Science Advances
AAAS
Manuscript Template

1 **FRONT MATTER**

2 **Title**

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

5 **Authors**

6 [1, 2, *]Stephenson Chea, [3]Jesse Kreger, [1, 2]Martha E. Lopez-Burks, [3]Adam L. MacLean, [1, 2]Arthur D.
7 Lander, [4, 1, 2, *]Anne L. Calof

8 **Affiliations**

9 [1]Department of Developmental and Cell Biology, School of Biological Sciences, University of
10 California Irvine, Irvine, CA 92697

11 [2]Center for Complex Biological Systems, University of California Irvine, Irvine, CA 92697

12 [3]Department of Quantitative and Computational Biology, Dornsife College of Letters, Arts, and
13 Sciences, University of Southern California, Los Angeles, CA

14 [4]Department of Anatomy and Neurobiology, School of Medicine, University of California Irvine,
15 Irvine, CA 92697

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

17 **Abstract**

18 In animal models, *Nipbl*-deficiency phenocopies gene expression changes and birth defects seen
19 in Cornelia de Lange Syndrome (CdLS), the most common cause of which is *Nipbl*-
20 haploinsufficiency. Previous studies in *Nipbl*$^{+/-}$ mice suggested that heart development is
21 abnormal as soon as cardiogenic tissue is formed. To investigate this, we performed single-cell
22 RNA-sequencing on wildtype (WT) and *Nipbl*$^{+/-}$ mouse embryos at gastrulation and early cardiac

crescent stages. *Nipbl*$^{+/-}$ embryos had fewer mesoderm cells than WT and altered proportions of mesodermal cell subpopulations. These findings were associated with underexpression of genes implicated in driving specific mesodermal lineages. In addition, *Nanog* was found to be overexpressed in all germ layers, and many gene expression changes observed in *Nipbl*$^{+/-}$ embryos could be attributed to *Nanog* overexpression. These findings establish a link between *Nipbl*-deficiency, *Nanog* overexpression, and gene expression dysregulation/ lineage misallocation, which ultimately manifest as birth defects in *Nipbl*$^{+/-}$ animals and CdLS.

Teaser

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

Short title

Gastrula-stage origins of a birth defects syndrome

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 major impact that birth defects have on infant 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 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

48 hands, coupled with issues in the visual, auditory, gastrointestinal, genitourinary, and
49 cardiopulmonary systems (*6*).

50 The majority of CdLS cases – more than 55% – are caused by heterozygous mutations in
51 the gene *Nipped-B-like* (*NIPBL*) (*7*), named for its homology to the Drosophila gene *Nipped-B*.
52 These mutations often produce a non-functional protein, suggesting CdLS arises from
53 haploinsufficiency (*8*). Remarkably, even a subtle 15% reduction in *NIPBL* gene expression can
54 produce a mild yet recognizable CdLS phenotype (*9*). Such observations highlight the importance
55 of precise *NIPBL* gene dosage in human development.

56 The *NIPBL* gene encodes a universally conserved protein that plays a role in loading
57 cohesin onto chromosomes (*10*). Cohesin, similarly conserved and ubiquitous, is a four-subunit
58 protein complex (Smc1, Smc3, Rad21, and either SA1 or SA2) essential for chromosome
59 organization and genome stability (*11*). Mutations in the cohesin subunits *SMC1* and *SMC3*
60 account for a small proportion of clinically mild CdLS (~5% and <1%, respectively) (*12-14*).
61 Additionally, mutations in *HDAC8*, which catalyzes cohesin release from chromatin during
62 mitosis, are found in a distinct subset of CdLS patients (*15*). Mutations in *RAD21* have also
63 surfaced in individuals exhibiting a CdLS-like phenotype but with substantially milder cognitive
64 impairment (*16*). Collectively, these findings reinforce the idea that impairment of cohesin
65 function contributes to CdLS.

66 Cohesin was initially identified for its role in sister chromatid cohesion during mitosis
67 (*11*). However, pronounced defects in sister chromatid cohesion or irregularities in mitosis have
68 not been observed in either CdLS patients or *Nipbl*-haploinsufficient (*Nipbl*$^{+/-}$) mice (*17, 18*),
69 suggesting that cohesin has additional functions. Studies in CdLS patient cells and *Nipbl*-
70 haploinsufficiency animal models indicate that cohesin is involved in transcriptional regulation
71 (*19*). Specifically, *Nipbl*-haploinsufficiency leads to alterations in the expression of many
72 hundreds to thousands of genes (*18*). Interestingly, many of the affected genes are regulated by

73 long-distance enhancers (*20*), aligning with the emerging concept of NIPBL and cohesin as
74 critical determinants in DNA-looping (*21*).

75 Interestingly, most gene expression changes in *Nipbl*-deficient animals are small, usually
76 less than 2-fold. Though likely to be inconsequential individually, such small changes can act
77 collectively to produce structural and functional defects. For example, in zebrafish, joint depletion
78 of two developmental genes downregulated by *nipbl*-deficiency produced a CdLS-like phenotype
79 (*22*), suggesting that altered gene expression is the ultimate cause of developmental and
80 physiological abnormalities. Thus, CdLS exemplifies a class of genetic disorders known as
81 "transcriptomopathies" (*23*), in which the cumulative or synergistic effects of minor disturbances
82 to gene expression lead to developmental abnormalities.

83 Previously, we reported that *Nipbl*-haploinsufficient mice displayed birth defects
84 phenocopying those in CdLS (*18*). These included CHDs (primarily atrial septal defects), in about
85 30% of *Nipbl*$^{+/-}$ mice. Subsequently, we used a conditional *Nipbl* allelic series to investigate the
86 role of *Nipbl* expression in the production of CHDs (*24*). That study showed that *Nipbl*$^{+/-}$ mice
87 exhibit heart abnormalities early in development: At embryonic day 13.5 (E13.5), 70% of *Nipbl*$^{+/-}$
88 mice displayed delays in ventricular septal fusion, whereas three days earlier, at E10.5, 100%
89 showed right ventricle hypoplasia (*24*). In situ hybridization experiments showed reduced
90 expression of two transcription factors crucial for early heart progenitor cell growth and
91 differentiation: *Nkx2-5* and *Mesp1* (*24*). That structural abnormalities may begin at the earliest
92 stages of heart development was suggested by results in in *nipbl*-morphant zebrafish, in which
93 defects in the initial migration of cardiogenic mesoderm were detected as early as 18 hours post-
94 fertilization (*22*). These results suggested that, at least for CHDs, causal events may occur as early
95 as gastrulation, when the three primary germ layers form and the earliest progenitor cells of major
96 tissues and organs begin to differentiate (*25*).

97 In this study, we sought to identify developmental alterations during gastrulation that
98 might account for birth defects in *Nipbl*$^{+/-}$ mice. We used single-cell RNA sequencing
99 (scRNAseq) to compare the cellular compositions, lineage trajectories, and transcriptional
100 landscapes of *Nipbl*$^{+/-}$ mouse embryos to their WT counterparts at two stages spanning the end of
101 gastrulation: the late bud- (LB) stage (E7.5) and cardiac crescent- (CC) stage (E7.75). Our
102 findings reveal that *Nipbl*$^{+/-}$ embryos have the same cell populations as WT embryos but display
103 subtle misallocation of specific mesodermal cell populations. Our evidence suggests this occurs as
104 a result of alterations in specific cell fate decisions, including the choice by mesoderm cells to
105 progress toward a non-cardiac versus cardiac fate. Our observations strongly suggest these events
106 cannot be attributed to changes in apoptosis, cell proliferation, global developmental delay, or the
107 structure of cell lineages. As in earlier research, we observed that most gene expression changes
108 in *Nipbl*-deficient tissues at these stages were small (*18*), but we also identified several key
109 developmental genes that are more markedly misexpressed in *Nipbl*$^{+/-}$ embryos, the most notable
110 of which was *Nanog*. *Nipbl*$^{+/-}$ embryos failed to downregulate *Nanog* at the end of gastrulation,
111 which normally occurs in all but germ cells. As a result, the misexpression of many *Nanog* target
112 genes was observed in all germ layers. We also saw substantial underexpression of *Hox* genes and
113 overexpression of Nodal signaling genes, which play roles in anterior-posterior and left-right
114 patterning, respectively. As a result, we propose a model in which birth defects in CdLS arise
115 from the prolonged overexpression of *Nanog* and dysregulation of developmental pathways
116 governing axial patterning, resulting in the misdirection of cell fate decisions, and misallocation
117 of specific progenitor cell populations.

118 **Results**

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

120 To investigate early factors influencing birth defects in CdLS, we generated WT and
121 *Nipbl*$^{+/-}$ littermate embryos by crossing *Nanog*$^{Cre/+}$ mice (*26*) with *Nipbl*$^{Flox/Flox}$ mice (*24*) as

122 described in Methods and shown in Fig. 1A. To ensure an accurate accounting of cell populations
123 and proportions at well-defined stages of early embryonic development, we generated a large
124 excess of embryos and sorted them into groups of narrowly defined stage based on morphological
125 criteria (*27*). Using the 10X Genomics Chromium Single Cell Expression platform (*28*), we
126 performed scRNAseq on 5 WT and 6 *Nipbl*$^{+/-}$ samples of late bud stage (LB) embryos (E7.5), and
127 8 WT and 8 *Nipbl*$^{+/-}$ samples of cardiac crescent stage (CC) embryos (E7.75) (Fig. 1B and fig.
128 S1). Given the low cell numbers in individual LB-stage embryos, and the 10X Genomics
129 Chromium Single Cell Expression system's capture efficiency, samples at the LB stage consisted
130 of pairs of embryos, whereas at the CC-stage they consisted of single embryos. Pairing LB-stage
131 embryos facilitated the capture of rare cell groups, such as primordial germ cells, which typically
132 comprise about 50 cells per embryo at this stage (*29*).

133 At LB-stage, we captured a median of 2,537 cells per sample, with a median of 17,301
134 RNA transcripts per cell (fig. S2 and table S1). At CC-stage, we captured a median of 4,116 cells
135 per sample, with a median of 14,997 RNA transcripts per cell (table S1). As expected, all *Nipbl*$^{+/-}$
136 embryos expressed both *Cre* and *Betageo*, which report an inactivated *Nipbl* allele (*24*), while
137 WT embryos did not (fig. S3, A and B). *Nipbl*$^{+/-}$ embryos across both stages expressed *Nipbl* at
138 levels that were approximately 50% lower than WT counterparts (fig. S3C).

139 We corrected for batch effects among embryos of identical stage and genotype using
140 Seurat's integration protocol (*30*) (Fig. 1C and fig. S4), which identifies anchors (cells of similar
141 gene expression across samples of the same biological condition) and uses them to align cells into
142 a space shared by all samples. Recognizing that all cells in *Nipbl*$^{+/-}$ tissues exhibit substantial gene
143 expression changes (*18*), and considering that clustering algorithms rely on differences in gene
144 expression among cells (*30*), we implemented measures to prevent our cell clustering from being
145 skewed by gene expression alterations attributable to *Nipbl*-haploinsufficiency. We did this by
146 first clustering cells from WT samples (Fig. 1C), using a robust iterative clustering method that

147 considered both intra-cluster stability and inter-cluster variation (fig. S5) to determine the optimal
148 number of clusters for cells at each developmental stage. WT cells were clustered into 18 cell
149 populations at LB-stage (Fig. 1D) and 26 cell populations at CC-stage (Fig. 1E). We then
150 projected *Nipbl*$^{+/-}$ cells onto these WT cell populations at corresponding stages (Fig. 1C). We
151 observed that all *Nipbl*$^{+/-}$ cells could be projected to WT cell populations, with each WT
152 population receiving some *Nipbl*$^{+/-}$ cells (Fig. 1, D and E). This pattern held true across individual
153 *Nipbl*$^{+/-}$ samples, as well as the total *Nipbl*$^{+/-}$ cells in aggregate at each stage (Fig. 1, F and G). To
154 ensure that any differences observed in cell populations were not the result of technical artifacts
155 introduced by the projection algorithm, we performed additional studies in which WT cells were
156 projected onto *Nipbl*$^{+/-}$ clusters (fig. S6 and fig. S15). In both cases, *Nipbl*$^{+/-}$ cells contributed to
157 all clusters, but as we describe later, in varying proportions (Fig. 3).

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

159 To assign biological identities to clustered cell populations, we used WT cells as a
160 reference. For each WT cell population, we performed differential gene expression analysis
161 (DGEA) by comparing it to all other cells within embryos of identical stage (Mann-Whitney U
162 test). We detected markers indicative of germ layer identity in every cluster (data S1, data S2, and
163 table S2), enabling us to associate each cluster with a specific germ layer at each developmental
164 stage (Fig. 2A). Figure 2B displays all cells from all samples, categorizing them into germ layers
165 at both LB- and CC-stages of development. Figures 2C and 2D further depict the assignment of
166 cells based on the expression of germ layer markers, at LB- and CC-stages, respectively. These
167 markers include those for ectoderm, such as *Utf1, Sox2,* and *Pou3f1* (*31-33*), mesoderm, including
168 *T, Hand1, Vim, Twist1,* and *Prrx2 (34-38)*, and endoderm, *Ttr (39)*.

169 To identify cell types within germ layers, we re-applied DGEA after clustering individual
170 germ layers. For the endoderm, this enabled us to identify definitive endoderm progenitors (DEP)
171 (*Sox4* and *Foxa2*) (*40, 41*), as well as extraembryonic endoderm (XEnd) (*Hnf4a* and *Tbx3) (42,*

172 *43)* at LB-stage (Fig. 2E, data S3, and table S3). In CC-stage embryos, we captured identical cell
173 types in the endoderm, and additionally, visceral endoderm (VE) (*Hnf4a* and *Esx1*) (*42, 44*)(Fig.
174 2F, data S4, and table S3). DEPs and XEnd give rise to the gut and placenta, respectively.
175 Within the ectoderm of LB-stage embryos, we identified early progenitor populations
176 including intraembryonic ectoderm progenitors (IEPs) (*Pou3f1* and *Otx2*) (*33, 45*), neural
177 ectoderm progenitors (NEPs) (*Cdx1* and *Nkx1-2) (46, 47)*, and surface ectoderm (SE) (*Id1, Msx2,*
178 and *Msx1*) (*48*) (Fig. 2, E and G, data S5, and table S4). Additionally, we detected a small portion
179 of extraembryonic ectoderm (XEct) (*Elf5* and *Tead4*) (*49, 50*), which was expected, as a small
180 quantity of XEct cells was included in our dissection. By CC-stage, IEPs, NEPs, and XEct were
181 no longer detectable, a phenomenon previously noted (*51*). Rather, we continued to detect surface
182 ectoderm (SE). We also detected six distinct neural populations (Fig. 2, F and H, data S6, and
183 table S4). These neural populations included: prosencephalic neural plate (*Six3* and *Hesx1*) (*52,*
184 *53*), mesencephalic neural plate (*En1* and *Pax2*) (*54, 55*), rhombencephalic neural plate (*Hoxa1*
185 and *Hes3*) (*56, 57*), midline neural plate (*Foxa2* and *Nkx2-9*) (*58, 59*), spinal neural plate (*Sox2*
186 and *Ezr*) (*60, 61*), and neural crest (*Pax3* and *Foxd3*) (*62, 63*) (data S6). Collectively, these
187 populations constitute the nascent progenitors of the brain, spinal cord, and peripheral nervous
188 system.
189 Within the mesoderm of LB-stage embryos, we identified six distinct groups of cells,
190 detailed in Fig. 2, E and G, data S7, and table S5. These consisted of: 1) neuromesodermal
191 progenitors (NMPs): expressing markers *T,* and *Nkx1-2 (34, 47)*; 2) derivatives of NMPs:
192 including *Mesp1*-expressing mesoderm progenitors (MMPs) (*Mesp1*) (*64*), two types of first heart
193 field cells (FHF1 and FHF2) marked by *Hand1 (35)*; cells of the second heart field (SHF) with
194 *Id2* and *Mef2c (65, 66)*; and three kinds of paraxial mesoderm cells, further categorized into
195 paraxial mesoderm progenitors (PMPs) (*Tbx6* and *Meis2) (67, 68)*, trunk paraxial mesoderm
196 (TPM) (*Foxc2* and *Tcf15*) (*69, 70*), and head paraxial mesoderm (HPM) (*Alx1* and *Tcf15*) (*70,*

197 *71*); 3) hematopoietic cells: including hemangioblasts (HMBs) (*Tal1* and *Etv2*) (*72, 73*) and
198 primitive erythrocytes (PEs) (*Gata1* and *Tal1*) (*72, 74*); and 4) two rare populations: comprising
199 cells of the primitive node (PN) (*Noto* and *Foxj1*) (*75, 76*), and primordial germ cells (PGCs)
200 (*Tfap2c* and *Msx1*) (*77, 78*). At CC-stage, we detected all mesodermal populations found earlier,
201 along with several that emerge later in development. These comprised lateral plate mesoderm
202 progenitors (LPMP) (*Tlx2* and *T*) (*34, 79*), additional populations of *Mesp1*-expressing mesoderm
203 progenitors (MMP1 and MMP2) (*Mesp1*) (*64*), cardiomyocytes (CDMs) (*Mef2c* and *Nkx2-5*) (*80,*
204 *81*), and presomitic mesoderm (PSM) (*Tcf15* and *Meox1*) (*70, 82*) (Fig. 2, F and H, data S8, and
205 table S5). Collectively, these progenitors are responsible for the formation of structures including
206 skeletal muscle, bone, blood, and heart.

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

209 We analyzed the cellular composition of embryos at both LB- and CC-stage, quantifying
210 the proportion of cells within each germ layer that contributed to the total cell count. LB-stage
211 mutants exhibited 13% fewer mesoderm cells (Fig. 3, A and B, data S9), while concurrently
212 displaying a greater proportion of endoderm and ectoderm cells. We sought to ascertain if this
213 decrease in mesoderm was generalized or restricted to specific subpopulations. Focusing on LB-
214 stage embryos, we merged mesodermal cell subpopulations of similar biological identity together
215 (FHF1 + FHF2 into FHF, and PSM + CPM into PM) (Fig. 3C) and calculated the percentage of
216 cells across these subpopulations, relative to the overall mesoderm. The findings (Fig. 3D)
217 revealed a 77% reduction in PEs $Nipbl^{+/-}$ embryos (P=0.06, T-Test, data S10). Additionally,
218 differences were observed in two mesodermal derivatives of the neuromesoderm: 1) Mutants had
219 24% fewer first heart field cells (FHF) (Fig. 3D); and 2) $Nipbl^{+/-}$ embryos showed a contrasting
220 pattern in paraxial mesoderm (PM), a 33% increase in PM cells compared to WT (Fig. 3D).

221 Since *Nipbl*$^{+/-}$ cells were projected onto WT cell populations, we wondered whether any
222 observed differences between *Nipbl*$^{+/-}$ and WT embryos might have occurred as a result of
223 misprojecting cells, i.e. assigning cells to incorrect clusters. To investigate this, we independently
224 clustered *Nipbl*$^{+/-}$ cells and projected WT cells onto those clusters; we refer to this as reverse
225 projection (fig. S6). If observed differences in the allocation of *Nipbl*$^{+/-}$ cells to particular clusters
226 had been a result of misclassification due to technical limitations of projecting cells, then the
227 same differences should not appear when reverse projection was done. Indeed, if misclassified
228 cells were simply distributed at random one should expect reverse projection to produce
229 differences opposite to those produced by the original, forward projection. This was not found to
230 be the case, however. Cluster assignments resulting from reverse projection quantitatively
231 resembled those of forward projection (fig. S6, A to D). Germ layer identities, which were based
232 on the expression of specific markers (table S6 and data S11), were generally the same in the
233 reverse projection analysis as well (fig. S6, E to H).
234 Finally, we quantified the proportion of cells within each germ layer and mesodermal
235 population following reverse projection. Consistent with the findings from forward projection
236 (Fig. 3B), *Nipbl*$^{+/-}$ embryos showed 13% fewer mesoderm cells than WT embryos, while
237 concurrently showing more endoderm and ectoderm cells (fig. S6M and data S13). Likewise, the
238 mesoderm of *Nipbl*$^{+/-}$ embryos showed fewer FHF cells (24% fewer) and more PM cells (22%
239 more) than WT embryos (Fig. 3D, fig. S6N, and data S14). (Because very few PEs are present in
240 *Nipbl*$^{+/-}$ embryos (Fig. 3D), it was not possible to cluster these cells into a distinct cell population
241 during reverse projection (fig. S6, K and L).) Altogether, these data indicate that differences in
242 cell population sizes observed in *Nipbl*$^{+/-}$ embryos represent true biological changes and are not a
243 consequence of technical issues, such as misprojection.
244 ***Nipbl*$^{+/-}$ embryos do not show changes in developmental timing**

245 Cellular composition changes very quickly during and after gastrulation; one possible
246 explanation for differences in proportions of cell types could be a change in the overall pace of
247 development, such that embryos of one genotype were slightly delayed or accelerated relative to
248 the other. To investigate whether this was the case for *Nipbl*$^{+/-}$ embryos, we employed
249 pseudotemporal ordering. To ensure that pseudotime accurately mirrored real developmental
250 time, we augmented our analysis with scRNAseq data from an additional two pairs of WT
251 embryos at the early bud (EB) stage and two pairs of WT embryos at the early head fold (EHF)
252 stage (Fig. 3E and fig. S7). EB-stage embryos are two hours younger and EHF-stage embryos are
253 two hours older than LB-stage embryos, so these supplementary samples provided the temporal
254 resolution necessary to discern even minor changes (less than 2 hours) in developmental timing
255 (*27*). We eliminated low-quality cells and doublets (Fig. 3F, fig. S8 and table S8) before
256 integrating cells of the same stage (fig. S9). We then clustered EB-stage cells into 8 distinct
257 populations and EHF-stage cells into 13 (fig. S10, A and B). Each population was subsequently
258 annotated as ectoderm (Ect), mesoderm (Mes), or endoderm (End) (Fig. 3G, and fig. S10, C and
259 D) based on the expression of specific germ layer markers (data S15 and S16, and table S9).
260 We ordered cells from each embryo at every developmental stage (EB, LB, and EHF)
261 using URD (*83*), which constructs a diffusion map of transition probabilities and, starting with an
262 assigned group of root cells, performs a probabilistic breadth-first graph search using the
263 transition probabilities. When we visualized the arrangement of cells from each stage based on
264 their density in pseudotime, we found that cells from WT embryos ordered in accordance to their
265 developmental stage: EB came first, followed by LB, and finally EHF (Fig. 3H, data S17), while
266 showing partial overlap between cells of different stages. This confirmed that the pseudotime
267 values we acquired were an accurate representation of actual developmental timing. When the
268 pseudotime orderings of WT and *Nipbl*$^{+/-}$ LB-stage embryos were compared with each other, we
269 found no statistically significant deviations (Kolmogorov-Smirnov test) (Fig. 3I). These findings

argue that the observed reduction in mesoderm cells in *Nipbl*$^{+/-}$ mice at LB-stage cannot be attributed to overall developmental delay or acceleration. The fact that *Nipbl*-haploinsufficiency does not result in global changes in developmental timing strongly suggests that birth defects in CdLS result from specific cellular or molecular irregularities within individual developmental pathways, and not a globally-altered developmental timeline.

