UC Irvine
UC Irvine Electronic Theses and Dissertations

Title
On the Origins of Birth Defects from Studies of a Mouse Model of Cornelia de Lange Syndrome

Permalink
https://escholarship.org/uc/item/3xw2w3g0

ISBN
9798173722782

Author
Chea, Stephenson

Publication Date
2024-11-03

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA,
IRVINE

On the Origins of Birth Defects from Studies of a Mouse Model of Cornelia de Lange Syndrome

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Stephenson Chea

Dissertation Committee:
Professor Anne L. Calof, Chair
Professor Arthur D. Lander
Professor Ken W.Y. Cho
Professor Qing Nie
Professor John S. Lowengrub
Professor Michael V. Zaragoza

2024

Chapter 3 & Appendix B © 2024 American Association for the Advancement of Science
All Other Materials © 2024 Stephenson Chea

TABLE OF CONTENTS

		Page
LIST OF FIGURES		v
LIST OF TABLES		vii
ACKNOWLEDGEMENTS		viii
VITA		ix
ABSTRACT OF DISSERTATION		xiii
CHAPTER 1:	Introduction	1
	A need for understanding the underlying causes of birth defects	1
	Challenges in developing animal models for birth defects with unknown causes	2
	The identification and recognition of Cornelia de Lange Syndrome	3
	Understanding CdLS as a syndrome with a spectrum of manifestations	4
	The spectrum of clinical manifestations in CdLS	4
	De novo mutations as the primary cause of CdLS	20
	Mutations in cohesin-related proteins cause CdLS	21
	The structure of the cohesin complex	23
	The function of the cohesin complex	24
	Nipbl-haploinsufficiency and its impact on global gene expression regulation	25
	A mouse model for the *Nipbl*-haploinsufficient form of CdLS	26
	Early murine heart development	27
	Insights into the origins of congenital heart defects from studies of *Nipbl*-haploinsufficient mice	28

CHAPTER 2: Reduced homeodomain gene expression in *Nipbl*$^{+/-}$ mice made haploinsufficient for *Isl1* reveals novel quantitative risk determinants for congenital heart defects 30

Abstract 30
Introduction 31
Results 34
Discussion 67
Materials and Methods 71
Acknowledgements 81

CHAPTER 3: Gastrulation-stage gene expression in *Nipbl*$^{+/-}$ mouse embryos foreshadows the development of syndromic birth defects 81

Abstract 82
Introduction 82
Results 86
Discussion 134
Materials and Methods 140
Acknowledgements 152

CHAPTER 4: Conclusions 154

A need for understanding the origins of birth defects 154

The *Nipbl*-haploinsufficient mouse model can be used to study birth defects 155

The role of *Isl1*-haploinsufficiency in the development of CHDs 156

Progenitor cell misallocation in *Nipbl*$^{+/-}$ embryos foreshadows birth defects in CdLS 159

Nanog overexpression precedes birth defects in *Nipbl*$^{+/-}$ mice 161

Impact of *Nanog* overexpression on developmental pathways in *Nipbl*$^{+/-}$ mice 162

Can reduction of Nanog rescue the phenotypes observed in *Nipbl*$^{+/-}$ mice? 162

Nipbl's effects on chromatin topology and gene expression in *Nipbl*$^{+/-}$ embryos 166

Nipbl-haploinsufficiency results in reduced chromatin accessibility at enhancers and CTCF sites 167

Why do birth defects develop? 183

REFERENCES 192

APPENDIX A: Supplementary Materials for Chapter 2 214

APPENDIX B: Supplementary Materials for Chapter 3 215

LIST OF FIGURES

		Page
Figure 2.1	Light sheet fluorescence microscopy detects a spectrum of CHDs in *Nipbl*$^{+/-}$ mice	37
Figure 2.2	*Nipbl* made deficient in the SHF domain using one Cre driver (*Isl1-Cre*), but not another (*Mef2c-Cre*), results in CHDs	41
Figure 2.3	*Mef2c-Cre* and *Isl1-Cre* recombine similarly within the heart, but *Isl1-Cre* recombines extensively in non-cardiac tissues	45
Figure 2.4	Reduced *Isl1* expression in *Isl1*$^{Cre/+}$ hearts results in CHDs	48
Figure 2.5	Mice doubly haploinsufficient for *Nipbl* and *Isl1* show fetal demise and reduced body size	51
Figure 2.6	*Isl1*-haploinsufficiency enhances the frequency and spectrum of CHDs in *Nipbl*-haploinsufficient mice	54
Figure 2.7	Gene expression in E10.5 hearts potentially contributing to CHDs in *Nipbl*$^{+/-}$, *Isl1*$^{+/-}$, and *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ mice	59
Figure 2.8	*Nipbl*$^{+/-}$ mice overexpress *Isl1* while *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ mice express *Isl1* at WT levels	63
Figure 2.9	*Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ mice fail to upregulate *Pitx1* to WT levels, and *Hoxc4* and *Pax6* above WT levels	66
Figure 3.1	*Nipbl*$^{+/-}$ mice do not lack any cell populations found in WT mice	89
Figure 3.2	scRNAseq captured cells from all germ layers, as well as progenitors of major tissues	94
Figure 3.3	*Nipbl*$^{+/-}$ embryos exhibit changes in the sizes of mesodermal subpopulations that are not accompanied by changes in developmental timing	99
Figure 3.4	*Nipbl*$^{+/-}$ mice misdirect mesoderm cells into paraxial mesoderm at the expense of the first heart field	103
Figure 3.5	*Nipbl*$^{+/-}$ mice underexpress genes predicted to drive the transition of mesoderm cells into first heart field	110

Figure 3.6	*Nipbl*$^{+/-}$ mice show large changes in the expression of major developmental regulators in all germ layers	114
Figure 3.7	*Nipbl*$^{+/-}$ mice overexpress *Nanog* during and after gastrulation1	119
Figure 3.8	LB-stage *Nipbl*$^{+/-}$ mice replicate the gene expression changes of *Nanog* overexpression	126
Figure 3.9	*Nipbl*$^{+/-}$ mice show a delay in the expression of anterior *Hox* genes	130
Figure 3.10	*Nipbl*$^{+/-}$ mice show anteriorization of thoracic vertebrae with left-right asymmetry	133
Figure 4.1	Mating *Stra8-iCre*; *Nipbl*$^{Flox/+}$ male mice with *Nanog* $^{GFP/+}$ female mice does not produce Mendelian genotypic frequencies	165
Figure 4.2	Single cell multiome (RNA + ATAC) sequencing of LB-stage WT and *Nipbl*$^{+/-}$ embryos captured nuclei from all germ layers	166
Figure 4.3	LB-stage *Nipbl*$^{+/-}$ embryos show reduced accessibility of intergenic chromatin	170
Figure 4.4	LB-stage *Nipbl*$^{+/-}$ embryos show reduced accessibility of intergenic enhancers	173
Figure 4.5	LB-stage *Nipbl*$^{+/-}$ embryos show reduced accessibility of intragenic and intergenic CTCF sites	175
Figure 4.6	The *Nanog* gene is located between two CTCF sites, with a predicted distal enhancer element that shows increased accessibility in LB-stage *Nipbl*$^{+/-}$ embryos	182

LIST OF TABLES

		Page
Table 4.1	Genes and enhancers make up the majority of accessible chromatin in LB-stage WT and *Nipbl*$^{+/-}$ embryos, with CTCF sites making up the minority	171

ACKNOWLEDGEMENTS

I would like to extend my deepest gratitude to a group of remarkable individuals whose expertise, guidance, and support have been invaluable during my doctoral journey.

First, I am profoundly grateful to Dr. Anne L. Calof, whose mentorship was instrumental throughout my years of study. Our frequent discussions and her passion for science not only enhanced my research but also deeply influenced my approach to data visualization and creativity in scientific inquiry.

I also wish to thank Dr. Arthur D. Lander for his invaluable discussions regarding statistical analysis methods and for his insightful interpretation of the results, which significantly shaped the direction of my research.

My sincere appreciation goes to my thesis committee members, Ken W.Y. Cho, Qing Nie, John S. Lowengrub, and Michael V. Zaragoza, for their consistent guidance and expert advice throughout my doctoral training.

Special thanks are due to Martha E. Lopez-Burks, Louise Villagomez, and Brian Bui. Your assistance with mouse husbandry and unwavering support in conducting experiments were crucial to the success of many projects.

I am also grateful to Christopher Wu, Tara Hosseini, Jaime A. Vasilev, Arianna G. Favela, Lorelle D. Meidt, Han Nguyen, Kimia Rezaeizadeh, Zahra Amiri, Daniel P. Gilchrist, Monica Sharda, Amanda Nyam, and Henry Chang for your enthusiasm and eagerness to learn as research assistants.

I would also like to acknowledge the financial support of several institutions. The Center for Complex Biological Systems, Cornelia de Lange Syndrome Foundation, and NSF-Simons Center for Multiscale Cell Fate provided critical early study support. The National Heart, Lung, and Blood Institute of the NIH was instrumental in funding the majority of this research (NHLBI-NIH R01HL138659). Additionally, my training was supported by a grant from the National Institute of Neurological Disorders and Stroke (NINDS-NIH T32NS082174).

Chapter 3 and Appendix B of this dissertation is a reprint of the material as it appears in Chea, Stephenson et al. “Gastrulation-stage gene expression in *Nipbl*$^{+/-}$ mouse embryos foreshadows the development of syndromic birth defects.” *Science Advances* vol. 10,12 (2024): eadl4239. doi:10.1126/sciadv.adl4239, used with permission from the American Association for the Advancement of Science. The co-authors listed in this publication are Jesse Kreger, Martha E. Lopez-Burks, Adam L. Maclean, Arthur D. Lander, and Anne L. Calof.

VITA

Stephenson Chea

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE	END DATE	FIELD OF STUDY
California State University, Long Beach	BS	12/2015	Microbiology
University of California, Irvine	PHD	06/2024	Biological Sciences
American Society for Clinical Pathology	Other training	12/2015	Technologist in Molecular Biology

POSITIONS AND SCIENTIFIC APPOINTMENTS

2018 -	Graduate Research/Teaching Assistant, University of California, Irvine
2017 - 2018	Clinical Molecular Laboratory Consultant, Quest Diagnostics
2017 - 2017	Clinical Molecular Laboratory Consultant, LabCorp
2016 - 2017	Clinical Molecular Laboratory Manager, Sonic Healthcare USA
2015 - 2016	Clinical Molecular Laboratory Technologist, West Pacific Medical Laboratory
2015 - 2015	Clinical Molecular Laboratory Technician, Vantage Point Laboratory
2014 - 2015	Undergraduate Teaching Assistant, California State University, Long Beach
2011 - 2015	Undergraduate Research Assistant, California State University, Long Beach
2009 - 2010	High School Research Assistant, Center for Energy and Environmental Research and Services

HONORS

2022 - 2024	Predoctoral Fellowship, Training Program in Stem Cell Translational Medicine for Neurological Disorders, National Institute of Neurological Disorders and Stroke, National Institutes of Health
2021 - 2022	Interdisciplinary Opportunity Award for Understanding the Organelle Landscape of Stem Cells in a Birth Defects Syndrome
	Interdisciplinary Opportunity Award for Investigating Cell Fate Determination in a Mouse Model of Cornelia de Lange Syndrome, NSF-Simons Center for Multiscale Cell Fate Research, University of California, Irvine

2019 - 2020	Research Grant for Investigating Cell Fate Misallocation as a Source of Developmental Defects in Cornelia de Lange Syndrome, Cornelia de Lange Syndrome Foundation
2017 - 2018	Graduate Dean's Recruitment Fellowship, Graduate Division, University of California, Irvine
2010 - 2015	President's List, California State University, Long Beach
2020	Edward A. Steinhaus Teaching Award, School of Biological Sciences, University of California, Irvine
2019	Travel Award for Presentation at 2019 Weinstein Cardiovascular Development & Regeneration Conference, University of California, Irvine
2019	Travel Award for Presentation at 2019 Weinstein Cardiovascular Development & Regeneration Conference, Weinstein Organizing Committee
2019	Presentation Award for 2019 Associated Graduate Students Symposium, Associated Graduate Students, University of California, Irvine
2018	Travel Award for Presentation at CdLS Foundation 8th Biennial Scientific & Educational Symposium, Cornelia de Lange Syndrome Foundation
2015	Bachelor of Science (Summa Cum Laude), California State University, Long Beach

PUBLICATIONS

1. Chea S, Kreger J, Lopez-Burks ME, MacLean AL, Lander AD, Calof AL. Gastrulation-stage gene expression in *Nipbl*$^{+/-}$ mouse embryos foreshadows the development of syndromic birth defects. Sci Adv. 2024 Mar 22;10(12):eadl4239. doi: 10.1126/sciadv.adl4239. Epub 2024 Mar 20. PMID: 38507484; PMCID: PMC10954218.
2. Oliver C, Groves L, Hansen BD, Salehi M, Kheradmand S, Carrico CS, Caudill P, Mattingly M, Dorsett D, Chea S, Singh VP, Krantz ID, Huisman S, Deardorff MA, Kline AD. Cornelia de Lange syndrome and the Cohesin complex: Abstracts from the 9th Biennial Scientific and Educational Virtual Symposium 2020. Am J Med Genet A. 2022 Mar;188(3):1005-1014. PubMed PMID: 34877788.
3. Kline AD, Krantz ID, Bando M, Shirahige K, Chea S, Sakata T, Rao S, Dorsett D, Singh VP, Gerton JL, Horsfield JA, Calof AL, Katz O, Grados M, Raible S, Barañano K, Lyon G, Musio A, Carrico CS, Clemens DK, Caudill P, Massa V, McGill BE, Dommestrup A, O'Connor J, Haaland RE. Cornelia de Lange syndrome, related disorders, and the Cohesin complex: Abstracts from the 8th biennial scientific and educational symposium 2018. Am J Med Genet A. 2019 Jun;179(6):1080-1090. PubMed PMID: 30874362.

PRESENTATIONS

1. Chea S, Lander AD, Calof AL. Using Single-Cell Transcriptomics to Trace the Origins of Birth Defects in a Mouse Model of Cornelia de Lange Syndrome. XII Cornelia de Lange Syndrome World Conference. 2023 November 29; Zaragoza, Spain.
2. Chea S, Scipioni L, Tedeschi G, Digman M, Calof AL. Understanding the Organelle Landscape of Stem cells in a Birth Defects Syndrome. NSF-Simons Center for Multiscale Cell Fate 2023 Early Career Researcher Symposium; 2023 April 29; Irvine, CA.
3. Chea S, Kreger J, Wu X, Xiong I, MacLean A, Lander AD, Calof AL. Investigating Cell Fate Determination in a Mouse Model of Cornelia de Lange Syndrome. NSF-Simons Center for Multiscale Cell Fate 2022 Annual Symposium on Multiscale Cell Fate; 2022 April 29; Irvine, CA.
4. Chea S, Favela AG, Lander AD, Calof AL. Tracing the Origins of Birth Defects in CdLS Using Single-Cell RNA Sequencing. CdLS Foundation 10th Biennial Scientific & Educational Symposium; 2022 June 21.
5. Chea S, Favela AG, Lander AD, Calof AL. Single-cell Analysis Implicates Early Dysregulation of Nanog in the Production of Syndromic Heart Defects. Weinstein Cardiovascular Development & Regeneration; 2022 May 12; Marseille, France.
6. Chea S, Kreger J, Wu X, Xiong I, MacLean A, Lander AD, Calof AL. Investigating the Origins of Birth Defects in Cornelia de Lange Syndrome Using Single-Cell RNA-Sequencing. NSF-Simons Center for Multiscale Cell Fate 2022 Early Career Researcher Symposium; 2022 April 29; Irvine, CA.
7. Chea S, Villagomez L, Bui B, Lopez-Burks ME, Lander AD, Calof AL. Origins of Structural Birth Defects in Cornelia de Lange Syndrome: Misexpression of Genes or Misallocation of Progenitors?. CdLS Foundation 9th Biennial Scientific & Educational Symposium; 2020 October.
8. Chea S, Gilchrist DP, Santos R, Villagomez L, Lopez ME, Bui B, Lander AD, Calof AL. Are Congenital Heart Defects in Cornelia de Lange Syndrome Due to Misallocation of Early Cardiac Progenitor Cells?. Weinstein 2019 Cardiovascular Development and Regeneration Meeting. 2019 May
9. Chea S, Lander AD, Calof AL. Understanding the Developmental Origins of Congenital Heart Defects in the Transcriptomopathy Known as Cornelia de Lange Syndrome. UCI AGS Graduate Research Symposium. 2019 April; Irvine, CA.
10. Chea S, Gilchrist DP, Lander AD, Calof AL. Is Altered Allocation of Cardiac Stem Cell Populations Responsible for Congenital Heart Defects in Cornelia de Lange Syndrome?. Center for Complex Biological Systems 2019 Annual Retreat. 2019 March; Pasadena, CA.
11. Chea S, Gilchrist D, Lander AD, Calof AL. Is Altered Allocation of Cardiac Stem Cell Populations Responsible for Congenital Heart Defects in Cornelia de Lange Syndrome?. Southern California Rare Disease Genetics Meeting. 2019 March; Irvine, CA.

12. Chea S, Daniel G, Lander AD, Calof AL. Is Altered Allocation of Cardiac Stem Cell Populations Responsible for Congenital Heart Defects in Cornelia de Lange Syndrome?. Center for Complex Biological Systems 7th Annual Southern California Systems Biology Conference. 2019 January; Irvine, CA.
13. Chea S, MacLean A, Nie Q, Lander AD, Calof AL. Single-Cell RNA Sequencing Approaches to Understanding the Developmental Origins of Heart Defects. NSF-Simons Center for Multiscale Cell Fate Research 1st Annual Symposium on Multiscale Cell Fate. 2018 October; Irvine, CA.
14. Chea S, Lander AD, Calof AL. Elucidating the Origins of Congenital Heart Defects in a Mouse Model of Cornelia de Lange Syndrome. CdLS Foundation 8th Biennial Scientific & Educational Symposium. 2018 May; Minneapolis, MN.
15. Chea S, Livingston BT. Characterization of Conserved Proteins in the Developing Echinoderm Skeleton at the Transcriptional Level. CSULB CNSM Student Research Symposium; 2015 September; Long Beach, CA.
16. Chea S, Flores R, Quinn B, Satterfield S, Livingston BT. Identification of Conserved Proteins in the Echinoderm Skeleton Using an Evolutionary Approach. CSU Program for Education and Research in Biotechnology; 2015 January; San Francisco, CA.
17. Chea S, Nguyen K, Rahai H. Mapping the 3D Distribution of Heat from a Cylindrical Horizontal Jet Flow in a Fluid Environment. Center for Energy and Environmental Research and Services; 2010 August; Long Beach, CA.

ABSTRACT OF THE DISSERTATION

On the Origins of Birth Defects from Studies

of a Mouse Model of Cornelia de Lange Syndrome

By

Stephenson Chea

Doctor of Philosophy in Biological Sciences

University of California, Irvine, 2024

Professor Anne L. Calof, Chair

Approximately 3% of newborns worldwide are affected by birth defects (*1*), yet the causes for about 70-80% of these cases remain unidentified (*2*), despite various known contributing factors. Cornelia de Lange Syndrome (CdLS), a disorder characterized by congenital malformations across multiple organs and tissues (*3*), is caused by mutations in genes associated with the cohesin complex or its interacting proteins (*4*). Most CdLS cases are caused by haploinsufficiency for the *NIPBL* gene (*5*), which plays a role in the loading of the cohesin complex onto chromatin (*6*). Originally recognized for its critical role in sister chromatid cohesion (*7*), the cohesin complex is now known to also play a role in regulating global gene expression (*8*). Studies using cell lines derived from CdLS patients and *Nipbl*-haploinsufficient tissues from animal models indicate that subtle, significant, and widespread changes in gene expression precede birth defects (*9-12*). Given that these changes are generally modest (less than 2-fold) (*9-12*), this suggests that the birth defects observed in CdLS and related animal models are likely not caused by alterations in any single gene. Rather, they are probably the result of cumulative changes in the expression of multiple genes. Thus, studies of CdLS and its animal models have shed light on the mechanisms underlying birth defects of unknown causes, that

broad shifts in gene expression rather than specific mutations in individual genes may drive the development of birth defects.

