, 2499-2505 (2012).

1201 111. A. N. Economides *et al.*, Conditionals by inversion provide a universal method for the
1202 generation of conditional alleles. *Proc Natl Acad Sci U S A* **110**, E3179-3188 (2013).

1203 112. S. Van Der Lee, L. M. Boot, Spontaneous pseudopregnancy in mice. *Acta Physiol*
1204 *Pharmacol Neerl* **4**, 442-444 (1955).

1205 113. W. K. Whitten, Modification of the oestrous cycle of the mouse by external stimuli
1206 associated with the male. *J Endocrinol* **13**, 399-404 (1956).

1207 114. K. Shea, N. Geijsen, Dissection of 6.5 dpc mouse embryos. *J Vis Exp*, 160 (2007).

1208 115. S. Das *et al.*, The Drosophila T-box transcription factor Midline functions within the
1209 Notch-Delta signaling pathway to specify sensory organ precursor cell fates and regulates
1210 cell survival within the eye imaginal disc. *Mech Dev* **130**, 577-601 (2013).

1211 116. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21
1212 (2013).

1213 117. A. T. L. Lun *et al.*, EmptyDrops: distinguishing cells from empty droplets in droplet-based
1214 single-cell RNA sequencing data. *Genome Biol* **20**, 63 (2019).

1215 118. T. Ilicic *et al.*, Classification of low quality cells from single-cell RNA-seq data. *Genome*
1216 *Biol* **17**, 29 (2016).

1217 119. C. Hafemeister, R. Satija, Normalization and variance stabilization of single-cell RNA-seq
1218 data using regularized negative binomial regression. *Genome Biol* **20**, 296 (2019).

1219 120. G. La Manno *et al.*, RNA velocity of single cells. *Nature* **560**, 494-498 (2018).

1220 121. T. Wu *et al.*, clusterProfiler 4.0: A universal enrichment tool for interpreting omics data.
1221 *Innovation (Camb)* **2**, 100141 (2021).

122. G. Korotkevich *et al.*, Fast gene set enrichment analysis. *bioRxiv*, 060012 (2021).

123. I. F. Davidson *et al.*, CTCF is a DNA-tension-dependent barrier to cohesin-mediated loop extrusion. *Nature* **616**, 822-827 (2023).

Acknowledgments

The authors are grateful to Henry Chang, Lorelle Meidt, Han Nguyen, Arianna Favela, Christopher Wu, Louise Villagomez, and Brian Bui for assistance with mouse husbandry, genotyping, and data analysis; to Melanie Oakes of the UCI Genomics and Research Technology Hub (GRTH) for 10x library preparation and sequencing; and to Qing Nie (UCI Depts of Mathematics and Developmental & Cell Biology) and Rosaysela Santos (Kaiser Permanente Bernard J. Tyson School of Medicine) for helpful discussions. SC acknowledges support from a training grant award (NINDS/NIH T32NS082174), and JK acknowledges support by a Momental Foundation Mistletoe Research Fellowship. ADL and ALC dedicate this paper to the memory of Isabel Eva Calof Lander.

Funding:

National Institutes of Health R01HL138659 (ALC, ADL)

National Institutes of Health R01DE019638 (ADL)

National Institutes of Health R35GM143019 (ALM)

NSF-Simons Center for Multiscale Cell Fate Research (SC, JK, ADL, ALC, ALM)

Author contributions:

Conceptualization: ALC, ADL, SC, ALM, JK

Methodology: ADL, SC, ALC, ALM

Software: SC, ADL, JK, ALM

Validation: ALC, ADL, SC, ALM

Formal Analysis: SC, ADL, ALC, JK, ALM

Investigation: SC, JK, MEL

Resources: ADL, ALC

Data Curation: SC

Writing – Original Draft: SC, ALC, ADL

Writing – Review & Editing: ADL, SC, ALC, ALM, JK

Visualization: SC, ADL, ALC

Supervision: ALC, ADL, ALM

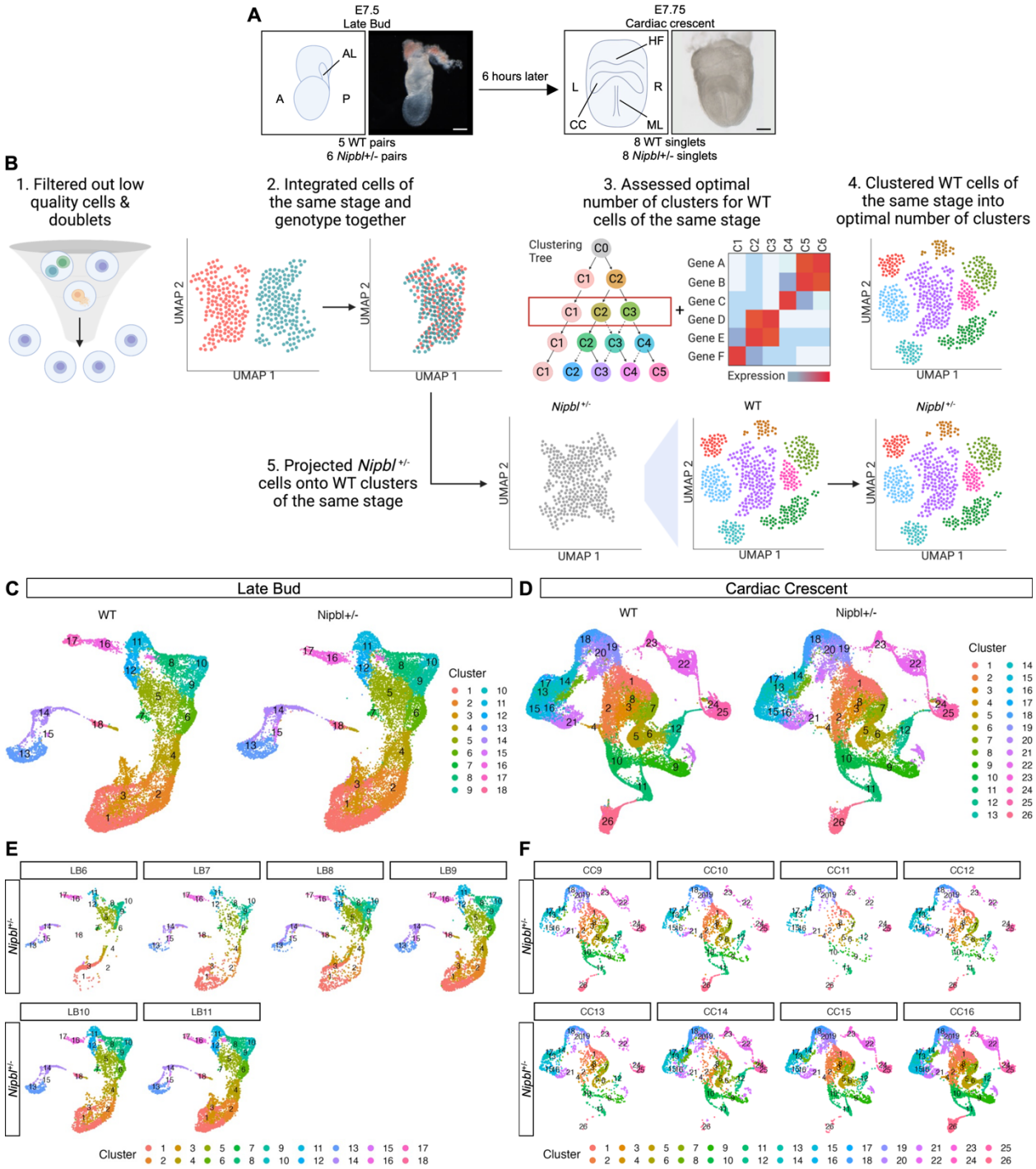
Project Administration: ALC, ADL

Funding Acquisition: ALC, ADL, SC, ALM, JK

Competing interests: All other authors declare they have no competing interests.

Data and materials availability: All data are available in the main text, supplementary materials, Gene Expression Omnibus (Accession: GSE151589), and Sequence Read Archive (Accession: SRP265522). All code is available in Zenodo (DOI: 10.5281/zenodo.10001570).