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

To discern whether alterations in the sizes of cell population in *Nipbl*$^{+/-}$ embryos were attributable to changes in the structures of cell lineages, we used scVelo (*84*) to compute the RNA velocity for all mesoderm cells from LB-stage embryos. RNA velocity provides a predictive metric for a cell's future transcriptional state, gauging the equilibrium between the synthesis of spliced mRNA from unspliced mRNA and mRNA degradation. scVelo was used to build cell lineage trajectories separately for WT and *Nipbl*$^{+/-}$ embryos at LB-stage (Fig. 4A). In WT mesoderm, five distinct lineages were discernible (Fig. 4B, data S18), with neuromesodermal progenitors (NMPs) giving rise to three terminal fates via four specific pathways. These pathways encompassed the differentiation of NMPs into FHF, SHF, and PM. It's intriguing to note that *Mesp1*-expressing cells have the versatility to differentiate into both cardiac and paraxial mesoderm. While *Mesp1* is typically recognized as a cardinal factor for cardiac specification (*85*), our findings suggest its influence may extend beyond this role into paraxial specification as well. As a result, the paraxial mesoderm (PM) has dual progenitors, PMPs and MMPs. We also observed a fifth lineage, independent of NMP lineages, where HMBs differentiate into PEs. Notably, PGCs and PN cells did not form part of any identified lineages. Examination of the mesoderm in *Nipbl*$^{+/-}$ embryos revealed lineage trajectories identical to those in WT (Fig. 4A, data S19). These observations suggest that variations in the overall structures of cell lineage pathways do not explain differences in the number of FHF and PM cells in *Nipbl*$^{+/-}$ embryos (Fig. 3, B and D).

295 ***Nipbl*$^{+/-}$ embryos misdirect mesoderm cells into paraxial mesoderm at the expense of the first**
296 **heart field**

297 To explore whether variations in cell fate decisions might be responsible for the observed
298 differences in the numbers of FHF and PM cells in *Nipbl*$^{+/-}$ embryos, we employed CellRank (*86*)
299 to compute the likelihood of cells from the mesoderm (excluding HMBs, PEs, PGCs, and PN
300 cells) transitioning terminally into cells of SHF, FHF, or PM fate (data S20 and data S21).
301 HMBs, PEs, PGCs, and PN were omitted from the calculation because our trajectories derived
302 from RNA velocity indicated that these populations do not ultimately transition into SHF, FHF, or
303 PM. In CellRank, the absorption probability defines the likelihood that a specific cell type will
304 make a transition to a terminal state, derived using RNA velocity-directed, random walks from
305 initial to terminal cell states (*86*). We represented these probabilities in UMAP space, with shades
306 of blue to signify low probabilities and red for high probabilities. Fig. 4, C and D illustrates that
307 the probabilities of NMPs, MMPs, and PMPs making a terminal transition solely into SHF do not
308 vary significantly between *Nipbl*$^{+/-}$ and WT embryos, as quantified in the adjacent violin plots.
309 However, the transition probabilities into FHF are 27-48% lower in *Nipbl*$^{+/-}$ embryos (Fig. 4, E
310 and F), while the probabilities for transitions into PM are 9-11% higher (Fig. 4, G and H). Taken
311 together, these results support the hypothesis that disparities in cell numbers in *Nipbl*$^{+/-}$ embryos
312 stem from a misdirection of cells from FHF pathway towards a PM fate (Fig. 4I).

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

314 Altered cell numbers in different mesodermal subpopulations of *Nipbl*$^{+/-}$ embryos could
315 conceivably also stem from modifications in rates of cell death or proliferation. Focusing on
316 apoptotic activity, we carried out differential gene expression analysis (DGEA) on WT and
317 *Nipbl*$^{+/-}$ embryos at LB-stage, using genes associated with apoptosis from two well-known gene
318 sets: Reactome Apoptosis and Hallmark Apoptosis (*87, 88*). We ranked each gene by minimum
319 Q-value, arranging them from lowest to highest, and visualized their expression across all germ

320 layers in *Nipbl*$^{+/-}$ embryos as heatmaps of the fold change relative to their WT counterparts (fig.
321 S11, A and B, and data S22). For both gene sets, we saw no visible pattern of differences in the
322 expression of genes with even the lowest Q-values between any of the germ layers in both WT
323 and *Nipbl*$^{+/-}$ embryos. Although genes from these gene sets tend to mark cells with the capacity
324 for apoptosis, their expression levels are not necessarily indicative of cells actively undergoing
325 apoptosis. We therefore also turned to a gene set identified in a study that compared 180
326 apoptosis-associated genes in hematopoietic cells that were either healthy or undergoing apoptosis
327 (*89*), and identified 93 apoptosis-associated genes that were differentially expressed. When we
328 examined these genes in the germ layers of *Nipbl*$^{+/-}$ embryos, we also saw no visible difference
329 with WT embryos (fig. S11C and data S22). These data suggest that changes in apoptotic activity
330 are unlikely to be a major contributor to differences in cell type proportions.

331 To examine cell proliferation, we again employed DGEA across all germ layers of both
332 WT and *Nipbl*$^{+/-}$ embryos at LB-stage, focusing on the meta-PCNA gene set, an ensemble of
333 genes that exhibit the strongest positive correlation with PCNA expression, a recognized
334 biomarker of cell proliferation (*90*). When we visualized the expression of these genes in the
335 same manner as the gene sets above, we found no discernible variations in the expression levels
336 of the meta-PCNA genes between the germ layers of WT and *Nipbl*$^{+/-}$ embryos (fig. S12A and
337 data S23). As another measure of cell proliferation, we used Seurat to assign cells from LB-stage
338 WT and *Nipbl*$^{+/-}$ embryos into phases (G1, S, and G2/M) of the cell cycle (*91*). Seurat does this by
339 calculating a cell cycle phase score based on the expression of canonical S phase and G2/M phase
340 markers. Seurat considers these marker sets to be anticorrelated in their expression, so when cells
341 express neither, they are considered to be in G1 phase. As expected, in all germ layers, cells from
342 LB-stage WT and *Nipbl*$^{+/-}$ embryos assigned into all phases of the cell cycle (fig. S12B and data
343 S24). In all germ layers, there was no statistically significant difference in the proportions of cells
344 in each cell cycle between WT and *Nipbl*$^{+/-}$ embryos (fig. S12C). We therefore concluded that

345 extensive alterations in cell proliferation are not likely responsible for disparities in cell
346 subpopulations.

347 **$Nipbl^{+/-}$ mice underexpress genes predicted to drive the transition of mesoderm cells into**
348 **first heart field**

349 Next, we sought to investigate whether the misdirection of mesoderm cells in $Nipbl^{+/-}$
350 embryos at the LB-stage could be attributed to the misexpression of genes that drive the transition
351 of mesoderm cells into either FHF or PM. To answer this question, we also employed CellRank to
352 identify potential driver genes of the FHF and PM fates, using those mesoderm cells that
353 specifically contribute to these developmental pathways (Fig. 5, A and D). CellRank achieves this
354 by calculating a correlation coefficient between the likelihood (absorption probabilities) of cells
355 progressing towards a particular lineage fate and the expression level of individual genes (*86*). In
356 this context, genes with positive correlation coefficients are considered drivers, as their
357 expression elevates alongside an increase in absorption probabilities. Conversely, genes with
358 negative correlation coefficients are deemed anti-drivers, as their expression diminishes with
359 increasing absorption probabilities. To minimize the risk of false discoveries, we set a threshold,
360 accepting genes with correlation coefficients greater than 0.25 as drivers and those less than -0.25
361 as anti-drivers.

362 CellRank identified 383 genes as drivers of FHF fate (Fig. 5B and data S25). 11% of them
363 were transcription factors. Among these were *Hand1* and *Tbx20*, transcription factors previously
364 recognized for their roles in promoting FHF development (*35, 92*). Several transcription factors
365 that have not been previously linked to FHF development were also predicted to be FHF drivers,
366 including *Msx1, Msx2, Foxf1, Hoxb6,* and *Pitx1*. Of these, *Hoxb6* and *Pitx1* are known to play
367 important roles in segmentation and patterning. Gene set over-representation analysis using
368 MSigDB's Hallmark gene sets (*87*), which characterize well-defined biological states and
369 processes, revealed that FHF drivers were significantly enriched in genes associated with TGF-

370 beta signaling (Fig. 5C), which has been previously described as important in cardiomyocyte
371 proliferation (*93*). Genes associated with epithelial to mesenchymal transition (EMT) were also
372 significantly enriched (see below). Conversely, CellRank predicted 225 genes to be anti-drivers of
373 FHF fate, with 16% being transcription factors. Among these were genes known to be involved in
374 pluripotency (*Pou5f1*), mesoderm specification/expansion (*T, Cdx1, Sp5,* and *Zic3*), anterior-
375 posterior patterning (*Hoxb1* and *Hoxa1*), and paraxial mesoderm development (*Hes7, Meis2,*
376 *Tbx6, Gbx2,* and *Foxb1*).

377 CellRank predicted 66 genes to be drivers and 115 genes to be anti-drivers of PM fate
378 (Fig. 5E and data S26). A substantial proportion (22%) of PM drivers were transcription factors.
379 These included *Tcf15* and *Foxc2*, transcription factors that have been previously recognized for
380 their role in regulating PM development (Fig. 5E) (*69, 94*). The *Irx* transcription factors *Irx5,*
381 *Irx1,* and *Irx3* – known for their roles in segmentation during development (*95*) – were also
382 identified as PM drivers. PM drivers did not exhibit significant over-representation of any
383 Hallmark gene sets (Fig. 5F). Of all PM anti-drivers, 20% were identified as transcription factors,
384 including *Hand1, Cdx2, Msx2, Msx1,* and *Cdx4*. Intriguingly, *Hand1, Msx2, and Msx1* had also
385 been predicted as drivers for FHF. This overlap prompted us to conduct a comparative analysis
386 between PM anti-drivers and FHF drivers. Our examination revealed that 40% of all PM anti-
387 drivers were concurrently FHF drivers (Fig. 5, G and H).

388 The confluence of expressed genes between FHF drivers and PM anti-drivers seems to
389 indicate that during normal development, increased expression of these shared genes by
390 mesoderm cells predisposes them to transition into FHF, while decreased expression of these
391 same genes steers them towards PM (Fig. 5I). This set of shared genes may thus be considered to
392 be a "lineage branching signature".

393 How does *Nipbl*-haploinsufficiency impact the expression of driver and anti-driver genes?
394 Examination of *Nipbl*$^{+/-}$ embryos revealed significant misexpression in NMP, MMP, and PMP

395 cells of numerous driver and anti-driver genes associated with both FHF (first heart field) and PM
396 (paraxial mesoderm) (Fig. 5J, fig. S13, data S27, and data S28). Interestingly, most affected driver
397 and anti-driver genes were downregulated. To ascertain whether this was more than coincidental,
398 we employed a Chi-square analysis, which revealed a striking pattern: *Nipbl*$^{+/-}$ embryos
399 underexpressed FHF drivers at a higher frequency than they overexpressed them (166 out of 174
400 differentially expressed genes) (Fig. 5J). Similarly, they commonly underexpressed PM anti-
401 drivers (40 out of 51) (fig. S13), aligning with expectations since many PM anti-drivers serve dual
402 roles as FHF drivers.

403 Strikingly, genes associated with EMT were highly enriched among FHF drivers that were
404 underexpressed in *Nipbl*$^{+/-}$ embryos (Fig. 5K). A total of 15 EMT genes was found to be
405 underexpressed, including the EMT transcription factors *Msx1* and *Snai2*, and signaling genes
406 such as *Igfbp4*, *Pdgfrb*, and *Tgfb1* (Fig. 5J). These results suggest that, in the context of *Nipbl*-
407 haploinsufficiency, skewed differentiation of mesoderm cells towards the PM lineage at the
408 expense of FHF lineage may be attributable to the underexpression of genes driving FHF fate.
409 These results are consistent with the idea that genes associated with EMT play a role in this
410 lineage misdirection.

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

413 Using DGEA to compare the germ layers of LB-stage WT and *Nipbl*$^{+/-}$ embryos (data
414 S29, data S30, and data S31), we found that *Nipbl*$^{+/-}$ embryos misexpressed hundreds of genes
415 across all germ layers. Underexpression was more common than overexpression in all cases (Fig.
416 6A), supporting a broad role for *Nipbl* in enhancing gene expression. The majority of gene
417 expression changes were subtle, i.e., less than 2-fold (Fig. 6A). This pattern of gene expression
418 changes agrees with earlier studies (*18*). Gene set enrichment analysis (GSEA) for MSigDB's
419 Hallmark gene sets found that *Nipbl*$^{+/-}$ embryos showed enrichment, and no de-enrichment, for 4

420 out of 40 gene sets: oxidative phosphorylation in the mesoderm; Mtorc1 signaling in the
421 ectoderm; and G2M checkpoint, Myc targets, and E2F targets in the endoderm (Fig. 6B). In these
422 cases, however, enrichment was driven by a relatively small number of the genes in these sets; the
423 vast majority were either not differentially expressed in *Nipbl*$^{+/-}$ embryos or showed changes
424 much smaller than 1.5-fold (Fig. 6C).

425 Genes exhibiting substantial changes in expression – greater than 2-fold upregulated or
426 downregulated – were nevertheless observed across every germ layer of *Nipbl*$^{+/-}$ embryos (Fig.
427 6A). Individual curation of these genes unveiled associations with developmental processes (Fig.
428 6D). Within the mesoderm, two- to six-fold upregulation of genes involved in Nodal signaling,
429 including *Tdgf1, Eomes, Lefty2, Lhx1*, was observed (*96-99*) . At the same time, three clusters of
430 genes were robustly downregulated: those associated with primitive erythropoiesis, including
431 *Gata1* and *Klf1* (2-3 fold) (*74, 100*); *Hox* genes such as *Hoxb8, Hoxb5,* and *Hoxa5* (2-fold); and
432 genes associated with Notch signaling, including *Hey2* and *Nr2f2* (2-fold) (*101, 102*) (Fig. 6D).

433 In the ectoderm, five transcription factor genes tied to the hedgehog signaling pathway
434 were downregulated more than 2-fold, including *Pax2, Sox9, Nkx6-1, Sox1,* and *Irx2* (*103-107*)
435 (Fig. 6D). In the endoderm, there was a distinct 2-3 fold upregulation of several adult hemoglobin
436 genes such as *Hbb-bt, Hba-a2, Hba-a1,* and *Hbb-bs* (Fig. 6D). Intriguingly, consistent patterns
437 were evident in both the ectoderm and endoderm, mirroring the mesoderm's trends. Genes related
438 to Nodal signaling were upregulated, and *Hox* genes were noticeably downregulated, across all
439 germ layers. The magnitude of changes in Nodal signaling genes and *Hox* genes, being greater
440 than 2-fold in both directions, and their critical roles in various germ layers, strongly suggest that
441 disruptions in these pathways could be major contributors to phenotypes observed in *Nipbl*$^{+/-}$
442 mice.

443 When we compared the sets of differentially expressed genes (DEGs) obtained from
444 forward projection (Fig. 6) with those obtained from reverse projection (fig. S14), we found

445 greater than 94% of DEGs in mesoderm and ectoderm were identical (fig. S14, A and B). For the
446 endoderm, reverse projection showed more than 75% of the same DEGs (fig. S14C). As was the
447 case for changes in the relative sizes of cell populations, we conclude that DEGs observed in
448 *Nipbl*$^{+/-}$ embryos represent true biological differences and are not a consequence of technical
449 issues in cell classification.

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

451 To uncover developmental processes that might be the most broadly impacted in *Nipbl*$^{+/-}$
452 embryos, we used the large, significant, changes in gene expression (greater than 2-fold up or
453 down) as input data for STRING (*108*), a database and algorithm that constructs a network of
454 potential gene interactions (data S35), in this case limiting predictions to those informed by
455 experiments demonstrating co-expression or protein-protein interactions. The outcome of this
456 analysis (Figs. 7, A and B) revealed that many genes with substantial expression changes in
457 *Nipbl*$^{+/-}$ embryos are predicted to interact either directly or indirectly with *Nanog*. Indeed, *Nanog*
458 itself was overexpressed across all three germ layers in *Nipbl*$^{+/-}$ embryos, with a more pronounced
459 overexpression, exceeding 2-fold, in the mesoderm and endoderm (Fig. 7B).

460 *Nanog*, which encodes a transcription factor, is notably expressed at two junctures during
461 mouse embryonic development. Initially, it appears at the blastocyst stage, where it plays an
462 essential role in sustaining the pluripotency of the inner cell mass cells (*109*). Subsequently,
463 during gastrulation, *Nanog* expression temporarily surges, only to be silenced as cells transition
464 out of the primitive streak (*110*). This expression pattern could be observed directly in our data: In
465 Fig. 7C, we ordered mesodermal cell populations from LB-stage embryos by their stage in
466 gastrulation, as inferred from their RNA velocity positions (Fig. 4A). In WT NMPs, *Nanog* is
467 expressed at a relatively high level. As cells move through successive stages of gastrulation
468 within WT embryos, *Nanog's* expression exhibits a consistent decline (Fig. 7C and data S36),
469 reaching near zero in cells of the first and second heart field (FHF and SHF). An exception to this

470 trend occurs in primordial germ cells, which are known to maintain elevated *Nanog* expression
471 (*111*). In contrast, in *Nipbl*$^{+/-}$ embryos (Fig. 7C), *Nanog* expression remained elevated across all
472 cell populations. Although expression still falls after the NMP stage in *Nipbl*$^{+/-}$ cells, *Nanog* never
473 falls to WT levels. This is evident even in primordial germ cells. Moreover, the overexpression of
474 *Nanog* in *Nipbl*$^{+/-}$ embryos is sustained up to the CC-stage (Fig. 7D and table S10), where it
475 reaches a 10-fold increase over WT levels. These observations indicate that *Nipbl*-
476 haploinsufficiency leads to a marked failure of *Nanog* downregulation.

477 Another transcription factor that plays a role in maintaining pluripotency is *Pou5f1* (Oct4).
478 From studies in embryonic stem cells, it is known that *Pou5f1* promotes the expression of *Nanog*,
479 and conversely, *Nanog* promotes the expression of *Pou5f1* (*112*). Given this, we were curious
480 about whether *Nipbl*$^{+/-}$ embryos might also overexpress of *Pou5f1*. Despite the fact that *Pou5f1* is
481 more highly expressed than *Nanog*, we did not observe any statistically significant difference in
482 *Pou5f1* expression across mesodermal cell types or throughout LB-stage *Nipbl*$^{+/-}$ embryos (Fig.
483 7E, Fig. 7F, data S37, and table S11). However, a more than 2-fold elevation in *Pou5f1*
484 expression was detected in *Nipbl*$^{+/-}$ embryos at CC-stage (Fig. 7F).

485 We sought to determine whether elevated *Nanog* expression in *Nipbl*$^{+/-}$ embryos during
486 gastrulation might reflect some sort of non-specific overactivity of the *Nanog* gene. To investigate
487 this hypothesis, we looked at *Nanog* expression in WT and *Nipbl*$^{+/-}$ embryonic stem cells (ESCs),
488 as ESCs are known to express *Nanog* (as do to the blastocyst inner cell mass cells from which
489 ESCs are derived). We generated WT and *Nipbl*$^{+/-}$ ESCs by treating *Nipbl*$^{Flex/+}$ (*Nipbl*$^{+/-}$) ESCs
490 with Flp recombinase (which inverts a gene trap in *Nipbl*-Flex ESCs in such a way that reverses
491 gene trapping) to produce *Nipbl*$^{Flrt/+}$ (WT) ESCs (Fig. 7G) (*24*). We assessed *Nanog* expression in
492 nine independent clones from each genotype, using quantitative reverse transcription polymerase
493 chain reaction (RT-qPCR). We observed no discernible difference in *Nanog* expression between
494 *Nipbl*$^{+/-}$ and WT ESCs, a pattern that was also observed for *Pou5f1* (Fig. 7H and data S38). As a

495 control, we showed that *Nipbl*$^{+/-}$ ESCs exhibit the expected reduced expression of *Nipbl*, as also
496 seen in *Nipbl*$^{+/-}$ embryos. Collectively, these findings support the conclusion that the
497 overexpression of *Nanog* in *Nipbl*$^{+/-}$ LB-stage and CC-stage embryos is result of a specific failure
498 to appropriately suppress *Nanog* following gastrulation, rather than unusually elevated *Nanog*
499 expression persisting from early embryogenesis.

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

501 As a critical regulator of pluripotency in early embryonic development, *Nanog* directly
502 influences the expression of a multitude of other genes. This relationship prompted us to ask how
503 many of the gene expression differences in *Nipbl*$^{+/-}$ embryos might be attributable to the
504 overexpression of *Nanog*. To address this question, we took advantage of data from a recent
505 study, by Tiana et al., 2022 (*113*), in which mice were engineered to express *Nanog* under
506 doxycycline-inducible (Dox) control. In that study, bulk RNA sequencing was performed on
507 untreated (*Nanog* Dox-) and Dox-treated (*Nanog* Dox+) embryos at stages E7.5 and E9.5, with
508 E7.5 corresponding closely to the LB-stage embryos analyzed here. The E7.5 *Nanog* Dox+
509 embryos that were sequenced had been treated with doxycycline from E4.5 to E7.5, while E9.5
510 *Nanog* Dox+ embryos were treated with doxycycline from E6.5 to E9.5. Differential gene
511 expression analysis, performed in that study, found hundreds of gene expression changes between
512 *Nanog* Dox- and *Nanog* Dox+ embryos at both stages (*113*). We subsequently analyzed the
513 results of their analysis in four ways.