In the United States, congenital heart defects (CHDs) affect nearly 1% of all newborns (*13*). They are also the primary cause of birth defect-related infant mortality and morbidity (*13-15*). The causal origins of CHDs are largely unknown, with mutations in single genes accounting for fewer than 20% of cases (*16*). About 30% of individuals with the *Nipbl*-haploinsufficient form of CdLS present with CHDs (*17*). Studies in $Nipbl^{+/-}$ mice, which exhibit a similar prevalence of CHDs, have primarily identified atrial septal defects (*18*) – among the most common types of CHDs observed in nonsyndromic cases in man (*19*). $Nipbl^{+/-}$ mice also display abnormal development in the ventricular septum and right ventricle (*18*). Intriguingly, these cardiac structures originate from a progenitor cell population known as the second heart field (SHF) (*20*), which points to the SHF as a potential source of CHDs in $Nipbl^{+/-}$ mice.

In Chapter 2, we explored the impact of *Nipbl*-haploinsufficiency within the SHF on the development of CHDs using light sheet fluorescence microscopy (LSFM) and two SHF-Cre recombinase-expressing mouse lines: *Mef2c-Cre* and $Isl1^{Cre/+}$. Remarkably, it was found that only $Isl1^{Cre/+}$ driven *Nipbl*-haploinsufficiency in the SHF led to CHDs. This occurrence was likely due to extensive *Isl1*-Cre recombination in non-cardiac tissues and/or a reduction in *Isl1* expression, the latter resulting from knock-in mediated haploinsufficiency. In surprising contrast, mice rendered globally haploinsufficient for both *Isl1* and *Nipbl* ($Nipbl^{+/-}$; $Isl1^{+/-}$) exhibited a wider range and higher frequency of CHDs compared to those haploinsufficient for either gene alone. RNA sequencing of E10.5 mouse hearts indicated that *Nipbl*-haploinsufficiency upregulates *Isl1* expression, seemingly providing a protective effect against the development of severe CHDs, as evidenced by fewer and less severe CHDs in $Nipbl^{+/-}$ mice compared to their

Nipbl$^{+/-}$*; Isl1*$^{+/-}$ counterparts. Furthermore, we observed significant alterations in the expression of homeodomain genes such as *Pitx1*, *Pax6*, and *Hoxc4*. *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts did not upregulate these genes to the same level as *Nipbl*$^{+/-}$ hearts. These results introduce new quantitative risk factors for the development of CHDs in both CdLS and nonsyndromic cases.

In addition to exhibiting ASDs, *Nipbl*$^{+/-}$ mice also demonstrate early abnormalities in other cardiac regions during embryonic development. By embryonic day 13.5 (E13.5), 70% of *Nipbl*$^{+/-}$ mice displayed delays in ventricular septal fusion, while three days earlier, at E10.5, all exhibited right ventricle hypoplasia (*18*). Even earlier, at E7.75 and E7.5, in situ hybridization experiments revealed diminished expression of two key transcription factors, *Nkx2-5* and *Mesp1 (18)*, which play roles in the growth and differentiation of early heart progenitor cells. These suggests that cardiac development abnormalities might originate at the very onset of heart formation. Supporting this hypothesis, studies on *nipbl*-morphant zebrafish have shown that defects in the initial migration of cardiogenic mesoderm can be detected as early as 18 hours post-fertilization (*11*). These findings imply that, at least for CHDs, causative events may occur as early as gastrulation – the stage when the three primary germ layers form and the earliest progenitor cells of major tissues and organs begin to differentiate.

To investigate the early developmental origins of CHDs associated with *Nipbl*-haploinsufficiency, in Chapter 3, we performed single-cell RNA-sequencing on both wildtype (WT) and *Nipbl*$^{+/-}$ mouse embryos during the late-gastrulation and early cardiac crescent stages. The results indicate that although *Nipbl*$^{+/-}$ embryos contain the same cell populations as WT embryos, they exhibit subtle yet significant misallocations of specific mesodermal cell populations. Our analysis confirmed that these discrepancies are not related to apoptosis, cell proliferation, global developmental delays, or changes in the overall structure of cell lineages.

Rather this phenomenon appears to stem from altered cell fate decisions in mesoderm cells' differentiation toward cardiac or non-cardiac fates.

As in previous studies, we observed that while most gene expression changes in *Nipbl*-deficient tissues at these stages were minor, certain key developmental genes, particularly *Nanog*, were significantly misexpressed. Specifically, *Nipbl*$^{+/-}$ embryos did not downregulate *Nanog* at the end of gastrulation – a process typically observed in all cells except germ cells. This persistent overexpression of *Nanog* leads to aberrant expression of many of its target genes across all germ layers. We additionally observed underexpression of *Hox* genes and overexpression of *Nodal* signaling genes, which play roles in anterior-posterior and left-right patterning, respectively. Based on these findings, we propose a model suggesting that birth defects in CdLS may arise from prolonged overexpression of *Nanog* and subsequent dysregulation of developmental pathways. These disruptions likely contribute to incorrect cell fate decisions and the misallocation of specific progenitor cell populations, ultimately leading to the developmental anomalies observed.

CHAPTER 1

Introduction

A need for understanding the underlying causes of birth defects

Approximately 3% of newborns globally are affected by birth defects (*1*). Birth defects are structural or functional anomalies that occur during the development of a baby and are present at birth. These defects can affect various body parts and systems. The most common birth defects are heart defects and neural tube defects (*1*). Birth defects are the leading cause of infant death in developed countries (*21*). Babies born with birth defects and who survive past infancy often face long-term health challenges, requiring ongoing medical care and support to manage physical, developmental, or cognitive disabilities. Additionally, only 82.51% of children born with a birth defect and who survive past infancy will survive to the age of 25 years (*22*). Birth defects not only impact individual mortality and morbidity but also place significant emotional and financial burdens on families and strain healthcare systems. Therefore, there is a critical need to develop therapies for not only mitigating the clinical severity of birth defects, but also for preventing their occurrence.

To effectively develop therapies for birth defects, we have to first comprehend the causal mechanisms behind these conditions. To acquire this understanding, we have to develop accurate animal models of birth defects. However, creating animal models requires an initial knowledge of the causes of birth defects.

Birth defects can be caused by environmental exposures, infections during pregnancy, nutritional deficiencies, genetic factors, and combinations of genetic and environmental factors. Exposure to certain teratogens such as alcohol, tobacco, or certain medications during pregnancy

can cause birth defects such as fetal alcohol spectrum disorders (*23*). Infections like rubella and cytomegalovirus can lead to serious congenital conditions, affecting the development of the baby (*24*). Deficiencies in essential nutrients, such as folic acid, are associated with neural tube defects such as spina bifida (*25*). Genetic factors also play a role, as single-gene mutations can cause the defects seen in congenital NAD deficiency disorder (*26*) and chromosomal abnormalities resulting in the defects seen in trisomy 21 or Down syndrome (*27*). Additionally, the interaction between genetic predispositions and environmental factors can precipitate birth defects, with both influencing the risk and severity of these conditions (*28*). While a variety of causes have been identified for birth defects, they only account for about 20-30% of birth defects, leaving the etiology of the remaining 70-80% of cases unknown (*2*).

Challenges in developing animal models for birth defects with unknown causes

Given the diverse and frequently isolated occurrence of birth defects, identifying common causal mechanisms presents significant challenges. Traditionally, researchers have studied birth defects individually to determine their causes. Yet, this method has yielded identifiable causes for less than 30% of birth defects (*2*). Even if we were to continue with this approach and succeed in pinpointing the causes for each defect, replicating them in animal models to study their underlying mechanisms is a time-consuming and costly process due to the extensive animal husbandry required. Suppose we manage to create these models for multiple defects. We would then need to investigate each one to uncover their causal mechanisms and attempt to identify commonalities – a complex and daunting task.

What if we adopted a different strategy? One that focuses on birth defect syndromes? Birth defect syndromes exhibit multiple concurrent defects that often mirror the characteristics of more common, isolated birth defects. They also typically have a single genetic basis, making

their replication in animal models feasible. Two of the most common syndromes that display birth defects are Down syndrome and Turner syndrome. For example, Down syndrome is caused by the presence of an extra copy of chromosome 21 (*29*), and Turner syndrome is caused by the complete or partial loss of one X chromosome, often presenting in mosaic karyotypes (*30*). Due to their singular genetic origins, successful animal models have been developed for both syndromes (*31, 32*). By creating a single animal model of a birth defects syndrome, we could simultaneously study several types of birth defects. This approach would not only be more efficient than the one discussed above, but it could also provide significant insights into the shared underlying mechanisms of various birth defects.

The identification and recognition of Cornelia de Lange Syndrome

One well studied birth defects syndrome for which an animal model has been developed is Cornelia de Lange Syndrome (CdLS). CdLS was first documented in 1916 by the German physician, W. Brachmann, who described a patient displaying distinctive facial features and limb abnormalities (*33*). However, Brachmann's findings initially failed to gain widespread recognition within the medical community. It was not until 1933 that Dutch pediatrician Cornelia de Lange provided a more detailed examination of CdLS in two children exhibiting similar distinctive features, including severe developmental delays, characteristic facial anomalies, and limb deformities (*3*). Her comprehensive description brought significant attention to the condition, leading to its naming after her. As CdLS became recognized as a medical syndrome, the cases initially documented by Brachmann were eventually reassessed in light of de Lange's findings. Consequently, CdLS is sometimes referred to as Brachmann-de Lange syndrome, acknowledging both contributors to its identification.

Understanding CdLS as a syndrome with a spectrum of manifestations

Initially, CdLS was characterized by severe developmental delays, distinctive facial features, and growth retardation. However, the medical community soon recognized significant variability in its presentation throughout the 1980s and 1990s. For example, in 1985 there was a report that documented the classic severe manifestations but also cases with milder developmental and physical anomalies (*34*), suggesting that CdLS could present in less severe forms than previously recognized. As research into CdLS advanced into the 2000s, its genetic basis began to be elucidated, revealing mutations in several genes associated with the cohesin complex. These studies linked specific mutations to varying degrees of severity. For example, mutations in the *NIPBL* gene were associated with a more severe form of CdLS (*35*), while mutations in other genes (*SMC1A, SMC3, HDAC8*, and *RAD21*) typically led to milder symptoms (*4*). This genetic diversity provided a framework for understanding why a broad clinical spectrum was presented among CdLS patients. By the 2010s, comprehensive reviews and meta-analyses of recorded cases solidified the understanding that CdLS is a spectrum disorder. A 2013 study, for instance, synthesized data from multiple studies and highlighted considerable variability in individual cases—from those with profound disabilities and recognizable features to those with subtler symptoms and better developmental outcomes (*36*). Together, these advancements in case reporting and genetic research have broadened the clinical and genetic understanding of CdLS, affirming its nature as a condition with a broad spectrum of manifestations rather than a uniform set of defects.

The spectrum of clinical manifestations in CdLS

Cornelia de Lange Syndrome (CdLS) is associated with a wide range of structural and physiological abnormalities, impacting growth, intellectual and behavioral development,

neurological function, and limb formation. Additionally, it can cause complications in the gastrointestinal, ophthalmologic, otolaryngologic, genitourinary, cardiovascular, immunological, and hematologic systems, among other health issues.

Growth

Prenatal onset: Growth restriction in CdLS often starts before birth, with many infants displaying low birth weights and lengths significantly below normal standards. A report highlights a case where fetal forearm defects were noted at 23 weeks, with subsequent findings of fetal growth restriction becoming evident, leading to intrauterine fetal death at 29 weeks. At birth, the baby's weight and length were markedly below the expected standards for gestational age (*37*).

Postnatal growth failure: Children with CdLS continue to exhibit growth failure after birth, often marked by a significant and persistent lag in growth when compared to typical developmental curves. Research has established that growth metrics like height and weight consistently remain below the fifth percentile throughout their lives, indicating severe and ongoing growth retardation (*38, 39*). This pattern includes slower pubertal growth and persistently low weight velocity until late adolescence.

Milder forms and variability: In cases of CdLS with milder clinical manifestations or mosaic pathogenic variants, the growth impairment is generally less pronounced. These individuals may display a near-normal growth pattern initially but usually remain smaller compared to their peers. Mutations in the *RAD21* gene, for example, are often associated with milder phenotypes in CdLS. Individuals with these mutations tend to have less severe growth deficits and fewer systemic anomalies (*40*).

Compounding factors: Several factors contribute to growth difficulties in CdLS. For instance, gastroesophageal reflux disease (GERD) is common among these patients and can significantly impact nutritional intake due to associated esophagitis and feeding aversions (*41*). GERD is a chronic condition where stomach acid frequently flows back into the tube connecting the mouth and stomach, causing symptoms like heartburn and indigestion. Furthermore, many children with CdLS encounter challenges with feeding due to oropharyngeal dysphagia, poor sucking reflexes, and sensory aversions to different food textures. Oropharyngeal dysphagia is a swallowing disorder that involves difficulty in starting the swallowing process, often due to abnormalities in the throat or esophagus. The presence of severe feeding difficulties, including problems with sucking and chewing, often contributes directly to inadequate nutritional intake, which further impairs growth (*42*).

Long-term outcomes: Despite various levels of intervention, the long-term growth prognosis for individuals with CdLS generally remains significantly below average. Growth metrics such as height and weight typically do not reach those of age-matched peers, persistently placing these individuals below the normal growth curves (*38*). Serial measurements of patients with CdLS demonstrate that height velocity is equal to the normal range, but there is slower pubertal growth, and weight velocity remains below the normal range until late adolescence.

Intellectual disability

Intellectual disability in CdLS is a central feature and exhibits a broad spectrum in terms of severity and functional impact. IQ levels among individuals with CdLS vary widely, from below 30 to as high as 102, although the average IQ is typically around 53 (*38, 43, 44*). Severe-to-profound intellectual disability is more commonly observed in individuals with classic

phenotypic features of CdLS, particularly in cases associated with *NIPBL*-haploinsufficiency (*45*).

Language development in CdLS is disproportionately affected, with expressive language skills usually more impaired than receptive language skills (*46*). A study examining gesture use during conversational interaction in people with CdLS found that, despite limited spoken language, individuals with CdLS used significantly more gestures compared to their peers (*47, 48*). This finding indicates that while their expressive verbal communication is notably reduced, non-verbal communication methods are frequently utilized to compensate. This study also suggests that while individuals with CdLS may understand more language and communication than they can express, their ability to verbally articulate thoughts and needs is significantly compromised.

Despite the general challenges with expressive language, cognitive functions such as problem-solving and adaptive behavior might be relatively better preserved (*46*). This indicates that intellectual disabilities in CdLS do not uniformly affect all areas of cognition and suggests a complex neurodevelopmental impact that varies widely among affected individuals.

Behavioral issues

Behavioral issues in individuals with CdLS are complex and multifaceted, often significantly impacting their daily functioning and interactions (*48*). Many of these challenges stem from inherent difficulties with communication and sensory processing, which are exacerbated by the syndrome's cognitive and developmental aspects.

Communication frustrations: The inability to effectively communicate is a primary source of frustration for many with CdLS. These individuals often have limited expressive language skills but a better understanding of receptive language, which can lead to significant

frustration when they cannot express their needs or desires. Only a minority of children with CdLS are able to use speech, with most relying on non-verbal means to express needs (*49*). This disparity can result in behaviors such as screaming, aggression, or even self-injury when they are unable to make their needs understood.

Sensory processing issues: Sensory processing anomalies are prevalent in CdLS and manifest as either hyposensitivity (under-responsiveness) or hypersensitivity (over-responsiveness) to sensory stimuli (*50*). Such sensory challenges can lead to disorientation, extreme fixation on particular sensory inputs, or withdrawal from stimuli, which can complicate interactions with the environment and social engagements. For instance, an individual might be overly fixated on specific textures or sounds, or conversely, may withdraw from touch or loud noises.

Autistic-like behaviors: A significant proportion of individuals with CdLS display behaviors commonly associated with autism spectrum disorders (ASDs), such as social avoidance, limited eye contact, and repetitive behaviors (*51*). These can include hand-flapping, rocking, and repetitive use of objects or speech (*52*). 78.6% of individuals with CdLS meet the criteria for ASD characteristics, showing behaviors such as hyperactivity and stereotyped actions that are atypical compared to other ASD profiles (*51*).

Self-injurious behavior: Clinically significant self-injurious behavior is reported in 56% of individuals with CdLS (*53*). This can include behaviors such as head-banging, biting, and scratching. These actions are often responses to sensory overload, anxiety, or as previously noted, frustrations due to communication difficulties.

Repetitive behaviors: Repetitive behaviors are a hallmark in CdLS and are often associated with both sensory processing issues and autistic traits (*54*). These behaviors may

provide a sense of control and predictability for the individual, serving as a coping mechanism for dealing with an environment that they find unpredictable or overwhelming.

Neurological issues

Neurological issues in individuals with CdLS are diverse, affecting both the central and peripheral nervous systems.

Seizure types: Approximately 25% of individuals with CdLS experience seizures, predominantly partial epilepsy (*55*). Partial epilepsy is a type of epilepsy where seizures originate from a specific area of the brain and affect a localized part of the body. In CdLS partial epilepsy originate from the temporal and parietal areas (*55*). These seizures typically manifest before two years of age, posing significant management challenges early in life. While partial seizures are the most common, other types of seizures, including generalized tonic-clonic seizures, have also been reported in this population. Generalized tonic-clonic seizures involve sudden muscle rigidity followed by convulsions and are characterized by a loss of consciousness, affecting the entire brain.

Structural brain anomalies: Neuroimaging studies have revealed that individuals with CdLS may exhibit various structural brain anomalies (*56*). These include cerebral atrophy, ventriculomegaly, and agenesis of the corpus callosum. A series of neuroimaging studies documented the presence of cerebral and brainstem volume loss, gyral simplification, and other abnormalities in brain structures such as the absence of the massa intermedia and basilar artery fenestration. Such abnormalities can contribute to the neurological and developmental challenges observed in CdLS, including cognitive impairment and motor delays.

Neuropathology and developmental delays: Neurological issues in CdLS extend beyond seizures. Many individuals exhibit significant global developmental delays, including

motor, cognitive, and speech delays. Mutations in *SMC1A*, which are present in some individuals with CdLS, are directly associated with these broad developmental delays (*57*).

Peripheral neuropathy: There is emerging evidence to suggest that peripheral neuropathy may also be a component of the neurological spectrum of CdLS. Peripheral neuropathy is a condition resulting from damage to the peripheral nerves, causing weakness, numbness, and pain, typically in the hands and feet. This can manifest as reduced pain sensitivity, which may explain the high tolerance for pain observed in some individuals with the syndrome. A study on the autonomic nervous system in CdLS identified significant dysfunction in pathological sympathetic skin response and sudomotor tests, while large fiber nerve function studies showed normal values (*58*).

Behavioral and psychiatric co-morbidities: Neurological issues are frequently accompanied by behavioral and psychiatric disorders in individuals with CdLS, including autism spectrum disorders and attention-deficit/hyperactivity disorder (ADHD). Individuals with CdLS often display a range of autistic traits, particularly repetitive behaviors and expressive communication deficits, which are more prominent than the core social deficits typically seen in idiopathic autism (*59*). These findings suggest that neurological complications in CdLS are intricately linked with its behavioral and psychiatric manifestations.