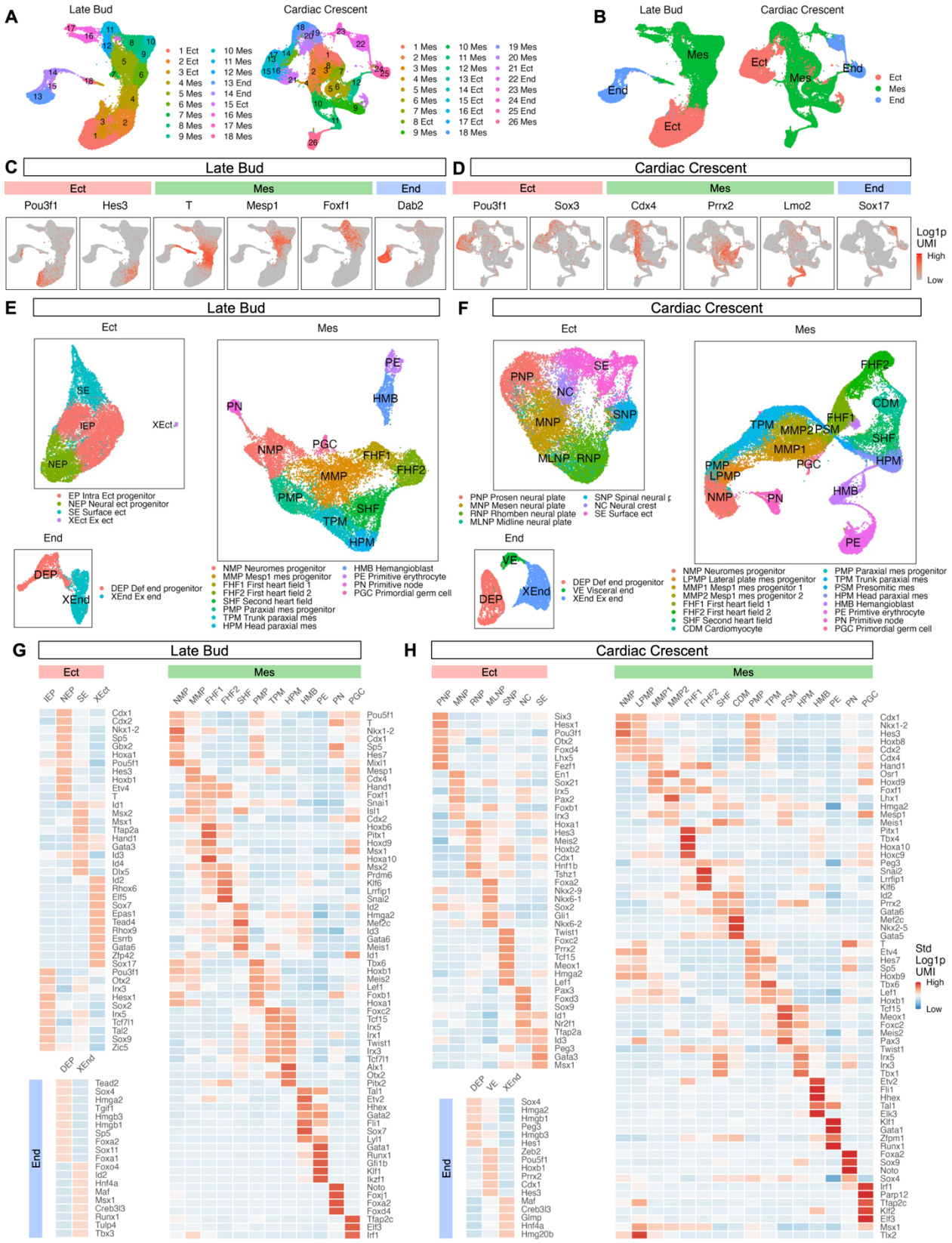
Figures and Tables


A
E7.5
Late Bud
AL
A
P
5 WT pairs
6 Nipbl+/- pairs
6 hours later
E7.75
Cardiac crescent
HF
L
R
CC
ML
8 WT singlets
8 Nipbl+/- singlets
B
1. Filtered out low quality cells & doublets
2. Integrated cells of the same stage and genotype together
3. Assessed optimal number of clusters for WT cells of the same stage
4. Clustered WT cells of the same stage into optimal number of clusters
5. Projected Nipbl +/- cells onto WT clusters of the same stage
UMAP 1
UMAP 2
Clustering Tree
Gene A
Gene B
Gene C
Gene D
Gene E
Gene F
Expression
Nipbl +/-
WT
C
Late Bud
D
Cardiac Crescent
Nipbl+/-
Cluster
E
LB6
LB7
LB8
LB9
LB10
LB11
F
CC9
CC10
CC11
CC12
CC13
CC14
CC15
CC16

1260

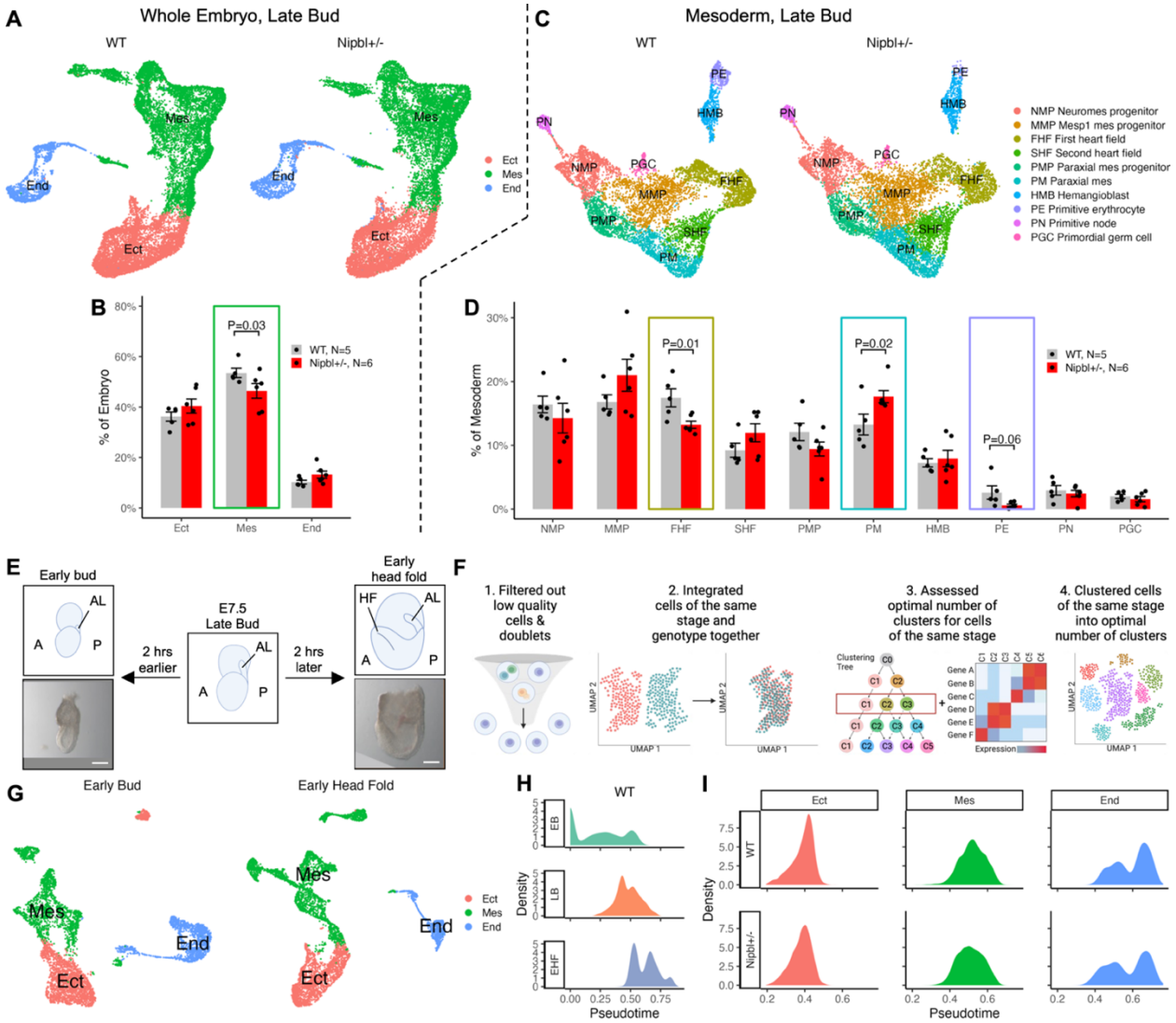
1261

1262 **Fig. 1. *Nipbl*$^{+/-}$ mice do not lack any cell populations found in WT mice.** (**A**) Lateral view of
1263 LB-stage and anterior view of CC-stage embryos subjected to scRNAseq. A = anterior, P =
1264 posterior, AL = allantois, L = left, R = right, HF = head fold, CC = cardiac crescent, ML =
1265 midline. Dashed line represents where embryonic tissue was separated from extraembryonic
1266 tissue. Scale bar = 100 um. (**B**) Workflow used to filter out low quality cells and doublets, cluster
1267 WT cells into optimal number of clusters at each stage, and project *Nipbl*$^{+/-}$ cells onto WT clusters
1268 of the same stage. Uniform manifold approximation and projection (UMAP) of clusters in WT
1269 and *Nipbl*$^{+/-}$ embryos at (**C**) LB- and (**D**) CC-stage. UMAP of clusters in each *Nipbl*$^{+/-}$ embryo (**E**)
1270 LB- and (**F**) CC-stage.