514 First, we plotted the fold change in expression of genes in *Nipbl*$^{+/-}$ embryos, as a whole,
515 from that of WT embryos, concentrating on LB-stage (Fig. 8A data S39). (CC-stage comparisons
516 are shown in fig. S15A, fig. S15B, and data S40.) In Fig. 8A, genes significantly overexpressed in
517 *Nipbl*$^{+/-}$ embryos are colored pink and those significantly underexpressed, light blue. DEGs in
518 *Nipbl*$^{+/-}$ embryos that are also significantly overexpressed or underexpressed in E7.5 *Nanog* Dox+
519 embryos, are colored red and blue, respectively. The results are summarized in Fig. 8B. At LB-

520 stage, 68.8% of overexpressed and 61.5% of underexpressed genes in *Nipbl*$^{+/-}$ embryos were also
521 overexpressed or underexpressed, respectively, in E7.5 *Nanog* Dox+ embryos. A binomial test
522 confirmed that *Nanog* Dox+ DEGs are highly over-represented in *Nipbl*$^{+/-}$ embryos (Fig. 8B).
523 Second, we again plotted the fold change in expression of genes in LB-stage *Nipbl*$^{+/-}$
524 embryos from that of WT embryos, but this time separately by germ layer, rather than across
525 whole embryos (Fig. 8C). Genes colored red and blue are DEGs in *Nipbl*$^{+/-}$ embryos that are also
526 overexpressed or underexpressed in *Nanog* Dox+ embryos. When we manually curated genes that
527 were expressed with more than 2-fold change (Fig. 8C), we found that all germ layers showed
528 upregulation of genes involved with Nodal signaling (*Tdgf1, Lefty2, Eomes, Lhx1,* and *Ffg5*) (*96-*
529 *99, 114*) and downregulation of *Hox* genes (*Hoxb8 and Hoxb6*). Additionally, the mesoderm of
530 *Nipbl*$^{+/-}$ embryos underexpressed genes associated with primitive erythropoiesis (*Hbb-bh1*, *Klf1*,
531 and *Gata1*) (*74, 100*) and hedgehog signaling (*Pax2* and *Sox9*) (*103, 104*). (CC-stage gene
532 expression changes are discussed in fig. S15C.) These findings are consistent with the idea that
533 many of the largest changes in gene expression that occur in *Nipbl*$^{+/-}$ embryos, as well as the
534 developmental pathways they regulate, could be attributable to overexpression of *Nanog*.
535 Third, we analyzed how closely *Nipbl*$^{+/-}$ embryos mirrored the gene expression alterations
536 of *Nanog* Dox+ embryos in terms of magnitude and direction. To do this, we generated a plot
537 illustrating the fold-changes in expression of individual genes in *Nipbl*$^{+/-}$ embryos compared to
538 their fold change in *Nanog* Dox+ embryos (Fig. 8D). We colored genes that were significantly
539 overexpressed in *Nanog* Dox+ embryos pink and those significantly underexpressed, light blue.
540 Additionally, we calculated the regression line depicting the relationship between fold changes in
541 gene expression in *Nipbl*$^{+/-}$ *vs. Nanog* Dox+ embryos. At LB-stage, the slope of the regression
542 line between *Nipbl*$^{+/-}$ and E7.5 *Nanog* Dox+ embryos was 0.8, implying that, on average, the fold
543 changes in gene expression in *Nipbl*$^{+/-}$ embryos are quantitatively similar to those in *Nanog* Dox+
544 embryos (Fig. 8D). Fig. 8D also depicts that, at LB-stage, *Nipbl*$^{+/-}$ embryos upregulate 46% of the

545 same genes that were overexpressed in E7.5 *Nanog* Dox+ embryos and downregulate 93% of the
546 same genes that were underexpressed in E7.5 *Nanog* Dox+ embryos. (Similar analysis was
547 performed at CC-stage, and is described in fig. S15D.) These findings highlight a remarkable
548 similarity in gene expression changes between *Nipbl*$^{+/-}$ and *Nanog* Dox+ embryos.

549 Fourth, we examined the magnitudes of fold changes in individual gene expression in
550 *Nipbl*$^{+/-}$ embryos and *Nanog* Dox+ embryos, focusing on the 50 most highly up- and down-
551 regulated genes in *Nanog* Dox+ embryos. As shown in Fig. 8E, we categorized genes into two
552 groups at each stage: those overexpressed in *Nanog* Dox+ embryos (ranked from top to middle)
553 and those underexpressed in *Nanog* Dox+ embryos (ranked from bottom to middle), ordering
554 them by ascending Q-value. Next, we constructed a heatmap that visually captures the fold
555 changes in gene expression within *Nanog* Dox+ whole embryos, contrasting them with the
556 corresponding fold changes in the individual germ layers of *Nipbl*$^{+/-}$ embryos (Fig. 8E). E7.5
557 *Nanog* Dox+ embryos overexpressed genes linked with the Nodal signaling pathway (*Tdgf1,*
558 *Eomes, Fgf5, Lhx1, Foxd3, Lefty2,* and *Nodal*) (*96-99, 114-116*) and those integral to maintaining
559 pluripotency (*Nanog* and *Pou5f1*) (Fig. 8E) (*112*). Concurrently, they underexpressed genes
560 associated with primitive erythropoiesis (*Hbb-bh1, Tal1,* and *Lmo2*) (*117, 118*) and specific *Hox*
561 genes (*Hoxa1, Hoxb3os,* and *Hoxb1*). Strikingly, this pattern of fold changes was mirrored, to a
562 remarkable degree, in the germ layers of LB-stage *Nipbl*$^{+/-}$ embryos (Fig. 8E). (The corresponding
563 analysis for CC-stage embryos is given in fig. S15E.)

564 **Differentially expressed genes from *Nipbl*$^{+/-}$ embryos are enriched for Nanog binding sites**

565 Given the large overlap between the gene expression changes observed in LB-stage
566 *Nipbl*$^{+/-}$ embryos and E7.5 *Nanog* Dox+ embryos, we wondered how many differentially
567 expressed genes in *Nipbl*$^{+/-}$ embryos might be direct targets of Nanog. Although genome-wide
568 patterns of Nanog binding have not been characterized for embryos at this stage, many groups
569 have used chromatin immunoprecipitation sequencing (ChIP-seq) to define Nanog binding sites in

570 mESCs. We turned to a recent study (*119*), in which Avsec et al. captured 52,456 Nanog binding
571 sites in the genome of mESCs, and reported that 3,645 genes showed one or more Nanog binding
572 sites located within 250 nt on either side of their transcription start sites (Fig. 8F). On comparing
573 these genes with those up- and down-regulated in LB-stage *Nipbl*$^{+/-}$ embryos, we found that 25%
574 of the overexpressed genes and 22% of the underexpressed genes match those genes with Nanog
575 binding sites in mESCs (Fig. 8, G and H). This represented significant enrichment for Nanog
576 binding sites in differentially vs. non-differentially expressed genes ($P<1x10^{-12}$ and $P<2x10^{-16}$,
577 Fisher's Exact Tests). These data suggest that a substantial portion of differentially expressed
578 genes in LB-stage *Nipbl*$^{+/-}$ embryos could indeed be direct targets of Nanog.

579 ***Nanog* overexpression may account for the downregulation of FHF drivers in *Nipbl*$^{+/-}$ mice**

580 As shown above in Fig. 3K and fig. S13, FHF drivers and PM antidrivers are
581 downregulated in LB-stage *Nipbl*$^{+/-}$ mice. Given the large overlap between downregulated genes
582 in LB-stage *Nipbl*$^{+/-}$ embryos and downregulated genes in E7.5 *Nanog* Dox+ embryos (Fig. 8, B
583 and D), we wondered how many of the downregulated FHF drivers and PM antidrivers in *Nipbl*$^{+/-}$
584 embryos could have their changes in expression attributed to the overexpression of *Nanog*. To
585 answer this question, we compared the FHF drivers and PM antidrivers downregulated in LB-
586 stage *Nipbl*$^{+/-}$ embryos to those genes exhibiting downregulation in E7.5 *Nanog* Dox+ embryos.
587 We found that approximately 40% of the FHF drivers and 38% of the PM antidrivers that were
588 downregulated in *Nipbl*$^{+/-}$ embryos were also downregulated in E7.5 *Nanog* Dox+ embryos (Fig.
589 8, I and J). This raises the possibility that overexpression of *Nanog* leads to the downregulation of
590 FHF drivers and PM antidrivers, and that this results in the misallocation of mesoderm cells to a
591 PM fate at the expense of FHF (Fig. 3 and Fig. 8K).

592 ***Nipbl*$^{+/-}$ mice exhibit delayed expression of anterior *Hox* genes**

593 We noticed that LB-stage *Nipbl*$^{+/-}$ embryos exhibited a significant underexpression,
594 exceeding 2-fold, in *Hox* genes across all germ layers, especially genes within the *Hoxb* cluster

595 (Fig. 6D and Fig. 8C). As *Hox* genes are integral to spatial patterning in development, we
596 compared the expression of all *Hox* genes in *Nipbl*$^{+/-}$ embryos with their WT counterparts. The
597 results, shown in Fig. 9, A and B, revealed extensive misexpression at both LB- and CC-stage. At
598 LB-stage, any *Hox* gene that was significantly differentially expressed was underexpressed in
599 *Nipbl*$^{+/-}$ embryos in every germ layer where expression was detectable (Fig. 9A). At CC-stage, we
600 observed a trend of underexpression across all germ layers for the majority of misexpressed *Hox*
601 genes (Fig. 9B), but exceptions were found in the mesoderm and ectoderm, where a select few
602 *Hox* genes were overexpressed. These included *Hoxb1* and *Hoxb2* in the mesoderm and *Hoxa1* in
603 the ectoderm, all of which are characterized as anterior *Hox* genes.

604 *Hox* genes are organized into chromosomal clusters that turn on expression in a wave-like
605 temporal pattern, producing highly structured and sequential activation during development.
606 Specifically, genes at the 5' end of the cluster, known as anterior *Hox* genes, are activated early in
607 development, and progressively silenced as genes further forward the 3' end (posterior *Hox* genes)
608 are turned on (*120*). To investigate whether our data reflected this known pattern, we examined
609 the *Hoxb* genes, as they were the most prominently expressed *Hox* genes at both LB- and CC-
610 stage. After ordering them according to their chromosomal positions, we plotted their expression
611 within the mesoderm and ectoderm of WT embryos at the corresponding developmental stages. In
612 the mesoderm of LB-stage embryos anterior *Hoxb* genes were actively expressed, while posterior
613 *Hoxb* genes were only faintly detectable, if at all (Fig. 9C and data S44). At CC-stage, the
614 expression of anterior *Hoxb* genes decreased from their LB-stage levels, accompanied by the
615 emergence of posterior *Hoxb* gene expression (Fig. 9C). This dynamic shift in expression aligns
616 closely with the established understanding of *Hox* gene regulation during development. In
617 contrast, the ectoderm exhibited a less pronounced pattern, with both anterior and posterior *Hoxb*
618 genes expressing at minimal levels (Fig. 9D and data S45). No discernible difference in the

619 expression of anterior *Hoxb* genes was detected in the ectoderm between LB- and CC-stage
620 embryos (Fig. 9D).

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

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

634 A previous study (*121*) showed that a knockout of *Hoxb1-Hoxb9* in mice led to the
635 anteriorization of the axial skeleton, characterized by an increased number of thoracic vertebrae
636 with ribs. Since *Nipbl*$^{+/-}$ mice underexpress *Hox* genes and show delayed regulation of anterior
637 *Hox* genes, we were curious if similar anteriorizations could be detected in *Nipbl*$^{+/-}$ mice (Fig. 9).
638 To explore this, we analyzed the vertebrae and ribs of 15 WT and 12 littermate *Nipbl*$^{+/-}$ embryos
639 at E18.5, using alizarin red and alcian blue stains. In WT embryos, we confirmed the presence of
640 the normal count of 13 ribs (Fig. 10A). In contrast, many *Nipbl*$^{+/-}$ embryos displayed 14 thoracic
641 vertebrae, the 14th of which bore a range of rib growths, often displaying left-right asymmetry
642 (Fig. 10A). We classified the growths of these 14th vertebrae based on severity, ranking them
643 from low (ss1) to high (ss5), and documented left-right differences (Fig. 7B). Three out of 7

644 *Nipbl*$^{+/-}$ embryos displayed partial growth of a 14th rib on the left side only (S1-S2). These
645 growths were either cartilage only or disconnected formation of rib nub with cartilage, but never
646 any whole ribs. Four out of 7 *Nipbl*$^{+/-}$ embryos showed bilateral growth (S3-S5). Partial growth
647 occurred on both the left and right sides, but complete formation a whole rib only ever occurred,
648 in one instance, on the right side (S5). In total, 58% of E18.5 *Nipbl*$^{+/-}$ embryos demonstrated some
649 form of rib growth from a 14th thoracic vertebra (Fig. 10C). These findings suggest that
650 anteriorization of the axial vertebrae does occur in *Nipbl*$^{+/-}$ embryos, possibly as a result of
651 delayed activation (and inactivation) of anterior *Hox* gene expression.

652 **Discussion**

653 The *Nipbl*$^{+/-}$ mouse, a model of the multisystem birth defects syndrome, Cornelia de
654 Lange Syndrome (CdLS) (*122*), offers a distinct lens through which to explore the genetic origins
655 of birth defects. *Nipbl*$^{+/-}$ mice display many of the same birth defects observed in human CdLS
656 and a key feature of *Nipbl*-deficiency across various organisms is quantitative alterations in gene
657 expression, including both upregulation and downregulation, affecting at least hundreds of genes
658 in every tissue (*18, 24, 123*). Previous research into animal models of *Nipbl*-deficiency suggested
659 that the root causes of birth defects in CdLS likely manifest during a period when progenitor cell
660 populations are being formed for all major tissue and organ lineages (*22, 24*). Therefore, to
661 elucidate how early, cell-type-specific changes in gene expression might contribute to the onset of
662 birth defects, we used scRNAseq to conduct a comparative analysis between *Nipbl*$^{+/-}$ embryos and
663 their WT littermates at key developmental stages, focusing on the conclusion of gastrulation, LB-
664 stage (approximately E7.5) to early CC-stage (approximately E7.75) (*25*).

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

667 In Fig. 1, we showed that *Nipbl*$^{+/-}$ embryos do not lack any cell populations found in WT
668 embryos. However, LB-stage *Nipbl*$^{+/-}$ embryos possess fewer total mesoderm cells, and within the

669 mesoderm, have fewer PEs and FHF cells, and more PM cells (Fig. 2). These changes, while not
670 always large, are likely to be of physiological importance. For example, the observation that PEs
671 are reduced in number is consistent with defects in blood formation and circulation observed in
672 *Nipbl*-morphant zebrafish (*22*), and thrombocytopenia in CdLS (*124*).

673 In addition, the reduction in FHF cell number could contribute to the presence of CHDs in
674 *Nipbl*$^{+/-}$ mice and individuals with CdLS (*24, 125*). In a prior study in *Nipbl*$^{+/-}$ mice, the majority
675 of CHDs observed were in heart regions (right ventricle, atrial/ventricular septa) considered to be
676 derivatives of the second heart field (SHF) (*24*). Although FHF cells are not thought to give rise
677 directly to "SHF structures", they likely exert indirect effects on the development of SHF
678 derivatives. This echoes the conclusions of Santos et al., 2016 (*24*), who found that complex
679 interactions between different cell types (even between non-cardiogenic and cardiogenic cell
680 types), influence the incidence of CHDs in *Nipbl*$^{+/-}$ mice.

681 The increased number of PM cells at LB-stage is intriguing, as it correlates with observed
682 defects in PM derivatives, including the axial skeleton (Fig. 10) and limb digits in *Nipbl*$^{+/-}$ mice
683 (*126*). Notably, both such deficits are features of CdLS (*127*). How an overabundance of PM cells
684 could contribute to limb defects is not obvious. Nevertheless, these findings suggest that birth
685 defects in *Nipbl*-deficient organisms arise not only from the absence of specific progenitor cell
686 populations, but also from progenitor cell misallocation events that alter the relative proportions
687 of cell types.

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

690 *Nipbl*$^{+/-}$ embryos at LB-stage did not exhibit global changes in apoptotic activity, cell
691 proliferation, developmental timing, or overall lineage trajectories (fig. S11, fig. S12, Fig. 3, and
692 Fig. 4). However, we produced evidence that nascent mesoderm cells in *Nipbl*$^{+/-}$ embryos
693 differentiated more frequently into PM than FHF, compared to WT (Fig. 4, C to I). This led us to

694 conclude that misallocation of mesoderm cells in *Nipbl*$^{+/-}$ embryos is driven by alterations in early
695 cell fate choices that occur along the main pathways of otherwise unaltered lineage trajectories.
696 These findings suggest that some types of structural birth defects, such as heart defects, arise from
697 quantitative alterations to cell fate determination, rather than a complete disruption of an entire
698 lineage pathway.

699 In Fig. 5, we presented evidence that reduced differentiation of mesoderm cells into FHF
700 in *Nipbl*$^{+/-}$ embryos could be attributed to the underexpression of genes predicted to drive FHF
701 differentiation. Interestingly, underexpressed FHF driver genes show strong enrichment for genes
702 associated with EMT (Fig. 5, J and K). EMT not only occurs during the emergence of
703 mesenchymal mesoderm and FHF cells during gastrulation; it also plays a role in the contribution
704 of later cell lineages (SHF cells, endocardial cells, and neural crest cells) to various cardiac
705 structures. SHF cells undergo EMT as they participate in the elongation of the heart tube and
706 outflow tract (*128*), and endocardial cells undergo EMT as they form structures such as the
707 cardiac cushion (*129*). Later in development, migratory neural crest cells, which arise from neural
708 epithelium by EMT, contribute to the aorticopulmonary septum and parts of the outflow tract
709 (*130*). Taken together, these observations suggest that dysregulation of EMT genes may play an
710 important role in the development of heart defects in *Nipbl*$^{+/-}$ mice.

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

712 In contrast to our prior studies of *Nipbl*$^{+/-}$ mice, in which differences in gene expression
713 were invariably found to be small (less than 2-fold) (*18*), the present study demonstrated that, in
714 gastrulation stage embryos, some genes show much larger shifts in expression (as high as eight-
715 fold) (Figs. 6D and 8H). One of the largest of such changes was in *Nanog,* which displayed
716 overexpression in LB-stage *Nipbl*$^{+/-}$ embryos, and which persisted to CC-stage (Figs. 7 and 8).
717 Using data from Tiana et al., 2022 (*113*), we demonstrated that many of the gene expression
718 changes in *Nipbl*$^{+/-}$ embryos mirror those seen in embryos with induced *Nanog* overexpression

719 (Fig. 8, A to E, and fig. S15, A to E). Moreover, many *Nipbl*$^{+/-}$ DEGs are likely to be direct Nanog
720 targets, as the corresponding genes display enrichment for Nanog binding sites (Fig. 8, F and H).
721 Comparisons of our data with those of Tiana et al., 2022 (*113*) also suggest that misallocation of
722 mesoderm cells to a PM fate at the expense of FHF, resulting from downregulation of FHF
723 drivers / PM antidrivers, may be a consequence of *Nanog* overexpression (Fig. 3 and Fig. 8K).

724 A question that arises from these observations is whether all gene expression alterations
725 observed in *Nipbl*$^{+/-}$ embryos might be due to *Nanog* overexpression. For a number of reasons, we
726 do not think that this is the case. First, for the largest gene expression shifts (greater than two-fold
727 over- or underexpressed), only 39% at LB-stage and 18% at CC-stage were misexpressed in the
728 same direction as DEGs in embryos in which *Nanog* was overexpressed (Fig. 8 and fig. S15)
729 (*113*). Second, although *Nanog* continues to be overexpressed in *Nipbl*$^{+/-}$ embryos from LB- to
730 CC-stage, the magnitude of its expression decreases substantially by CC-stage (Fig. 7). Third,
731 *Nipbl*-haploinsufficiency itself causes significant, large expression changes in other genes not
732 implicated as consequences of *Nanog* overexpression. For instance, we observed reduced
733 expression of genes linked to the Notch signaling pathway in the mesoderm, and increased
734 expression of adult hemoglobin genes in the endoderm (Fig. 6). Therefore, we conclude that the
735 extensive gene expression changes in *Nipbl*$^{+/-}$ embryos are unlikely to be due to *Nanog*
736 overexpression alone.

737 **Skeletal anomalies in *Nipbl*$^{+/-}$ mice reflect earlier delays in anterior *Hox* gene expression**

738 In Figs. 4D and 8G, we showed that large decreases in *Hox* gene expression occur across
739 all germ layers of LB-stage *Nipbl*$^{+/-}$ embryos. While only a subset of *Hox* genes is underexpressed
740 by more than two-fold, many *Hox* genes across all germ layers show small (less than two-fold)
741 but significant decreases in expression (Fig. 9A). We also provide evidence for a temporal
742 disturbance in anterior *Hox* gene expression in *Nipbl*$^{+/-}$ embryos, whereby *Nipbl*-
743 haploinsufficiency postpones the onset of anterior *Hox* gene expression at LB-stage and delays

744 their subsequent deactivation at CC-stage (Fig. 9, C to F). These shifts in *Hox* gene expression
745 dynamics mirror those in a previous study of *nipbl*-morphant zebrafish, in which genome-
746 location-specific misexpression of *hox* genes was also observed (*131*). In Fig. 10, we showed that
747 anteriorization of thoracic vertebrae occurs in E18.5 *Nipbl*$^{+/-}$ mice (Fig. 9). Such alterations in
748 axial skeletal development parallel skeletal anomalies that have been identified in individuals
749 with CdLS, including fused, absent, or misshapen ribs (*132*). Since *Nipbl*-deficiency is the
750 dominant form of CdLS, it is interesting to speculate that these skeletal anomalies might originate
751 from underexpression of and delayed anterior *Hox* gene activity during gastrulation.