Limb abnormalities

Limb abnormalities are a defining characteristic of CdLS, affecting a significant proportion of individuals with this syndrome. These abnormalities are most pronounced in the upper extremities but also involve the lower limbs to a lesser extent.

Upper extremities

Severity and variety: Approximately one-third of individuals with CdLS have severe upper limb abnormalities (*60, 61*). These can range from the complete absence of forearms, known as phocomelia, to various forms of oligodactyly (missing digits) and adactyly (complete absence of digits). Such anomalies significantly impact the functionality and independence of affected individuals.

Common features: Nearly all individuals exhibit some form of micromelia, characterized by unusually short and small bones in the hands and arms (*61, 62*). Proximally placed thumbs, which are positioned closer to the wrist than usual, and clinodactyly, the curvature of fingers typically towards the radial side, are also common. These features can affect fine motor skills and hand function.

Radioulnar synostosis: Radioulnar synostosis has been observed in individuals with CdLS (*63*). This condition involves the fusion of the radius and ulna, the two long bones of the forearm, which can lead to limited range of motion and elbow flexion contractures. This fusion impedes normal arm movement and can be bilateral or unilateral.

Lower Extremities

While the lower limbs in CdLS are typically less affected than the upper limbs, abnormalities do occur. 80% of individuals exhibit syndactyly of the second and third toes, where two or more toes are fused together (*60, 61*). This does not typically impair walking as severely as the upper limb abnormalities impair manual abilities. Other less frequent but notable abnormalities in the lower extremities include talipes equinovarus (clubfoot), which requires early orthopedic consultation. Small foot size and anomalies in foot shape can also be observed, which may affect balance and gait.

Gastrointestinal complications

Gastrointestinal (GI) complications are prevalent in individuals with CdLS and range from common to complex disorders, significantly affecting their health and quality of life.

GERD and associated complications

GERD has been observed in individuals with CdLS. It is a frequent medical complication of CdLS, affecting almost all patients to varying degrees (*64, 65*). GERD can lead to severe complications if not properly managed, including esophagitis, which is the inflammation of the esophagus lining, and aspiration, where stomach contents are inhaled into the lungs.

Structural GI anomalies

Pyloric stenosis: Pyloric stenosis has been observed in 4% of individuals with CdLS, contributing to the complex gastrointestinal issues frequently seen in this condition (*64, 65*). Pyloric stenosis involves the thickening of the pyloric muscle, which blocks food from entering the small intestine and can lead to severe vomiting and dehydration.

Intestinal malrotation: Intestinal malrotation has been observed in 2% of individuals with CdLS (*65, 66*). Malrotation is a defect in the intestines' positioning during fetal development, which can lead to volvulus (twisting of the intestine) and potentially life-threatening bowel obstructions or ischemia.

Congenital diaphragmatic hernia: Congenital diaphragmatic hernia (CDH) has been observed in 1% of individuals with CdLS (*65*). Congenital diaphragmatic hernia is a defect where abdominal organs herniate into the chest cavity through an opening in the diaphragm, affecting lung development and function. Individuals with CDH often had a lower birth weight and a poorer prognosis compared to other CdLS patients.

Other gastrointestinal concerns

Gastrointestinal dysmotility: Gastrointestinal dysmotility has been observed in individuals with CdLS, although its prevalence is not high (*64*). Gastrointestinal dysmotility refers to a range of conditions in which the muscles of the digestive system become impaired, leading to slowed or irregular movement of food and waste through the digestive tract. Poor motility in various parts of the GI tract can lead to symptoms ranging from gastroesophageal reflux to severe constipation and nutritional deficiencies.

Ophthalmologic issues

Ptosis: Drooping of one or both eyelids, known as ptosis, is a common finding in CdLS, about 60% (*67, 68*). This condition can affect one or both eyes and may contribute to visual impairment if severe enough to cover the pupil.

Myopia and nystagmus: Many individuals with CdLS (60%) are nearsighted, or myopic, which means they see nearby objects clearly, but distant objects appear blurred (*68, 69*). Many, approximately 37%, also show nystagmus (*68*), which is involuntary eye movements, often appearing as a shuddering or flickering of the eyes. Both conditions can significantly affect visual acuity and quality of life.

Nasolacrimal duct stenosis: Nasolacrimal duct stenosis is frequently observed in individuals with CdLS (*69*). This condition involves narrowing or obstruction of the nasolacrimal duct, which can lead to tear duct infections and chronic tearfulness or epiphora.

Glaucoma: Increased intraocular pressure leading to glaucoma is another serious concern in CdLS. Glaucoma is a group of eye conditions that damage the optic nerve, often caused by abnormally high pressure in the eye. A case of successfully managed congenital glaucoma in a

patient with CdLS has been reported (*69*). Glaucoma can cause damage to the optic nerve, leading to progressive, permanent vision loss if not treated promptly.

Microcornea, astigmatism, and optic atrophy: Microcornea, astigmatism, and optic atrophy have been observed in individuals with CdLS (*69*). Microcornea refers to an abnormally small cornea, which can complicate the optical focus of the eye. Astigmatism, involving an irregular curvature of the cornea or lens, distorts vision at all distances. Optic atrophy, the damage to the optic nerve, results in diminished visual acuity and can severely impair the affected individual's ability to see clearly.

Coloboma: Coloboma has been observed in individuals with CdLS (*69*). A coloboma is a keyhole-shaped defect in one of the structures of the eye, such as the iris, choroid, or optic nerve. It results from incomplete closure of the eye's embryonic fissure during gestation. This defect can affect depth perception and visual clarity.

Strabismus and proptosis: Strabismus, or misalignment of the eyes, commonly occurs in CdLS and can lead to amblyopia if not corrected early (*69*). Proptosis, the abnormal protrusion of one or both eyes, can be seen in more severe cases and may require surgical management to prevent further complications (*69*).

Otolaryngologic issues

Hearing impairment: Hearing loss in CdLS is a significant concern, with about 40% of affected children experiencing sensorineural hearing loss, which originates from problems in the inner ear or neural pathways (*70*). Another 60% exhibit conductive hearing loss, resulting from abnormalities in the external ear canal, eardrum, or middle ear that impede sound transmission (*71*). Interestingly, some individuals show improvement in hearing capabilities over time, possibly due to changes in middle ear pathology or effective management.

Chronic otitis media: Frequent middle ear infections, or otitis media, are common in individuals with CdLS (*72*). These can lead to chronic otitis media with effusion, which exacerbates conductive hearing loss. The eustachian tube dysfunction, often due to anatomical abnormalities in CdLS, contributes significantly to these recurrent infections.

Airway anomalies: Structural anomalies in the airway, including laryngomalacia, tracheomalacia, and bronchomalacia, are noted in some cases (*73*). These conditions can lead to stridor, wheezing, and difficulties in breathing, especially in young children with CdLS.

Speech and language delay: Due to hearing impairment combined with cognitive delays, children with CdLS often experience significant speech and language development delays (*74*). These are compounded by anatomical factors, such as palatal abnormalities and hypotonia, affecting articulation and vocal quality.

Dental and maxillofacial issues: Dental anomalies, including delayed tooth eruption, malocclusion, and high-arched palate, frequently occur in CdLS (*75*). These dental issues can affect chewing, speech, and overall oral health, necessitating early dental interventions.

Nasal anomalies: Anomalies such as choanal atresia or stenosis are occasionally seen in CdLS (*73*). These conditions can obstruct nasal passages, leading to breathing challenges, chronic nasal congestion, and sinusitis.

Genitourinary issues

Renal abnormalities

Vesicoureteral reflux, where urine flows backward from the bladder into the ureters, has been reported as a renal issue in 12% of individuals with CdLS (*76*). This condition increases the risk of urinary tract infections and can lead to kidney damage if untreated. Additional renal anomalies such as renal agenesis (absence of one or both kidneys) and renal dysplasia

(malformed kidneys) are also observed, contributing to the spectrum of urinary system challenges in these patients.

Male reproductive issues

Cryptorchidism: This condition, affecting approximately 73% of males with CdLS, involves the failure of one or both testes to descend into the scrotum (*61*). Cryptorchidism can complicate fertility and increases the risk of testicular cancer if not properly managed.

Hypoplastic genitalia and hypospadias: 57% of males with CdLS may also exhibit underdeveloped genitalia and 9% may exhibit hypospadias, where the opening of the urethra is on the underside of the penis rather than at the tip (*61*). These anomalies can affect urinary function and sexual development.

Female Reproductive Issues

Bicornuate uterus: Observed in a subset of 5 out of 20 female patients, a bicornuate uterus has two separate uterine cavities, which can lead to complications in pregnancy such as increased risk of miscarriage and preterm labor (*77*). This malformation can also cause significant abdominal discomfort.

Other malformations: Females may also experience other anatomical variations such as vaginal atresia (where the vagina is abnormally closed or absent) or anomalies in the development of the ovaries and fallopian tubes (*78*), which can impact reproductive health and menstrual function.

Cardiovascular

Prevalence of congenital heart defects

Approximately 30% of individuals with CdLS are affected by congenital heart defects (*79*). These defects can significantly impact the overall health and quality of life of these individuals, necessitating careful monitoring and management from infancy.

Types of heart defects

Pulmonic stenosis: Pulmonary stenosis has reported it as the second most common type of heart defect in these patients, after septal defects (*79*). This involves the narrowing of the pulmonary valve or pulmonary artery that slows the blood flow from the heart to the lungs.

Tetralogy of Fallot: There is a documented case involving a 21-year-old male with CdLS who had an unrepaired Tetralogy of Fallot (TOF) (*80*). TOF is a complex heart defect that combines four heart problems: a ventricular septal defect, pulmonary stenosis, right ventricular hypertrophy, and an overriding aorta. TOF requires surgical intervention often soon after birth to correct the defects and improve oxygenation.

Ventricular septal defect and atrial septal defect: Both ventricular septal defects (VSD) and atrial septal defects (ASDs) have been observed in individuals with CdLS. These involve openings between the right and left chambers of the heart (ventricles for VSD and atria for ASD), leading to blood mixing between the systemic and pulmonary circulations. These are among the most common congenital defects and can range from mild to severe. An analysis of congenital heart defects in a cohort of 87 patients with CdLS identified a significant incidence of VSD and other structural congenital heart defects (*81*). Another study involving 149 patients with CdLS found that 34.9% had congenital heart defects, with septal defects (both ASD and VSD) comprising a significant portion of these cases (*79*).

Coarctation of the aorta: Coarctation of the aorta has been observed in individuals with CdLS. This is a narrowing of the aorta that can impede the flow of blood to the lower part of the body and increase blood pressure in the upper body. A documented case involves a preterm infant presenting with coarctation of the aorta and a perimembranous ventricular septal defect (*82*).

Aortic valve anomalies: Aortic valve anomalies have been observed in individuals with CdLS. Abnormalities in the aortic valve, such as bicuspid aortic valve, can affect blood flow from the heart into the aorta and onward to the rest of the body. One documented case describes a patient with CdLS who had a severely calcified bicuspid aortic valve and suggests a possible link between altered Notch signaling, which is implicated in CdLS, and the development of calcific bicuspid aortic valve disease (*83*).

Double-outlet right ventricle: Double outlet right ventricle (DORV) has been observed in individuals with CdLS. In this condition, both the aorta and the pulmonary artery emerge from the right ventricle, which can disrupt the normal flow of blood through the heart. One case reports a patient with CdLS who presented with both DORV and an aortopulmonary window, marking a unique combination of congenital heart defects in this syndrome (*84*).

Immunological issues

Individuals with CdLS frequently exhibit various types of immune deficiencies, which can range from mild to severe. These deficiencies often manifest as decreased immune response capabilities, leading to increased susceptibility to infections, particularly respiratory and ear infections. The immune system's compromised ability to fight off pathogens results in frequent infections, which can become chronic and recurrent, complicating the overall health picture of these individuals.

Cellular immunity issues: Patients with CdLS often experience recurrent infections, indicating potential underlying immunodeficiency. Some individuals may have defects in cellular immunity, particularly involving T-cells, which can affect the body's ability to respond to viral infections and certain fungi and protozoa. One study assessed infectious histories and conducted full immune evaluations in patients with CdLS, revealing decreased percentages of T regulatory cells and T follicular helper cells (*85*).

Chronic respiratory issues: The frequent respiratory infections that arise from immune deficiencies can lead to chronic lung problems such as bronchitis or pneumonia (*85*). These conditions can be particularly severe in CdLS due to the structural abnormalities of the airways that some individuals may have, compounding the difficulty in managing respiratory health. A 13-month-old boy with CdLS experienced repeated lower respiratory tract infections and eventually succumbed to bronchopneumonia at 15 months of age (*86*). Another case reported respiratory complications postoperatively in a 19-month-old female with CdLS who presented for palatoplasty (*87*). The complications included respiratory distress.

Hematologic issues

Thrombocytopenia: Thrombocytopenia, characterized by an abnormally low number of platelets, is a common hematologic issue in CdLS (*88, 89*). Platelets are critical for blood clotting, and their deficiency can lead to increased bleeding and bruising. While thrombocytopenia often resolves after infancy, it can persist in some individuals, leading to ongoing risks of hemorrhage and requiring careful monitoring and management.

Other features

Low-pitched cry: A low-pitched cry is observed in 75% of children with CdLS, typically indicating more severe manifestations of the syndrome (*61*). Multiple cases, including one

describing the cry as low-pitched, weak, and growling (*90*), and another noting its low pitch(*91*), underscore its common occurrence in these patients.

Hypoplastic nipples and umbilicus: About 50% of individuals with CdLS exhibit underdeveloped (hypoplastic) nipples and a poorly formed umbilicus (belly button) (*61, 92*).

De novo mutations as the primary cause of CdLS

The sporadic occurrence of CdLS in families suggested to clinicians and geneticists that de novo mutations – genetic alterations that appear for the first time in a single family member – might be the primary cause. With the advent of more sophisticated genetic technologies in the late 1990s and early 2000s, researchers gained the ability to precisely identify and characterize the mutations associated with CdLS. By comparing the DNA of children diagnosed with CdLS with that of their unaffected parents, geneticists consistently identified mutations linked to CdLS that were only present in the affected children (*36*). This strongly reinforced the hypothesis that most CdLS cases are the result of de novo mutations, most likely originating in a parent's germ cell or in the fertilized egg itself.

While most cases of CdLS arise from de novo mutations, there is a subset that demonstrates familial inheritance, though these instances are comparatively rare. One of the earliest documented cases of familial occurrence of CdLS was in 1974, involving two siblings both diagnosed with the syndrome (*93*). This early case suggested a potential genetic basis for CdLS, possibly indicative of an autosomal dominant pattern, although the precise genetic mechanisms remained unclear at the time. Further evidence of familial transmission emerged in 1985 when a study reported the direct familial transmission of CdLS in a mother and her daughter (*34*), marking the first documented instance of such direct lineage transmission. Another significant report from the same year documented CdLS across multiple generations

within the same family, highlighting variable expression of the syndrome's features among family members (*94*). The repeated observation of CdLS in successive generations brought to light the likelihood that milder cases may frequently be underdiagnosed or entirely overlooked, particularly when symptoms are subtle or atypical.

These examples indicate that familial inheritance of CdLS may occur when a parent carries a mild form of the syndrome or a germline mosaic mutation. Such mutations might not manifest prominently in the parent but can be passed on to the offspring. In instances where a parent displays a subclinical or milder phenotype – their symptoms are too subtle to be easily recognized – and they may unknowingly transmit the mutation to their children. The resulting more pronounced symptoms in the offspring confirm that the mode of familial inheritance is autosomal dominant, as the children display characteristic features of CdLS having inherited only one mutated copy of the relevant gene.

Mutations in cohesin-related proteins cause CdLS

The discovery of genetic mutations associated CdLS has progressed significantly over the years, with several key findings made incrementally. The first major breakthrough occurred in 2004 when it was identified that mutations in the *NIPBL* gene as a primary cause of CdLS (*5*). This discovery was pivotal as it demonstrated that approximately 50% of all CdLS cases (and the severest cases) were due to mutations in this gene. Collectively, these mutations typically result in the production of a functionally null protein, supporting the theory that the majority of CdLS cases are caused by *NIPBL*-haploinsufficiency (*95*). This deficiency is significant because *NIPBL* encodes a protein that plays a role in the loading of the cohesin complex onto DNA (*6*). It does this by forming a complex with another protein called MAU2. Together, they facilitate the

attachment of the cohesin complex's rings around chromatin, a process that enables the cohesin complex to maintain sister chromatid cohesion and regulate gene expression.

Over a few years following the initial discovery of mutations in *NIPBL*, mutations in several other genes were identified as causes of CdLS. These included, in chronological order, mutations in *SMC1A* (*4*)*, SMC3* (*4*)*, HDAC8 (96),* and *RAD21 (97)*. Each of these genes accounts for less than 5% of CdLS cases each and are typically associated with milder manifestations compared to mutations in *NIPBL*. Like the mutations found in *NIPBL*, mutations in these genes often result in a functionally null protein, pointing to haploinsufficiency as a likely mechanism underlying CdLS.

Interestingly, like *NIPBL*, all these genes are either components of or interact with the cohesin complex. For instance, *SMC1A* and *SMC3* encode core parts of the cohesin complex (*7*). They connect with each other via their hinge domains at one end and interact with a kleisin subunit at their head domains to form the cohesin complex's ring-like structure that encircles DNA strands. HDAC8 plays a regulatory role by deacetylating SMC3, which controls the cohesin complex's retention on chromatin during the cell cycle and aids in its release post-mitosis *(98)*. *RAD21*, which encodes the kleisin subunit of the cohesin complex, connects the head domains of SMC1 and SMC3, completing the cohesin complex's ring structure (*97*). This gene also regulates the attachment and detachment of the cohesin complex from chromatin. When RAD21 is cleaved at the onset of anaphase it results in the opening of the cohesin ring and the release of sister chromatids. Together, mutations in *NIPBL, SMC1A, SMC3, HDAC8*, and *RAD21* reveal a genetic landscape where disruptions in various components of the cohesin pathway contribute to the clinical features of CdLS.

The structure of the cohesin complex

The cohesin complex is composed of four proteins: two Structural Maintenance of Chromosomes (SMC) proteins, SMC1A and SMC3; one kleisin protein, RAD21; and one of three variant stromal antigen (SA) proteins – SA1, SA2, or SA3 – each with specific roles in different cellular contexts (*7*). SMC1A and SMC3 are characterized by their long, antiparallel coiled-coil arms, with a hinge domain at one end and an ATPase head domain at the other (*99*). Together, SMC1A and SMC3 form a V-shaped structure at their hinge domains, with their ATPase head domains at their open ends (*100*). These head domains come together by connecting with RAD21, thereby completing the ring structure of the cohesin complex. The SA proteins (SA1, SA2, or SA3) interact with the cohesin complex by interacting with RAD21 (*101*). They facilitate the integration of the cohesin complex at specific chromosomal locations and times like telomeres (SA1), centromeres (SA2 and SA3), or during meiosis (SA3), respectively (*102-104*).

The process by which cohesin initially encircles chromatin is still under investigation, but two main models have been proposed: the topological loading model and the handcuff model. In the topological loading model, the cohesin ring opens at the interface between the SMC heads and RAD21 during DNA replication, when chromatin is more accessible, allowing it to encircle the DNA (*105*). Alternatively, the handcuff model suggests that two separate cohesin complexes may initially bind to DNA and subsequently link via their RAD21 subunits to form a larger ring that can encompass chromatin (*106*). It is thought that the NIPBL-MAU2 complex induces conformational changes at the SMC3-RAD21 interface to transiently open the cohesin ring (*107*). This complex not only facilitates the opening but also maintains the ring in this state during loading. After cohesin is loaded onto DNA, the dissociation of NIPBL-MAU2 allows the

ring to close (*108*), securely encircling the DNA and ensuring proper chromosomal cohesion and separation during cell division.