A
Late Bud
Cardiac Crescent
1 Ect
2 Ect
3 Ect
4 Mes
5 Mes
6 Mes
7 Mes
8 Mes
9 Mes
10 Mes
11 Mes
12 Mes
13 End
14 End
15 Ect
16 Mes
17 Mes
18 Mes
19 Mes
20 Mes
21 Ect
22 End
23 Mes
24 End
25 End
26 Mes
B
Mes
End
Ect
Ect
Mes
End
C
Late Bud
Ect
Mes
End
Pou3f1
Hes3
T
Mesp1
Foxf1
Dab2
D
Cardiac Crescent
Pou3f1
Sox3
Cdx4
Prrx2
Lmo2
Sox17
Log1p UMI
High
Low
E
Late Bud
Ect
SE
IEP
XEct
NEP
EP Intra Ect progenitor
NEP Neural ect progenitor
SE Surface ect
XEct Ex ect
Mes
PE
HMB
PN
NMP
PGC
FHF1
FHF2
MMP
PMP
SHF
TPM
HPM
NMP Neuromes progenitor
MMP Mesp1 mes progenitor
FHF1 First heart field 1
FHF2 First heart field 2
SHF Second heart field
PMP Paraxial mes progenitor
TPM Trunk paraxial mes
HPM Head paraxial mes
HMB Hemangioblast
PE Primitive erythrocyte
PN Primitive node
PGC Primordial germ cell
End
DEP
XEnd
DEP Def end progenitor
XEnd Ex end
F
Cardiac Crescent
Ect
PNP
SE
NC
MNP
SNP
MLNP
RNP
PNP Prosen neural plate
MNP Mesen neural plate
RNP Rhomben neural plate
MLNP Midline neural plate
SNP Spinal neural p
NC Neural crest
SE Surface ect
Mes
FHF2
CDM
TPM
FHF1
MMP2
PSM
SHF
MMP1
PGC
HPM
PMP
LPMP
NMP
PN
HMB
PE
NMP Neuromes progenitor
LPMP Lateral plate mes progenitor
MMP1 Mesp1 mes progenitor 1
MMP2 Mesp1 mes progenitor 2
FHF1 First heart field 1
FHF2 First heart field 2
SHF Second heart field
CDM Cardiomyocyte
PMP Paraxial mes progenitor
TPM Trunk paraxial mes
PSM Presomitic mes
HPM Head paraxial mes
HMB Hemangioblast
PE Primitive erythrocyte
PN Primitive node
PGC Primordial germ cell
End
VE
XEnd
DEP
DEP Def end progenitor
VE Visceral end
XEnd Ex end
G
Late Bud
Ect
Mes
IEP
NEP
SE
XEct
NMP
MMP
FHF1
FHF2
SHF
PMP
TPM
HPM
HMB
PE
PN
PGC
Cdx1
Cdx2
Nkx1-2
Sp5
Gbx2
Hoxa1
Pou5f1
Hes3
Hoxb1
Etv4
T
Id1
Msx2
Msx1
Tfap2a
Hand1
Gata3
Id3
Id4
Dlx5
Id2
Rhox6
Elf5
Sox7
Epas1
Tead4
Rhox9
Esrrb
Gata6
Zfp42
Sox17
Pou3f1
Otx2
Irx3
Hesx1
Sox2
Irx5
Tcf7l1
Tal2
Sox9
Zic5
Pou5f1
T
Nkx1-2
Cdx1
Sp5
Hes7
Mixl1
Mesp1
Cdx4
Hand1
Foxf1
Snai1
Isl1
Cdx2
Hoxb6
Pitx1
Hoxd9
Msx1
Hoxa10
Msx2
Prdm6
Klf6
Lrrfip1
Snai2
Id2
Hmga2
Mef2c
Id3
Gata6
Meis1
Id1
Tbx6
Hoxb1
Meis2
Lef1
Foxb1
Hoxa1
Foxc2
Tcf15
Irx5
Irx1
Twist1
Irx3
Tcf7l1
Alx1
Otx2
Pitx2
Tal1
Etv2
Hhex
Gata2
Fli1
Sox7
Lyl1
Gata1
Runx1
Gfi1b
Klf1
Ikzf1
Noto
Foxj1
Foxa2
Foxd4
Tfap2c
Elf3
Irf1
DEP
XEnd
End
Tead2
Sox4
Hmga2
Tgif1
Hmgb3
Hmgb1
Sp5
Foxa2
Sox11
Foxa1
Foxo4
Id2
Hnf4a
Maf
Msx1
Creb3l3
Runx1
Tulp4
Tbx3
H
Cardiac Crescent
Ect
PNP
MNP
RNP
MLNP
SNP
NC
SE
Mes
NMP
LPMP
MMP1
MMP2
FHF1
FHF2
SHF
CDM
PMP
TPM
PSM
HPM
HMB
PE
PN
PGC
Six3
Hesx1
Pou3f1
Otx2
Foxd4
Lhx5
Fezf1
En1
Sox21
Irx5
Pax2
Foxb1
Irx3
Hoxa1
Hes3
Meis2
Hoxb2
Cdx1
Hnf1b
Tshz1
Foxa2
Nkx2-9
Nkx6-1
Sox2
Gli1
Nkx6-2
Twist1
Foxc2
Prrx2
Tcf15
Meox1
Hmga2
Lef1
Pax3
Foxd3
Sox9
Id1
Nr2f1
Tfap2a
Id3
Peg3
Gata3
Msx1
Cdx1
Nkx1-2
Hes3
Hoxb8
Cdx2
Cdx4
Hand1
Osr1
Hoxd9
Foxf1
Lhx1
Hmga2
Mesp1
Meis1
Pitx1
Tbx4
Hoxa10
Hoxc9
Peg3
Snai2
Lrrfip1
Klf6
Id2
Prrx2
Gata6
Mef2c
Nkx2-5
Gata5
T
Etv4
Hes7
Sp5
Hoxb9
Tbx6
Lef1
Hoxb1
Tcf15
Meox1
Foxc2
Meis2
Pax3
Twist1
Irx5
Irx3
Tbx1
Etv2
Fli1
Hhex
Tal1
Elk3
Klf1
Gata1
Zfpm1
Runx1
Foxa2
Sox9
Noto
Sox4
Irf1
Parp12
Tfap2c
Klf2
Elf3
Msx1
Tlx2
Std Log1p UMI
High
Low
DEP
VE
XEnd
End
Sox4
Hmga2
Hmgb1
Peg3
Hmgb3
Hes1
Zeb2
Pou5f1
Hoxb1
Prrx2
Cdx1
Hes3
Maf
Creb3l3
Glmp
Hnf4a
Hmg20b

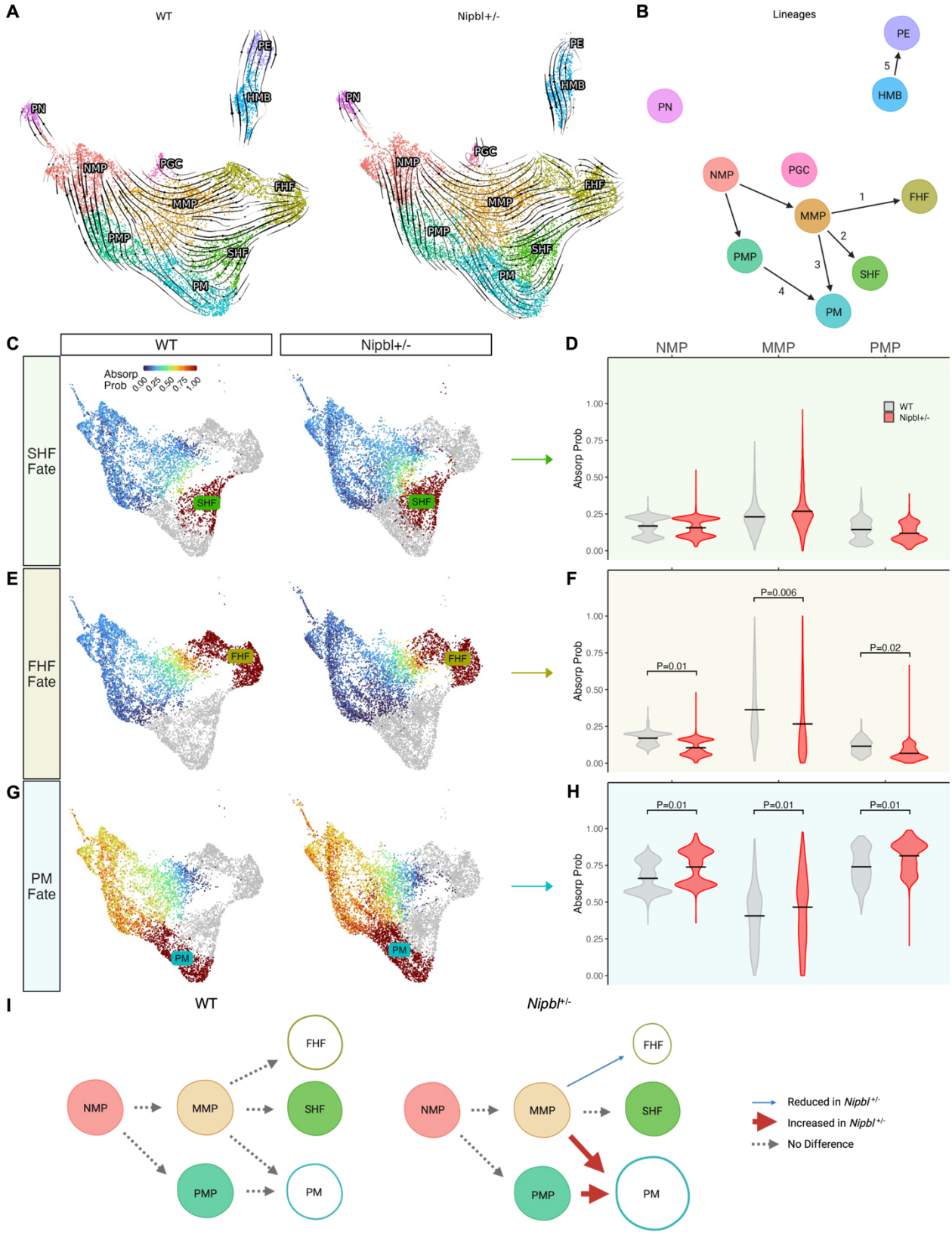
1271

1272 **Fig. 2. scRNAseq captured cells from all germ layers, as well as progenitors of major tissues.**
1273 (**A**) UMAP of clusters assigned to germ layers in LB- and CC-stage embryos. (**B**) UMAP of germ
1274 layers in LB- and CC-stage embryos. Expression of genes marking germ layers of (**C**) LB- and
1275 (**D**) CC-stage embryos in UMAP. UMAP of cell populations in germ layers of (**E**) LB- and (**F**)
1276 CC-stage embryos. Heatmap of fold change in expression of the most differentially expressed
1277 transcription factor genes (lowest *Q*-values from Mann-Whitney U Test) (*69*) of cell populations
1278 from all other cell populations in germ layers of (**G**) LB- and (**H**) CC-stage embryos.