752 **Overexpression of Nodal signaling pathway genes may contribute to left-right patterning**
753 **defects in CdLS**

754 In Fig. 4D, we showed that *Nipbl*$^{+/-}$ embryos at LB-stage show large overexpression of
755 genes associated with the Nodal signaling pathway across all germ layers. Nodal signaling,
756 pivotal for left-right patterning (*133*), ensures the appropriate positioning and morphogenesis of
757 the musculoskeletal system and internal organs (*134*). In Fig. 10, we showed that growth of a 14th
758 vertebra in E18.5 *Nipbl*$^{+/-}$ embryos displayed left-right asymmetry, with the growth of a whole
759 14th rib only occurring on the right side of the embryo. Such patterning alterations are consistent
760 with the heightened incidence of right-sided skeletal anomalies in CdLS (*127*). Changes in left-
761 right patterning are also consistently associated with intestinal malrotation and some types of
762 cardiac anomalies observed in *Nipbl*$^{+/-}$ mice, *nipbl*-morphant zebrafish, and CdLS (*18, 24, 135*).
763 These data suggest that left-right patterning defects in CdLS might be due to overexpression of
764 Nodal signaling genes as early as gastrulation.

765 **Gene expression changes observed in early *Nipbl*$^{+/-}$ embryos are likely driven by Nipbl's**
766 **effects on chromatin topology**

767 The results of this study, together with prior studies of mouse and zebrafish models,
768 reinforce the notion that the chromosomal location of genes is an important factor in their

769 sensitivity to variations in *Nipbl* levels. For example, our prior studies of *Nipbl*$^{+/-}$ embryonic brain
770 demonstrated that the largest expression changes among genes in the 22-gene protocadherin beta
771 (*Pcdhb*) cluster were found in those genes situated at the 5' and 3' ends of the cluster (*18*), where
772 CTCF sites are located (*136*). In developing pectoral fins of *nipbl*-morphant zebrafish, *hox* genes
773 located near the 3' end of three separate *hox* clusters (*hoxa*, *hoxc*, and *hoxd*) show a position-
774 dependent pattern of overexpression (*131*). Similarly, in the present study, we showed that
775 anterior *Hoxb* genes – located at the 3' end of the *Hoxb* cluster – are preferentially underexpressed
776 in the mesoderm of *Nipbl*$^{+/-}$ embryos at gastrulation (Fig. 8E). Thus, we consistently observe a
777 strong influence of chromosomal location on gene sensitivity to *Nipbl* levels, particularly within
778 gene clusters, and this influence is already apparent at gastrulation.

779 Recent studies suggest that changes in gene expression due to *Nipbl*-deficiency may be the
780 result of global alterations to chromosomal structure and organization. For instance, drastic
781 reduction of *Nipbl* expression in mouse hepatocytes resulted in genome-wide depletion of
782 topologically associated domains (TADs) and Hi-C interaction peaks (*137*), both of which require
783 chromatin looping (*138*). In fact, recent studies have demonstrated that both the formation and
784 rate of chromatin loop extrusion were reduced *in vitro* when a CdLS pathogenic mutation was
785 introduced into *NIPBL* (*139*).

786 Chromatin looping within TADs is thought to bring distant cis-regulatory elements, such
787 as enhancers, into proximity with target promoters. Interestingly, Nipbl has been reported to
788 preferentially bind to enhancers and promoters (*140*), and ChiP-seq studies show reduced
789 enhancer-promoter interactions in *Nipbl*$^{+/-}$ mouse embryonic fibroblasts (*141*). Since chromatin
790 looping is essential for gene activation, impaired loop extrusion as a consequence of reduced
791 Nipbl levels may provide an explanation for why *Nipbl*-haploinsufficiency causes more genes to
792 be downregulated than upregulated in gastrula-stage mouse embryos (e.g., >80% of mesodermal
793 genes, Fig. 6A), a trend that has been observed in studies of other tissues (*141*).

794 In light of such observations, it is interesting that overexpression of *Nanog* and its target
795 genes emerged from the present study as particularly likely to play important roles in causing
796 CdLS phenotypes. During normal gastrulation, *Nanog* expression falls to very low levels in most
797 cell types, so what occurs in *Nipbl*$^{+/-}$ embryos is perhaps best viewed as a failure of gene
798 repression. Although it is possible that *Nipbl*$^{+/-}$ haploinsufficiency leads to downregulation of a
799 gene encoding a *Nanog* repressor, no obvious candidate stands out within the gene expression
800 changes identified in the present study. We speculate, instead, that *Nipbl*-sensitive loop formation
801 is required for the silencing of *Nanog* gene expression during gastrulation, and thus that changes
802 in chromatin architecture play a key role in timing the termination of this critical embryonic
803 event.

804 **Materials and Methods**

805 **Statement on Care and Use of Animals**

806 All animals were handled in accordance with approved procedures as defined by the
807 National Institutes of Health, and all animal work was approved by the Institutional Animal Care
808 and Use Committee of the University of California, Irvine. For collection of mouse tissues,
809 pregnant dams were humanely killed by CO_2 anesthesia followed by cervical dislocation.

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

811 WT and *Nipbl*$^{+/-}$ mouse littermates were generated by mating *Nanog*$^{Cre/+}$ mice (*26*) and
812 *Nipbl*$^{Flox/Flox}$ mice (*24*). *Nipbl*$^{Flox/Flox}$ mice have an inverted gene trap cassette encoding *β–geo* that
813 is flanked by *Cre* recombinase target sites in intron 1 of *Nipbl* alleles (Fig. 1A) (*24*). In this
814 inverted orientation, referred herein as Flox, there is no trapping of the *Nipbl* gene and *Nipbl* is
815 expressed normally. However, when this cassette is exposed to Cre recombinase, the gene trap
816 cassette gets inverted into a non-inverted orientation which we call FIN (Fig. 1A). In this non-
817 inverted orientation, trapping of the *Nipbl* gene occurs and *β–geo* is expressed as a reporter of
818 successful gene trapping. Therefore, the *Nipbl* FIN allele is a null allele. *Nanog*$^{Cre/+}$ mice carry a

819 transgene encoding a *Cre* recombinase downstream of a promoter of the *Nanog* gene and initiate
820 recombination in the earliest cells of the embryo (*26*). Consequently, mating *Nanog*$^{Cre/+}$ mice with
821 *Nipbl*$^{Flox/Flox}$ mice results in littermates that are either *Nipbl*$^{Flox/+}$ or *Nipbl*$^{FIN/+}$, entirely. *Nipbl*$^{Flox/+}$
822 mice express *Nipbl* at wildtype levels (Fig. 1A) and show no defects, making them essentially WT
823 (*24*). *Nipbl*$^{FIN/+}$ mice express *Nipbl* at levels ~50% lower than WT (fig. S3), and show defects
824 similar to those observed in CdLS, making them essentially *Nipbl*$^{+/-}$ (*24*).

825 **Timing of Mouse Pregnancies**

826 To generate female mice that were pregnant on the same day, male mice were singly
827 housed and female mice were group-housed in groups of five for a minimum of one week to
828 synchronize their estrous cycles, thus taking advantage of the Lee-Boot effect (*142*). At the
829 beginning of the night cycle, the bedding from female cages were discarded and the bedding from
830 at least two male cages were transferred into each of them. At the beginning of the third night
831 cycle after which the females were exposed to male bedding, all females were transferred into
832 male cages, resulting in two females per male, thus taking advantage of the Whitten effect (*143*).
833 At the end of the third night cycle, the females were inspected for vaginal plugs. Those that had
834 vaginal plugs were considered potentially pregnant. To time the dissection of embryos from
835 potentially pregnant females, we considered the end of the 12-hour night cycle after which the
836 vaginal plug was discovered as E0 and dissected the embryos at the following times post E0: E7.5
837 (7 days + 12 hours) for late bud stage, E7.75 (7 days + 18 hours) for cardiac crescent stage, E7.41
838 (7 days + 10 hours) for early bud stage, E7.58 (7 days + 14 hours) for early head fold stage
839 embryos, and at E18.5 (18 days + 12 hours).

840 **Dissection of Mouse Embryos**

841 Pregnant female mice were euthanized by CO_2 inhalation followed by cervical dislocation.
842 The uterine horns were dissected out of their abdomens with dissection forceps and scissors and
843 placed in a petri dish with 1X DEPC PBS on ice. Individual deciduae were separated from one

844 another and transferred into their own petri dishes with 1X DEPC PBS on ice. Embryos were
845 dissected out of each deciduae under a dissection microscope as described in (*144*). The
846 Reichert's membrane was removed from each embryo. The ectoplacental cone was separated
847 from embryos collected for scRNAseq, and transferred by forcep into microcentrifuge tubes,
848 where they were kept on ice or stored at 20°C as tissue for polymerase chain reaction (PCR)
849 genotyping. The exocoelom was separated from embryos collected for scRNAseq and transferred
850 into a welled plate by wide bore pipette tip with fixative on ice as tissue for genotyping by X-gal
851 stain. Embryos for scRNAseq were transferred by wide bore pipette tip with 1X DEPC PBS into
852 microcentrifuge tubes and kept on ice until dissociation. Tails were separated from E18.5
853 embryos collected for Alcian Blue – Alizarin Red staining. Embryos for Alcian Blue – Alizarin
854 Red staining were transferred by forcep into 10% neutral buffered formalin in scintillation vials
855 and kept at 4°C for 24 hours.

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

857 ***X-gal Stain***

858 Since *Nipbl*$^{FIN/+}$ (*Nipbl*$^{+/-}$) mouse embryos express *β–geo* (beta-galactosidase), their tissues
859 will turn blue when they are treated with X-gal, a substrate that releases a blue chromophore when
860 enzymatically acted on by beta-galactosidase. The exocoeloms from embryos were transferred
861 into welled plates containing fixative (0.2% Glutaraldehyde, 5mM EGTA, and 2mM MgCl2 in
862 0.1M phosphate buffer, pH 7.5) and kept on ice for a minimum of 15 min. After fixation, the
863 exocoeloms were rinsed with a detergent rinse (0.02% Igepal, 0.01% Sodium Deoxycholate, and
864 2mM MgCl2 in 0.1M phosphate buffer, pH 7.5) before they were treated with X-gal stain (0.02%
865 Igepal, 0.01% Sodium Deoxycholate, 5mM Potassium Ferricyanide, 5mM Potassium
866 Ferrocyanide, and 2mM MgCl2 in 0.1M phosphate buffer, pH 7.5). The exocoeloms were
867 incubated in X-gal stain at 37°C in the dark for a minimum of 1 hour after which they were

868 inspected for coloration under a dissection microscope. Those that turned blue were considered
869 *Nipbl*$^{+/-}$ (*Nipbl*$^{FIN/+}$) and those that did not were considered WT.

870 ***Polymerase Chain Reaction***

871 Ectoplacental cones were treated with 50 ul 60 ug/ml proteinase K in PBND (50 mM KCl,
872 10 mM Tris-HCl, 2.5 mM MgCl2, 0.1 mg/ml gelatin, 0.45% Nonident P40, and 0.45% Tween
873 20), for 1 hour at 55°C to extract DNA for PCR. To deactivate proteinase K so that it does not
874 interfere with DNA amplification, they were then incubated at 95°C for 10 min. To genotype
875 mouse embryos as either *Nipbl*$^{Flox/+}$ (WT) or *Nipbl*$^{FIN/+}$ (*Nipbl*$^{+/-}$) and ES cells as either *Nipbl*$^{Flrt/+}$
876 (WT) or *Nipbl*$^{FLEX/+}$ (*Nipbl*$^{+/-}$), standard Taq PCR was performed on the extracted DNA using the
877 primers and thermocyling protocol previously described in (*24*). Primer 1: 5'-CTCCGC
878 CTCCTCTTCCTCCATC-3'. Primer 2: 5'-CCTCCCCCGTGCCTTCCTTGAC-3'. Primer 3: 5'-
879 TTTGAGGGGACGACGACAGTCT-3'. Thermocycling conditions: 1 cycle a 95°C for 30 sec,
880 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
881 hold at 4°C. PCR products were treated with loading dye and electrophoresed in agarose gel
882 stained with SYBER Safe DNA Gel Stain (Invitrogen: S33102) without cleanup and visualized
883 under UV light. Flox conformation: 782 base pairs (bp). FIN conformation 518 bp. Flrt
884 conformation: 735 bp. FLEX conformation: 652 bp.

885 **Single Cell RNA Sequencing**

886 Embryos were transferred by wide bore pipette tip in 20 ul 1X DEPC PBS into
887 microcentrifuge tubes. 200 ul of 1X TrypLE Express Enzyme with phenol red (Gibco: 12605010)
888 prewarmed to 37°C was added to each embryo. Embryos were triturated 4x with a wide bore
889 pipette tip. Embryos were incubated at 37°C and triturated 4x with a wide bore pipette tip every
890 minute (4-8 min) until no tissue aggregates were visible under a dissection microscope. TrypLE
891 Expres Enzyme activity was inactivated with the addition of pre-ice-chilled 200 ul 0.04% (m/v)
892 non-acetylated BSA in DPBS (Sigma-Aldrich: B6917-100MG, Gibco: 14190144). The resulting

893 cell suspension was underlayed with 200 ul of 1% non-acetylated BSA in HBSS (Sigma: 55021C-
894 1000ML) using a gel loading pipette tip. The suspension was centrifuged in a swing bucket
895 centrifuge at 300 rcf for 5 min at 4°C. 600 ul of supernatant was removed without disturbing the
896 cell pellet. 600 ul of HBSS was added to the cell pellet and the cell pellet was resuspended in it
897 using a wide bore pipette tip. The resulting cell suspension was centrifuged in a swing bucket
898 centrifuge at 300 rcf for 5 min at 4°C. 600 ul of supernatant was removed without disturbing the
899 cell pellet. The cell pellet was resuspending in the remaining 20 ul of supernatant using a wide
900 bore pipette tip.

901 20 ul single cell suspensions were submitted to the Genomics High Throughput Facility
902 (GHTF) at the University of California, Irvine (*145*) for scRNAseq using 10X Genomics'
903 Chromium Next GEM Single Cell 3' Kit v3.1 (10x Genomics: 1000268), Chromium Next GEM
904 Chip G Single Cell Kit (10x Genomics: 1000120), and Chromium Controller. Passage through the
905 Chromium Controller resulted in sample, cell, and transcript barcoded cDNA, which GHTF
906 amplified by PCR. GHTF assessed the quality and quantity of the amplified cDNA by
907 electrophoresis using Agilent's Agilent High Sensitivity DNA Kit (Agilent: 5067-4626) and
908 Bioanalyzer before library construction. Constructed libraries, representing embryonic samples,
909 were multiplexed and sequenced by GHTF on the Illumina HiSeq 4000 to a minimum depth of 20
910 million read pairs per cell. GHTF demultiplexed Illumina's raw base call (BCL) files and returned
911 FASTQ files as deliverables.

912 **Read Mapping and Cell Calling**

913 Cell Ranger v3.0 was used to map reads onto the GRCm38/mm10 C57BL/6J Mus
914 musculus genome/transcriptome assembly and call cells. Cell Ranger does this by using a read
915 mapper called STAR (*146*), which performs splicing-aware mapping of reads to the genome.
916 Reads are considered confidentially mapped to the genome with a MAPQ of 255. Exonic reads
917 are further mapped to annotated transcripts. A read that is compatible with the exons of an

918 annotated transcript, and aligned to the same strand, is considered mapped to the transcriptome.
919 CellRanger called cells using the EmptyDrops method described in (*147*).

920 **Normalization of Library Depth**

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

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

925 Cells exceeding three median absolute deviations in any one of the following criteria
926 among cells of the same stage were considered either low-quality cells or doublets (*148*) and
927 removed: 1) percentage of mitochondrial genes expressed, 2) number of genes expressed, and/or
928 3) number of transcripts detected (fig. S2 and fig. S8).

929 **Normalization of Cell Depth**

930 Seurat v3.0 was used to normalize the read depth between cells of the same stage using the
931 SCtransform method, which is described in (*149*). SCtransform does this by modeling the read
932 counts in a regularized negative binomial model to determine the variation due to read depth and
933 then adjusting that variance according to genes of similar abundances. SCtransform was also
934 used to normalize the read depth between cells when cells were subset from the whole embryo
935 into germ layers and clusters.

936 **Batch Effect Correction**

937 Seurat v3.0 was used to correct for batch effects among libraries of the same stage and
938 genotype. Seurat does this by identifying a set of shared variable genes among the libraries being
939 considered, and using these genes, identifies pairs of cells between any two libraries whose
940 expression of these genes are similar to other. These pairs of cells act as anchors between libraries
941 for batch effect correction and integration (*30*).

942 **Clustering of WT Cells**

943 Seurat v3.0 was used to cluster WT cells. Seurat does this by first performing principal
944 component analysis on the shared variable genes identified during batch effect correction and
945 integration. Principal components whose explained variances exceeded 2 median absolute
946 deviations were used to calculate k-nearest neighbors and construct a shared nearest-neighbor
947 graph. Clusters were determined by optimization of the modularity function using the Louvain
948 algorithm. The number of clusters was controlled by modulating the resolution function.

949 The optimal number of clusters was determined by clustering cells at increasing
950 consecutive numbers of clusters and generating a clustering tree (fig. S5) visualizing how cell
951 cluster identities change as the number of clusters consecutively increase. Clusters are stable
952 when a large proportion of cells are derived from a single preceding cluster rather than multiple
953 preceding clusters. We adopted Shannon Entropy as a measure of these proportions as a measure
954 of intra-cluster stability. A low Shannon Entropy represents high intra-cluster stability. We
955 visually inspected the clustering tree to determine which number of clusters maximized the
956 number of clusters while at the same time minimizing the total Shannon Entropy across all
957 clusters at that number of clusters.

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

962 **Differential Gene Expression Analysis Between Clusters**

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

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

969 Seurat v3.0 was used to project *Nipbl*$^{+/-}$ cells onto WT clusters. Seurat does this by
970 identifying a set of shared variable genes among WT samples, and using these genes, identifies
971 pairs of cells between WT cells and *Nipbl*$^{+/-}$ cells whose expression of these genes are similar to
972 other. These pairs of cells act as anchors between WT cells and *Nipbl*$^{+/-}$ cells for projecting and
973 sorting *Nipbl*$^{+/-}$ cells into WT clusters (*30*).

974 **Reverse Projection**

975 Using the same method that was used to cluster WT cells (see Clustering of WT Cells),
976 cells from LB-stage *Nipbl*$^{+/-}$ embryos were first clustered and annotated (fig. S6A). Cells from
977 LB-stage WT embryos were then projected onto the *Nipbl*$^{+/-}$ clusters (fig. S6, B to D) (see
978 Projection of *Nipbl*$^{+/-}$ Cells onto WT Clusters).

979 **Differential Gene Expression Analysis Between Genotypes**

980 Seurat v3.0 was used to perform differential gene expression analysis between WT and
981 *Nipbl*$^{+/-}$ cells. For each cluster, Seurat performs the Mann-Whitney U Test between WT and
982 *Nipbl*$^{+/-}$ cells using normalized read counts. P-values were corrected for false discovery using the
983 Bonferroni correction method. Genes with Q-values less than 0.05 were considered statistically
984 significant and differentially expressed.

985 **Pseudotime**

986 URD was used to calculate the pseudotime of cells from EB-, LB-, and EHF-stage
987 embryos (*83*). URD does this by constructing a diffusion map of transition probabilities and
988 starting with an assigned group of root cells, performs a probabilistic breadth-first graph search
989 using the transition probabilities. This moves step-wise outward from the root cells, until the
990 entire graph is visited. Several simulations are run, and then pseudotime is calculated as the
991 average iteration that visited each cell.

992 **Construction of Lineage Trajectories**

993 Velocyto (*150*) was used to count the numbers of spliced and unspliced transcripts per
994 gene, using default parameters. Reads aligning to exonic regions were counted as spliced. Reads
995 aligning to intronic regions were counted as unspliced. Reads aligning to exon-intron boundaries
996 were considered ambiguous and excluded from downstream analyses. scVelo (*84*) was used
997 estimate RNA velocities. Counts were normalized using the pp.filter_and_normalize() function,
998 moments of unspliced vs spliced abundances were computed using the pp.moments() function,
999 and velocities were computed using the tl.velocity() function, all using default parameters.
1000 Lineage trajectories were visually inferred from stream plots of computed RNA velocities.

1001 **Calculating Fate Probabilities**

1002 CellRank was used to calculate the fate probabilities of mesoderm cells (*86*). CellRank
1003 does this by performing RNA velocity-directed random walks from initial cell states to terminal
1004 cell states. Fate probabilities correspond to the fraction of walks in which a cell was a part of that
1005 that terminated in a particular terminal cell state. Terminal states were set with the
1006 *set_terminal_states()* function, absorption probabilities were computed with the
1007 *compute_absorption_probabilities()* function, and driver genes were computed with the
1008 *compute_lineage_drivers()* function, all with default parameters.

1009 **Cell Cycle Phase Assignment**

1010 Seurat was used to assign cells into G1, S, or G2/M phase based on the expression of
1011 markers of S phase and G2/M phase provided by Seurat. Using the CellCycleScoring() function,
1012 Seurat calculated scores for the expression of S phase and G2/M phase markers, while
1013 considering the expression of these marker genes to be anticorrelated to one another. When cells
1014 express neither, they are considered to be in G1 phase.

1015 **Identification of Drivers and Anti-drivers**

1016 CellRank was used to identify the drivers and anti-drivers of FHF and PM fates. CellRank
1017 does this by calculating a correlation coefficient between the fate probabilities of cells towards

1018 their lineage fate and the expression of their genes. Those genes with positive correlation
1019 coefficients are considered drivers, since their expression increases as absorption probabilities
1020 increase and those with negative correlation coefficients are considered anti-drivers, since their
1021 expression decreases as absorption probabilities decrease. To reduce the likelihood of false
1022 discovery, we considered those genes with correlation coefficients greater than 0.25 as drivers and
1023 those less than -0.25 as anti-drivers.

1024 **Gene Set Over-Representation Analysis**

1025 clusterProfiler (*151*) was used to perform gene set over-representation analysis. It does
1026 this by performing a Fisher's Exact Test on a contingency table of the genes in a gene set that
1027 match or do not match the genes of interest. P-values were corrected for false discovery using the
1028 Bonferroni correction method. Gene sets with Q-values less than 0.05 were considered
1029 statistically significant and over represented by the genes of interest.