The function of the cohesin complex

Sister chromatid cohesion

Cohesin was originally recognized for its role in sister chromatid cohesion during mitosis (*7*). It begins to associate with chromatin in the G1 phase of the cell cycle and its association extends into the S-phase, where it plays a role in holding replicated DNA together as it condenses into sister chromatids (*109*). During prophase, the first stage of mitosis, cohesin starts to dissociate from the arms of the sister chromatids, a process that is facilitated by the phosphorylation of RAD21 and/or SA1 (*7*). Centromere cohesin remains protected by ECO1, PDS5, WALP1, and Sororin until the onset of anaphase (*110*). At anaphase, the enzyme separase (ESP1) is activated and cleaves RAD21, leading to cohesin's degradation and the consequent separation of sister chromatids (*111*). Intriguingly, despite its pivotal role in cell division, significant defects in sister chromatid cohesion or abnormalities in mitosis have not been observed in either CdLS patients or animal models of CdLS, including *Nipbl*-haploinsufficient mice (*10, 112*). This suggests that cohesin serves additional functions beyond its well-known role in sister chromatid cohesion.

Global transcriptional regulation

Studies on cell lines derived from individuals with CdLS and animal models lacking proteins involved in the cohesin complex has yielded compelling evidence for their role in gene regulation. Initial insights into the involvement of the cohesin complex in gene expression emerged from studies on gene silencing in yeast, and on long-range gene activation in Drosophila. In yeast, cohesin was found to establish the boundaries defining silenced chromatin

domains (*8*), while in Drosophila, *Nipped-B* was shown to be required for long-range activation of homeobox protein genes (*113*). Subsequent investigations revealed that cohesin binds a subset of DNase I hypersensitive sites, many of which contain sequence motifs resembling those recognized by CTCF, a DNA-binding protein implicated in enhancer blocking function and boundary-element activity (*114*). Further elucidating this, another study identified an ESC specific binding pattern that is characterized by CTCF independent co-localization of the cohesin complex with pluripotency related transcription factors (*115*). Preceding this, a study demonstrated a physical and functional connection between mediator (a transcriptional coactivator), cohesin, and the enhancers and core promoters of active genes in murine embryonic stem cells (*116*). Collectively, these studies progressively highlight cohesin's multifaceted roles in gene expression regulation, likely mediated through its influence on chromatin structure and dynamics.

Nipbl-haploinsufficiency and its impact on global gene expression regulation

Nipbl-haploinsufficiency, the most common form of CdLs, leads to alterations in the expression of many hundreds to thousands of genes (*9-12*). Interestingly, many of the affected genes are regulated by long-distance enhancers (*117*), aligning with the emerging concept of NIPBL and cohesin as critical determinants in DNA-looping (*116, 118-120*). Most gene expression changes in *Nipbl*-deficient animals are small, usually less than 2-fold. Though likely to be inconsequential individually, such small changes can act collectively to produce structural and functional defects. For example, in zebrafish, joint depletion of two developmental genes downregulated by *nipbl*-deficiency produced a CdLS-like phenotype (*11*), suggesting that altered gene expression is the ultimate cause of developmental and physiological abnormalities. Thus, CdLS exemplifies a class of genetic disorders known as "transcriptomopathies" (*121*), in which

the cumulative or synergistic effects of minor disturbances to gene expression lead to developmental abnormalities.

A mouse model for the *Nipbl*-haploinsufficient form of CdLS

In 2009, a significant milestone in comprehending the molecular underpinnings of CdLS was achieved with the creation of a mouse model of *Nipbl*-haploinsufficiency (*10*). This was accomplished by generating mice heterozygous for a gene-trap mutation upstream of the first coding exon of *Nipbl*. RNase protection assays revealed that *Nipbl* levels in mutant mice were approximately 72–82% of wildtype levels in adult liver and around 70% in embryonic brain tissue. Phenotypically, $Nipbl^{+/-}$ mice mirrored numerous aspects of human CdLS, exhibiting prenatal and postnatal growth retardation, cardiac septal defects, delayed bone development, a lean body phenotype, microbrachycephaly accompanied by characteristic craniofacial alterations, behavioral anomalies, ophthalmological irregularities, cerebellar hypoplasia, and hearing impairments. In some cases, the $Nipbl^{+/-}$ mouse exhibited striking parallels with CdLS, including the fluctuating growth rates during early postnatal life (*122*) and the distinctive upturned nose (*123*). However, certain common CdLS features, such as upper limb reduction and fusion abnormalities, observed in 30–50% of affected children (*57, 60*), were absent in the mouse model. Furthermore, while atrial septal defects were evident in the $Nipbl^{+/-}$ mouse heart, ventricular septal defects, equally prevalent in CdLS (*81*), were absent. Nevertheless, the $Nipbl^{+/-}$ mouse emerges as a robust animal model that captures many CdLS traits. Cellular and molecular analyses of $Nipbl^{+/-}$ cells and tissues also captured subtle yet widespread alterations in gene expression, particularly in genomic regions where transcription is known to be governed by long-range chromosomal interactions. These data underscore the idea that the collective impact of

numerous minor transcriptional changes likely contributes to the developmental abnormalities observed in CdLS.

Early murine heart development

The development of the heart is a highly intricate and coordinated process that arises from multiple tissue lineages. The heart is primarily derived from the mesoderm, one of the three germ layers in early embryogenesis (*124*). Within the mesoderm, the first cardiac cells to differentiate form the initial cardiac crescent and are termed the first heart field (FHF) (*20*). As development proceeds, cells from the second heart field (SHF) contribute to the cardiac crescent in the elongation and looping of the heart tube, giving rise to the right ventricle (but not the left), outflow tract, atria, and atrial/ventricular septa (*20*). Neural crest cells, which originate from the ectoderm, also play a role in the development of the outflow tract and the division of the primitive heart tube into the aorta and pulmonary trunk (*125*). Collectively, these diverse tissue lineages orchestrate heart morphogenesis.

Before the onset of gastrulation, the first heart progenitors locate to the intraembryonic epiblast at E6.5. Here, they are not yet committed to a cardiac fate, as they are still clonally related to the endoderm (*126*). When early gastrulation begins, these heart progenitors ingress at the mid-streak stage and migrate to the anterior region of the primitive streak (*127*). During this time, they express the basic helix-loop-helix transcription factor *Mesp1.* Genetic tracing has revealed that these *Mesp1*-expressing cells will give rise to more than 70% of all the cells of the heart (*128*). Clonal analysis of the same cell population has revealed that two distinct cell lineages contribute to the early embryonic heart: the first cell lineage forms the future left ventricle as well as other compartments of the heart; and the second cell lineage forms the future right ventricle, outflow tract, and atria (*129*).

Segregation of these two cardiac lineages occurs during late gastrulation. Both ingress sequentially from the *Mesp1*-expressing heart progenitors in the primitive streak into the lateral plate mesoderm: cells of the first lineage ingress between E6.25-E6.75 and cells of the second lineage ingress between E6.75-E7.25 (*130*). As gastrulation comes to an end, the cells of these two lineages move anteriorly and locate to both sides of the embryo midline under the head folds at E7.5, where they will coalesce into the first morphologically visible structure of the heart, the cardiac crescent at E7.75 (*131*). Cells of the first lineage differentiate into the epithelial structure of the cardiac crescent and cells of the second lineage differentiate into the medial and posterior regions of the crescent (*132*). Since these regions are derived from the first and second cell lineages, they have been subsequently referred to as the first and second heart fields (*133*). Cells from both of these fields contribute cells to almost all of the structures of the developing heart in accordance with contributions from the first and second cardiac cell lineages.

Insights into the origins of congenital heart defects from studies of *Nipbl*-haploinsufficient mice

One study of *Nipbl*-haploinsufficiency in mice asked in which cell lineages *Nipbl*-haploinsufficiency is necessary and sufficient to cause congenital heart defects (CHDs) similar to those observed in entirely *Nipbl*-haploinsufficient mice (*18*). In that study, *Nipbl*-haploinsufficiency in neural crest cells and cells of the cardiac crescent. *Nipbl*-haploinsufficiency in neural crest cells alone did not result in CHDs (*18*). However, *Nipbl*-haploinsufficiency in cardiac crescent cells, which consist of both FHF and SHF cells, did result in CHDs (*18*). Interestingly, the heart abnormalities observed in $Nipbl^{+/-}$ mice (which include ASDs, incomplete fusion of the ventricular septum, and right ventricle hypoplasia) are found in tissues derived from the SHF. In addition, other studies have demonstrated that disruptions in the second heart field

can result in outflow tract defects (*134*), atrial and ventricular septal defects (*133*), and right ventricular hypoplasia (*135*). Collectively, these findings suggest that the SHF may be a potential source of CHDs in *Nipbl*$^{+/-}$ mice.

Studies of CHDs in *Nipbl*-haploinsufficient mice also suggest that the cause(s) of CHDs may occur much earlier in development than expected. 30% of *Nipbl*$^{+/-}$ mice display CHDs, primarily atrial septal defects. *Nipbl*$^{+/-}$ mice also exhibit heart abnormalities at earlier developmental stages: At embryonic day 13.5 (E13.5), 70% of *Nipbl*$^{+/-}$ mice display delays in ventricular septal fusion, whereas three days earlier, at E10.5, an overwhelming 100% exhibit right ventricle hypoplasia (*18*). Even earlier, these mice display reduced expression of transcription factors crucial for early heart progenitor cell growth and differentiation: specifically, *Nkx2-5* in the second heart field at E7.75, and *Mesp1* in mesoderm progenitors at E7.5 (*18*). In *nipbl*-morphant zebrafish, defects in the initial migration of cardiogenic mesoderm can be detected as early as 18 hours post-fertilization (*11*). These findings suggest the importance of focusing on the earliest stages of development – particularly during gastrulation, when the three primary germ layers form and the earliest progenitor cells of major tissues and organs begin to differentiate (*136*) – to investigate the origins of CHDs and potentially other birth defects.

These insights prompted the studies discussed in Chapters 2 and 3.

CHAPTER 2

Reduced homeodomain gene expression in *Nipbl*$^{+/-}$ mice made haploinsufficient for *Isl1* reveals novel quantitative risk determinants for congenital heart defects

Abstract

Approximately 30% of individuals with the *Nipbl*-haploinsufficient form of Cornelia de Lange Syndrome (CdLS), exhibit congenital heart defects (CHDs). In this study, we assessed the role of *Nipbl*-haploinsufficiency in the second heart field (SHF) on the development of CHDs using light sheet fluorescence microscopy (LSFM) and two SHF-Cre recombinase-expressing mouse lines: *Mef2c-Cre* and *Isl1*$^{Cre/+}$. Remarkably, only *Isl1*$^{Cre/+}$ driven *Nipbl*-haploinsufficiency in SHF resulted in CHDs, likely due to extensive *Isl1-Cre* recombination in non-cardiac tissues and/or reduced *Isl1* expression, the latter caused by knock-in mediated haploinsufficiency. Surprisingly, mice made globally haploinsufficient for both *Isl1* and *Nipbl* (*Nipbl*$^{+/-}$*; Isl1*$^{+/-}$) displayed a broader spectrum and higher frequency of CHDs than mice haploinsufficient for either gene alone. RNA sequencing of E10.5 hearts revealed that *Nipbl*-haploinsufficiency alone upregulates *Isl1*, and that this appears to provide a protective effect against the development of more severe CHDs, as demonstrated by fewer and less severe CHDs in *Nipbl*$^{+/-}$ mice compared to *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ mice. Additionally, significant changes in the expression of homeodomain genes, *Pitx1*, *Pax6*, and *Hoxc4* were observed; *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts failed to upregulate these genes to the same extent as *Nipbl*$^{+/-}$ hearts. These findings suggest new quantitative risk determinants for the development of CHDs in both CdLS and nonsyndromic cases.

Introduction

Congenital heart defects (CHD) are structural abnormalities of the heart that are present at birth. In the United States, CHDs affect nearly 1% of all newborns and are the leading cause of birth defect-associated infant mortality and morbidity (*13-15*). The causes of CHDs are largely unknown, with mutations to single genes accounting for less than 20% of cases (*16*). The majority of CHDs are thought to arise from multifactorial and/or polygenic origins (*137, 138*), which has complicated the creation of animal models for studying CHDs.

Studies of Cornelia de Lange Syndrome (CdLS) have shed some light on the causal factors behind CHDs. CDLS is a multisystem birth defects disorder that occurs in approximately 1 in 10,000 to 1 in 30,000 live births (*36*). Individuals with CdLS exhibit defects including intellectual disability, distinctive facial features, growth retardation, abnormalities of the upper limbs, gastrointestinal malformations, genitourinary anomalies, and cardiac defects (*64*). About 30% of individuals with CdLS have CHDs, with ventricular septum defects (VSD) and atrial septum defects (ASD) being the most common (*17*).

Most CdLS cases - more than 55% - are caused by haploinsufficiency for the gene *Nipped-B-like* (*NIPBL*) (*95*). The *NIPBL* gene encodes a universally conserved protein that plays a role in loading the cohesin complex onto chromosomes (*139-142*). While cohesin is widely recognized for its role in maintaining sister chromatid cohesion during mitosis (*7*), sister chromatid cohesion is largely unaffected in CdLS. In recent years, the cohesin complex has been increasingly acknowledged for its role in regulating global gene expression. In support of this view, both CdLS patient cells and *Nipbl*-deficient animal models exhibit alterations in the expression of hundreds to thousands of genes in all examined tissues (*9-12*).

Interestingly, most of these gene expression changes are small, typically less than 2-fold change. Though likely inconsequential individually, studies in *Nipbl*-deficient mice and zebrafish indicate that these changes contribute to the structural and functional defects seen in CdLS (*11*). Thus, CdLS exemplifies an emerging class of genetic disorders – recently coined "transcriptomopathies" – in which quantitative variations in multiple gene product levels lie at the root of developmental abnormalities. Thus, such disorders provide a window into the kinds of multifactorial interactions potentially underlying polygenic traits, including that of CHDs.

Among the features of CdLS that are phenocopied in *Nipbl*-deficient animal models are CHDs. In $Nipbl^{+/-}$ mice, these include atrial septal defects (ASDs) (*18*), which occur at the same frequency as that of CdLS patients (30%) (*17*) and which are also among the most common kind of CHD in non-syndromic cases in humans (*19*). Thus, the $Nipbl^{+/-}$ mouse serves as an excellent model for studying the developmental origins of CHDs. Intriguingly, recent studies in $Nipbl^{+/-}$ mice have found that ASDs are preceded by heart abnormalities at earlier stages of development: these include incomplete fusion of the ventricular septum with the cardiac cushion in 77% of hearts at E13.5, smaller right ventricles in 100% of hearts at E10.5, and reduced expression of early cardiac transcription factors, *Nkx2.5* at E7.57 and *Mesp1* at E7.5 (*18*). Together, these findings suggest that the cause(s) of CHDs in $Nipbl^{+/-}$ mice may originate very early in heart development.

The development of the heart is a highly intricate and coordinated process that arises from multiple tissue lineages. The heart is primarily derived from the mesoderm, one of the three germ layers in early embryogenesis (*124*). Within the mesoderm, the first cardiac cells to differentiate form the initial cardiac crescent and are termed the first heart field (FHF) (*20*). As development proceeds, cells from the second heart field (SHF) contribute to the cardiac crescent in the elongation and looping of the heart tube, giving rise to the right ventricle (but not the left),

outflow tract, atria, and atrial/ventricular septa (*20*). Neural crest cells, which originate from the ectoderm, also play a role in the development of the outflow tract and the division of the primitive heart tube into the aorta and pulmonary trunk (*125*). Collectively, these diverse tissue lineages orchestrate heart morphogenesis.

In a previous study, we asked in which cell lineages *Nipbl*-haploinsufficiency is necessary and sufficient to cause CHDs similar to those observed in entirely *Nipbl*-haploinsufficient mice (*18*). In that study, we used a *Nipbl* allelic series containing a conditional/invertible gene trap and Cre recombinase expressing mouse lines, to create *Nipbl*-haploinsufficiency in neural crest cells (*Wnt-Cre*) and cells of the cardiac crescent (*Nkx2-5-Cre*). *Nipbl*-haploinsufficiency in neural crest cells alone did not result in CHDs (*18*). However, *Nipbl*-haploinsufficiency in cardiac crescent cells, which consist of both FHF and SHF cells, did result in CHDs (*18*). Interestingly, the heart abnormalities observed in $Nipbl^{+/-}$ mice (which include ASDs, incomplete fusion of the ventricular septum, and right ventricle hypoplasia) are found in tissues derived from the SHF. In addition, other studies have demonstrated that disruptions in the second heart field can result in outflow tract defects (*134*), atrial and ventricular septal defects (*133*), and right ventricular hypoplasia (*135*). Collectively, these findings suggest that the SHF may be a potential source of CHDs in $Nipbl^{+/-}$ mice.

In this study, we used light sheet fluorescence microscopy (LSFM) to look for CHDs in mice made *Nipbl*-haploinsufficient in the SHF using two SHF-Cre recombinase expressing mouse lines: *Mef2c-Cre* (*143*) and $Isl1^{Cre/+}$ (*144*). Surprisingly, only $Isl1^{Cre/+}$ driven *Nipbl*-haploinsufficiency in the SHF resulted in CHDs, whereas *Mef2c-Cre* did not. We report that this difference could be due to extensive *Isl1-Cre* recombination in non-cardiac tissues and reduced expression of *Isl1* in $Isl1^{Cre/+}$ hearts, the latter occurring as a result of knock-in mediated

haploinsufficiency. These data led us to investigate the role of *Isl1*-haploinsufficiency in heart development in the context of global *Nipbl*-haploinsufficiency. When mice were made globally haploinsufficient for both *Isl1* and *Nipbl (Nipbl*$^{+/-}$*; Isl1*$^{+/-}$*)*, the spectrum and frequency of CHDs were enhanced beyond those observed in mice haploinsufficient for either *Isl* or *Nipbl* alone. RNA sequencing of wildtype (WT), *Nipbl*$^{+/-}$*, Isl1*$^{+/-}$*,* and *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ E10.5 hearts found that *Nipbl*-haploinsufficiency alone results in the upregulation of *Isl1*. However, double haploinsufficiency for both *Nipbl* and *Isl1* had no effect on *Isl1* expression. Since *Isl1* is only upregulated in *Nipbl*$^{+/-}$ hearts and *Nipbl*$^{+/-}$ mice show fewer and less severe CHDs than *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ mice, we suggest from these data that upregulation of *Isl1* and its homeodomain target genes involved in specifying and patterning tissues (*Pitx1, Hoxc4, and Pax6*), act as a protective mechanism against the development CHDs to the extent observed in *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ mice. Together, these results identify new quantitative risk factors for the development of CHDs in both CdLS and nonsyndromic cases.