A
Whole Embryo, Late Bud
WT
Nipbl+/-
Mes
End
Ect
Ect
Mes
End
B
% of Embryo
P=0.03
WT, N=5
Nipbl+/-, N=6
Ect
Mes
End
C
Mesoderm, Late Bud
WT
Nipbl+/-
PE
HMB
PN
NMP
PGC
MMP
FHF
PMP
SHF
PM
NMP Neuromes progenitor
MMP Mesp1 mes progenitor
FHF First heart field
SHF Second heart field
PMP Paraxial mes progenitor
PM Paraxial mes
HMB Hemangioblast
PE Primitive erythrocyte
PN Primitive node
PGC Primordial germ cell
D
% of Mesoderm
P=0.01
P=0.02
P=0.06
WT, N=5
Nipbl+/-, N=6
NMP
MMP
FHF
SHF
PMP
PM
HMB
PE
PN
PGC
E
Early bud
AL
A
P
2 hrs earlier
E7.5 Late Bud
2 hrs later
Early head fold
HF
F
1. Filtered out low quality cells & doublets
2. Integrated cells of the same stage and genotype together
3. Assessed optimal number of clusters for cells of the same stage
4. Clustered cells of the same stage into optimal number of clusters
UMAP 1
UMAP 2
Clustering Tree
Gene A
Gene B
Gene C
Gene D
Gene E
Gene F
Expression
G
Early Bud
Early Head Fold
Mes
End
Ect
H
WT
EB
LB
EHF
Density
Pseudotime
I
Ect
Mes
End
WT
Nipbl+/-
Density
Pseudotime

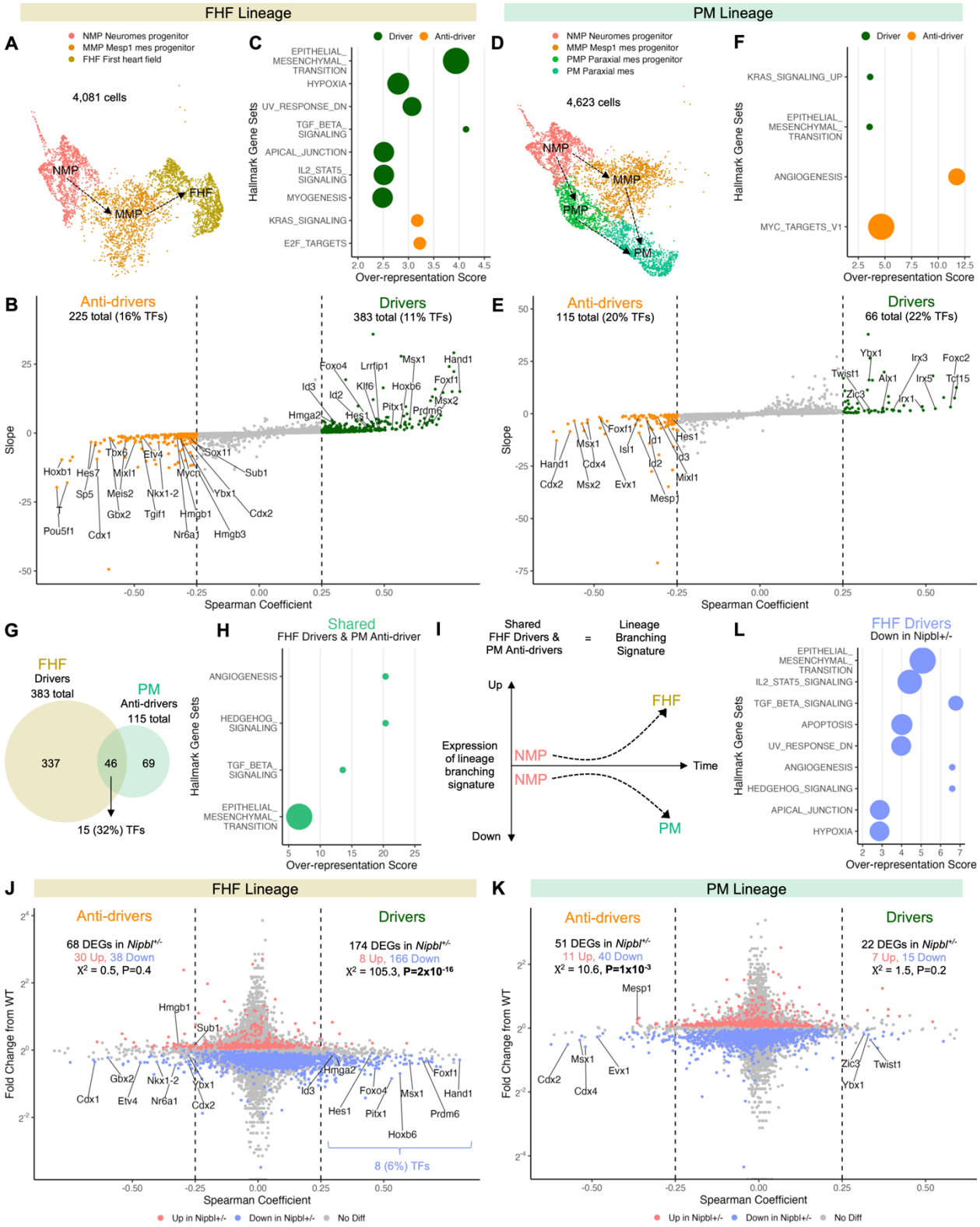
1279

1280 **Fig. 3. *Nipbl*$^{+/-}$ embryos exhibit changes in the sizes of mesodermal subpopulations that are**
1281 **not accompanied by changes in developmental timing.** (**A**) UMAP of germ layers in LB-stage
1282 WT and *Nipbl*$^{+/-}$ embryos. (**B**) Percentage of cells in germ layers from all cells in LB-stage WT
1283 and *Nipbl*$^{+/-}$ embryos. Error bars show standard error of the mean (*123*). *P*-values from T-Test.
1284 (**C**) UMAP of cell populations in mesoderm of LB-stage WT and *Nipbl*$^{+/-}$ embryos. (**D**)
1285 Percentage of cells in mesodermal cell populations from all cells in LB-stage WT and *Nipbl*$^{+/-}$
1286 embryos (bottom). Error bars show SEM. *P*-values from T-Test. (**E**) Lateral view of WT EB- and
1287 EHF-stage embryos subjected to scRNAseq. A = anterior, P = posterior, AL = allantois, L = left,
1288 R = right, HF = head fold. Dashed line represents where embryonic tissue was separated from
1289 extraembryonic tissue. Scale bar = 100 um. (**F**) Workflow used to filter out low quality cells and
1290 doublets and cluster cells at EB- and EHF-stage into optimal number of clusters. (**G**) UMAP of
1291 clusters assigned to germ layers in EB- and EHF-stage embryos. (**H**) Density of cells from WT
1292 EB-, LB-, and EHF-stage embryos along pseudotime (calculated using URD). (**I**) Density of cells
1293 in germ layers of LB-stage WT and *Nipbl*$^{+/-}$ embryos along pseudotime.


A
WT
Nipbl+/-
PE
HMB
PN
NMP
PGC
MMP
FHF
PMP
SHF
PM
B
Lineages
5
1
2
3
4
C
D
E
F
G
H
SHF Fate
FHF Fate
PM Fate
Absorp Prob
0.00
0.25
0.50
0.75
1.00
NMP
MMP
PMP
WT
Nipbl+/-
P=0.006
P=0.02
P=0.01
I
Nipbl+/-
Reduced in Nipbl+/-
Increased in Nipbl+/-
No Difference