1030 **Gene Set Enrichment Analysis**

1031 FGSEA (*152*) was used to perform gene set enrichment analysis.

1032 **Generation of *Nipbl*$^{FLEX/+}$ and *Nipbl*$^{Flrt/+}$ ESCs**

1033 ES cells were grown in Glasgow's MEM, 15% heat-inactivated FBS (ES cell qualified,
1034 Hyclone SH30071.03E), 1X Glutamine, 1X Pen-Strep, 5mM mercaptoethanol, and 1000 U/ml
1035 leukemia inhibitory factor (LIF: ESGRO Millipore). Following limited dilution single cell
1036 cloning, a clone of EUC313f02 *Nipbl*$^{FLEX/+}$ ES cells (European Conditional Mouse Mutagenesis
1037 Program) was transfected with pCAG-Cre:GFP (Addgene #13776) using Lipofectamine 2000
1038 (Invitrogen) to convert the *Nipbl*^{FLEX} allele to the *Nipbl*^{Flrt} conformation in vitro. 48 hours post-
1039 transfection, cells were plated at 1 cell/well into several 96 well plates. Clonal colonies were
1040 isolated, and the clones were stained for β-galactosidase (lacZ) activity using X-gal. For LacZ
1041 staining, the ES cells were fixed for 5 minutes in 2 mM MgCl2, 0.5% glutaraldehyde in 1X PBS,
1042 followed by 3 washes with 1xPBS at room temperature. X-gal staining (5 mM K3Fe(CN)6, 5 mM

1043 K4Fe(CN)6, 2 mM MgCl2, 1 mg/mL X-gal in 1X PBS) was performed at 37°C until the blue
1044 precipitate was detected. Colonies positive for X-gal staining, *Nipbl*$^{FLEX/+}$ ES cells, and negative
1045 for X-gal staining, *Nipbl*$^{Flrt/+}$, were verified by PCR genotyping, described above.

1046 **Reverse Transcription-Quantitative Polymerase Chain Reaction**

1047 Clones for both *Nipbl*$^{FLEX/+}$ and *Nipbl*$^{Flrt/+}$ were expanded, 9 each, and RNA was extracted
1048 using the Monarch Total RNA Miniprep Kit (New England Biolabs). cDNA was made using
1049 iSCRIPT reverse transcriptase (BioRad). qRT-PCR for *Nipbl*, *Nanog*, and *Pouf5f1* was performed
1050 using iTaq SYBR green (BioRad) as per manufacturer's instructions, *Rpl4* was used as the
1051 housekeeping gene. *Rpl4* primer 1: 5'-ATCTGGACGGAGAGTGCTTT-3'. *Rpl4* primer 2: 5'-
1052 GGTCGGTGTTCATCATCTTG-3'. *Nipbl* primer 1: 5'-AGTCCATATGCCCCACAGAG-3'.
1053 *Nipbl* primer 2: 5'-ACCGGCAACAATAGGACTTG-3'. *Nanog* primer 1: 5'-
1054 AAATCCCTTCCCTCGCCATC-3'. *Nanog* primer 2: 5'-GCCCTGACTTTAAGCCCAGA-3'.
1055 *Pou5f1* primer 1: 5'-CACCCTGGGCGTTCTCTTT-3'. *Pou5f1* primer 2: 5'-
1056 GTCTCCGATTTGCATATCTCCTG-3'.

1057 **Identification of Genes with Nanog Binding Sites**

1058 In Avsec et al., 2021 (*119*), ChIP-seq was performed for Nanog in mESCs and 56,459
1059 peaks were called (Fig. 8F). We considered the called peaks from that study as representing
1060 Nanog binding sites. Using the narrowPeak file from that study (which contains the called peaks),
1061 the GRCm38/mm10 C57BL/6J Mus musculus genome annotation, and ChIPseeker v3, we
1062 identified which genes had a peak +/- 250 nt of their TSS. We identified a total of 3,646 genes
1063 with a Nanog binding site at their TSS (Fig. 8F).

1064 **Alcian Blue – Alizarin Red Staining**

1065 E18.5 embryos were fixed in 10% neutral buffered formalin in scintillation vials at 4°C
1066 overnight. Embryos were washed with H_2O twice over 2 days at room temperature. Embryos
1067 were immersed in 95% EtOH for 1 week. Skin was removed from embryos with forceps under

1068 dissection microscope. Embryos were stained with 0.02% Alcian Blue (10 mg Alcian Blue + 80
1069 ml EtOH + 20 ml glacial acetic acid) over three days. Embryos were washed with 70% EtOH
1070 twice in 1 day, 40% EtOH overnight, 15% EtOH in 1 day, and H_2O overnight. Embryos were
1071 washed with 1% KOH twice over 3 days until they became translucent. Embryos were stained
1072 with 0.015% Alizarin Red (15 mg Alizarin Red + 100 ml 1% KOH) over 3 days. Embryos were
1073 washed with 1% KOH three times in 1 day. Fat pads and internal organs were removed from
1074 embryos with forceps under dissection microscope. Embryos were washed with 20% glycerol +
1075 1% KOH overnight, 50% glycerol + 1% KOH overnight, and 80% glycerol + 1% KOH overnight.
1076 Embryos were stored in 100% glycerol at room temperature.

1077 **References**

1078 1. C. Centers for Disease, Prevention, Update on overall prevalence of major birth defects--
1079 Atlanta, Georgia, 1978-2005. *MMWR Morb Mortal Wkly Rep* **57**, 1-5 (2008).
1080 2. C. T. Mai *et al.*, National population-based estimates for major birth defects, 2010-2014.
1081 *Birth Defects Res* **111**, 1420-1435 (2019).
1082 3. R. Sun, M. Liu, L. Lu, Y. Zheng, P. Zhang, Congenital Heart Disease: Causes, Diagnosis,
1083 Symptoms, and Treatments. *Cell Biochem Biophys* **72**, 857-860 (2015).
1084 4. D. M. Ely, A. K. Driscoll, Infant Mortality in the United States, 2020: Data From the
1085 Period Linked Birth/Infant Death File. *Natl Vital Stat Rep* **71**, 1-18 (2022).
1086 5. L. Mannini, F. Cucco, V. Quarantotti, I. D. Krantz, A. Musio, Mutation spectrum and
1087 genotype-phenotype correlation in Cornelia de Lange syndrome. *Hum Mutat* **34**, 1589-
1088 1596 (2013).
1089 6. A. D. Kline *et al.*, Cornelia de Lange syndrome: clinical review, diagnostic and scoring
1090 systems, and anticipatory guidance. *Am J Med Genet A* **143A**, 1287-1296 (2007).
1091 7. Z. A. Bhuiyan *et al.*, Genotype-phenotype correlations of 39 patients with Cornelia De
1092 Lange syndrome: the Dutch experience. *J Med Genet* **43**, 568-575 (2006).

1093 8. L. A. Gillis *et al.*, NIPBL mutational analysis in 120 individuals with Cornelia de Lange
1094 syndrome and evaluation of genotype-phenotype correlations. *Am J Hum Genet* **75**, 610-
1095 623 (2004).
1096 9. G. Borck *et al.*, Father-to-daughter transmission of Cornelia de Lange syndrome caused by
1097 a mutation in the 5' untranslated region of the NIPBL Gene. *Hum Mutat* **27**, 731-735
1098 (2006).
1099 10. R. Ciosk *et al.*, Cohesin's binding to chromosomes depends on a separate complex
1100 consisting of Scc2 and Scc4 proteins. *Mol Cell* **5**, 243-254 (2000).
1101 11. K. Nasmyth, C. H. Haering, Cohesin: its roles and mechanisms. *Annu Rev Genet* **43**, 525-
1102 558 (2009).
1103 12. A. Musio *et al.*, X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. *Nat*
1104 *Genet* **38**, 528-530 (2006).
1105 13. M. A. Deardorff *et al.*, Mutations in cohesin complex members SMC3 and SMC1A cause
1106 a mild variant of cornelia de Lange syndrome with predominant mental retardation. *Am J*
1107 *Hum Genet* **80**, 485-494 (2007).
1108 14. L. Mannini *et al.*, SMC1A codon 496 mutations affect the cellular response to genotoxic
1109 treatments. *Am J Med Genet A* **158A**, 224-228 (2012).
1110 15. M. A. Deardorff *et al.*, HDAC8 mutations in Cornelia de Lange syndrome affect the
1111 cohesin acetylation cycle. *Nature* **489**, 313-317 (2012).
1112 16. M. A. Deardorff *et al.*, RAD21 mutations cause a human cohesinopathy. *Am J Hum Genet*
1113 **90**, 1014-1027 (2012).
1114 17. M. Kaur *et al.*, Precocious sister chromatid separation (PSCS) in Cornelia de Lange
1115 syndrome. *Am J Med Genet A* **138**, 27-31 (2005).

1116 18. S. Kawauchi *et al.*, Multiple organ system defects and transcriptional dysregulation in the
1117 Nipbl(+/-) mouse, a model of Cornelia de Lange Syndrome. *PLoS Genet* **5**, e1000650
1118 (2009).
1119 19. D. Dorsett, I. D. Krantz, On the molecular etiology of Cornelia de Lange syndrome. *Ann N*
1120 *Y Acad Sci* **1151**, 22-37 (2009).
1121 20. R. Chien *et al.*, Cohesin mediates chromatin interactions that regulate mammalian beta-
1122 globin expression. *J Biol Chem* **286**, 17870-17878 (2011).
1123 21. Y. Guo *et al.*, CTCF/cohesin-mediated DNA looping is required for protocadherin alpha
1124 promoter choice. *Proc Natl Acad Sci U S A* **109**, 21081-21086 (2012).
1125 22. A. Muto, A. L. Calof, A. D. Lander, T. F. Schilling, Multifactorial origins of heart and gut
1126 defects in nipbl-deficient zebrafish, a model of Cornelia de Lange Syndrome. *PLoS Biol* **9**,
1127 e1001181 (2011).
1128 23. B. Yuan *et al.*, Global transcriptional disturbances underlie Cornelia de Lange syndrome
1129 and related phenotypes. *J Clin Invest* **125**, 636-651 (2015).
1130 24. R. Santos *et al.*, Conditional Creation and Rescue of Nipbl-Deficiency in Mice Reveals
1131 Multiple Determinants of Risk for Congenital Heart Defects. *PLoS Biol* **14**, e2000197
1132 (2016).
1133 25. M. A. Mole, A. Weberling, M. Zernicka-Goetz, Comparative analysis of human and
1134 mouse development: From zygote to pre-gastrulation. *Curr Top Dev Biol* **136**, 113-138
1135 (2020).
1136 26. A. N. Economides *et al.*, Conditionals by inversion provide a universal method for the
1137 generation of conditional alleles. *Proc Natl Acad Sci U S A* **110**, E3179-3188 (2013).
1138 27. K. M. Downs, T. Davies, Staging of gastrulating mouse embryos by morphological
1139 landmarks in the dissecting microscope. *Development* **118**, 1255-1266 (1993).

1140 28. G. X. Zheng *et al.*, Massively parallel digital transcriptional profiling of single cells. *Nat*
1141 *Commun* **8**, 14049 (2017).
1142 29. M. Ginsburg, M. H. Snow, A. McLaren, Primordial germ cells in the mouse embryo
1143 during gastrulation. *Development* **110**, 521-528 (1990).
1144 30. T. Stuart *et al.*, Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902
1145 e1821 (2019).
1146 31. A. Okuda *et al.*, UTF1, a novel transcriptional coactivator expressed in pluripotent
1147 embryonic stem cells and extra-embryonic cells. *EMBO J* **17**, 2019-2032 (1998).
1148 32. S. Zhang, W. Cui, Sox2, a key factor in the regulation of pluripotency and neural
1149 differentiation. *World J Stem Cells* **6**, 305-311 (2014).
1150 33. Q. Zhu *et al.*, The transcription factor Pou3f1 promotes neural fate commitment via
1151 activation of neural lineage genes and inhibition of external signaling pathways. *Elife* **3**,
1152 (2014).
1153 34. V. Wilson, L. Manson, W. C. Skarnes, R. S. Beddington, The T gene is necessary for
1154 normal mesodermal morphogenetic cell movements during gastrulation. *Development*
1155 **121**, 877-886 (1995).
1156 35. R. M. Barnes, B. A. Firulli, S. J. Conway, J. W. Vincentz, A. B. Firulli, Analysis of the
1157 Hand1 cell lineage reveals novel contributions to cardiovascular, neural crest, extra-
1158 embryonic, and lateral mesoderm derivatives. *Dev Dyn* **239**, 3086-3097 (2010).
1159 36. B. Saykali *et al.*, Distinct mesoderm migration phenotypes in extra-embryonic and
1160 embryonic regions of the early mouse embryo. *Elife* **8**, (2019).
1161 37. H. Bildsoe *et al.*, Transcriptional targets of TWIST1 in the cranial mesoderm regulate cell-
1162 matrix interactions and mesenchyme maintenance. *Dev Biol* **418**, 189-203 (2016).

1163 38. B. Leussink *et al.*, Expression patterns of the paired-related homeobox genes MHox/Prx1
1164 and S8/Prx2 suggest roles in development of the heart and the forebrain. *Mech Dev* **52**, 51-
1165 64 (1995).
1166 39. G. S. Kwon, A. K. Hadjantonakis, Transthyretin mouse transgenes direct RFP expression
1167 or Cre-mediated recombination throughout the visceral endoderm. *Genesis* **47**, 447-455
1168 (2009).
1169 40. L. F. Chu *et al.*, Single-cell RNA-seq reveals novel regulators of human embryonic stem
1170 cell differentiation to definitive endoderm. *Genome Biol* **17**, 173 (2016).
1171 41. I. Burtscher, H. Lickert, Foxa2 regulates polarity and epithelialization in the endoderm
1172 germ layer of the mouse embryo. *Development* **136**, 1029-1038 (2009).
1173 42. W. S. Chen *et al.*, Disruption of the HNF-4 gene, expressed in visceral endoderm, leads to
1174 cell death in embryonic ectoderm and impaired gastrulation of mouse embryos. *Genes Dev*
1175 **8**, 2466-2477 (1994).
1176 43. S. F. Khan *et al.*, The roles and regulation of TBX3 in development and disease. *Gene*
1177 **726**, 144223 (2020).
1178 44. Y. Li, P. Lemaire, R. R. Behringer, Esx1, a novel X chromosome-linked homeobox gene
1179 expressed in mouse extraembryonic tissues and male germ cells. *Dev Biol* **188**, 85-95
1180 (1997).
1181 45. A. Simeone *et al.*, Two vertebrate homeobox genes related to the Drosophila empty
1182 spiracles gene are expressed in the embryonic cerebral cortex. *EMBO J* **11**, 2541-2550
1183 (1992).
1184 46. B. I. Meyer, P. Gruss, Mouse Cdx-1 expression during gastrulation. *Development* **117**,
1185 191-203 (1993).

1186 47. A. Rodrigo Albors, P. A. Halley, K. G. Storey, Lineage tracing of axial progenitors using
1187 Nkx1-2CreER(T2) mice defines their trunk and tail contributions. *Development* **145**,
1188 (2018).
1189 48. Y. Qu *et al.*, Transcriptome and proteome characterization of surface ectoderm cells
1190 differentiated from human iPSCs. *Sci Rep* **6**, 32007 (2016).
1191 49. M. Donnison *et al.*, Loss of the extraembryonic ectoderm in Elf5 mutants leads to defects
1192 in embryonic patterning. *Development* **132**, 2299-2308 (2005).
1193 50. N. Nishioka *et al.*, Tead4 is required for specification of trophectoderm in pre-
1194 implantation mouse embryos. *Mech Dev* **125**, 270-283 (2008).
1195 51. B. Pijuan-Sala *et al.*, A single-cell molecular map of mouse gastrulation and early
1196 organogenesis. *Nature* **566**, 490-495 (2019).
1197 52. G. Gestri *et al.*, Six3 functions in anterior neural plate specification by promoting cell
1198 proliferation and inhibiting Bmp4 expression. *Development* **132**, 2401-2413 (2005).
1199 53. C. L. Andoniadou *et al.*, Lack of the murine homeobox gene Hesx1 leads to a posterior
1200 transformation of the anterior forebrain. *Development* **134**, 1499-1508 (2007).
1201 54. M. T. Alves dos Santos, M. P. Smidt, En1 and Wnt signaling in midbrain dopaminergic
1202 neuronal development. *Neural Dev* **6**, 23 (2011).
1203 55. D. H. Rowitch, A. P. McMahon, Pax-2 expression in the murine neural plate precedes and
1204 encompasses the expression domains of Wnt-1 and En-1. *Mech Dev* **52**, 3-8 (1995).
1205 56. A. Gavalas, C. Ruhrberg, J. Livet, C. E. Henderson, R. Krumlauf, Neuronal defects in the
1206 hindbrain of Hoxa1, Hoxb1 and Hoxb2 mutants reflect regulatory interactions among
1207 these Hox genes. *Development* **130**, 5663-5679 (2003).
1208 57. C. S. Hong, J. P. Saint-Jeannet, The b-HLH transcription factor Hes3 participates in neural
1209 plate border formation by interfering with Wnt/beta-catenin signaling. *Dev Biol* **442**, 162-
1210 172 (2018).

1211 58. G. Cho, Y. Lim, I. T. Cho, J. C. Simonet, J. A. Golden, Arx together with FoxA2,
1212 regulates Shh floor plate expression. *Dev Biol* **393**, 137-148 (2014).
1213 59. A. Holz *et al.*, The transcription factors Nkx2.2 and Nkx2.9 play a novel role in floor plate
1214 development and commissural axon guidance. *Development* **137**, 4249-4260 (2010).
1215 60. C. Papanayotou *et al.*, A mechanism regulating the onset of Sox2 expression in the
1216 embryonic neural plate. *PLoS Biol* **6**, e2 (2008).
1217 61. J. D. Lee, N. F. Silva-Gagliardi, U. Tepass, C. J. McGlade, K. V. Anderson, The FERM
1218 protein Epb4.1l5 is required for organization of the neural plate and for the epithelial-
1219 mesenchymal transition at the primitive streak of the mouse embryo. *Development* **134**,
1220 2007-2016 (2007).
1221 62. C. Milet, F. Maczkowiak, D. D. Roche, A. H. Monsoro-Burq, Pax3 and Zic1 drive
1222 induction and differentiation of multipotent, migratory, and functional neural crest in
1223 Xenopus embryos. *Proc Natl Acad Sci U S A* **110**, 5528-5533 (2013).
1224 63. L. Teng, N. A. Mundell, A. Y. Frist, Q. Wang, P. A. Labosky, Requirement for Foxd3 in
1225 the maintenance of neural crest progenitors. *Development* **135**, 1615-1624 (2008).
1226 64. S. S. Chan *et al.*, Mesp1 patterns mesoderm into cardiac, hematopoietic, or skeletal
1227 myogenic progenitors in a context-dependent manner. *Cell Stem Cell* **12**, 587-601 (2013).
1228 65. M. R. Jongbloed *et al.*, Expression of Id2 in the second heart field and cardiac defects in
1229 Id2 knock-out mice. *Dev Dyn* **240**, 2561-2577 (2011).
1230 66. E. Dodou, M. P. Verzi, J. P. Anderson, S. M. Xu, B. L. Black, Mef2c is a direct
1231 transcriptional target of ISL1 and GATA factors in the anterior heart field during mouse
1232 embryonic development. *Development* **131**, 3931-3942 (2004).
1233 67. T. Sadahiro *et al.*, Tbx6 Induces Nascent Mesoderm from Pluripotent Stem Cells and
1234 Temporally Controls Cardiac versus Somite Lineage Diversification. *Cell Stem Cell* **23**,
1235 382-395 e385 (2018).

1236 68. F. Cecconi, G. Proetzel, G. Alvarez-Bolado, D. Jay, P. Gruss, Expression of Meis2, a
1237 Knotted-related murine homeobox gene, indicates a role in the differentiation of the
1238 forebrain and the somitic mesoderm. *Dev Dyn* **210**, 184-190 (1997).
1239 69. B. Wilm, R. G. James, T. M. Schultheiss, B. L. Hogan, The forkhead genes, Foxc1 and
1240 Foxc2, regulate paraxial versus intermediate mesoderm cell fate. *Dev Biol* **271**, 176-189
1241 (2004).
1242 70. O. R. Davies *et al.*, Tcf15 primes pluripotent cells for differentiation. *Cell Rep* **3**, 472-484
1243 (2013).
1244 71. J. M. Khor, C. A. Ettensohn, Transcription Factors of the Alx Family: Evolutionarily
1245 Conserved Regulators of Deuterostome Skeletogenesis. *Front Genet* **11**, 569314 (2020).
1246 72. M. Gering, Y. Yamada, T. H. Rabbitts, R. K. Patient, Lmo2 and Scl/Tal1 convert non-
1247 axial mesoderm into haemangioblasts which differentiate into endothelial cells in the
1248 absence of Gata1. *Development* **130**, 6187-6199 (2003).
1249 73. N. Koyano-Nakagawa, D. J. Garry, Etv2 as an essential regulator of mesodermal lineage
1250 development. *Cardiovasc Res* **113**, 1294-1306 (2017).
1251 74. K. Briegel *et al.*, Regulation and function of transcription factor GATA-1 during red blood
1252 cell differentiation. *Development* **122**, 3839-3850 (1996).
1253 75. A. Beckers, L. Alten, C. Viebahn, P. Andre, A. Gossler, The mouse homeobox gene Noto
1254 regulates node morphogenesis, notochordal ciliogenesis, and left right patterning. *Proc*
1255 *Natl Acad Sci U S A* **104**, 15765-15770 (2007).
1256 76. J. L. Stubbs, I. Oishi, J. C. Izpisua Belmonte, C. Kintner, The forkhead protein Foxj1
1257 specifies node-like cilia in Xenopus and zebrafish embryos. *Nat Genet* **40**, 1454-1460
1258 (2008).