Results

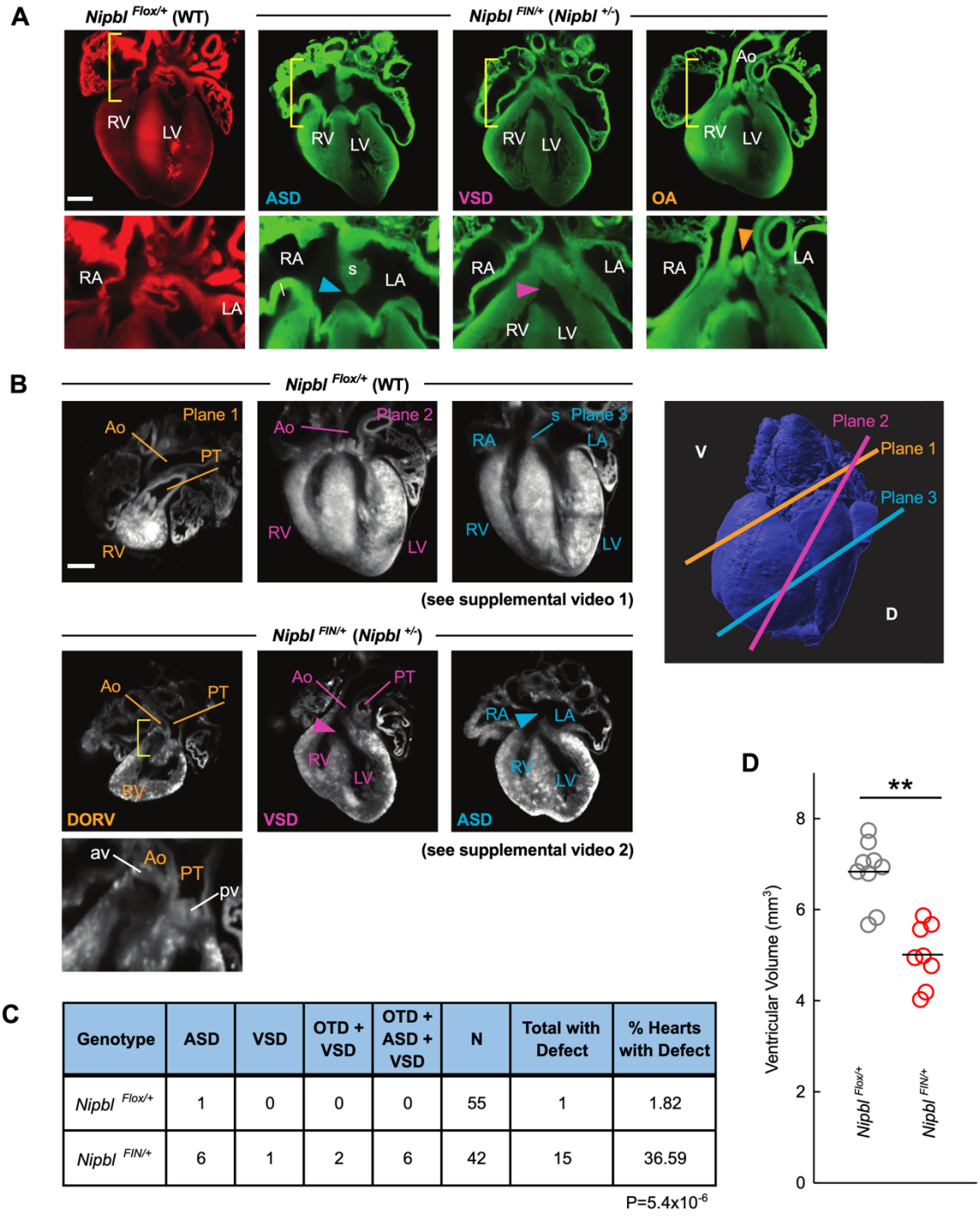
Light sheet fluorescence microscopy detects a spectrum of CHDs in *Nipbl*$^{+/-}$ mice

The traditional method of looking for CHDs in mouse hearts is histological analysis, which involves serial sectioning and staining (hematoxylin and eosin, H&E) to examine heart morphology. However, histological analysis is invasive and requires tissue sectioning, precluding repeated studies on the same specimen. In a previous study, we introduced micro-magnetic resonance imaging (μMRI) as an alternative method for CHD detection (*18*), but a limitation of μMRI its lower spatial resolution compared to histological analysis, which might overlook extremely fine anatomical details. To overcome the limitations of both methodologies, we introduce LSFM for the detection of CHDs in mice. LSFM enables rapid, high-resolution

three-dimensional imaging of biological samples with minimal phototoxicity, preserving sample integrity. Bypassing the need for invasive tissue sectioning, LSFM avoids potential artifacts and provides an almost 10-fold improvement in spatial resolution over that of μMRI (LSFM: 6-7 μm resolution voxel, μMRI: 50-70 μm), making it more adept at capturing fine anatomical details. Furthermore, when combined with fluorescent reporter lines, LSFM also offers insights into tissue-specific dynamics that contribute to the development of CHDs.

Using a *Nipbl* allelic series containing a conditional/invertible gene trap we previously developed (*18*), we generated WT ($Nipbl^{Flox/+}$) and $Nipbl^{+/-}$ ($Nipbl^{FIN/+}$) mouse littermates on a *Td-tomato-EGFP* fluorescent reporter background in which *Cre*-mediated *Nipbl*-haploinsufficiency is reported by enhanced green fluorescent protein (EGFP) expression. *Nipbl*-haploinsufficiency was mediated by *Nanog-Cre* recombination, which occurs in the earliest cells of the developing embryo – the inner cell mass (*145*). We also generated WT and $Nipbl^{+/-}$ littermates on a background without EGFP reporter expression, since LSFM can also image tissues using autofluorescence. Using LSFM, we looked for CHDs in these mice at E17.5. We detected a single ASD in WT mice and a spectrum of CHDs in $Nipbl^{+/-}$ mice, including ASDs, VSDs, and outflow tract defects (OTDs), such as overriding aorta (OA) (Fig. 2.1A), double outlet right ventricle (DORV) (Fig. 2.1B), and transposition of the great arteries (TGA) (not shown). These OTDs always occurred in conjunction with either an ASD or both an ASD and VSD (Fig. 2.1C). Previously, only ASDs were detected in $Nipbl^{+/-}$ mice by histological analysis or μMRI at the same stage (*18*). However, ASDs remain the most common CHD as detected by LSFM – 80% of $Nipbl^{+/-}$ mice with CHDs showed ASDs. In total, LSFM detected CHDs in 36.59% of $Nipbl^{+/-}$ hearts, substantially higher than what was detected in WT hearts (<2%) (Fig. 2.1C), and almost 7% higher than what was previously detected by histological analysis and

μMRI (~30%) (*18*). Since we previously observed that *Nipbl*$^{+/-}$ mice have smaller hearts than WT mice (*18*), we also used LSFM to measure the ventricular volumes of hearts from E17.5 WT and *Nipbl*$^{+/-}$ mice. *Nipbl*$^{+/-}$ hearts were 28% smaller than WT hearts (Fig. 2.1D). Thus, LSFM detects a broader spectrum of CHDs in *Nipbl*$^{+/-}$ mice than previously reported, highlighting its superior sensitivity over histological analysis and μMRI.



Genotype	ASD	VSD	OTD + VSD	OTD + ASD + VSD	N	Total with Defect	% Hearts with Defect
Nipbl $^{Flox/+}$	1	0	0	0	55	1	1.82
Nipbl $^{FIN/+}$	6	1	2	6	42	15	36.59

P=5.4x10^{-6}

Fig. 2.1. Light sheet fluorescence microscopy detects a spectrum of CHDs in *Nipbl*$^{+/-}$ mice.

(A) LSFM images of one representative *Nipbl*$^{Flox/+}$ (WT) E17.5 heart and three different *Nipbl*$^{FIN/+}$ (*Nipbl*$^{+/-}$) E17.5 hearts on a *Td-tomato-EGFP* fluorescent reporter background. *Nanog-Cre*-mediated recombination is reported by EGFP expression. *Nipbl*$^{Flox/+}$ mice showed no

recombination. All tissues showed recombination in *Nipbl*$^{FIN/+}$ mice. *Nipbl*$^{FIN/+}$ (*Nipbl*$^{+/-}$) hearts display a spectrum of heart defects. Panels in bottom row show higher magnification of areas indicated by yellow brackets in panels in top row. In high magnification panels, blue arrowhead points to atrial septal defect (ASD), pink arrowhead points to ventricular septal defect (VSD), and orange arrowhead points to overriding aorta (OA). RV, right ventricle; LV, left ventricle; Ao, aorta; RA, right atrium; LA, left atrium; S, atrial septum; PT, pulmonary trunk. Scale bar = 500 um. (B) LSFM images of one representative *Nipbl*$^{Flox/+}$ E17.5 heart and one representative *Nipbl*$^{FIN/+}$ E17.5 heart without fluorescence reporting of *Cre*-mediated recombination. Images are of three planes of one *Nipbl*$^{Flox/+}$ and one *Nipbl*$^{FIN/+}$ heart. The positions of these planes in the heart are shown in the schematic to the right. From left to right, images show the plane progressing from ventral to dorsal in the heart. The *Nipbl*$^{FIN/+}$ heart displays a spectrum of heart defects. Bottom left panel shows higher magnification of area indicated by yellow bracket in panel above it, revealing double outlet right ventricle (DORV). (C) Table of frequency of heart defects in E17.5 *Nipbl*$^{Flox/+}$ and *Nipbl*$^{FIN/+}$ hearts as detected by LSFM. ASD, atrial septal defect; VSD, ventricular septal defect. Hearts showing outflow tract defects (OTD) include one or more of the following: overriding aorta (OA), double outlet right ventricle (DORV), transposition of the great arteries (TGA). P-value from Fisher's Exact Test. (D) Ventricular volumes of E17.5 *Nipbl*$^{Flox/+}$ (N=8) and *Nipbl*$^{FIN/+}$ (N=9) hearts. Ventricular volumes were measured in ImageJ using LSFM images of hearts. Bars show means. ** represents $P \leq 0.01$. P-value from T-Test.

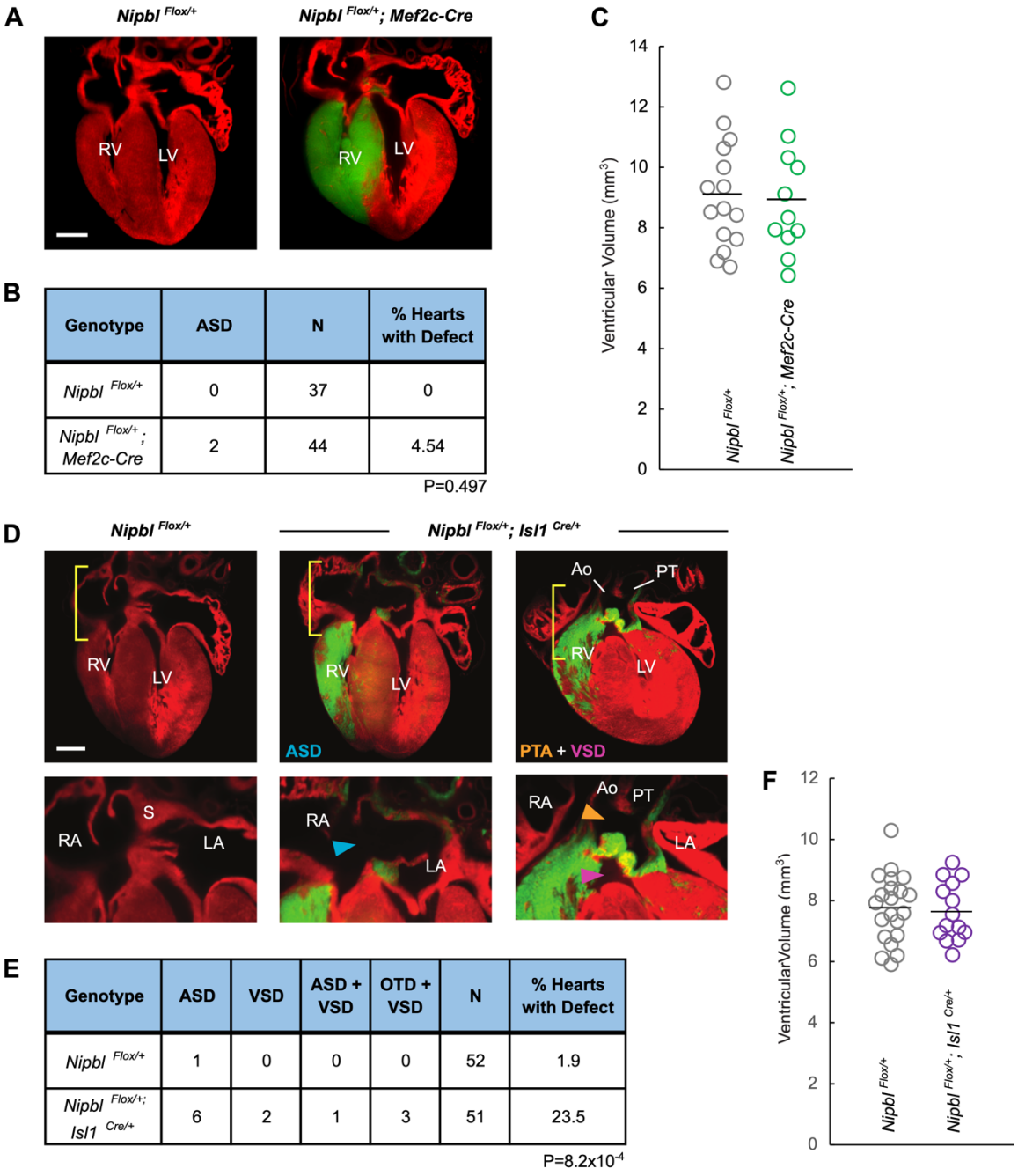
***Nipbl*-haploinsufficiency in the SHF *Mef2c-Cre* domain does not result in CHDs**

To investigate whether *Nipbl*-haploinsuffiency in the SHF alone could result in the CHDs observed in globally *Nipbl*-haploinsufficient mice, we generated E17.5 mice *Nipbl*-haploinsufficient in the SHF on a *Td-tomato-EGFP* fluorescent reporter background using the *Mef2c-Cre* mouse line so that *Mef2c-Cre* mediated *Nipbl*-haploinsufficiency is reported by EGFP expression. As expected, *Mef2c-Cre* mediated *Nipbl*-haploinsufficiency occurred in SHF derivatives, including the right ventricle and parts of the atria (Fig. 2.2A). Surprisingly, less than 5% of $Nipbl^{Flox/+}$*; Mef2c-Cre* mice showed a CHD, and when they did, they only showed ASDs, resulting in no significant difference from that of WT ($Nipbl^{Flox/+}$) mice (Fig. 2.2B). The majority of $Nipbl^{Flox/+}$*; Mef2c-Cre* mice showed normal heart development, similar to WT mice (Fig. 2.2, A and B). They also had hearts with the same size as WT mice (Fig. 2.2C). Thus, while *Mef2c-Cre* mediated *Nipbl*-haploinsufficiency in the SHF can lead to some CHDs, its impact appears limited, compared to CHD presentation in globally *Nipbl*-haploinsufficient mice.

***Nipbl*-haploinsufficiency in the SHF *Isl1-Cre* domain results in a spectrum of CHDs**

We also generated E17.5 mice *Nipbl*-haploinsufficient in the SHF using the $Isl1^{Cre/+}$ mouse line, which also occurred on a *Td-tomato-EGFP* fluorescent reporter background so that *Isl1-Cre* mediated *Nipbl*-haploinsufficiency is reported by EGFP expression. As expected, *Isl1-Cre* mediated *Nipbl*-haploinsufficiency, occurred in SHF derivatives, including the right ventricle and parts of the atria (Fig. 2.2D). $Nipbl^{Flox/+}$*;* $Isl1^{Cre/+}$ mice developed a significant number of CHDs (23.5%) compared to WT ($Nipbl^{Flox/+}$) mice, which did not develop any CHDs (Fig. 2.2E). Although the frequency of CHDs displayed by $Nipbl^{Flox/+}$*;* $Isl1^{Cre/+}$ mice was lower than that of globally *Nipbl*-haploinsufficient ($Nipbl^{+/-}$) mice (~30%), the spectrum of defects was similar, which included: ASDs, VSDs, and OTDs, including OA and persistent truncus arteriosis

(PTA) (Fig. 2.2D). Like *Nipbl*$^{+/-}$ mice (Fig. 2.1C), OTDs in *Nipbl*$^{Flox/+}$*; Isl1*$^{Cre/+}$ mice occurred in conjunction with VSDs (Fig. 2.2E). Unlike, *Nipbl*$^{+/-}$ mice, however, which showed smaller hearts than WT (Fig. 2.1D), heart size was unaffected in *Nipbl*$^{Flox/+}$*; Isl1*$^{Cre/+}$ mice compared to WT (*Nipbl*$^{Flox/+}$) (Figure 2F). Consequently, *Isl1-Cre* mediated *Nipbl*-haploinsufficiency in the SHF leads to a spectrum of CHDs reminiscent of globally *Nipbl*-haploinsufficient mice.



Genotype	ASD	N	% Hearts with Defect
Nipbl $^{Flox/+}$	0	37	0
Nipbl $^{Flox/+}$; *Mef2c-Cre*	2	44	4.54

Genotype	ASD	VSD	ASD + VSD	OTD + VSD	N	% Hearts with Defect
Nipbl $^{Flox/+}$	1	0	0	0	52	1.9
Nipbl $^{Flox/+}$; *Isl1* $^{Cre/+}$	6	2	1	3	51	23.5

P=8.2x10^{-4}

Fig. 2.2. *Nipbl* made deficient in the SHF domain using one Cre driver (*Isl1-Cre*), but not another (*Mef2c-Cre*), results in CHDs. (A) LSFM image of representative E17.5 *Nipbl*$^{Flox/+}$ and *Nipbl*$^{Flox/+}$; *Mef2c-Cre* heart on a *Td-tomato-EGFP* fluorescent reporter background. *Mef2c-Cre* mediated recombination is reported by EGFP expression. RV, right ventricle; LV, left ventricle. Scale bar = 500 uM. (B) Table of frequency of atrial septal defects (ASDs) in E17.5 *Nipbl*$^{Flox/+}$

(N=37) and *Nipbl*$^{Flox/+}$; *Mef2c-Cre* (N=44) hearts, as detected by LSFM (N=20, *Nipbl*$^{Flox/+}$, N=21, *Nipbl*$^{Flox/+}$; *Mef2c-Cre*) and histological analysis (N=17, *Nipbl*$^{Flox/+}$, N=23, *Nipbl*$^{Flox/+}$; *Mef2c-Cre*). P-value from Fisher's Exact Test. (C) Ventricular volumes of E17.5 *Nipbl*$^{Flox/+}$ (N=15) and *Nipbl*$^{Flox/+}$; *Mef2c-Cre* (N=11) hearts. Ventricular volumes were measured in ImageJ using LSFM images of hearts. Bars show means. (D) LSFM images of one representative *Nipbl*$^{Flox/+}$ E17.5 heart and two different *Nipbl*$^{Flox/+}$; *Isl1*$^{Cre/+}$ E17.5 hearts on a *Td-tomato-EGFP* fluorescent reporter background. *Isl1-Cre* mediated recombination is reported by EGFP expression. *Nipbl*$^{Flox/+}$; *Isl1*$^{Cre/+}$ hearts display a spectrum of heart defects. Panels in bottom row show higher magnification of areas indicated by yellow brackets in panels in top row. In high magnification panels, blue arrowhead points to atrial septal defect (ASD), pink arrowhead points to ventricular septal defect (VSD), and orange arrowhead points to overriding aorta (OA). RV, right ventricle; LV, left ventricle; Ao, aorta; PT, pulmonary trunk; RA, right atrium; LA, left atrium; S, atrial septum. Scale bar = 500 uM. (E) Table of frequency of heart defects in E17.5 *Nipbl*$^{Flox/+}$ (N=52) and *Nipbl*$^{Flox/+}$; *Isl1*$^{Cre/+}$ (N=51) hearts, as detected by LSFM. Of the three *Nipbl*$^{Flox/+}$; *Isl1*$^{Cre/+}$ hearts showing both an outflow tract defect and ventricular septal defect (OTD + VSD), two showed an overriding aorta OTD and one showed a persistent truncus arteriosis OTD. ASD, atrial septal defect. P-value from Fisher's Exact Test. (F) Ventricular volumes of E17.5 *Nipbl*$^{Flox/+}$ (N=21) and *Nipbl*$^{Flox/+}$; *Isl1*$^{Cre/+}$ (N=14) hearts. Ventricular volumes were measured in ImageJ using LSFM images of hearts. Bars show means.