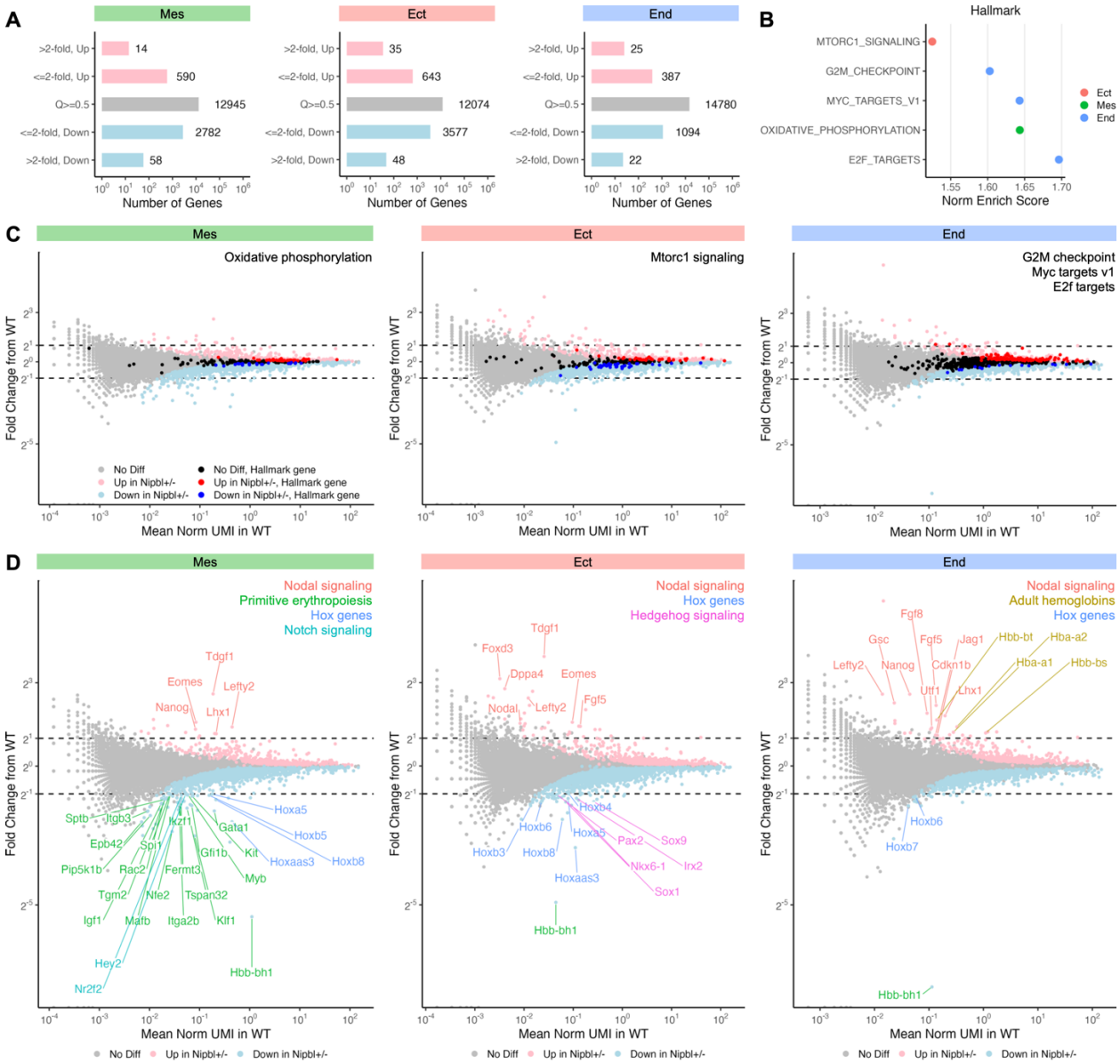
1294

1295 **Fig. 4. *Nipbl*$^{+/-}$ mice misdirect mesoderm cells into paraxial mesoderm at the expense of the**
1296 **first heart field.** (**A**) RNA velocities, calculated by scVelo of mesoderm cells from LB-stage WT
1297 and *Nipbl*$^{+/-}$ embryos in UMAP. (**B**) Cell lineage trajectories in mesoderm of LB-stage WT and
1298 *Nipbl*$^{+/-}$ embryos. Probability of mesoderm cells from LB-stage WT and *Nipbl*$^{+/-}$ embryos
1299 terminally transitioning (absorption probabilities calculated using CellRank) into (C) SHF, (E)
1300 FHF, or (G) PM fates in UMAP. Violin plot of probability of NMPs, MMPs, and PMPs from LB-
1301 stage WT and *Nipbl*$^{+/-}$ embryos terminally transitioning (absorption probabilities) into (D) SHF,
1302 (F) FHF, or (H) PM fates. Lines show means. *P*-values from T-Test. (**I**) Schematic illustrating
1303 how mesoderm cells in *Nipbl*$^{+/-}$ embryos are misdirected into PM fate at the expense of FHF fate.


FHF Lineage
PM Lineage
A
NMP Neuromes progenitor
MMP Mesp1 mes progenitor
FHF First heart field
4,081 cells
NMP
MMP
FHF
C
Driver
Anti-driver
EPITHELIAL_MESENCHYMAL_TRANSITION
HYPOXIA
UV_RESPONSE_DN
TGF_BETA_SIGNALING
APICAL_JUNCTION
IL2_STAT5_SIGNALING
MYOGENESIS
KRAS_SIGNALING
E2F_TARGETS
Hallmark Gene Sets
Over-representation Score
D
NMP Neuromes progenitor
MMP Mesp1 mes progenitor
PMP Paraxial mes progenitor
PM Paraxial mes
4,623 cells
NMP
MMP
PMP
PM
F
Driver
Anti-driver
KRAS_SIGNALING_UP
EPITHELIAL_MESENCHYMAL_TRANSITION
ANGIOGENESIS
MYC_TARGETS_V1
Hallmark Gene Sets
Over-representation Score
B
Anti-drivers
225 total (16% TFs)
Drivers
383 total (11% TFs)
Slope
Spearman Coefficient
E
Anti-drivers
115 total (20% TFs)
Drivers
66 total (22% TFs)
Slope
Spearman Coefficient
G
FHF
Drivers
383 total
PM
Anti-drivers
115 total
337
46
69
15 (32%) TFs
H
Shared
FHF Drivers & PM Anti-driver
ANGIOGENESIS
HEDGEHOG_SIGNALING
TGF_BETA_SIGNALING
EPITHELIAL_MESENCHYMAL_TRANSITION
Hallmark Gene Sets
Over-representation Score
I
Shared FHF Drivers & PM Anti-drivers
=
Lineage Branching Signature
Up
Down
Expression of lineage branching signature
NMP
NMP
FHF
PM
Time
L
FHF Drivers
Down in Nipbl+/-
EPITHELIAL_MESENCHYMAL_TRANSITION
IL2_STAT5_SIGNALING
TGF_BETA_SIGNALING
APOPTOSIS
UV_RESPONSE_DN
ANGIOGENESIS
HEDGEHOG_SIGNALING
APICAL_JUNCTION
HYPOXIA
Hallmark Gene Sets
Over-representation Score
J
FHF Lineage
Anti-drivers
68 DEGs in Nipbl+/-
30 Up, 38 Down
X² = 0.5, P=0.4
Drivers
174 DEGs in Nipbl+/-
8 Up, 166 Down
X² = 105.3, P=2x10-16
8 (6%) TFs
Fold Change from WT
Spearman Coefficient
Up in Nipbl+/-
Down in Nipbl+/-
No Diff
K
PM Lineage
Anti-drivers
51 DEGs in Nipbl+/-
11 Up, 40 Down
X² = 10.6, P=1x10-3
Drivers
22 DEGs in Nipbl+/-
7 Up, 15 Down
X² = 1.5, P=0.2
Fold Change from WT
Spearman Coefficient
Up in Nipbl+/-
Down in Nipbl+/-
No Diff