1259 77. J. Schemmer *et al.*, Transcription factor TFAP2C regulates major programs required for
1260 murine fetal germ cell maintenance and haploinsufficiency predisposes to teratomas in
1261 male mice. *PLoS One* **8**, e71113 (2013).
1262 78. J. Sun, M. C. Ting, M. Ishii, R. Maxson, Msx1 and Msx2 function together in the
1263 regulation of primordial germ cell migration in the mouse. *Dev Biol* **417**, 11-24 (2016).
1264 79. E. L. Maska *et al.*, A Tlx2-Cre mouse line uncovers essential roles for hand1 in
1265 extraembryonic and lateral mesoderm. *Genesis* **48**, 479-484 (2010).
1266 80. C. A. Desjardins, F. J. Naya, The Function of the MEF2 Family of Transcription Factors
1267 in Cardiac Development, Cardiogenomics, and Direct Reprogramming. *J Cardiovasc Dev*
1268 *Dis* **3**, (2016).
1269 81. V. Serpooshan *et al.*, Nkx2.5+ Cardiomyoblasts Contribute to Cardiomyogenesis in the
1270 Neonatal Heart. *Sci Rep* **7**, 12590 (2017).
1271 82. T. Jukkola *et al.*, Meox1Cre: a mouse line expressing Cre recombinase in somitic
1272 mesoderm. *Genesis* **43**, 148-153 (2005).
1273 83. J. A. Farrell *et al.*, Single-cell reconstruction of developmental trajectories during
1274 zebrafish embryogenesis. *Science* **360**, (2018).
1275 84. V. Bergen, M. Lange, S. Peidli, F. A. Wolf, F. J. Theis, Generalizing RNA velocity to
1276 transient cell states through dynamical modeling. *Nat Biotechnol* **38**, 1408-1414 (2020).
1277 85. Y. Saga *et al.*, MesP1 is expressed in the heart precursor cells and required for the
1278 formation of a single heart tube. *Development* **126**, 3437-3447 (1999).
1279 86. M. Lange *et al.*, CellRank for directed single-cell fate mapping. *Nat Methods* **19**, 159-170
1280 (2022).
1281 87. A. Liberzon *et al.*, The Molecular Signatures Database (MSigDB) hallmark gene set
1282 collection. *Cell Syst* **1**, 417-425 (2015).

1283 88. M. Milacic *et al.*, The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Res* **52**,
1284 D672-D678 (2024).
1285 89. S. M. Langemeijer *et al.*, Apoptosis-Related Gene Expression Profiling in Hematopoietic
1286 Cell Fractions of MDS Patients. *PLoS One* **11**, e0165582 (2016).
1287 90. D. Venet, J. E. Dumont, V. Detours, Most random gene expression signatures are
1288 significantly associated with breast cancer outcome. *PLoS Comput Biol* **7**, e1002240
1289 (2011).
1290 91. S. Nestorowa *et al.*, A single-cell resolution map of mouse hematopoietic stem and
1291 progenitor cell differentiation. *Blood* **128**, e20-31 (2016).
1292 92. Y. Chen *et al.*, The Role of Tbx20 in Cardiovascular Development and Function. *Front*
1293 *Cell Dev Biol* **9**, 638542 (2021).
1294 93. D. W. Sorensen, J. H. van Berlo, The Role of TGF-beta Signaling in Cardiomyocyte
1295 Proliferation. *Curr Heart Fail Rep* **17**, 225-233 (2020).
1296 94. C. Guibentif *et al.*, Diverse Routes toward Early Somites in the Mouse Embryo. *Dev Cell*
1297 **56**, 141-153 e146 (2021).
1298 95. D. T. Farmer, P. Patel, R. Choi, C. Y. Liu, J. G. Crump, A comprehensive series of Irx
1299 cluster mutants reveals diverse roles in facial cartilage development. *Development* **148**,
1300 (2021).
1301 96. H. Preiss *et al.*, Regulation of Nodal signaling propagation by receptor interactions and
1302 positive feedback. *Elife* **11**, (2022).
1303 97. S. Nowotschin *et al.*, The T-box transcription factor Eomesodermin is essential for AVE
1304 induction in the mouse embryo. *Genes Dev* **27**, 997-1002 (2013).
1305 98. C. Meno *et al.*, Diffusion of nodal signaling activity in the absence of the feedback
1306 inhibitor Lefty2. *Dev Cell* **1**, 127-138 (2001).

1307 99. I. Costello *et al.*, Lhx1 functions together with Otx2, Foxa2, and Ldb1 to govern anterior
1308 mesendoderm, node, and midline development. *Genes Dev* **29**, 2108-2122 (2015).
1309 100. M. Siatecka, J. J. Bieker, The multifunctional role of EKLF/KLF1 during erythropoiesis.
1310 *Blood* **118**, 2044-2054 (2011).
1311 101. W. K. Shen *et al.*, AnimalTFDB 4.0: a comprehensive animal transcription factor database
1312 updated with variation and expression annotations. *Nucleic Acids Res* **51**, D39-D45
1313 (2023).
1314 102. S. J. Bray, Notch signalling: a simple pathway becomes complex. *Nat Rev Mol Cell Biol*
1315 **7**, 678-689 (2006).
1316 103. R. Macdonald *et al.*, Midline signalling is required for Pax gene regulation and patterning
1317 of the eyes. *Development* **121**, 3267-3278 (1995).
1318 104. J. Park *et al.*, Regulation of Sox9 by Sonic Hedgehog (Shh) is essential for patterning and
1319 formation of tracheal cartilage. *Dev Dyn* **239**, 514-526 (2010).
1320 105. P. H. Su *et al.*, NKX6-1 mediates cancer stem-like properties and regulates sonic
1321 hedgehog signaling in leiomyosarcoma. *J Biomed Sci* **28**, 32 (2021).
1322 106. C. Cai, J. Thorne, L. Grabel, Hedgehog serves as a mitogen and survival factor during
1323 embryonic stem cell neurogenesis. *Stem Cells* **26**, 1097-1108 (2008).
1324 107. S. W. Choy *et al.*, A cascade of irx1a and irx2a controls shh expression during
1325 retinogenesis. *Dev Dyn* **239**, 3204-3214 (2010).
1326 108. D. Szklarczyk *et al.*, The STRING database in 2023: protein-protein association networks
1327 and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids*
1328 *Res* **51**, D638-D646 (2023).
1329 109. K. Mitsui *et al.*, The homeoprotein Nanog is required for maintenance of pluripotency in
1330 mouse epiblast and ES cells. *Cell* **113**, 631-642 (2003).

1331 110. A. H. Hart, L. Hartley, M. Ibrahim, L. Robb, Identification, cloning and expression
1332 analysis of the pluripotency promoting Nanog genes in mouse and human. *Dev Dyn* **230**,
1333 187-198 (2004).
1334 111. M. Zhang *et al.*, Esrrb Complementation Rescues Development of Nanog-Null Germ
1335 Cells. *Cell Rep* **22**, 332-339 (2018).
1336 112. G. Shi, Y. Jin, Role of Oct4 in maintaining and regaining stem cell pluripotency. *Stem*
1337 *Cell Res Ther* **1**, 39 (2010).
1338 113. M. Tiana *et al.*, Pluripotency factors regulate the onset of Hox cluster activation in the
1339 early embryo. *Sci Adv* **8**, eabo3583 (2022).
1340 114. M. Mossahebi-Mohammadi, M. Quan, J. S. Zhang, X. Li, FGF Signaling Pathway: A Key
1341 Regulator of Stem Cell Pluripotency. *Front Cell Dev Biol* **8**, 79 (2020).
1342 115. L. L. Chang, D. S. Kessler, Foxd3 is an essential Nodal-dependent regulator of zebrafish
1343 dorsal mesoderm development. *Dev Biol* **342**, 39-50 (2010).
1344 116. M. M. Shen, Nodal signaling: developmental roles and regulation. *Development* **134**,
1345 1023-1034 (2007).
1346 117. C. Porcher, H. Chagraoui, M. S. Kristiansen, SCL/TAL1: a multifaceted regulator from
1347 blood development to disease. *Blood* **129**, 2051-2060 (2017).
1348 118. V. S. Stanulovic, P. Cauchy, S. A. Assi, M. Hoogenkamp, LMO2 is required for TAL1
1349 DNA binding activity and initiation of definitive haematopoiesis at the haemangioblast
1350 stage. *Nucleic Acids Res* **45**, 9874-9888 (2017).
1351 119. Z. Avsec *et al.*, Base-resolution models of transcription-factor binding reveal soft motif
1352 syntax. *Nat Genet* **53**, 354-366 (2021).
1353 120. K. A. Hubert, D. M. Wellik, Hox genes in development and beyond. *Development* **150**,
1354 (2023).

1355 121. O. Medina-Martinez, A. Bradley, R. Ramirez-Solis, A large targeted deletion of Hoxb1-
1356 Hoxb9 produces a series of single-segment anterior homeotic transformations. *Dev Biol*
1357 **222**, 71-83 (2000).
1358 122. M. I. Boyle, C. Jespersgaard, K. Brondum-Nielsen, A. M. Bisgaard, Z. Tumer, Cornelia de
1359 Lange syndrome. *Clin Genet* **88**, 1-12 (2015).
1360 123. S. Kawauchi *et al.*, Using mouse and zebrafish models to understand the etiology of
1361 developmental defects in Cornelia de Lange Syndrome. *Am J Med Genet C Semin Med*
1362 *Genet* **172**, 138-145 (2016).
1363 124. M. P. Lambert *et al.*, The incidence of thrombocytopenia in children with Cornelia de
1364 Lange syndrome. *Am J Med Genet A* **155A**, 33-37 (2011).
1365 125. A. Ayerza Casas *et al.*, Cornelia de Lange syndrome: Congenital heart disease in 149
1366 patients. *Med Clin (Barc)* **149**, 300-302 (2017).
1367 126. M. E. Lopez-Burks, R. Santos, S. Kawauchi, A. L. Calof, A. D. Lander, Genetic
1368 enhancement of limb defects in a mouse model of Cornelia de Lange syndrome. *Am J Med*
1369 *Genet C Semin Med Genet* **172**, 146-154 (2016).
1370 127. D. Mehta *et al.*, Characterization of limb differences in children with Cornelia de Lange
1371 Syndrome. *Am J Med Genet C Semin Med Genet* **172**, 155-162 (2016).
1372 128. C. Cortes, A. Francou, C. De Bono, R. G. Kelly, Epithelial Properties of the Second Heart
1373 Field. *Circ Res* **122**, 142-154 (2018).
1374 129. A. von Gise, W. T. Pu, Endocardial and epicardial epithelial to mesenchymal transitions in
1375 heart development and disease. *Circ Res* **110**, 1628-1645 (2012).
1376 130. R. M. George, G. Maldonado-Velez, A. B. Firulli, The heart of the neural crest: cardiac
1377 neural crest cells in development and regeneration. *Development* **147**, (2020).
1378 131. A. Muto *et al.*, Nipbl and mediator cooperatively regulate gene expression to control limb
1379 development. *PLoS Genet* **10**, e1004671 (2014).

1380 132. F. A. Lee, F. M. Kenny, Skeletal changes in the Cornelia de Lange syndrome. *Am J*
1381 *Roentgenol Radium Ther Nucl Med* **100**, 27-39 (1967).
1382 133. J. Brennan, D. P. Norris, E. J. Robertson, Nodal activity in the node governs left-right
1383 asymmetry. *Genes Dev* **16**, 2339-2344 (2002).
1384 134. B. Mohapatra *et al.*, Identification and functional characterization of NODAL rare variants
1385 in heterotaxy and isolated cardiovascular malformations. *Hum Mol Genet* **18**, 861-871
1386 (2009).
1387 135. K. C. Chatfield *et al.*, Congenital heart disease in Cornelia de Lange syndrome: phenotype
1388 and genotype analysis. *Am J Med Genet A* **158A**, 2499-2505 (2012).
1389 136. T. H. Kim *et al.*, Analysis of the vertebrate insulator protein CTCF-binding sites in the
1390 human genome. *Cell* **128**, 1231-1245 (2007).
1391 137. W. Schwarzer *et al.*, Two independent modes of chromatin organization revealed by
1392 cohesin removal. *Nature* **551**, 51-56 (2017).
1393 138. A. S. Hansen, C. Cattoglio, X. Darzacq, R. Tjian, Recent evidence that TADs and
1394 chromatin loops are dynamic structures. *Nucleus* **9**, 20-32 (2018).
1395 139. M. Panarotto, I. F. Davidson, G. Litos, A. Schleiffer, J. M. Peters, Cornelia de Lange
1396 syndrome mutations in NIPBL can impair cohesin-mediated DNA loop extrusion. *Proc*
1397 *Natl Acad Sci U S A* **119**, e2201029119 (2022).
1398 140. D. Hnisz *et al.*, Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-
1399 947 (2013).
1400 141. D. A. Newkirk *et al.*, The effect of Nipped-B-like (Nipbl) haploinsufficiency on genome-
1401 wide cohesin binding and target gene expression: modeling Cornelia de Lange syndrome.
1402 *Clin Epigenetics* **9**, 89 (2017).
1403 142. S. Van Der Lee, L. M. Boot, Spontaneous pseudopregnancy in mice. *Acta Physiol*
1404 *Pharmacol Neerl* **4**, 442-444 (1955).

1405 143. W. K. Whitten, Modification of the oestrous cycle of the mouse by external stimuli
1406 associated with the male. *J Endocrinol* **13**, 399-404 (1956).
1407 144. K. Shea, N. Geijsen, Dissection of 6.5 dpc mouse embryos. *J Vis Exp*, 160 (2007).
1408 145. S. Das *et al.*, The Drosophila T-box transcription factor Midline functions within the
1409 Notch-Delta signaling pathway to specify sensory organ precursor cell fates and regulates
1410 cell survival within the eye imaginal disc. *Mech Dev* **130**, 577-601 (2013).
1411 146. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21
1412 (2013).
1413 147. A. T. L. Lun *et al.*, EmptyDrops: distinguishing cells from empty droplets in droplet-based
1414 single-cell RNA sequencing data. *Genome Biol* **20**, 63 (2019).
1415 148. T. Ilicic *et al.*, Classification of low quality cells from single-cell RNA-seq data. *Genome*
1416 *Biol* **17**, 29 (2016).
1417 149. C. Hafemeister, R. Satija, Normalization and variance stabilization of single-cell RNA-seq
1418 data using regularized negative binomial regression. *Genome Biol* **20**, 296 (2019).
1419 150. G. La Manno *et al.*, RNA velocity of single cells. *Nature* **560**, 494-498 (2018).
1420 151. T. Wu *et al.*, clusterProfiler 4.0: A universal enrichment tool for interpreting omics data.
1421 *Innovation (Camb)* **2**, 100141 (2021).
1422 152. G. Korotkevich *et al.*, Fast gene set enrichment analysis. *bioRxiv*, 060012 (2021).

1423 **Acknowledgments**

1424 The authors are grateful to Henry Chang, Lorelle Meidt, Han Nguyen, Arianna Favela,
1425 Christopher Wu, Louise Villagomez, and Brian Bui for assistance with mouse husbandry,
1426 genotyping, and data analysis; to Melanie Oakes of the UCI Genomics and Research Technology
1427 Hub (GRTH) for 10x library preparation and sequencing; and to Qing Nie (UCI Depts of
1428 Mathematics and Developmental & Cell Biology) and Rosaysela Santos (Kaiser Permanente
1429 Bernard J. Tyson School of Medicine) for helpful discussions. The authors are grateful to the UCI

Center for Complex Biological Systems and the NSF-Simons Center for Multiscale Cell Fate Research for support of early pilot studies in this project. 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)

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, ALM

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

1455 **Data and materials availability:** All raw and processed data are available in the main text,
1456 Sequence Read Archive (Accession: SRP265522) and Gene expression Omnibus
1457 (Accession: GSE151589). All Supplementary Figures and Tables cited in the main text have been
1458 uploaded under Supplementary Material as a single PDF file. All Supplementary Data cited in the
1459 main text have been uploaded under Auxiliary Supplementary Materials as a single ZIP file. No
1460 restrictions have been made on the availability of the data. All code is available in Zenodo
1461 (DOI: 10.5281/zenodo.10001570). *Nipbl*$^{Flox/+}$ mice will be provided by ALC and ADL pending
1462 scientific review and a completed Material Transfers Agreement.

1463 **Figures and Tables**

1464

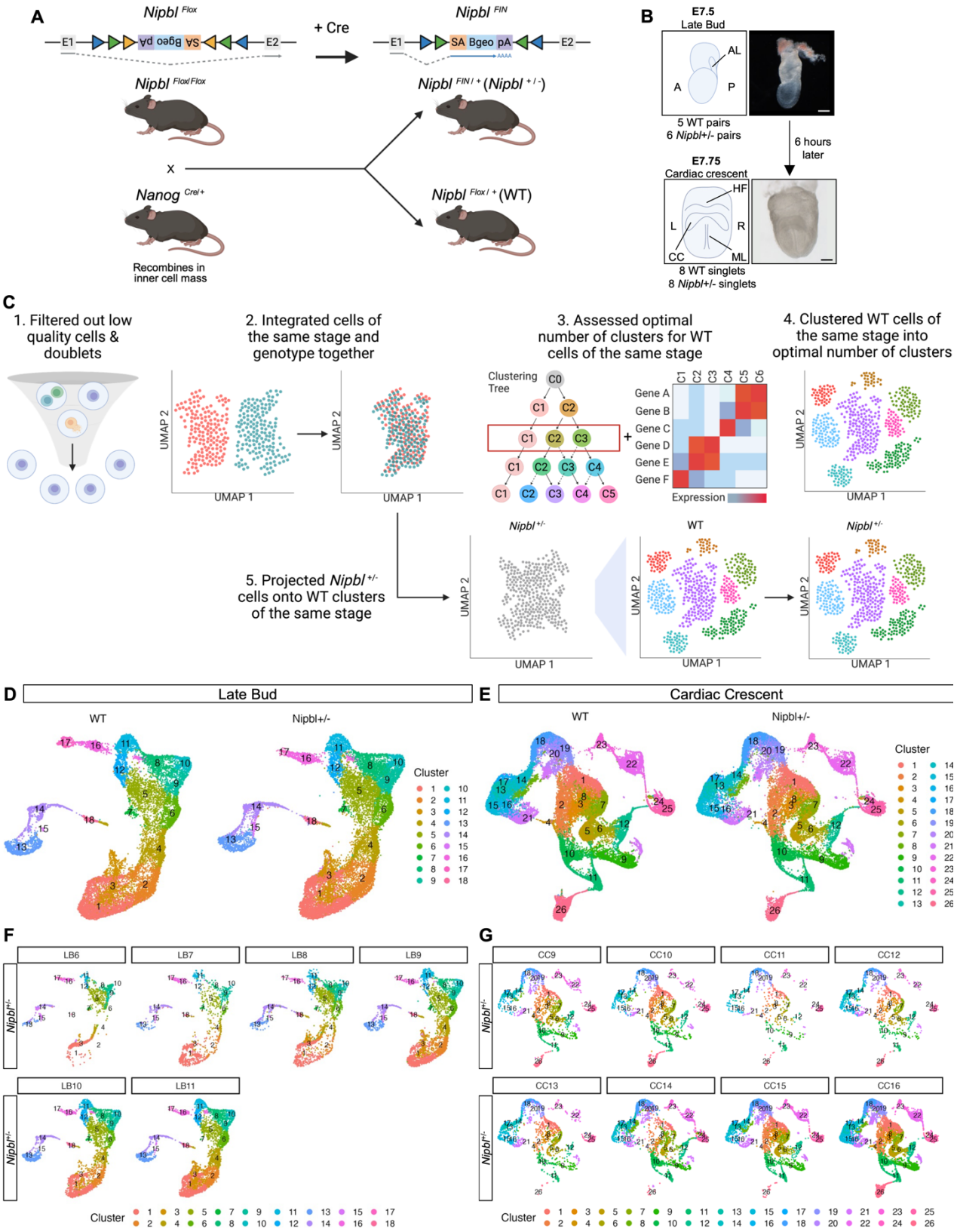


Fig. 1. *Nipbl*$^{+/-}$ mice do not lack any cell populations found in WT mice. (**A**) *Nipbl* alleles used in this study. *Nipbl*^{Flox} contains an inverted gene trap cassette encoding *β–geo* flanked by *Cre* recombinase target sites in intron 1 of *Nipbl* gene alleles (*24*). In this inverted orientation, there is no trapping of the *Nipbl* gene and *Nipbl* is expressed normally. However, when this cassette is exposed to Cre recombinase, the gene trap cassette gets inverted producing the *Nipbl*^{FIN} allele. In

this orientation, trapping of the *Nipbl* gene occurs and *β–geo* is expressed as a reporter of successful gene trapping. When *Nanog*$^{Cre/+}$ mice are mated with *Nipbl*$^{Flox/Flox}$ mice, the resulting littermates are either entirely *Nipbl*$^{Flox/+}$ or entirely *Nipbl*$^{FIN/+}$, as *Nanog*$^{Cre/+}$ mice carry a transgene that expresses *Cre* recombinase in the earliest cells of the developing embryo (*26*). (**B**) Lateral view of LB-stage and anterior view of CC-stage embryos subjected to scRNAseq. A = anterior, P = posterior, AL = allantois, L = left, R = right, HF = head fold, CC = cardiac crescent, ML = midline. Dashed line represents where embryonic tissue was separated from extraembryonic tissue. Scale bar = 100 microns. (**C**) Workflow used to filter out low quality cells and doublets, cluster WT cells into optimal number of clusters at each stage, and project *Nipbl*$^{+/-}$ cells onto WT clusters of the same stage. Uniform manifold approximation and projection (UMAP) of clusters in WT and *Nipbl*$^{+/-}$ embryos at (**D**) LB- and (**E**) CC-stage. UMAP of clusters in each *Nipbl*$^{+/-}$ embryo at (**F**) LB- and (**G**) CC-stage.