***Mef2c-Cre* and *Isl1-Cre* recombine similarly within the heart, but *Isl1-Cre* recombines extensively in non-cardiac tissues**

We wondered whether the development of CHDs in *Nipbl*$^{Flox/+}$*; Isl1*$^{Cre/+}$ mice and their limited appearance in *Nipbl*$^{Flox/+}$*; Mef2c-Cre* mice could be due to differences in the expression domains of *Mef2c-Cre* and *Isl1-Cre* during heart development. To answer this question, we visualized their expression domains on a *Td-Tomato-EGFP* reporter background from E8.5 to E17.5, so that *Cre*-mediated recombination is reported by EGFP expression. *Mef2c-Cre* and *Isl1-Cre* recombined in SHF derivates at all stages: at E8.5, cardiac crescent; at E10.5, developing right ventricle and outflow tract, and at E15.5-E17.5, ventricular/atrial septa, aorta, and pulmonary trunk (Fig. 2.3, A & B). With exception to the cardiac crescent at E8.5, *Mef2c-Cre* appeared to recombine more extensively in all stages of the developing heart than *Isl1-Cre.* Due to the slight expansion of the Cre-expressing domain observed in *Mef2c-Cre* hearts compared to *Isl1*$^{Cre/+}$ hearts, one might expect *Nipbl*$^{Flox/+}$*; Mef2c-Cre* hearts to develop CHDs as in *Nipbl*$^{Flox/+}$*; Isl1*$^{Cre/+}$ hearts. However, as we showed in Fig. 2, this is not what we observed.

In a previous study, we showed that *Nipbl*-haploinsufficiency in non-cardiac tissues can contribute to the development of CHDs (*18*). We therefore analyzed the recombination domains of Cre-mediated *Nipbl*-haploinsufficiency across the whole bodies of *Nipbl*$^{Flox/+}$*; Mef2c-Cre* and *Nipbl*$^{Flox/+}$*; Isl1*$^{Cre/+}$ mouse embryos E10.5, the earliest stage before which CHDs were observed in mice of either genotype. In these mice of both genotypes, Cre-recombination is reported by lacZ expression, which turns tissues blue when treated with X-gal. No recombination was observed in WT (*Nipbl*$^{Flox/+}$) mice (Fig. 2.3C). In *Nipbl*$^{Flox/+}$*; Mef2c-Cre* mice, *Cre* recombination was observed in SHF derivates (right ventricle and outflow tract), as expected, as well as outside the heart in the pharyngeal arches (Fig. 2.3D). In *Nipbl*$^{Flox/+}$*; Isl1*$^{Cre/+}$ mice, Cre

recombination also occurred in the same SHF derivates as *Nipbl*$^{Flox/+}$*; Mef2c-Cre* mice (Fig. 2.3E). However, *Isl1-Cre* recombination occurred extensively outside the heart, occurring in tissues including the hindlimbs, dorsal root ganglia, pharyngeal arches, trigeminal ganglia, and diencephalon. Thus, our analyses reveal that while both Cre lines target SHF derivatives, the *Isl1-Cre* line exhibits a broader recombination pattern, particularly outside the heart, which could account for the greater incidence of CHDs in *Nipbl*$^{Flox/+}$*; Isl1*$^{Cre/+}$ mice compared to their *Nipbl*$^{Flox/+}$*; Mef2c-Cre* counterparts.

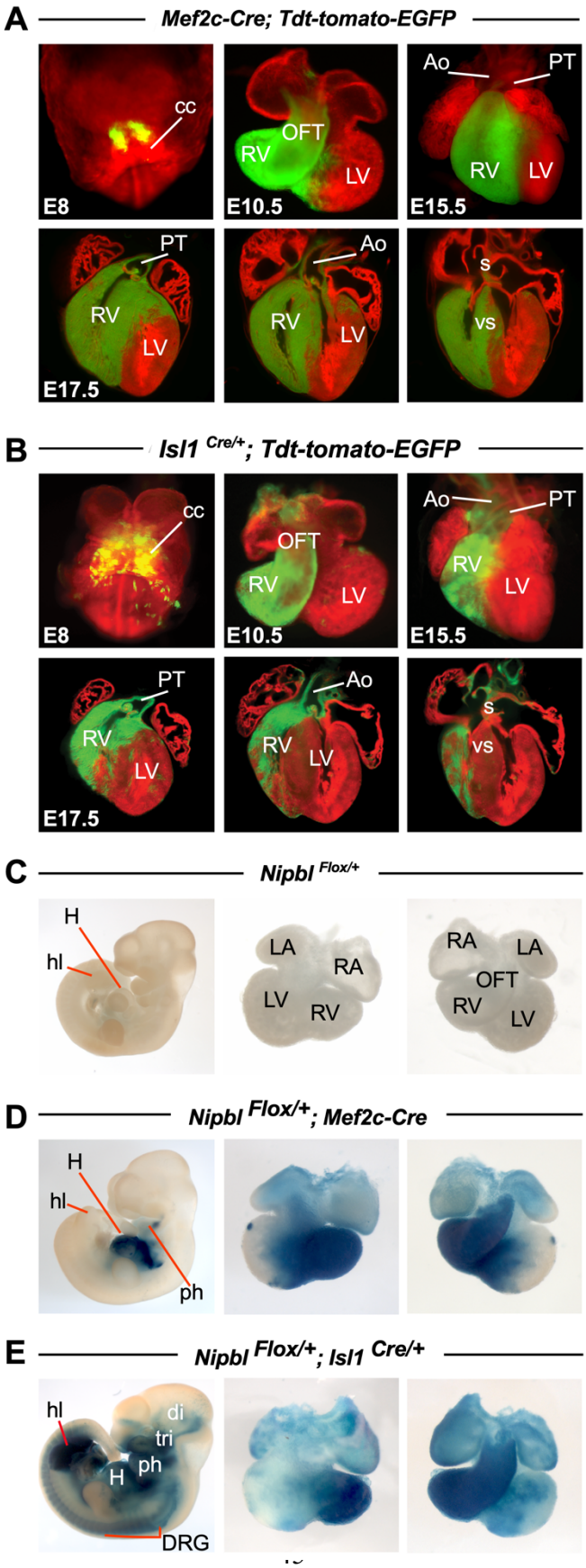
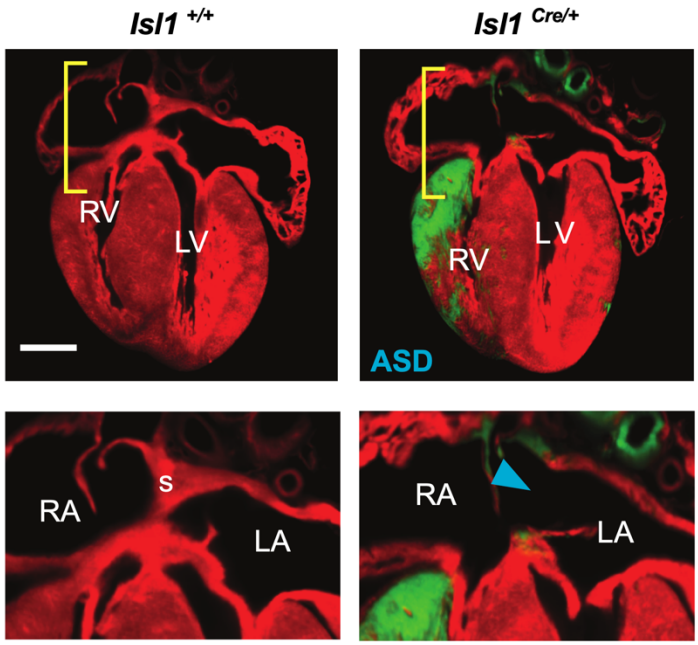

A
Mef2c-Cre; Tdt-tomato-EGFP
cc
E8
OFT
RV
LV
E10.5
Ao
PT
RV
LV
E15.5
PT
RV
LV
E17.5
Ao
RV
LV
s
vs
B
Isl1 Cre/+; Tdt-tomato-EGFP
cc
E8
OFT
RV
LV
E10.5
Ao
PT
RV
LV
E15.5
PT
RV
LV
E17.5
Ao
RV
LV
s
vs
C
Nipbl Flox/+
H
hl
LA
RA
LV
RV
RA
LA
OFT
RV
LV
D
Nipbl Flox/+; Mef2c-Cre
H
hl
ph
E
Nipbl Flox/+; Isl1 Cre/+
hl
di
tri
ph
H
DRG

Fig. 2.3. *Mef2c-Cre* and *Isl1-Cre* recombine similarly within the heart, but *Isl1-Cre* recombines extensively in non-cardiac tissues. (A) LSFM images of representative *Mef2c-Cre* E8.5 whole embryos, E10.5, E15.5, and E17.5 hearts on a *Td-tomato-EGFP* fluorescent reporter background. *Mef2c-Cre* mediated recombination is reported by EGFP expression. Bottom row shows three different planes of one E17.5 heart progressing from ventral to dorsal from left to right. CC, cardiac crescent; RV, right ventricle; LV, left ventricle; OFT, outflow tract; Ao, aorta; PT, pulmonary trunk; S, atrial septum; VS, ventricular septum. (B) LSFM images of representative $Isl1^{Cre/+}$ E8.5 whole embryos, E10.5, E15.5, and E17.5 hearts on a *Td-tomato-EGFP* fluorescent reporter background. *Isl1-Cre* mediated recombination is reported by EGFP expression. Bottom row shows three different planes of one E17.5 heart progressing from ventral to dorsal from left to right. CC, cardiac crescent; RV, right ventricle; LV, left ventricle; OFT, outflow tract; Ao, aorta; PT, pulmonary trunk; S, atrial septum; VS, ventricular septum. X-gal stains of representative E10.5 (C) $Nipbl^{Flox/+}$, (D) $Nipbl^{Flox/+}$; *Mef2c-Cre*, and (E) $Nipbl^{Flox/+}$; $Isl1^{Cre/+}$ embryos and hearts. H, heart; hl, hindlimb; LA, left atrium; RA, right atrium; LV, left ventricle; RV, right ventricle; OFT, outflow tract; DRG, dorsal root ganglia; ph, pharyngeal arches; tri, trigeminal ganglia; di, diencephalon.

Isl1-haploinsufficiency alone results in CHDs

A major discrepancy between the *Mef2c-Cre* and *Isl1*$^{Cre/+}$ lines lies in their generation. The *Mef2c-Cre* line was developed using a traditional bacterial artificial chromosome transgenic approach (*143*), whereas the *Isl1*$^{Cre/+}$ line is a knock-in transgenic that targeted the start codon of the *Isl1* locus (*144*). Thus, *Isl1*$^{Cre/+}$ mice are haploinsufficient for *Isl1*. Indeed, qRT-PCR for *Isl1* in E10.5 WT (*Isl1*$^{+/+}$) and *Isl1*$^{Cre/+}$ hearts found that *Isl1*$^{Cre/+}$ hearts expressed *Isl1* at levels more than 60% lower than that of WT (*Isl1*$^{+/+}$) mice (Fig. 2.4D). This led us to wonder whether haploinsufficiency for *Isl1* alone influences heart development. When we looked for CHDs in WT (*Isl1*$^{+/+}$) and *Isl1*$^{Cre/+}$ mice at E17.5, CHDs were observed in 22% of *Isl1*$^{Cre/+}$ mice, significantly higher than the 2% of CHDs observed in WT (*Isl1*$^{+/+}$) mice (Fig. 2.4B). Furthermore, ASDs were the only type of defect observed in *Isl1*$^{Cre/+}$ mice. Heart size was unaffected in *Isl1*$^{Cre/+}$ mice compared to WT (*Isl1*$^{+/+}$) mice (Fig. 2.4C). Thus, *Isl1*-haploinsufficiency alone results in CHDs, highlighting the critical role of *Isl1* dosage in heart development.

A

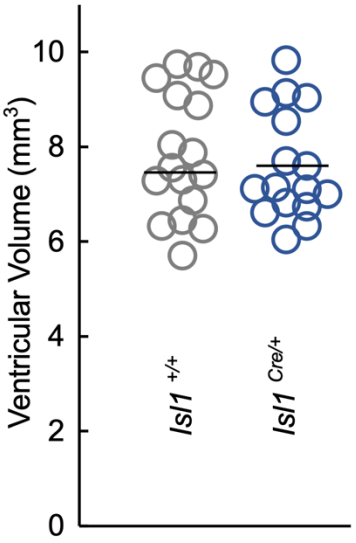


B

Genotype	ASD	N	% Hearts with Defect
Isl1 $^{+/+}$	1	50	2
Isl1 $^{Cre/+}$	9	41	21.95

P=0.0045

C



D

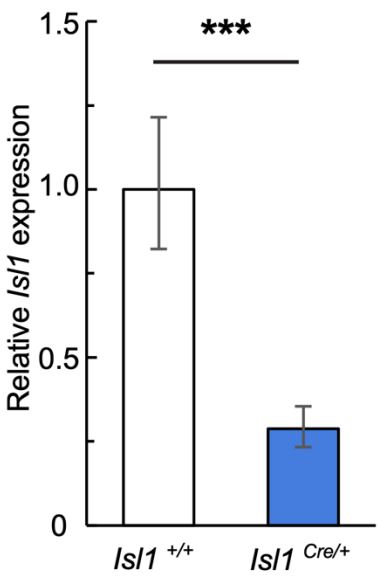


Fig. 2.4. Reduced *Isl1* expression in *Isl1*$^{Cre/+}$ hearts results in CHDs. (A) LSFM images of representative E17.5 *Isl1*$^{+/+}$ and *Isl1*$^{Cre/+}$ hearts on a *Td-tomato-EGFP* fluorescent reporter background. *Isl1-Cre* mediated recombination is reported by EGFP expression. Panels in bottom row show higher magnification of areas indicated by yellow brackets in panels in top row. In high magnification panels, blue arrowhead points to atrial septal defect (ASD). RV, right ventricle; LV, left ventricle; RA, right atrium; LA, left atrium; S, atrial septum. (B) Table of frequency of atrial septal defects (ASDs) in E17.5 *Isl1*$^{+/+}$ (N=50) and *Isl1*$^{Cre/+}$ (N=41) hearts, as detected by LSFM. P-value from Fisher's Exact Test. (C) Ventricular volumes of E17.5 *Isl1*$^{+/+}$ (N=17) and *Isl1*$^{Cre/+}$ (N=16) hearts. Volumes of both left and right ventricles were measured in ImageJ using LSFM images of hearts and summed. Bars show means. (D) Relative expression of *Isl1* (normalized to *Rpl4* expression) in E10.5 *Isl1*$^{+/+}$ (N=9) and *Isl1*$^{Cre/+}$ (N=9) hearts as measured by quantitative reverse transcription PCR. Error bars show standard error of the mean. *** represents $P \leq 0.001$. P-value from T-Test.

Mice doubly haploinsufficient for *Nipbl* and *Isl1* show fetal demise and reduced body size

Since *Nipbl*-haploinsuffiency in the *Isl1-Cre* expression domain results in CHDs and *Isl1*-haploinsufficiency alone also results in CHDs, we sought to understand the impact *Isl1*-haploinsufficiency on heart development in the context of global *Nipbl*-haploinsufficiency. To investigate this, we turned to *Stra8-iCre*; *Nipbl*$^{Flox/+}$ male mice, which make sperm that are either *Nipbl* $^{+}$ or *Nipbl* $^{-}$ (see Materials and Methods). We mated *Stra8-iCre*; *Nipbl*$^{Flox/+}$ male mice with *Isl1*$^{Cre/+}$ female mice and generated littermates that were either WT (*Nipbl*$^{+/+}$; *Isl1*$^{+/+}$), *Isl1*-haploinsufficient (*Isl1*$^{+/-}$), *Nipbl*-haploinsufficient (*Nipbl*$^{+/-}$), or doubly haploinsufficient for *Nipbl* and *Isl1* (*Nipbl*$^{+/-}$; *Isl1*$^{+/-}$) (Fig. 2.5A). Having shown in Fig. 2.4D that *Isl1*$^{Cre/+}$ mice are haploinsufficient for *Isl1*, we will now refer to mice of this genotype as *Isl1*$^{+/-}$. Upon examining embryos of these genotypes (WT, *Isl1*$^{+/-}$, *Nipbl*$^{+/-}$, and *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$) at E17.5, we found, unexpectedly, that *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ embryos showed signs of fetal demise, notably resorptions and hemorrhaging (Fig. 2.5B), with only 17% of these embryos at this stage being viable (Fig. 2.5C). This prompted us to investigate the viability of *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ embryos at an earlier stage, E15.5, at which point, the majority were found to be viable (Fig. 2.5D). Interestingly, we noticed at both E17.5 and E15.5, that both *Nipbl*$^{+/-}$ and *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ embryos were smaller than their WT littermates (Fig. 2.5, B & D). Measurements of crown-rump lengths at E15.5, found that, as in previous studies (*10, 18*), *Nipbl*$^{+/-}$ embryos were 11% smaller than WT embryos (Fig. 2.5E). Surprisingly, *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ embryos were also 11% smaller than WT embryos. There was no difference between the crown-rump lengths of *Isl1*$^{+/-}$ and WT embryos. These results indicate that haploinsufficiency for *Nipbl* in the entire body and haploinsufficiency for *Isl1* significantly affect overall development, leading to fetal demise and reduced body size.