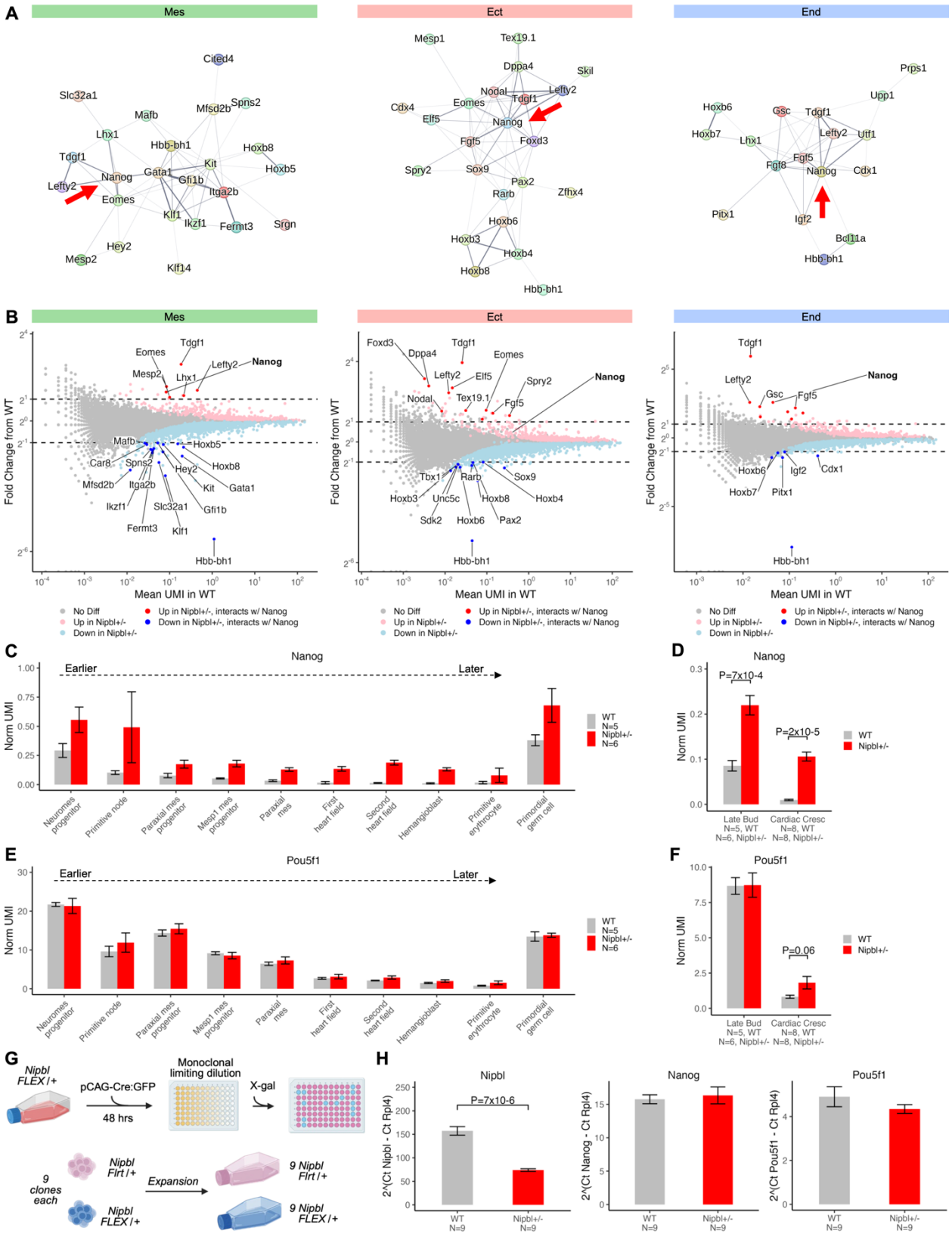
1304

1305 **Fig. 5. *Nipbl*$^{+/-}$ mice underexpress genes predicted to drive the transition of mesoderm cells**
1306 **into first heart field.** UMAP of mesoderm cells in (**A**) FHF and (**D**) PM lineages of WT LB-
1307 stage embryos. Genes whose expression are positively correlated (drivers) or negatively
1308 correlated (anti-drivers) with the transition (absorption probabilities from Fig. 3J) of mesoderm
1309 cells from WT LB-stage embryos into (**B**) FHF or (**E**) PM fates. Correlation coefficients
1310 calculated using Spearman's Rank Correlation. Genes with correlation coefficient greater than
1311 0.25 or less than -0.25 were considered drivers and anti-drivers, respectively. Slope refers to
1312 change in gene expression along absorption probability. Over-representation score of MSigDB's
1313 Hallmark gene sets among drivers and anti-drivers of (**C**) FHF and (**F**) PM fate. (**G**) Venn
1314 diagram of genes shared between PM anti-drivers and FHF drivers. (**H**) Over-representation score
1315 of MSigDB's Hallmark gene sets among shared FHF drivers and PM anti-drivers. (**I**) Expression
1316 of lineage branching signature as NMPs transition into FHF or PM fates. Fold change in
1317 expression of genes in mesoderm cells of (**J**) FHF and (**K**) PM lineages of LB-stage *Nipbl*$^{+/-}$
1318 embryos differentially expressed ($Q < 0.05$, Mann-Whitney U Test) from that of WT embryos
1319 along Spearman's Rank Correlation coefficient from Fig. 5, B and E. (**L**) Over-representation
1320 score of MSigDB's Hallmark gene sets among shared FHF drivers downregulated in FHF lineage
1321 of *Nipbl*$^{+/-}$ mice.


A
Mes
Ect
End
>2-fold, Up
<=2-fold, Up
Q>=0.5
<=2-fold, Down
>2-fold, Down
14
590
12945
2782
58
35
643
12074
3577
48
25
387
14780
1094
22
Number of Genes
B
Hallmark
MTORC1_SIGNALING
G2M_CHECKPOINT
MYC_TARGETS_V1
OXIDATIVE_PHOSPHORYLATION
E2F_TARGETS
Ect
Mes
End
Norm Enrich Score
C
Oxidative phosphorylation
Mtorc1 signaling
G2M checkpoint
Myc targets v1
E2f targets
Fold Change from WT
Mean Norm UMI in WT
No Diff
Up in Nipbl+/-
Down in Nipbl+/-
No Diff, Hallmark gene
Up in Nipbl+/-, Hallmark gene
Down in Nipbl+/-, Hallmark gene
D
Nodal signaling
Primitive erythropoiesis
Hox genes
Notch signaling
Hedgehog signaling
Adult hemoglobins
Tdgf1
Eomes
Lefty2
Nanog
Lhx1
Foxd3
Dppa4
Fgf5
Fgf8
Gsc
Jag1
Hbb-bt
Hba-a2
Hba-a1
Hbb-bs
Cdkn1b
Utf1
Sptb
Itgb3
Kzf1
Gata1
Epb42
Spi1
Gfi1b
Kit
Pip5k1b
Rac2
Fermt3
Myb
Tgm2
Nfe2
Tspan32
Igf1
Mafb
Itga2b
Klf1
Hey2
Nr2f2
Hbb-bh1
Hoxa5
Hoxb5
Hoxaas3
Hoxb8
Hoxb4
Hoxb6
Hoxb3
Pax2
Sox9
Nkx6-1
Irx2
Sox1
Hoxb7
No Diff
Up in Nipbl+/-
Down in Nipbl+/-

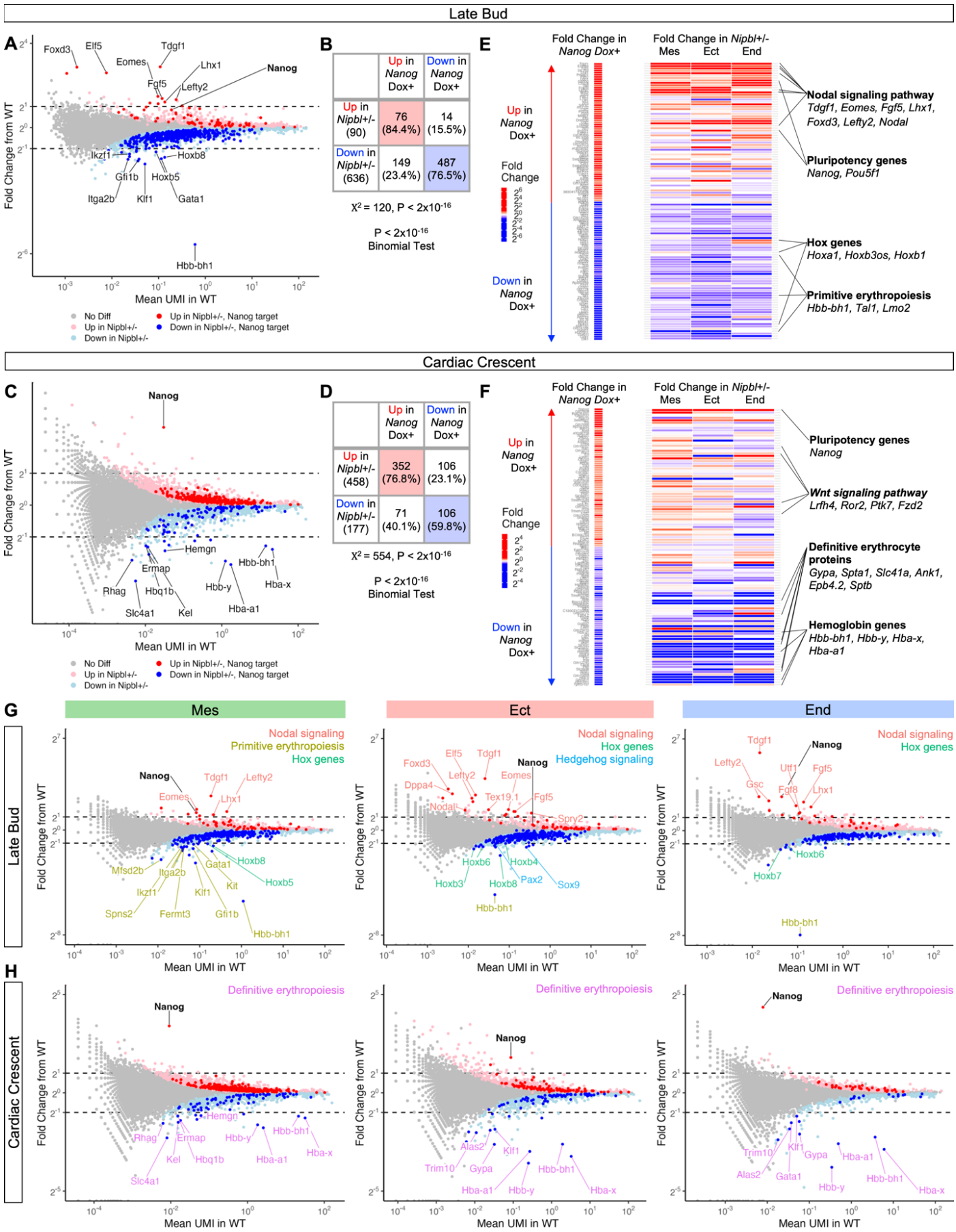
1322

1323 **Fig. 6. *Nipbl*$^{+/-}$ mice show large changes in the expression of major developmental regulators**
1324 **in all germ layers.** (**A**) Number of differentially expressed genes ($Q < 0.05$, Mann-Whitney U
1325 Test) in germ layers of LB-stage *Nipbl*$^{+/-}$ embryos showing small (≤ 2-fold) or large (> 2-fold)
1326 changes in expression from that of WT embryos. (**B**) Normalized enrichment score of statistically
1327 significant ($Q < 0.05$, FGSEA) Hallmark genes sets in germ layers of LB-stage *Nipbl*$^{+/-}$ embryos
1328 from that of WT embryos. (**C**) Fold change in expression of Hallmark gene sets from Fig. 5B in
1329 germ layers of LB-stage *Nipbl*$^{+/-}$ embryos from that of WT embryos along their average
1330 expression in WT embryos. (**D**) Fold change in expression of differentially expressed genes ($Q <$
1331 0.05, Mann-Whitney U Test) in germ layers of LB-stage *Nipbl*$^{+/-}$ embryos showing large changes
1332 in expression (> 2-fold) from that of WT embryos along their average expression in WT
1333 embryos.