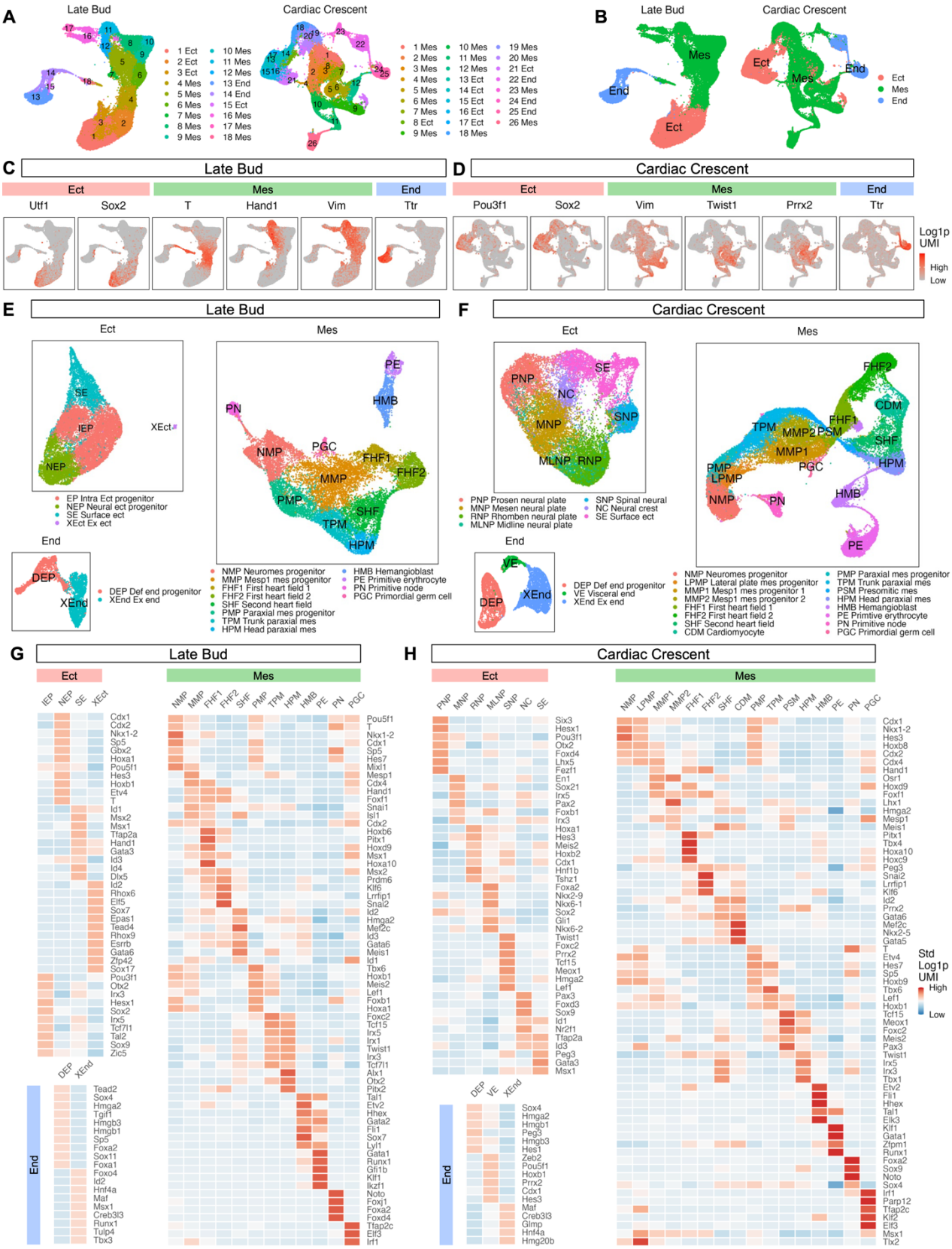
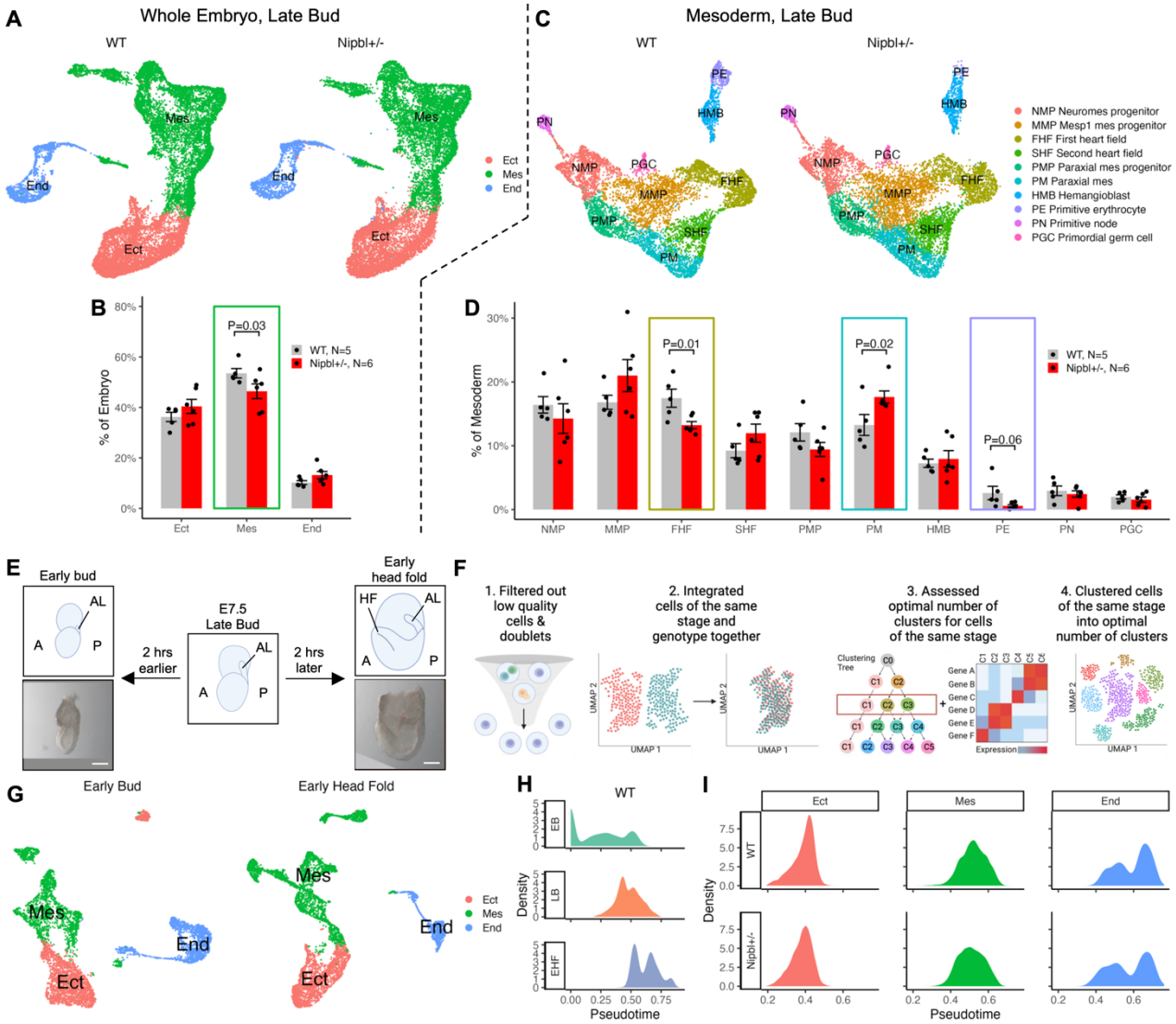


Fig. 2. scRNAseq captured cells from all germ layers, as well as progenitors of major tissues. (**A**) UMAP of clusters assigned to germ layers in LB- and CC-stage embryos. (**B**) UMAP of germ

1487 layers in LB- and CC-stage embryos. Expression of genes marking germ layers of (**C**) LB- and
1488 (**D**) CC-stage embryos in UMAP. UMAP of cell populations in germ layers of (**E**) LB- and (**F**)
1489 CC-stage embryos. Heatmap of fold change in expression of the most differentially expressed
1490 transcription factor genes (lowest *Q*-values from Mann-Whitney U Test) (*101*) of cell populations
1491 from all other cell populations in germ layers of (**G**) LB- and (**H**) CC-stage embryos.
1492



1493

1494 **Fig. 3. *Nipbl*$^{+/-}$ embryos exhibit changes in the sizes of mesodermal subpopulations that are**

1495 **not accompanied by changes in developmental timing.** (**A**) UMAP of germ layers in LB-stage

1496 WT and *Nipbl*$^{+/-}$ embryos. (**B**) Percentage of cells in germ layers from all cells in LB-stage WT

1497 and *Nipbl*$^{+/-}$ embryos. Error bars show standard error of the mean. *P*-values from T-Test. (**C**)

1498 UMAP of cell populations in mesoderm of LB-stage WT and *Nipbl*$^{+/-}$ embryos. (**D**) Percentage of

1499 cells in mesodermal cell populations from all cells in LB-stage WT and *Nipbl*$^{+/-}$ embryos

1500 (bottom). Error bars show SEM. *P*-values from T-Test. (**E**) Lateral view of WT EB- and EHF-

1501 stage embryos subjected to scRNAseq. A = anterior, P = posterior, AL = allantois, L = left, R =

1502 right, HF = head fold. Dashed line represents where embryonic tissue was separated from

1503 extraembryonic tissue. Scale bar = 100 um. (**F**) Workflow used to filter out low quality cells and

1504 doublets and cluster cells at EB- and EHF-stage into optimal number of clusters. (**G**) UMAP of
1505 clusters assigned to germ layers in EB- and EHF-stage embryos. (**H**) Density of cells from WT
1506 EB-, LB-, and EHF-stage embryos along pseudotime (calculated using URD). (**I**) Density of cells
1507 in germ layers of LB-stage WT and *Nipbl*$^{+/-}$ embryos along pseudotime.
1508

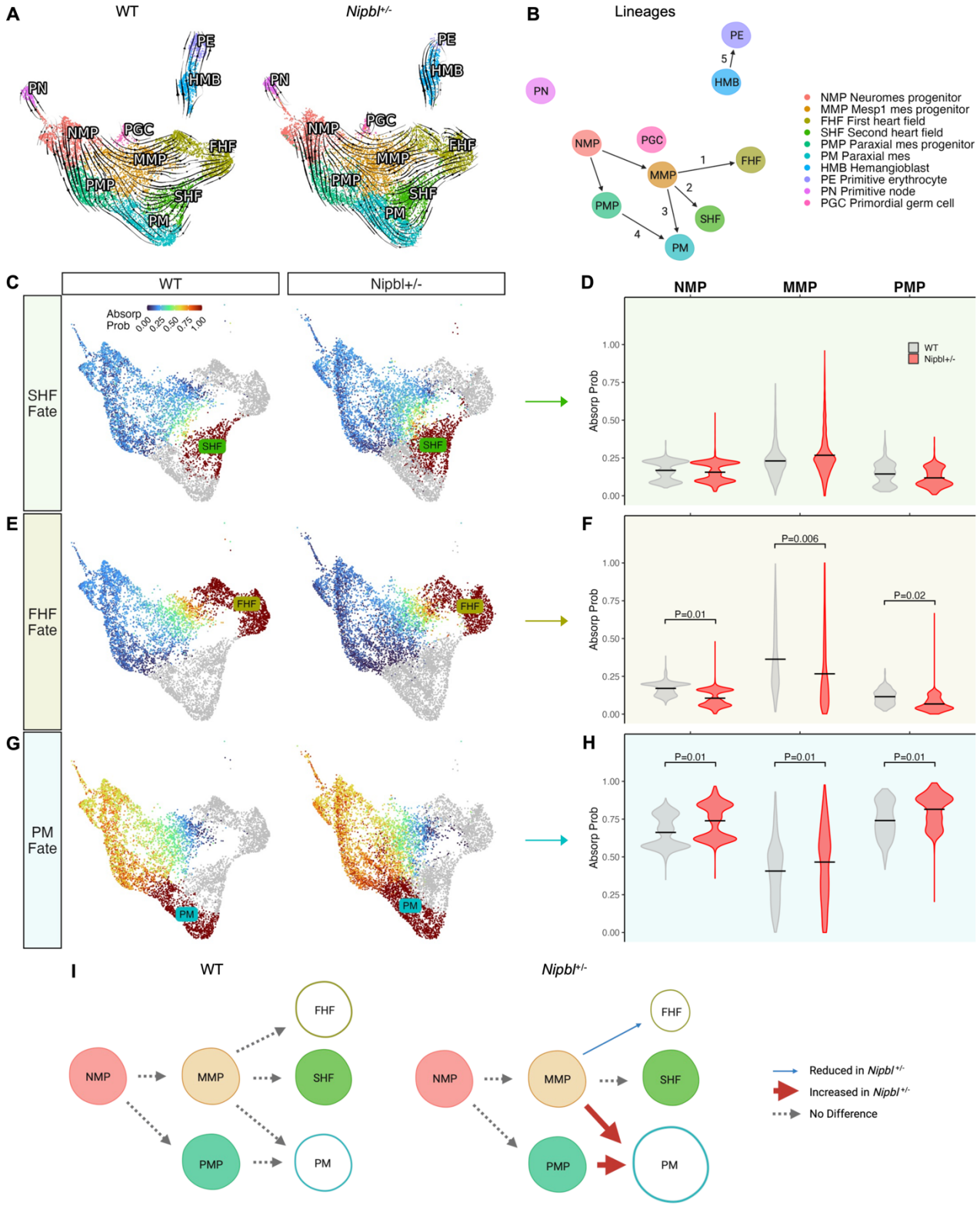


Fig. 4. *Nipbl*$^{+/-}$ mice misdirect mesoderm cells into paraxial mesoderm at the expense of the first heart field. (**A**) RNA velocities, calculated by scVelo of mesoderm cells from LB-stage WT and *Nipbl*$^{+/-}$ embryos in UMAP. (**B**) Cell lineage trajectories in mesoderm of LB-stage WT and

1513 *Nipbl*$^{+/-}$ embryos. Probability of mesoderm cells from LB-stage WT and *Nipbl*$^{+/-}$ embryos
1514 terminally transitioning (absorption probabilities calculated using CellRank) into (**C**) SHF, (**E**)
1515 FHF, or (**G**) PM fates in UMAP. Violin plot of probability of NMPs, MMPs, and PMPs from LB-
1516 stage WT and *Nipbl*$^{+/-}$ embryos terminally transitioning (absorption probabilities) into (**D**) SHF,
1517 (**F**) FHF, or (**H**) PM fates. Lines show means. *P*-values from T-Test. (**I)** Schematic illustrating
1518 how mesoderm cells in *Nipbl*$^{+/-}$ embryos are misdirected into PM fate at the expense of FHF fate.
1519

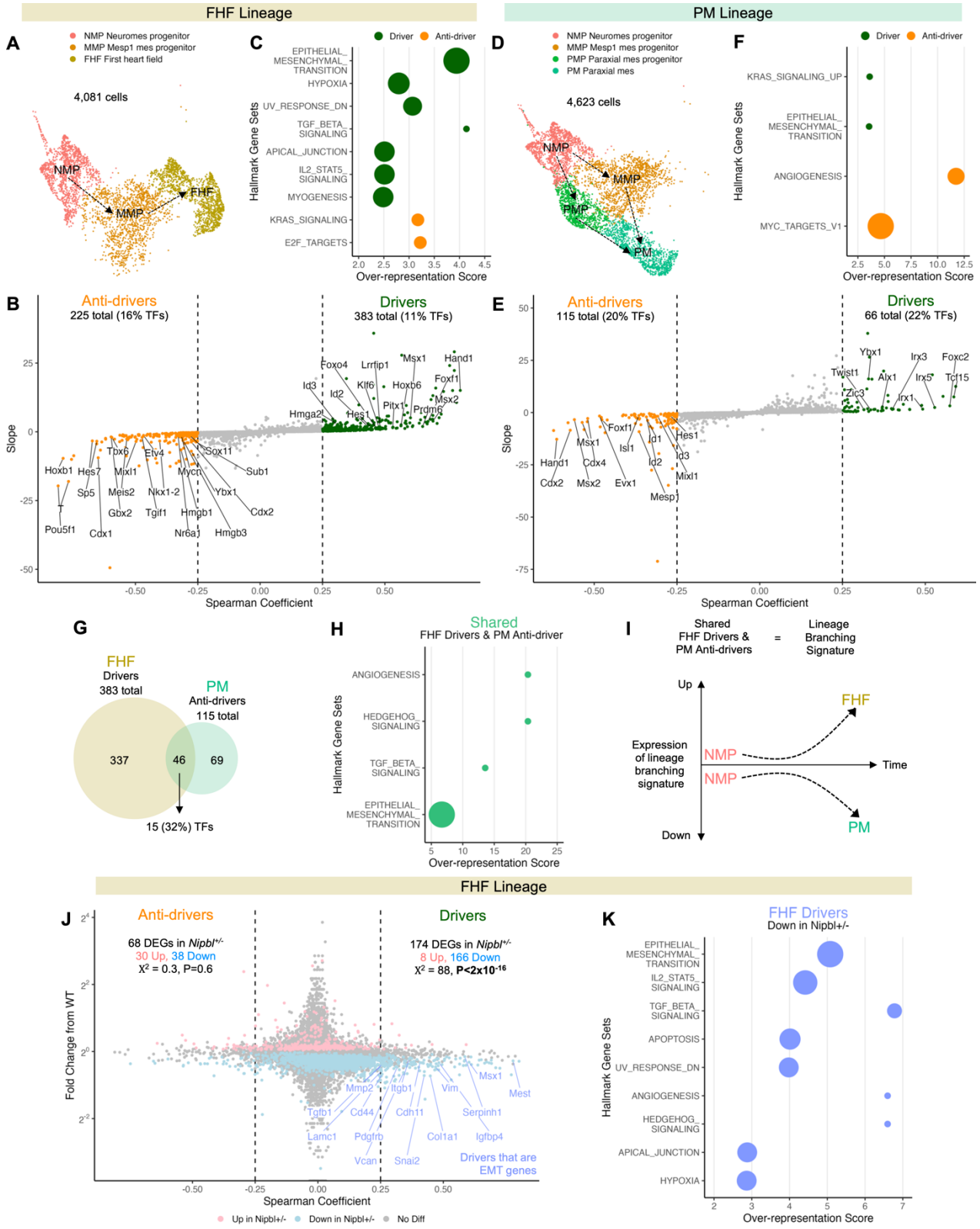


Fig. 5. *Nipbl*$^{+/-}$ mice underexpress genes predicted to drive the transition of mesoderm cells into first heart field. UMAP of mesoderm cells in (**A**) FHF and (**D**) PM lineages of WT LB-stage embryos. Genes whose expression are positively correlated (drivers) or negatively

1524 correlated (anti-drivers) with the transition (absorption probabilities from Fig. 3J) of mesoderm
1525 cells from WT LB-stage embryos into (**B**) FHF or (**E**) PM fates. Correlation coefficients
1526 calculated using Spearman's Rank Correlation. Genes with correlation coefficient greater than
1527 0.25 or less than -0.25 were considered drivers and anti-drivers, respectively. Slope refers to
1528 change in gene expression along absorption probability. Over-representation score of MSigDB's
1529 Hallmark gene sets among drivers and anti-drivers of (**C**) FHF and (**F**) PM fate. (**G**) Venn
1530 diagram of genes shared between PM anti-drivers and FHF drivers. (**H**) Over-representation score
1531 of MSigDB's Hallmark gene sets among shared FHF drivers and PM anti-drivers. (**I**) Expression
1532 of lineage branching signature as NMPs transition into FHF or PM fates. (**J**) Fold change in
1533 expression of genes in mesoderm cells of FHF lineage of LB-stage *Nipbl*$^{+/-}$ embryos differentially
1534 expressed ($Q < 0.05$, Mann-Whitney U Test) from that of WT embryos along Spearman's Rank
1535 Correlation coefficient from Fig. 5B. Genes associated with EMT are labeled. (**K**) Over-
1536 representation score of MSigDB's Hallmark gene sets among shared FHF drivers downregulated
1537 in FHF lineage of *Nipbl*$^{+/-}$ mice.

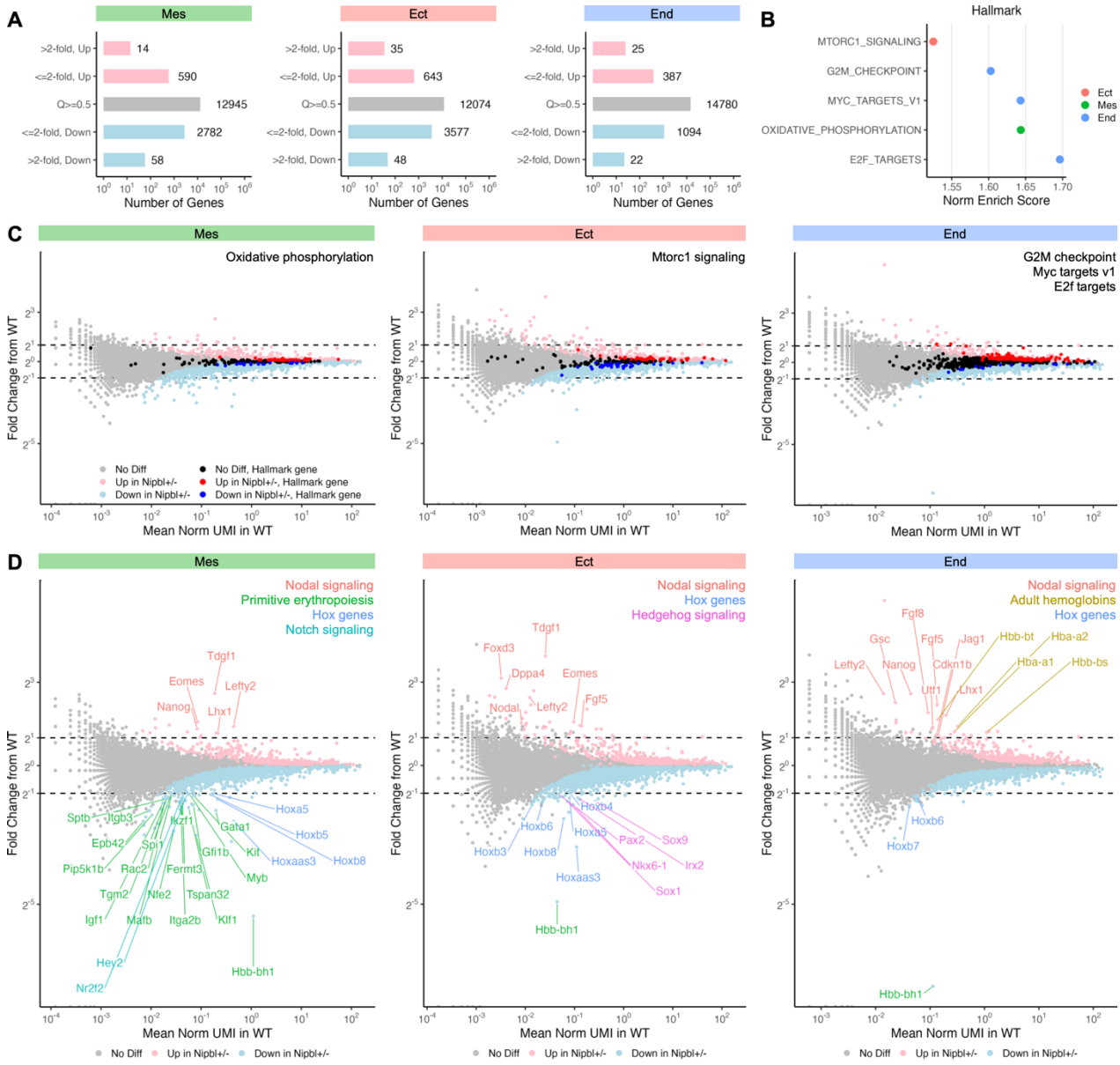


Fig. 6. *Nipbl*$^{+/-}$ mice show large changes in the expression of major developmental regulators in all germ layers. (**A**) Number of differentially expressed genes ($Q < 0.05$, Mann-Whitney U Test) in germ layers of LB-stage *Nipbl*$^{+/-}$ embryos showing small (≤ 2-fold) or large (> 2-fold) changes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

1547 0.05, Mann-Whitney U Test) in germ layers of LB-stage *Nipbl*$^{+/-}$ embryos showing large changes

1548 in expression (> 2-fold) from that of WT embryos along their average expression in WT

1549 embryos.