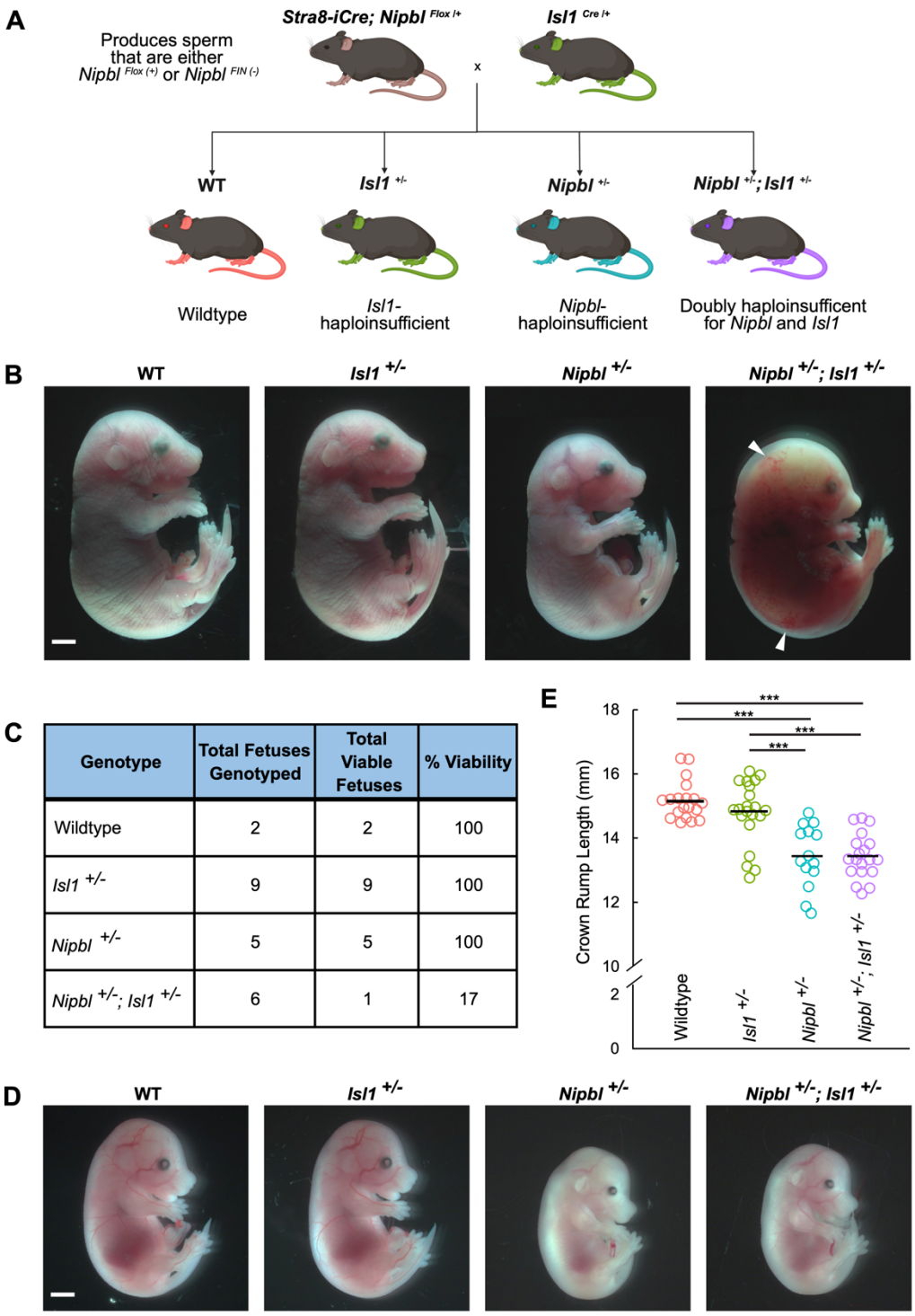
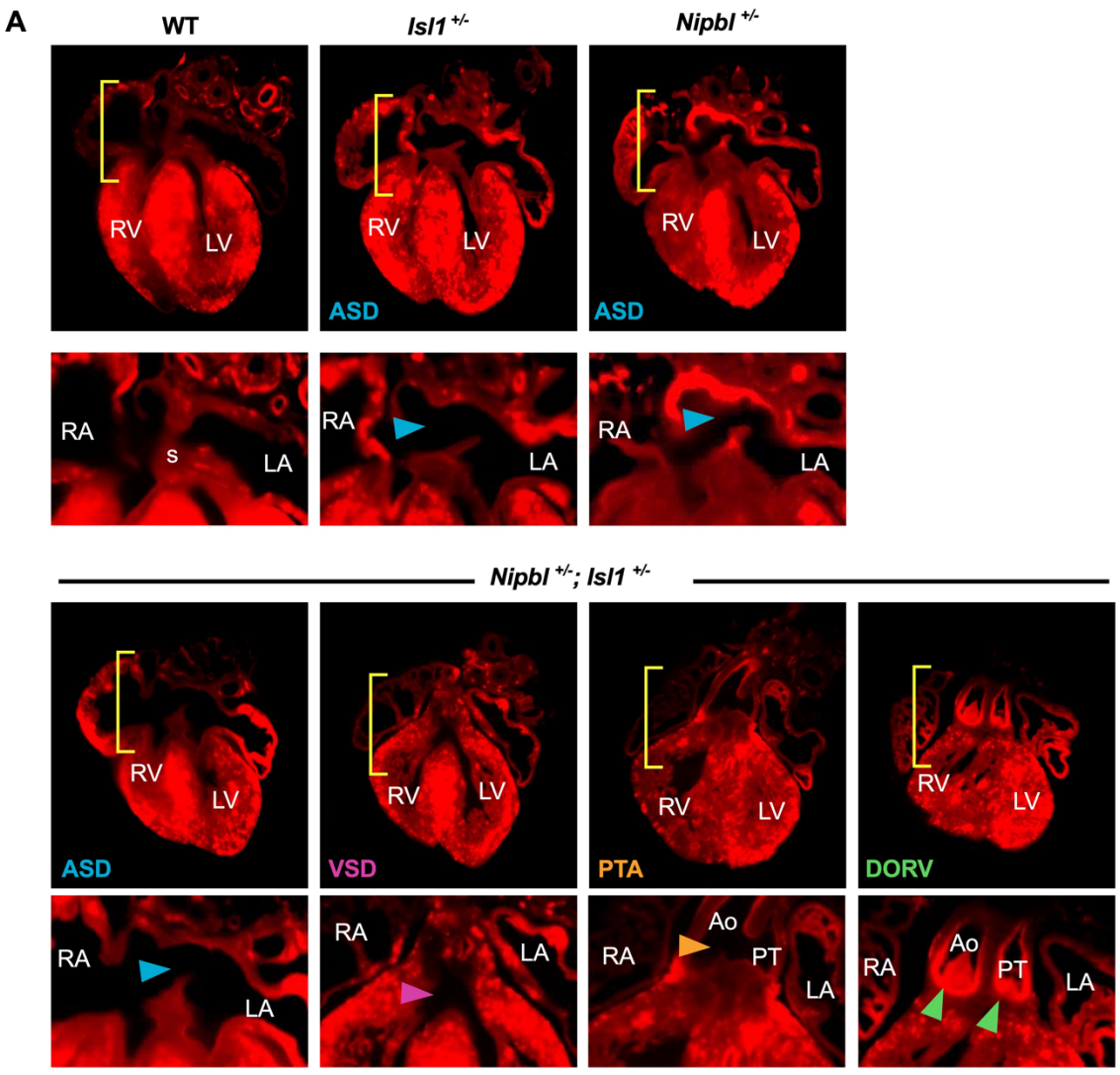


Fig. 2.5. Mice doubly haploinsufficient for *Nipbl* and *Isl1* show fetal demise and reduced body size. (A) *Stra8-iCre*; *Nipbl*$^{Flox/+}$ male mice make sperm that are either *Nipbl* $^{+}$ or *Nipbl* $^{-}$ (see Materials and Methods). Mating *Stra8-iCre*; *Nipbl*$^{Flox/+}$ male mice with *Isl1*$^{Cre/+}$ female mice

produces littermates that are either WT (*Nipbl*$^{+/+}$; *Isl1*$^{+/+}$), *Isl1*-haploinsufficient (*Isl1*$^{+/-}$), *Nipbl*-haploinsufficient (*Nipbl*$^{+/-}$), or doubly haploinsufficient for *Nipbl* and *Isl1* (*Nipbl*$^{+/-}$; *Isl1*$^{+/-}$). (B) Representative images of E17.5 WT, *Isl1*$^{+/-}$, *Nipbl*$^{+/-}$, and *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ embryos. One *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ embryo shows hemorrhaging, a sign of fetal demise. Arrows point to hemorrhaging in the back of the head and rump of one *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ embryo. Scale bar = 500 uM. (C) Table of frequency of viability in WT (N=2), *Isl1*$^{+/-}$ (N=9), *Nipbl*$^{+/-}$ (N=5), and *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ (N=6) embryos. (D) Representative images of E15.5 WT, *Isl1*$^{+/-}$, *Nipbl*$^{+/-}$, and *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ embryos. (E) Crown rump length (CRL) of E15.5 WT (N=19), *Isl1*$^{+/-}$ (N=21), *Nipbl*$^{+/-}$ (N=14), and *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ (N=19) embryos. CRL refers to the length of an embryo from the top of its head (the crown) to the bottom of its torso (the rump). *** represents $P \leq 0.001$. P-values from Tukey-Kramer Test.

Isl1-haploinsufficiency enhances the frequency and spectrum of CHDs in *Nipbl*-haploinsufficient mice

Since *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ mouse embryos died at E17.5, we looked for CHDs in the hearts of WT, *Nipbl*$^{+/-}$, *Isl1*$^{+/-}$ and *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ mice at E15.5 (Fig. 2.6A), when no deaths were observed among embryos of any genotype. As expected, we found that both *Isl1*$^{+/-}$ mice and *Nipbl*$^{+/-}$ mice showed CHDs (9.09% and 20.00%, respectively) (Fig. 2.6B). Surprisingly, we found that 66.67% of *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ mice showed CHDs, far higher than what was observed for mice haploinsufficient for either *Nipbl* or *Isl1* alone (Fig. 2.6B). As in *Nipbl*$^{+/-}$ mice, *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ mice also showed ASDs, VSDs, and OTDs (Fig. 2.6, A & B), with OTDs always occurring in conjunction with either a VSD or both an ASD and VSD. Among the OTDs displayed by *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ mice were DORVs and PTAs (Fig. 2.6B), the latter of which was exclusively observed in *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ mice. Also exclusively observed in *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ mice were CHDs exhibiting a combination of both an ASD and VSD, which occurred in one *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ mouse (Fig. 2.6B). Since *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ appeared to have smaller hearts that WT (Fig. S1A), we also measured the ventricular volumes of hearts across all genotypes (Fig. S1B). Indeed, *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ hearts were 31% smaller than WT hearts, similar to the 28% reduction in ventricular volumes observed in *Nipbl*$^{+/-}$ hearts in Fig. 2.1D. Since there was no significant difference between the ventricular volumes of *Nipbl*$^{+/-}$ and *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ hearts, these results demonstrate that *Isl1*-haploinsufficiency enhances the frequency and spectrum of CHDs when mice are also *Nipbl*-haploinsufficient in the entire body, without further reducing heart size.



B

Genotype	ASD	VSD	ASD+ VSD	OTD+ ASD	OTD+ VSD	OTD+ VSD+ ASD	N	Total with Defect	% Hearts With Defect
Wildtype	0	0	0	0	0	0	23	0	0.00
Isl1 +/-	1	1	0	0	0	0	22	2	9.09
Nipbl +/-	1	1	1	1	0	0	20	4	20.00
Nipbl +/-; *Isl1* +/-	6	0	1	0	5	2	21	14	66.67

P = $2x10^{-1}$
P = $6x10^{-2}$
P = $6x10^{-6}$
P = $3x10^{-4}$
P = $8x10^{-3}$

Fig. 2.6. *Isl1*-haploinsufficiency enhances the frequency and spectrum of CHDs in *Nipbl*-haploinsufficient mice. (A) LSFM images of anterior view of coronal section of representative E15.5 WT, *Isl1*$^{+/-}$, *Nipbl*$^{+/-}$, and *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ hearts. *Isl1*$^{+/-}$ and *Nipbl*$^{+/-}$ hearts shown show

atrial septal defects (ASD, blue arrow). *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts show atrial septal defects (ASD, blue arrow), ventricular septal defects (VSD, purple arrow), persistent truncus arteriosus (PTA, yellow arrow), and/or double outlet right ventricle (DORV, green arrows). RV = right ventricle, LV =left ventricle, RA = right atrium, LA = left atrium, Ao = aorta, PT = pulmonary tract. (B) Table of frequencies of heart defects observed in WT, *Isl1*$^{+/-}$, *Nipbl*$^{+/-}$, and *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ E15.5 hearts. Hearts with PTA or DORV were classified as outflow tract defects (OTD). P-values from Benjamini-Hochberg corrected Fisher's Exact Tests.

Gene expression in E10.5 hearts potentially contributing to CHDs in *Nipbl*$^{+/-}$, *Isl1*$^{+/-}$, and *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ mice

To identify gene expression changes potentially linked to the development of CHDs, we performed RNA sequencing on E10.5 hearts from WT, *Nipbl*$^{+/-}$, *Isl1*$^{+/-}$, and *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ mice, since E10.5 was the earliest stage before which the morphological abnormalities observed in mice of any genotype (ASDs, VSDs, OTDs) could be detected. When we performed differential gene expression analysis (DGEA) between *Nipbl*$^{+/-}$, *Isl1*$^{+/-}$, *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ and WT hearts, we found that *Nipbl*$^{+/-}$ hearts misexpressed 442 genes: 391 overexpressed and 51 underexpressed (Fig. 2.7A). As in previous studies in which it was shown that *Nipbl*-haploinsufficiency results in hundreds of small (less than two-fold) gene expression changes (*10, 18*), 79% of the gene expression changes that occurred in *Nipbl*$^{+/-}$ hearts occurred with less than two-fold change. Since the majority of these gene expression changes were small, we performed gene set enrichment analysis on them to see whether they were enriched or de-enriched for any Hallmark gene sets representing well curated biological processes. We found that the gene expression of *Nipbl*$^{+/-}$ hearts was enriched for 8 gene sets. One of these was IL6 JAK STAT3 signaling, which promotes angiogenesis, the formation of blood vessels (*146*). *Nipbl*$^{+/-}$ hearts were de-enriched for one gene set, Myc target genes, which also plays a role in angiogenesis, as well as cell growth/proliferation (*147, 148*).

When we manually curated the gene expression changes observed in *Nipbl*$^{+/-}$ hearts, we found that *Nipbl*$^{+/-}$ hearts overexpressed several *Hox* genes (*Hoxa3, Hoxa2, Hoxc4, Hoxa1, Hoxb2*). *Hox* genes play roles in anterior-posterior patterning, outflow tract patterning, and angiogenesis (*149-151*). Additionally, *Nipbl*$^{+/-}$ hearts overexpressed genes associated with sarcoplasmic/endoplasmic reticulum Ca2+ transportation (SERCA) (*Sln, Atpa21, Mybphl*),

which play roles in muscle relaxation (*152*). Overexpression was also observed in genes associated with chromatin and cohesin (*Hist3h2ba, Stag1*) (*153*), genes associated with protease activity (*Plat, Serpinb6b*), which play roles in vascular hemostasis (Fig. 2.7C) (*154*), and cadherin genes (*Cdh13, Cdh10*), which play roles in epithelial mesenchymal transition, heart tube formation, and chamber formation (*155, 156*). Conversely, *Nipbl*$^{+/-}$ hearts underexpressed genes associated with muscle contraction (*Myoz1, Tnni2*) (*157, 158*) and protocadherin beta genes (*Pcdhb18, Pcdhb19, Pcdhb16, Pcdhb20, Pcdhb22*), which play roles in establishing neural circuits and synapses (*159*).

Isl1$^{+/-}$ hearts misexpressed 26 genes: 24 of which were overexpressed, and 2 of which were underexpressed (Fig. 2.7D). These gene expression changes were enriched for 5 Hallmark gene sets (Fig. 2.7E). One such set was epithelial mesenchymal transition, genes involved in heart septa formation (*160*). Conversely, there was de-enrichment of 3 gene sets, one of which was Myc target genes. Manual curation of overexpressed genes in *Isl1*$^{+/-}$ hearts found a *Hox* gene (*Hoxa3*), a SERCA-associated gene (*Sln*), and *Foxo6* (a member of the *Foxo* transcription factor family), the latter of which plays a role in cardiomyocyte proliferation (Fig. 2.7F) (*161*). On the other hand, the transcription factor *Pax6,* known for its roles in eye (*162*) and brain development, was underexpressed. Specifically, in brain development, *Pax6* delineates boundaries between various brain regions, ensuring their proper development (*163*).

Nipbl$^{+/-}$; *Isl1*$^{+/-}$ hearts misexpressed 427 genes: 361 of which were overexpressed and 66 of which were underexpressed (Fig. 2.7G). These gene expression changes were enriched for 10 Hallmark gene sets, notably KRAS signaling (involved in cell differentiation), epithelial mesenchymal transition, IL6 JAK STAT3 signaling (Fig. 2.7H). Conversely, they were de-enriched for 7 gene sets. Notable among these were Myc target genes, E2F target genes

(important for cell cycle regulation) (*164*), and genes associated with MTORC1 signaling, which play roles in cell growth and proliferation (Fig. 2.7H) (*165*). Manual curation of overexpressed genes in *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ hearts found a diverse array of genes. These included: SERCA genes (*Sln, Atpa21, Mybpc2, Actn3*), cadherin genes (*Cdh10, Cdh13*), genes associated with TAM receptor family of tyrosine kinases (*Axl, Gas6*), known for their role in cardiac protection and remodeling (*166*), *Hox* genes (*Hoxa2, Hoxa3, Hoxa1*), genes associated with Wnt signaling (*Fzd7, Wnt9a, Wnt2b*), which plays roles in cardiac mesoderm formation, heart tube development, chamber differentiation, and outflow tract formation (*167-170*), genes associated with protease activity (*Plat, Serpinb6b*), genes associated with phospholipid signaling (*Plce1, Pik3c2g*), implicated in cardiac protection (*171, 172*), and genes associated with chromatin and cohesin (*Hist3h2ba, Stag1*) (Fig. 2.7I). Conversely, underexpressed genes in these hearts were associated with cardiac muscle contraction (*Myoz1, Tnni2*) and protocadherin beta gene cluster (*Pcdhb18, Pcdhb19, Pcdhb16, Pcdhb20, Pcdhb22*).

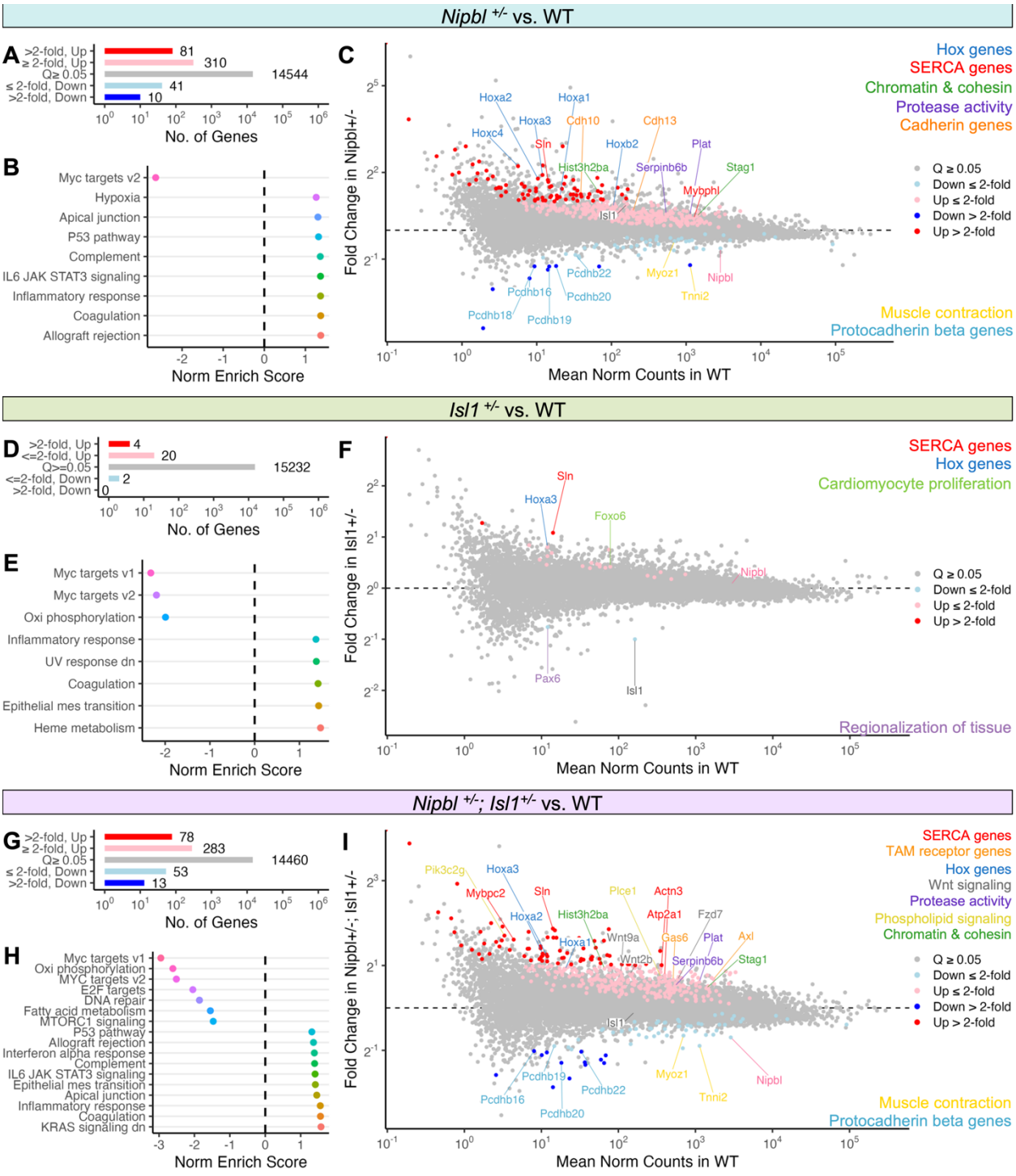


Fig. 2.7. Gene expression in E10.5 hearts potentially contributing to CHDs in *Nipbl*$^{+/-}$, *Isl1*$^{+/-}$, and *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ mice. Number of genes differentially expressed in (A) *Nipbl*$^{+/-}$, (D) *Isl1*$^{+/-}$, or (G) *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ E10.5 hearts relative to WT. Differentially expressed genes are binned

according to whether they were differentially expressed less than or more than two-fold up or down. Differential gene expression analysis was performed using the ANOVA and Tukey-Kramer tests. Genes with Benjamini-Hochberg corrected p-values less than alpha 0.05 were considered differentially expressed. Hallmark gene sets enriched or de-enriched in their expression in (B) *Nipbl*$^{+/-}$, (E) *Isl1*$^{+/-}$, or (H) *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ E10.5 hearts relative to WT. Gene set enrichment analysis was performed using FGSEA. Fold change in expression of genes in (C) *Nipbl*$^{+/-}$, (F) *Isl1*$^{+/-}$, or (I) *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ E10.5 hearts along their average expression in WT hearts. Genes are colored according to whether they were not differentially expressed (gray, $Q \geq 0.05$), overexpressed less than or equal to 2-fold (pink, Up ≤ 2-fold), overexpressed more than 2-fold (red, Up > 2-fold), underexpressed less than or equal to 2-fold (light blue, Down ≤ 2-fold), or underexpressed more than 2-fold (Down > 2-fold).