A
Mes
Ect
End
B
Fold Change from WT
Mean UMI in WT
No Diff
Up in Nipbl+/-
Down in Nipbl+/-
Up in Nipbl+/-, interacts w/ Nanog
Down in Nipbl+/-, interacts w/ Nanog
C
Nanog
Earlier
Later
Norm UMI
WT N=5
Nipbl+/- N=6
Neuromes progenitor
Primitive node
Paraxial mes progenitor
Mesp1 mes progenitor
Paraxial mes
First heart field
Second heart field
Hemangioblast
Primitive erythrocyte
Primordial germ cell
D
P=7x10-4
P=2x10-5
Late Bud N=5, WT N=6, Nipbl+/-
Cardiac Cresc N=8, WT N=8, Nipbl+/-
E
Pou5f1
F
P=0.06
G
Nipbl FLEX/+
pCAG-Cre:GFP
48 hrs
Monoclonal limiting dilution
X-gal
9 clones each
Nipbl Flrt/+
Nipbl FLEX/+
Expansion
9 Nipbl Flrt/+
9 Nipbl FLEX/+
H
Nipbl
P=7x10-6
2^(Ct Nipbl - Ct Rpl4)
2^(Ct Nanog - Ct Rpl4)
2^(Ct Pou5f1 - Ct Rpl4)
WT N=9
Nipbl+/- N=9

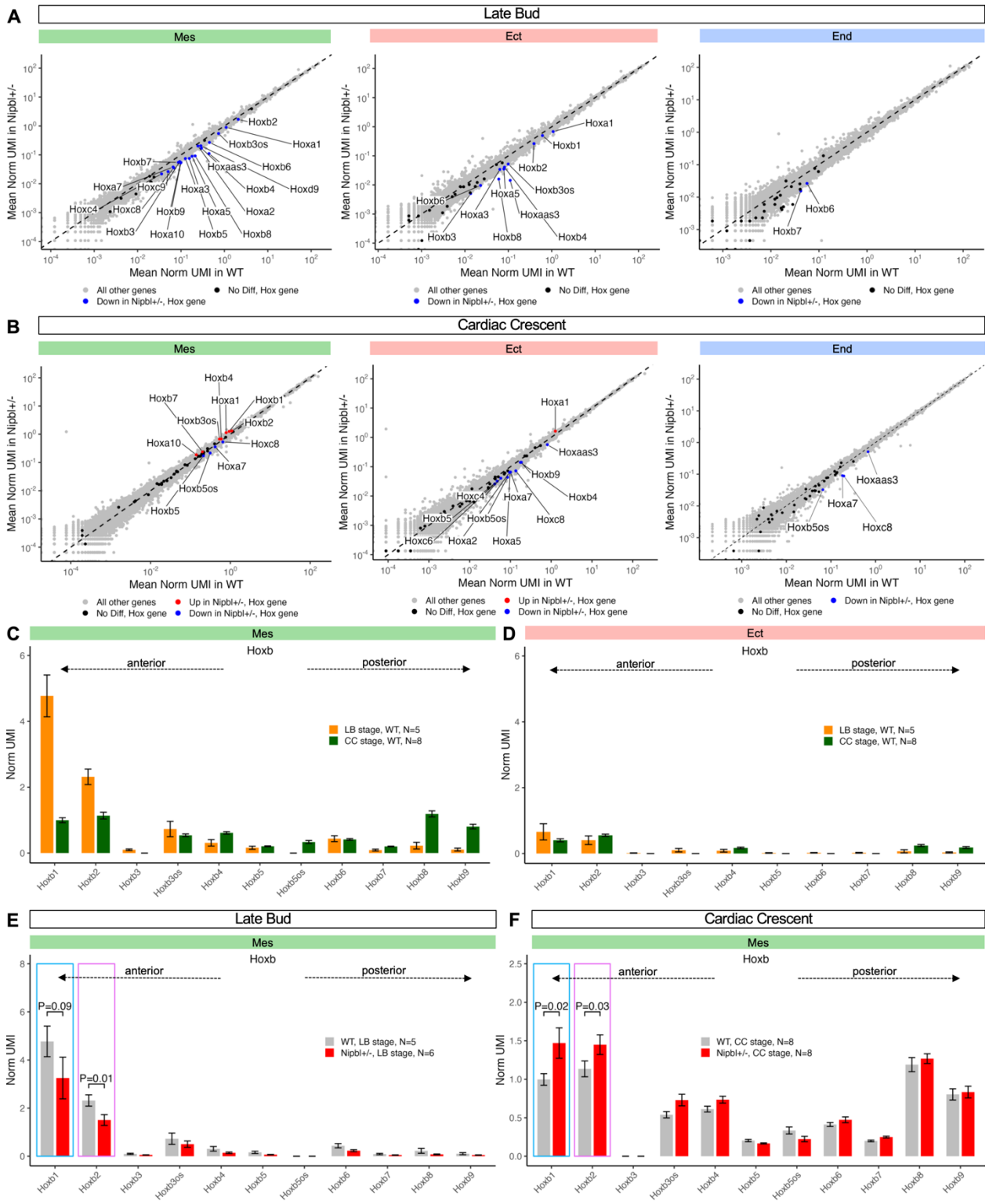
1334

1335 **Fig. 7. *Nipbl*$^{+/-}$ mice overexpress *Nanog* during and after gastrulation.** (**A**) Differentially
1336 expressed genes (greater than 2-fold upregulated or downregulated) between the germ layers of
1337 WT and *Nipbl*$^{+/-}$ embryos predicted by STRING to interact with each other. (**B**) Fold change in
1338 expression of differentially expressed genes ($Q < 0.05$, Mann-Whitney U Test) in germ layers of
1339 LB-stage *Nipbl*$^{+/-}$ embryos showing large changes in expression (> 2-fold) from that of WT
1340 embryos along their average expression in WT embryos that are predicted by STRING to interact
1341 with *Nanog*. (**C**) Expression of *Nanog* in mesodermal cell populations of LB-stage WT and
1342 *Nipbl*$^{+/-}$ embryos ordered from left (earlier) to right (later) by their RNA velocity positions in Fig.
1343 4A. Error bars show SEM. (**D**) Expression of *Nanog* in WT and *Nipbl*$^{+/-}$ embryos from LB- to
1344 CC-stage. Error bars show SEM. *P*-values from T-Test. € Expression of *Pou5f1* in mesodermal
1345 cell populations of LB-stage WT and *Nipbl*$^{+/-}$ embryos ordered from left (earlier) to right (later)
1346 by their RNA velocity positions in Fig. 4A. Error bars show SEM. (**F**) Expression of *Pou5f1* in
1347 WT and *Nipbl*$^{+/-}$ embryos from LB- to CC-stage. Error bars show SEM. *P*-values from T-Test.
1348 (**G**) Monoclonal generation of *Nipbl*$^{Flrt/+}$ (WT) and *Nipbl*$^{Flex/+}$ (*Nipbl*$^{+/-}$) ESCs. (**H**) Expression of
1349 *Nipbl*, *Nanog*, and *Pou5f1* in WT and *Nipbl*$^{+/-}$ embryonic stem cells as measured by RT-qPCR and
1350 normalized to the housekeeping gene, *Rpl4*. Error bars show SEM. *P*-values from T-Test.


Late Bud
A
Foxd3
Elf5
Tdgf1
Eomes
Lhx1
Nanog
Fgf5
Lefty2
Ikzf1
Hoxb8
Gfi1b
Hoxb5
Itga2b
Klf1
Gata1
Hbb-bh1
Fold Change from WT
Mean UMI in WT
No Diff
Up in Nipbl+/-
Down in Nipbl+/-
Up in Nipbl+/-, Nanog target
Down in Nipbl+/-, Nanog target
B
Up in Nanog Dox+
Down in Nanog Dox+
Up in Nipbl+/- (90)
76 (84.4%)
14 (15.5%)
Down in Nipbl+/- (636)
149 (23.4%)
487 (76.5%)
X² = 120, P < 2x10-16
P < 2x10-16
Binomial Test
E
Fold Change in Nanog Dox+
Fold Change in Nipbl+/-
Mes
Ect
End
Up in Nanog Dox+
Fold Change
Down in Nanog Dox+
Nodal signaling pathway
Tdgf1, Eomes, Fgf5, Lhx1, Foxd3, Lefty2, Nodal
Pluripotency genes
Nanog, Pou5f1
Hox genes
Hoxa1, Hoxb3os, Hoxb1
Primitive erythropoiesis
Hbb-bh1, Tal1, Lmo2
Cardiac Crescent
C
Nanog
Hemgn
Ermap
Hbb-bh1
Rhag
Hbq1b
Hbb-y
Hba-x
Slc4a1
Kel
Hba-a1
Fold Change from WT
Mean UMI in WT
No Diff
Up in Nipbl+/-
Down in Nipbl+/-
Up in Nipbl+/-, Nanog target
Down in Nipbl+/-, Nanog target
D
Up in Nanog Dox+
Down in Nanog Dox+
Up in Nipbl+/- (458)
352 (76.8%)
106 (23.1%)
Down in Nipbl+/- (177)
71 (40.1%)
106 (59.8%)
X² = 554, P < 2x10-16
P < 2x10-16
Binomial Test
F
Fold Change in Nanog Dox+
Fold Change in Nipbl+/-
Mes
Ect
End
Up in Nanog Dox+
Fold Change
Down in Nanog Dox+
Pluripotency genes
Nanog
Wnt signaling pathway
Lrfh4, Ror2, Ptk7, Fzd2
Definitive erythrocyte proteins
Gypa, Spta1, Slc41a, Ank1, Epb4.2, Sptb
Hemoglobin genes
Hbb-bh1, Hbb-y, Hba-x, Hba-a1
G
Mes
Ect
End
Late Bud
Nodal signaling
Primitive erythropoiesis
Hox genes
Hedgehog signaling
Nanog
Tdgf1
Lefty2
Eomes
Lhx1
Mfsd2b
Itga2b
Gata1
Hoxb8
Hoxb5
Ikzf1
Klf1
Kit
Spns2
Fermt3
Gfi1b
Hbb-bh1
Elf5
Foxd3
Dppa4
Nodal
Tex19.1
Fgf5
Spry2
Hoxb6
Hoxb4
Hoxb3
Hoxb8
Pax2
Sox9
Utf1
Gsc
Fgf8
Hoxb7
Fold Change from WT
Mean UMI in WT
H
Cardiac Crescent
Definitive erythropoiesis
Nanog
Hemgn
Rhag
Ermap
Hbb-y
Hbb-bh1
Hba-x
Kel
Hbq1b
Hba-a1
Slc4a1
Alas2
Klf1
Trim10
Gypa
Gata1
Fold Change from WT
Mean UMI in WT