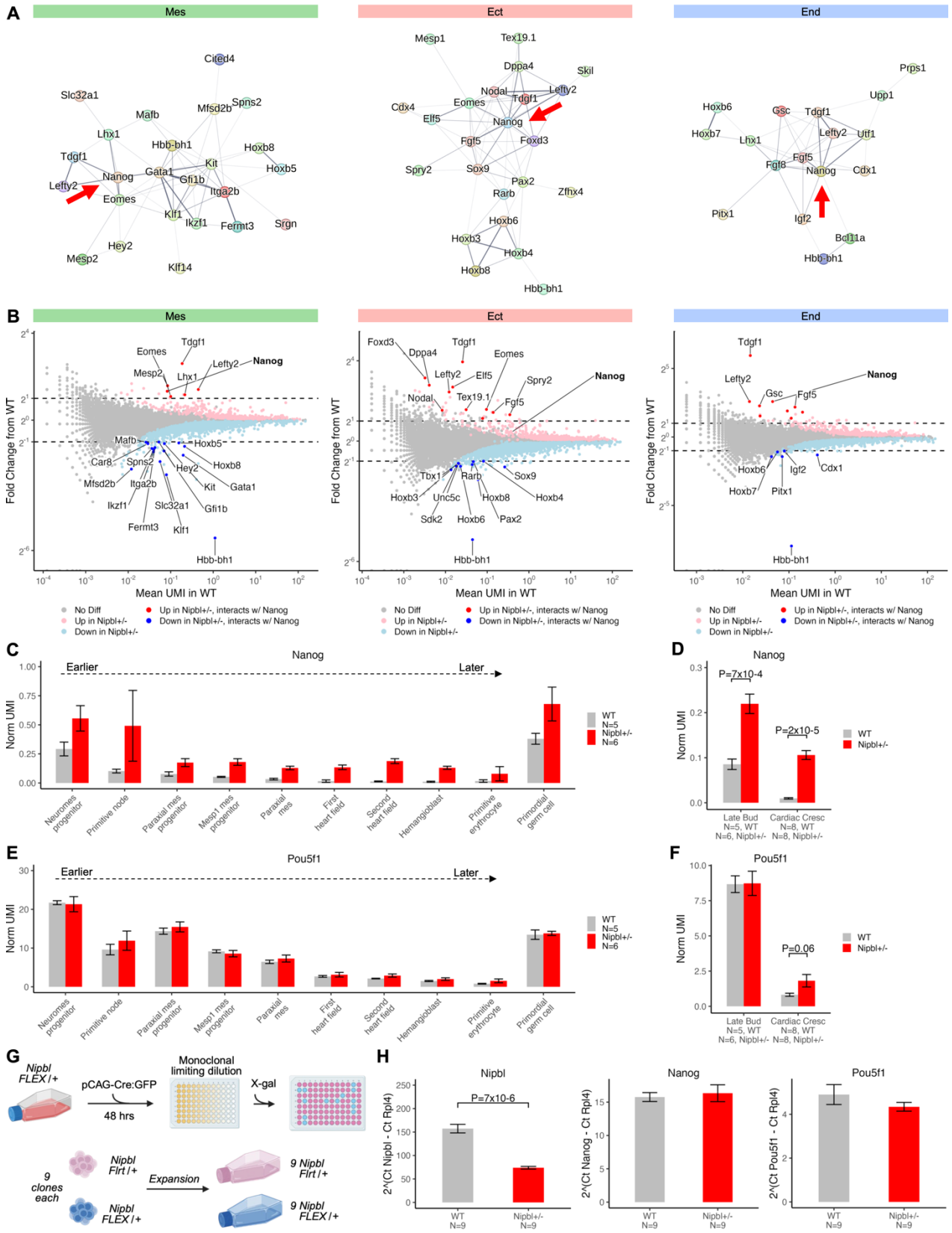


Fig. 7. *Nipbl*$^{+/-}$ mice overexpress *Nanog* during and after gastrulation. (**A**) Differentially expressed genes (greater than 2-fold upregulated or downregulated) between the germ layers of

1553 WT and *Nipbl*$^{+/-}$ embryos predicted by STRING to interact with each other. (**B**) Fold change in
1554 expression of differentially expressed genes ($Q < 0.05$, Mann-Whitney U Test) in germ layers of
1555 LB-stage *Nipbl*$^{+/-}$ embryos showing large changes in expression (> 2-fold) from that of WT
1556 embryos along their average expression in WT embryos that are predicted by STRING to interact
1557 with *Nanog*. (**C**) Expression of *Nanog* in mesodermal cell populations of LB-stage WT and
1558 *Nipbl*$^{+/-}$ embryos ordered from left (earlier) to right (later) by their RNA velocity positions in Fig.
1559 4A. Error bars show SEM. (**D**) Expression of *Nanog* in WT and *Nipbl*$^{+/-}$ embryos from LB- to
1560 CC-stage. Error bars show SEM. *P*-values from T-Test. € Expression of *Pou5f1* in mesodermal
1561 cell populations of LB-stage WT and *Nipbl*$^{+/-}$ embryos ordered from left (earlier) to right (later)
1562 by their RNA velocity positions in Fig. 4A. Error bars show SEM. (**F**) Expression of *Pou5f1* in
1563 WT and *Nipbl*$^{+/-}$ embryos from LB- to CC-stage. Error bars show SEM. *P*-values from T-Test.
1564 (**G**) Monoclonal generation of *Nipbl*$^{Flrt/+}$ (WT) and *Nipbl*$^{Flex/+}$ (*Nipbl*$^{+/-}$) ESCs. (**H**) Expression of
1565 *Nipbl*, *Nanog*, and *Pou5f1* in WT and *Nipbl*$^{+/-}$ embryonic stem cells as measured by RT-qPCR and
1566 normalized to the housekeeping gene, *Rpl4*. Error bars show SEM. *P*-values from T-Test.

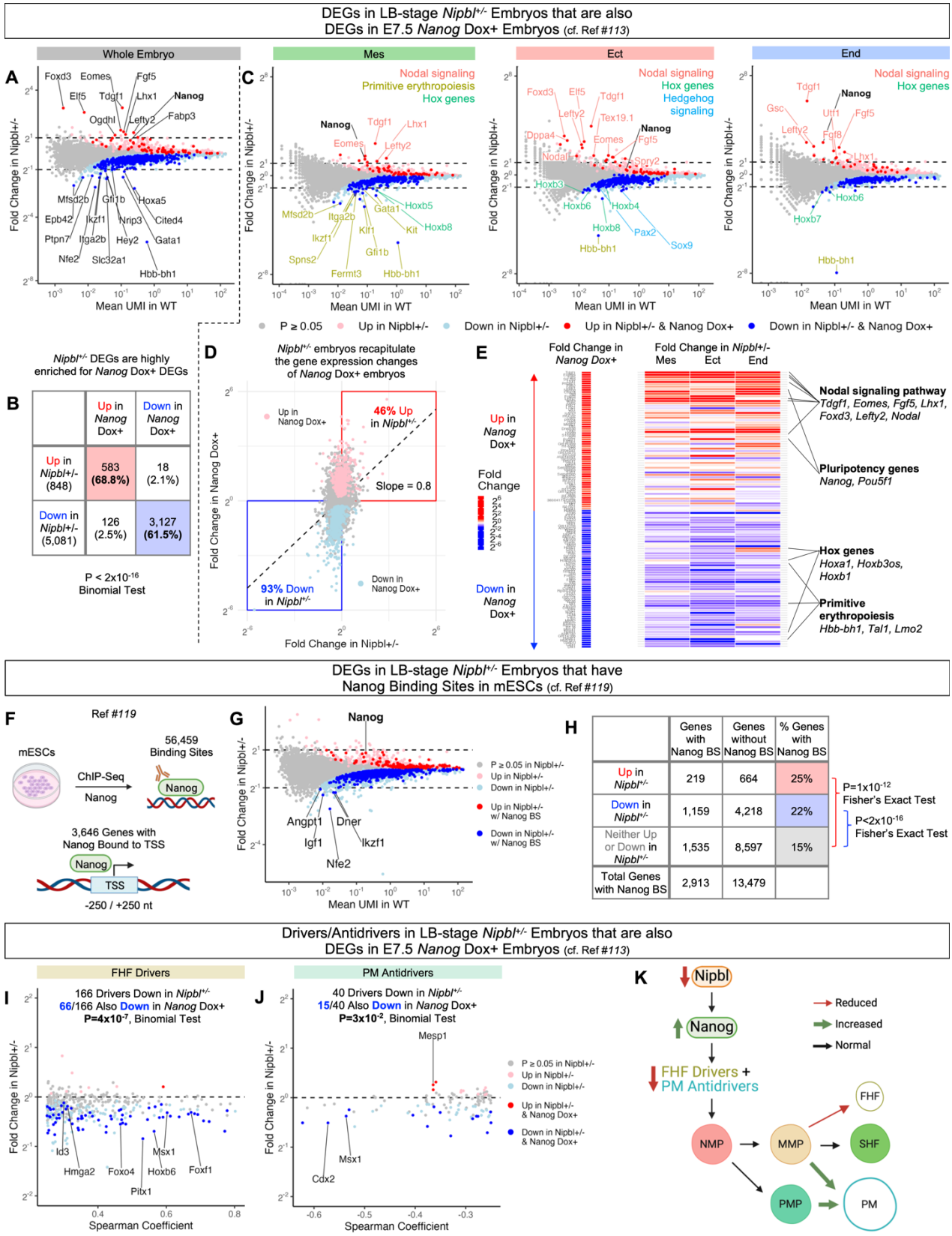


Fig. 8. LB-stage *Nipbl*$^{+/-}$ mice replicate the gene expression changes of *Nanog* overexpression. (**A**) Fold change in expression of DEGs in LB-stage *Nipbl*$^{+/-}$ embryos ($Q < 0.05$,

1570 Mann-Whitney U Test) that are also DEGs in E7.5 *Nanog* Dox+ embryos ($Q < 0.05$, T-Test)
1571 (*113*). (**B**) Percentage of DEGs in LB-stage *Nipbl*$^{+/-}$ embryos that are also DEGs in E7.5 *Nanog*
1572 Dox+ embryos (*113*). (**C**) Fold change in expression of DEGs in the germ layers of LB-stage
1573 *Nipbl*$^{+/-}$ embryos that are also DEGs in E7.5 *Nanog* Dox+ embryos (*113*). (**D**) Fold change in
1574 expression of genes in LB-stage *Nipbl*$^{+/-}$ embryos versus their fold change in E7.5 *Nanog* Dox+
1575 embryos. DEGs in E7.5 *Nanog* Dox+ embryos (*113*) are colored pink and light blue. (**E**) Heatmap
1576 of the fold change in expression of DEGs from E7.5 *Nanog* Dox+ embryos (lowest *Q*-values from
1577 T-Test) (*113*) in the germ layers of LB-stage *Nipbl*$^{+/-}$ embryos. (**F**) ChIP sequencing for Nanog in
1578 mESCs from (*119*) identified 3,646 genes with one or more Nanog binding sites within +/- 250 nt
1579 of their TSSs. (**G**) Fold change in expression of DEGs in LB-stage *Nipbl*$^{+/-}$ embryos with one or
1580 more Nanog binding sites from (*119*). (**H**) Percentage of DEGs in LB-stage *Nipbl*$^{+/-}$ embryos with
1581 one or more Nanog binding sites from (*119*). Fold change in expression of DEGs in mesoderm
1582 cells of (**I**) FHF and (**J**) PM lineages of LB-stage *Nipbl*$^{+/-}$ embryos that are also DEGs in E7.5
1583 *Nanog* Dox+ embryos (*113*). Transcription factor genes are labeled. (**K**) Reduction of *Nipbl* levels
1584 leads to the upregulation of *Nanog* in LB-stage *Nipbl*$^{+/-}$ embryos and downregulation of FHF
1585 drivers and PM antidrivers, resulting in the misallocation of mesoderm cells to PM at the expense
1586 of FHF.

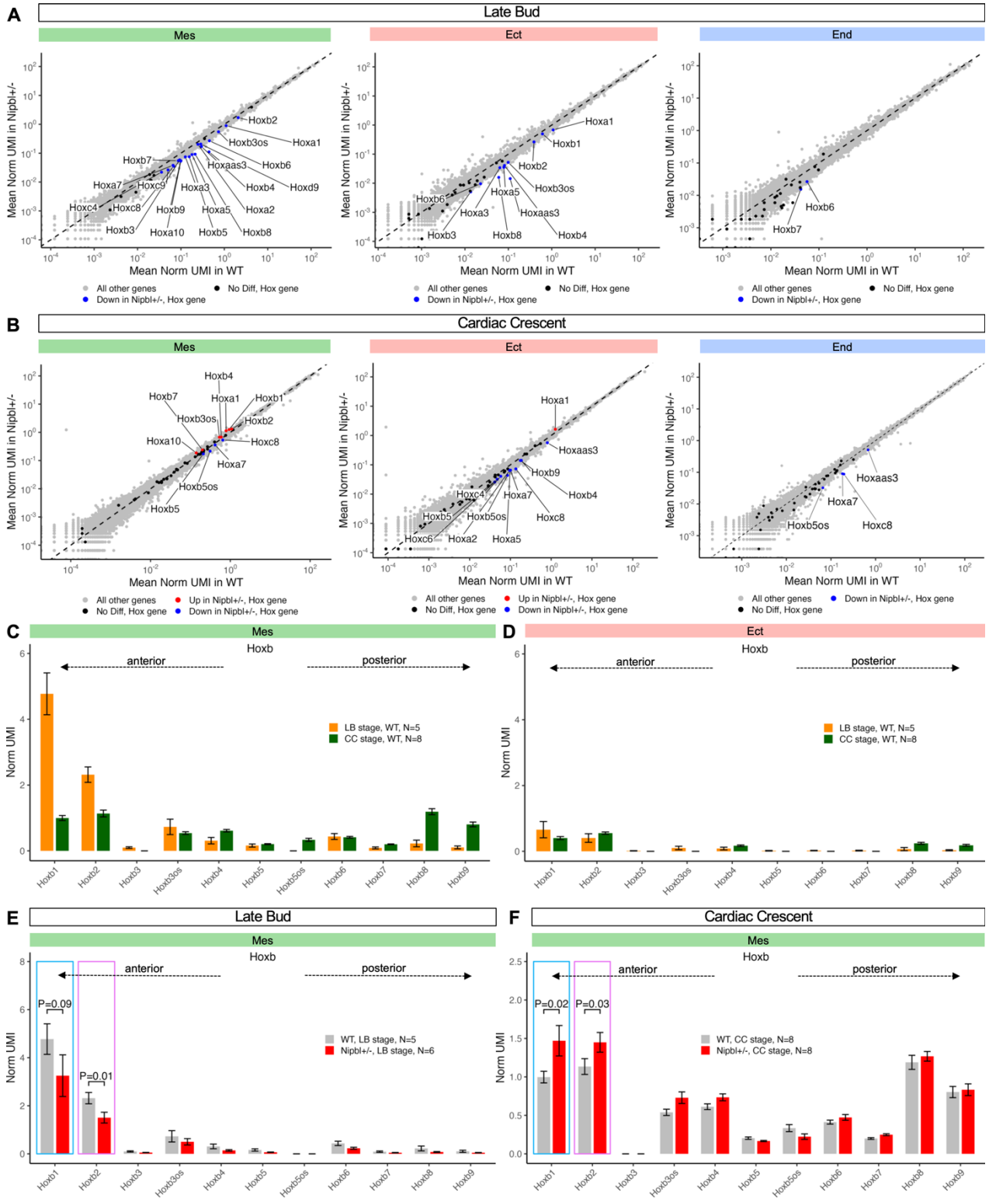
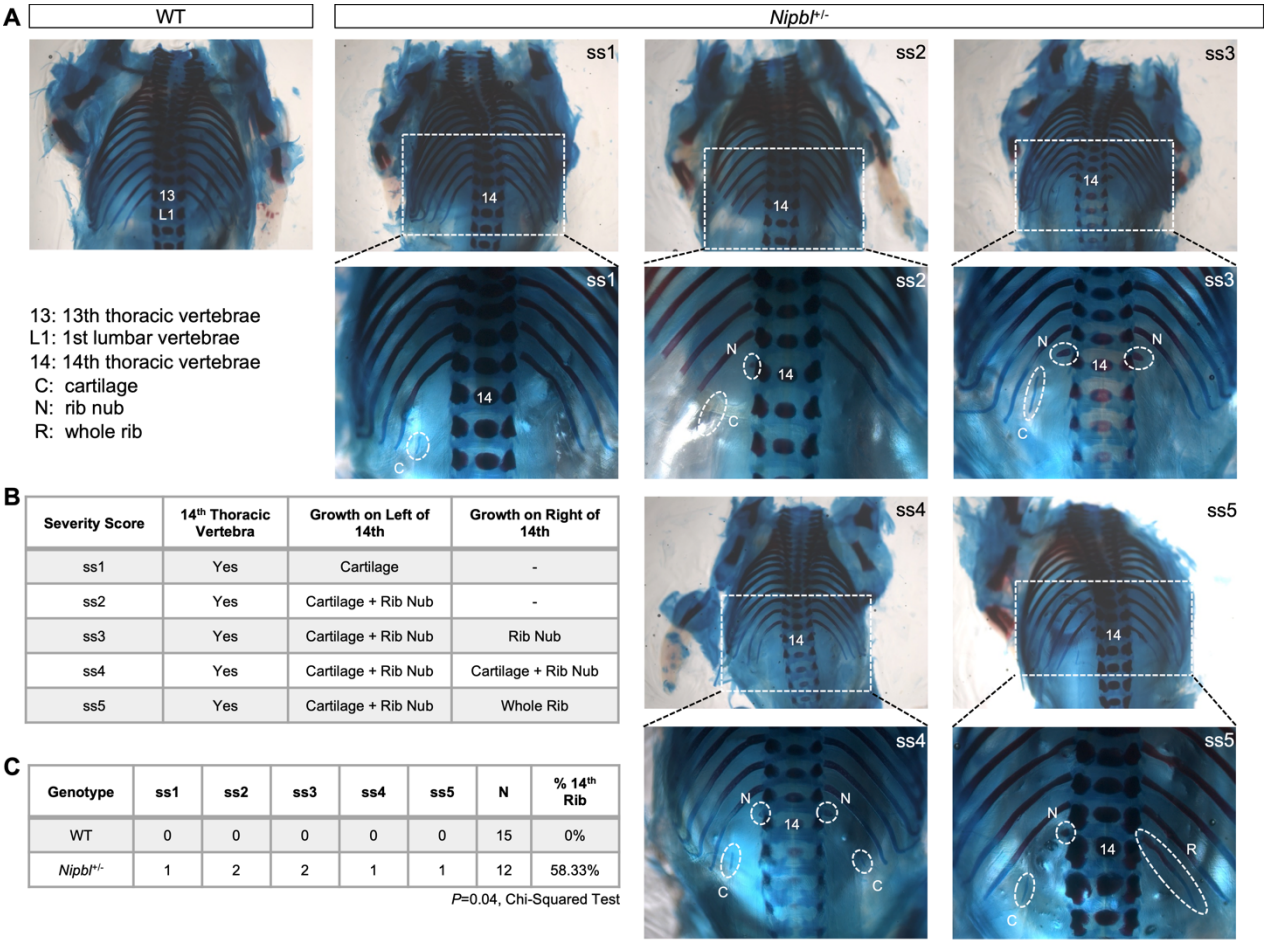


Fig. 9. *Nipbl*$^{+/-}$ mice show a delay in the expression of anterior *Hox* genes. Expression of *Hox* genes in germ layers of WT and *Nipbl*$^{+/-}$ embryos at (**A**) LB- and (**B**) CC-stage. Expression of *Hoxb* genes in mesoderm (**C**) and ectoderm (**D**) of WT embryos from LB- to CC-stage ordered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

1592 of WT embryos at (**E**) LB- and (**F**) CC-stage ordered from left (5') to right (3') by their

1593 chromosomal position.



B

Severity Score	14th Thoracic Vertebra	Growth on Left of 14th	Growth on Right of 14th
ss1	Yes	Cartilage	-
ss2	Yes	Cartilage + Rib Nub	-
ss3	Yes	Cartilage + Rib Nub	Rib Nub
ss4	Yes	Cartilage + Rib Nub	Cartilage + Rib Nub
ss5	Yes	Cartilage + Rib Nub	Whole Rib

C

Genotype	ss1	ss2	ss3	ss4	ss5	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. ss1-ss5 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 = ss1, high = ss5). (**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.