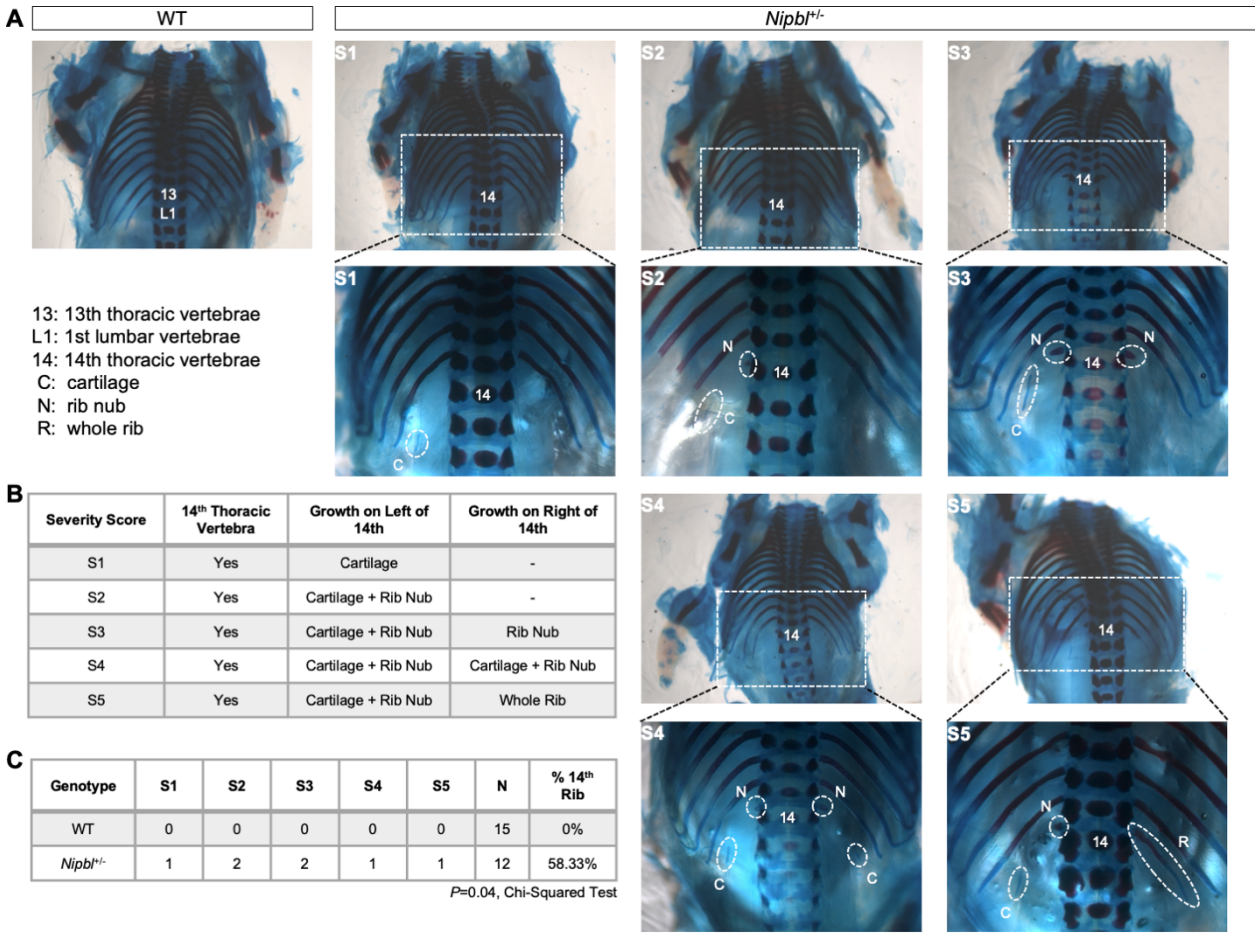
1351

1352 **Fig. 8. *Nipbl*$^{+/-}$ mice replicate the gene expression changes of *Nanog* overexpression.** Fold
1353 change in expression of gene targets of *Nanog* overexpression in (**A**) LB- and (**C**) CC-stage
1354 *Nipbl*$^{+/-}$ embryos differentially expressed ($Q < 0.05$, Mann-Whitney U Test) from that of WT
1355 embryos along their average expression in WT embryos. Genes differentially expressed ($Q <$
1356 0.05, T-Test) in E7.5 and E9.5 mouse embryos transgenic for doxycycline inducible expression of
1357 *Nanog* (*Nanog* Dox+) from that of WT embryos in Tiana et al., 2022 were considered targets of
1358 *Nanog* overexpression for LB- and CC-stage embryos, respectively. Percentage of genes ($Q <$
1359 0.05, Mann-Whitney U Test) in (**B**) LB- and (**D**) CC-stage *Nipbl*$^{+/-}$ embryos differentially
1360 expressed from that of WT embryos matching those differentially expressed in E7.5 and E9.5
1361 *Nanog* Dox+ embryos from that of WT embryos, respectively. Heatmap of fold change in
1362 expression in germ layers of (**E**) LB- and (**F**) CC-stage *Nipbl*$^{+/-}$ embryos from that of WT
1363 embryos of the most differentially expressed genes (lowest Q-values from T-Test) from E7.5 and
1364 E9.5 *Nanog* Dox+ embryos from that of WT embryos, respectively. Genes ordered from top to
1365 middle (overexpressed in *Nanog* Dox+ embryos) and bottom to middle (underexpressed in *Nanog*
1366 Dox+ embryos) by lowest to higher Q-value. Fold change in expression of genes in the germ
1367 layers of (**G**) LB- and (**H**) CC-stage *Nipbl*$^{+/-}$ embryos differentially expressed ($Q < 0.05$, Mann-
1368 Whitney U Test) from that of WT embryos along their average expression in WT embryos. Genes
1369 differentially expressed ($Q < 0.05$, T-Test) in E7.5 and E9.5 mouse embryos transgenic for
1370 doxycycline inducible expression of *Nanog* (*Nanog* Dox+) from that of WT embryos in Tiana et
1371 al., 2022 were considered targets of *Nanog* overexpression for LB- and CC-stage embryos,
1372 respectively, and are colored bright red or bright blue.


A
Late Bud
Mes
Ect
End
Mean Norm UMI in Nipbl+/-
Mean Norm UMI in WT
All other genes
No Diff, Hox gene
Down in Nipbl+/-, Hox gene
B
Cardiac Crescent
Up in Nipbl+/-, Hox gene
C
D
Hoxb
anterior
posterior
Norm UMI
LB stage, WT, N=5
CC stage, WT, N=8
E
F
P=0.09
P=0.01
P=0.02
P=0.03
WT, LB stage, N=5
Nipbl+/-, LB stage, N=6
WT, CC stage, N=8
Nipbl+/-, CC stage, N=8

1373

1374 **Fig. 9. *Nipbl*$^{+/-}$ mice show a delay in the expression of anterior *Hox* genes.** Expression of *Hox*
1375 genes in germ layers of WT and *Nipbl*$^{+/-}$ embryos at (**A**) LB- and (**B**) CC-stage. Expression of
1376 *Hoxb* genes in mesoderm (**C**) and ectoderm (**D**) of WT embryos from LB- to CC-stage ordered
1377 from left (5') to right (3') by their chromosomal position. Expression of *Hoxb* genes in mesoderm
1378 of WT embryos at (**E**) LB- and (**F**) CC-stage ordered from left (5') to right (3') by their
1379 chromosomal position.



B

Severity Score	14th Thoracic Vertebra	Growth on Left of 14th	Growth on Right of 14th
S1	Yes	Cartilage	-
S2	Yes	Cartilage + Rib Nub	-
S3	Yes	Cartilage + Rib Nub	Rib Nub
S4	Yes	Cartilage + Rib Nub	Cartilage + Rib Nub
S5	Yes	Cartilage + Rib Nub	Whole Rib

C

Genotype	S1	S2	S3	S4	S5	N	% 14th Rib
WT	0	0	0	0	0	15	0%
Nipbl+/-	1	2	2	1	1	12	58.33%

P=0.04, Chi-Squared Test

Fig. 10. *Nipbl*$^{+/-}$ mice show anteriorization of thoracic vertebrae with left-right asymmetry. (**A**) Dorsal view of bone (alizarin red) and cartilage (alcian blue) stained rib cage of E18.5 WT and *Nipbl*$^{+/-}$ embryos. WT embryos only show 13 ribs. *Nipbl*$^{+/-}$ embryos show incomplete asymmetric growth of 14th rib. S1-S5 refers to severity score in Fig. 10B. 13 = 13th thoracic vertebrae, L1 = 1st lumbar vertebrae, C = cartilage, N = rib nub, R = whole rib. (**B**) Table categorizing range of incomplete asymmetric growth of 14th rib observed in E18.5 *Nipbl*$^{+/-}$ embryos and ranking them by their severity (low = S1, high = S5). (**C**) Table quantifying numbers of WT and *Nipbl*$^{+/-}$ embryos in which incomplete asymmetric growth of 14th rib was observed per severity score. *P*-value from Chi-Squared Test.