$Nipbl^{+/-}$ mice overexpress *Isl1* while $Nipbl^{+/-}$; $Isl1^{+/-}$ mice express *Isl1* at WT levels

To see how similar or dissimilar $Nipbl^{+/-}$, $Isl1^{+/-}$, and $Nipbl^{+/-}$; $Isl1^{+/-}$ hearts were to each other in gene expression, we performed principal component analysis (PCA) on them and included WT hearts in the analysis. We found that the first 4 components explained 89% of the variance (Fig. 2.8A). When we plotted the position of each heart along these components, we found that they separated by genotype along Component 3 into two groups (Fig. 2.8B). In Group 1, $Isl1^{+/-}$ hearts separated to the right with WT hearts (red outline), and in Group 2, $Nipbl^{+/-}$; $Isl1^{+/-}$ hearts separated to the left with $Nipbl^{+/-}$ hearts (blue outline). These results suggest that, in terms of gene expression, $Isl1^{+/-}$ hearts are similar to WT hearts, $Nipbl^{+/-}$; $Isl1^{+/-}$ hearts are similar to $Nipbl^{+/-}$ hearts, while, as pairs, WT and $Isl1^{+/-}$ hearts are dissimilar to $Nipbl^{+/-}$ and $Nipbl^{+/-}$; $Isl1^{+/-}$ hearts.

These results echo the results of Fig. 2.7, in which $Nipbl^{+/-}$ and $Nipbl^{+/-}$; $Isl1^{+/-}$ hearts misexpressed a similarly large number of genes (442 and 427 genes, respectively) (Fig. 2.7, A & G), while $Isl1^{+/-}$ hearts misexpressed few genes (26 genes) (Fig. 2.7D). We also observed in Fig. 2.7 that $Nipbl^{+/-}$ and $Nipbl^{+/-}$; $Isl1^{+/-}$ hearts misexpressed genes associated with many of the same biological processes, whereas $Isl1^{+/-}$ hearts misexpressed genes with only a few of the same biological processes. For example, $Nipbl^{+/-}$ and $Nipbl^{+/-}$; $Isl1^{+/-}$ hearts similarly underexpressed protocadherin beta genes and genes associated with muscle contraction, and overexpressed SERCA genes, *Hox* genes, genes associated with chromatin and cohesin, genes associated with protease activity, and cadherin genes (Fig. 2.7, C & I). $Isl1^{+/-}$ hearts similarly overexpressed only a *Hox* gene and a SERCA gene (Fig. 2.7F).

We wondered exactly how many differentially expressed genes (DEGs) were shared among $Nipbl^{+/-}$, and $Nipbl^{+/-}$; $Isl1^{+/-}$ hearts and quantified them. We considered DEGS to be

shared when they were differentially expressed in the same direction (overexpressed or underexpressed relative to WT) in hearts of different genotypes. We found that *Nipbl*$^{+/-}$ and *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts shared 373 DEGs (Fig. 2.8C), 44 of which were downregulated genes (Fig. 2.8D) and 329 of which were upregulated (Fig. 2.8E). *Isl1*$^{+/-}$ hearts only shared 23 DEGs with both *Nipbl*$^{+/-}$ and *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts (Fig. 2.8C), all of which were upregulated (Fig. 2.8E) and none of which were downregulated (Fig. 2.8D).

When we looked at the DEGS that were shared and downregulated among the three genotypes, we found that *Nipbl* expression was downregulated in *Nipbl*$^{+/-}$ and *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts (Fig. 2.8, D & F). In *Isl1*$^{+/-}$ hearts, *Nipbl* was expressed at WT levels (Fig. 2.8, E & F). We also found that *Isl1* expression was downregulated in *Isl1*$^{+/-}$ hearts, as expected, but it was only downregulated in *Isl1*$^{+/-}$ hearts (Fig. 2.8, D & G). Surprisingly, in *Nipbl*$^{+/-}$ hearts, *Isl1* expression was upregulated, and in *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts, *Isl1* was expressed at WT levels, even though *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts are hemizygous for *Isl1* and *Nipbl* (Fig. 2.8, D & G). Thus, *Isl1* expression is upregulated in both *Nipbl*$^{+/-}$ and to *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts: in *Nipbl*$^{+/-}$ hearts, <u>above</u> WT levels, and in *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts, <u>to</u> WT levels. These results suggest that upregulation of *Isl1* expression is a compensatory developmental response to *Nipbl*-haploinsufficiency in the heart. Additionally, this upregulation appears to prevent *Nipbl*$^{+/-}$ mice from developing CHDs to the same extent as *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ mice, since *Nipbl*$^{+/-}$ mice have fewer and less severe CHDs than *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ mice, implying a protective role of *Isl1* overexpression against CHDs in the context of *Nipbl*-haploinsufficiency.

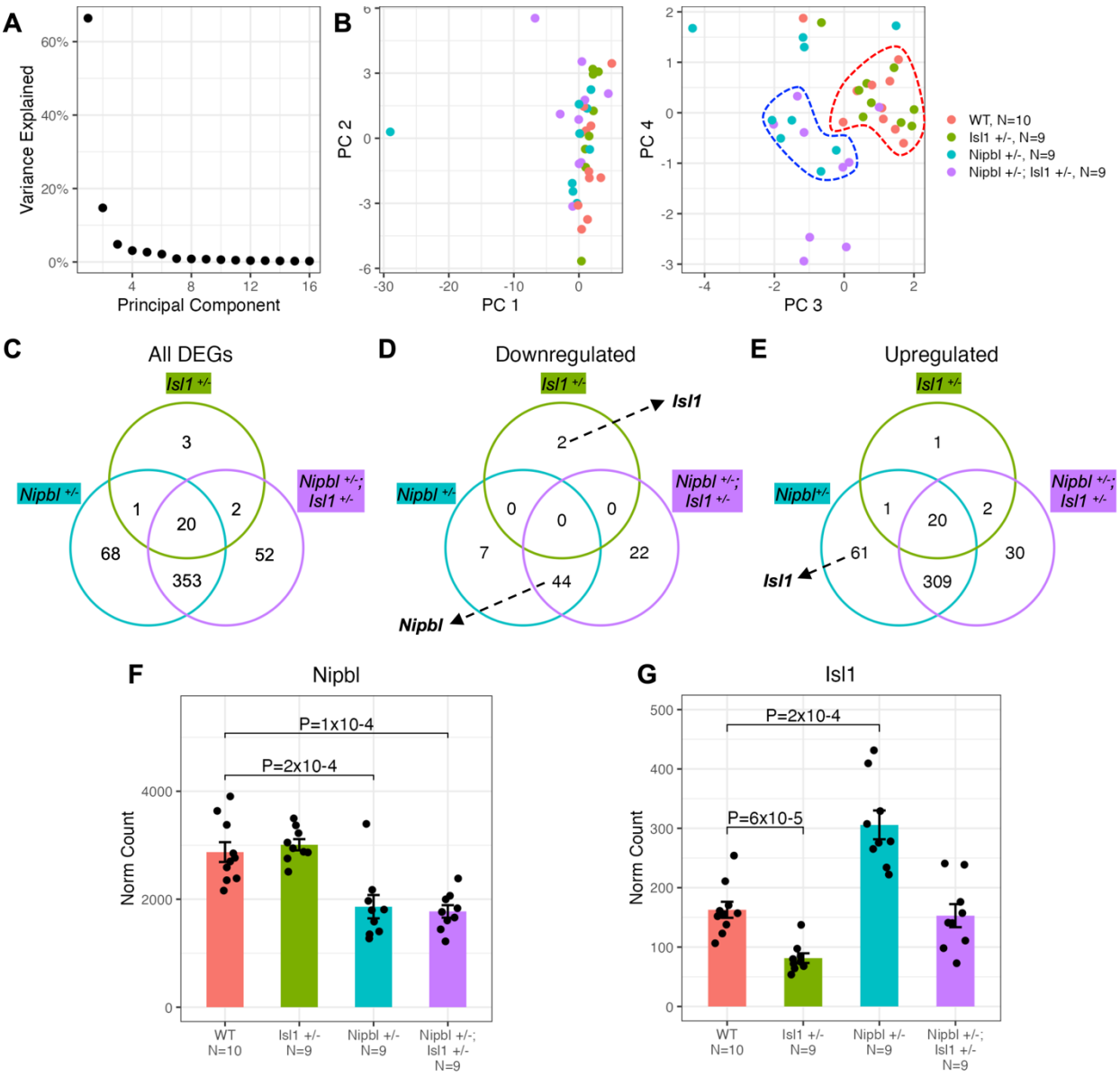


Fig. 2.8. *Nipbl*$^{+/-}$ mice overexpress *Isl1* while *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ mice express *Isl1* at WT levels.

(A) Variance explained by first 16 principal components among E10.5 WT, *Isl1*$^{+/-}$, *Nipbl*$^{+/-}$, and *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts. (B) Position of E10.5 WT, *Isl1*$^{+/-}$, *Nipbl*$^{+/-}$, and *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts along Principal Components 1-4. *Isl1*$^{+/-}$ hearts separated along Component 3 to the right with WT hearts (red outline). *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts separated along Component 3 to the left with *Nipbl*$^{+/-}$ hearts (blue outline). Number of genes (C) differentially expressed, (D) downregulated, or upregulated (E) in E10.5 *Isl1*$^{+/-}$, *Nipbl*$^{+/-}$, and *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts relative to WT hearts that

are shared and/or unique to each genotype. DEGS were considered shared when they were differentially expressed in the same direction (overexpressed or underexpressed relative to WT). Expression of (F) *Nipbl* and (G) *Isl1* across E10.5 WT, *Nipbl*$^{+/-}$, *Isl1*$^{+/-}$, and *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ hearts. Error bars show standard error of the mean. P-values from Tukey-Kramer test.

A small number of developmental regulatory genes shows unique expression patterns in *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ mice

To investigate which gene expression differences between *Nipbl*$^{+/-}$ hearts and *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ hearts might be responsible for the enhancement of CHDs observed in *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ mice, we plotted the fold-change in expression of every DEG in *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ hearts versus the fold-change in expression of every DEG in *Nipbl*$^{+/-}$ hearts (Fig. 2.9A). When we colored these genes according to their standard deviation away from a 1:1 correlation (Z-score), and manually annotated genes that were greater than 3 standard deviations away, all developmentally significant genes, which were three, occurred in one quadrant, where *Isl1* also was (Fig. 2.9A). Interestingly, all three of these genes encode homeodomain transcription factors, and all play major roles in patterning and organ size determination (*149, 173, 174*) (Fig. 2.3A). In *Nipbl*$^{+/-}$ hearts, *Pitx1* is expressed at WT levels, whereas it is expressed below WT levels in *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ hearts (Fig. 2.9B). In *Nipbl*$^{+/-}$ hearts, *Hoxc4* and *Pax6* are over-expressed above WT levels, whereas in *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ hearts these genes are expressed only at WT levels (Fig. 2.9, C & D). These results suggest that the hearts of *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ mice fail to upregulate a small number of developmentally significant genes to the same levels as observed in *Nipbl*$^{+/-}$, hearts, and that this contributes to the increased frequency and severity of CHDs observed in *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ mice. Furthermore, since *Isl1* is upregulated above WT levels in *Nipbl*$^{+/-}$ hearts, and *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ hearts fail to upregulate *Isl1* above WT levels (Fig. 2.8), failure to upregulate *Pitx1*, *Hoxc4*, and *Pax6* in *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ hearts is likely due to the failure to upregulate *Isl1* above WT levels.

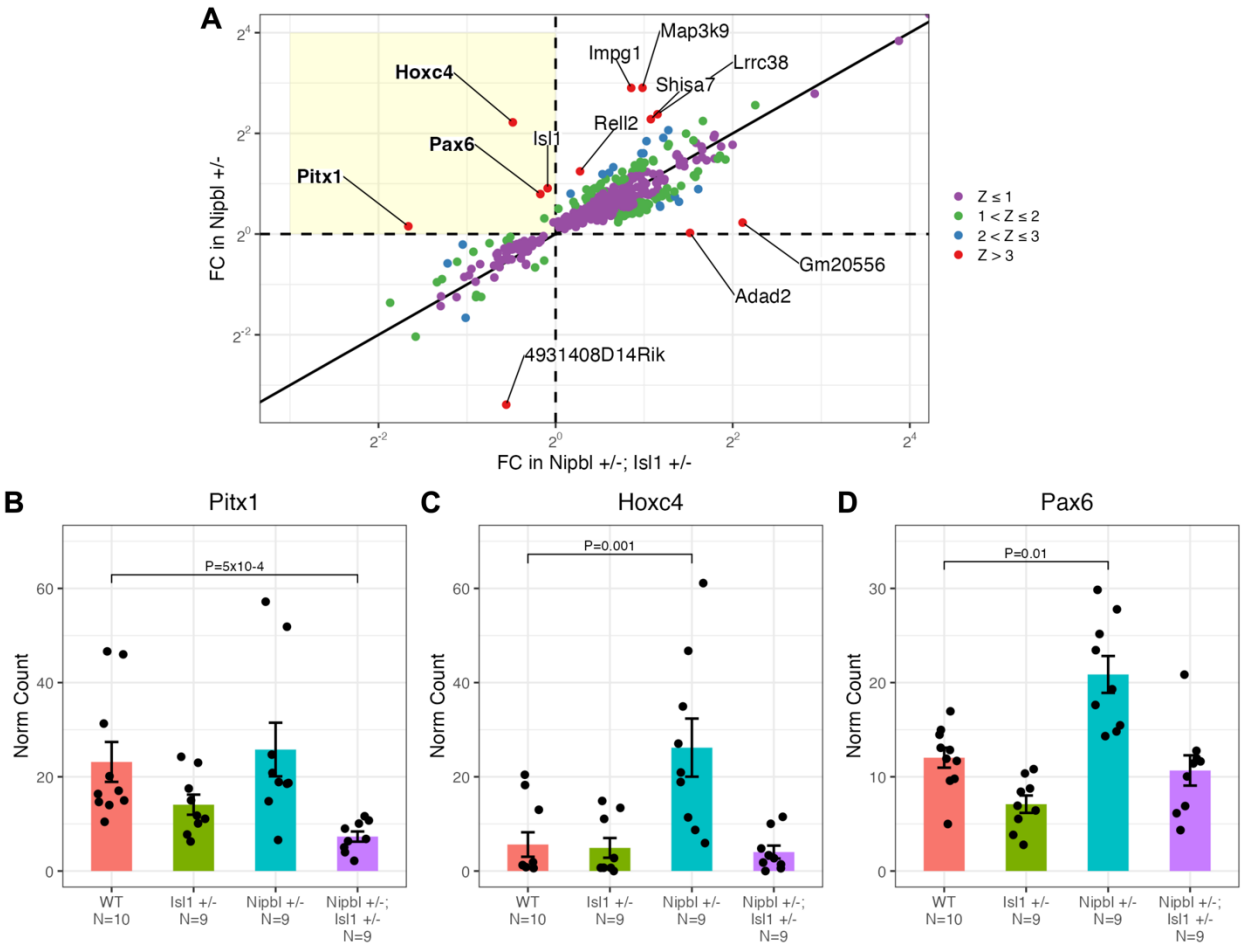


Fig. 2.9. *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ mice fail to upregulate *Pitx1* to WT levels, and *Hoxc4* and *Pax6* above WT levels. (A) Fold change in expression of genes in *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ E10.5 hearts relative to WT against that of *Nipbl*$^{+/-}$ hearts relative to WT. Solid line represents a 1:1 correlation in fold change in expression between the two genotypes. Genes are colored by standard deviation away from line of 1:1 correlation (Z-score). Genes with a Z-score greater than 3 and developmentally significant occupied the top left quadrant (yellow shading). Expression of (B) *Pitx1,* (C) *Hoxc4, and* (D) *Pax6* across WT, *Nipbl*$^{+/-}$*, Isl1*$^{+/-}$, and *Nipbl+/; Isl1*$^{+/-}$ E10.5 hearts. Error bars show standard error of the mean. P-values from Tukey-Kramer test.

Discussion

LSFM detects a spectrum of CHDs in *Nipbl*$^{+/-}$ mice

The *Nipbl*$^{+/-}$ mouse, a model of the birth defects syndrome, CdLS, presents an excellent model for studying the developmental origins of multifactorial and/or polygenic CHDs, because global transcriptional dysregulation in *Nipbl*$^{+/-}$ mice gives rise to CHDs similar to those observed in nonsyndromic cases in man. In this study, we used LSFM, which captures high resolution 3D images of tissues, to look for CHDs in E17.5 *Nipbl*$^{+/-}$ mouse hearts. Whereas prior studies, using μMRI and histological analysis only documented ASDs in E17.5 *Nipbl*$^{+/-}$ hearts (*18*), LSFM found a spectrum of CHDs in *Nipbl*$^{+/-}$ hearts (ASDs, VSDs, and OTDs), with OTDs in *Nipbl*$^{+/-}$ hearts always occurring alongside an ASD or both an ASD and VSD (Fig. 2.1). Thus, LSFM detects a broader spectrum of CHDs in *Nipbl*$^{+/-}$ mice than previously detected by histological analysis and μMRI.

Nipbl-haploinsufficiency in the *Isl1-Cre* domain results in CHDs

Since previous studies suggested that the SHF could be a potential source of CHDs in *Nipbl*$^{+/-}$ mice (*18*), we used LSFM to investigate this hypothesis. We created *Nipbl*-haploinsufficiency in the SHF using two Cre recombinase-expressing mouse lines: *Mef2c-Cre* and *Isl1*$^{Cre/+}$. While *Mef2c-Cre*-mediated *Nipbl*-haploinsufficiency in the SHF did not result in CHDs (Fig. 2), *Isl1*$^{Cre/+}$ did, resulting in a spectrum of CHDs strikingly similar to those observed in *Nipbl*$^{+/-}$ hearts (Fig. 2), with the addition of PTAs. A possible explanation for why *Isl1*$^{Cre/+}$ driven *Nipbl*-haploinsufficiency in the SHF resulted in CHDs, whereas *Mef2c-Cre* did not, is extensive recombination of *Isl1-Cre* in non-cardiac tissues (Fig. 2.3). This is consistent with our earlier findings showing that *Nipbl*-haploinsufficiency in non-cardiac tissues contributes to the development of CHDs (*18*). Another potential explanation is significant reduction of *Isl1*

expression in *Isl1*$^{Cre/+}$ hearts, over 60% lower compared to WT (Fig. 2.4D). In Fig. 2.4, we showed that reduced *Isl1* expression alone results in the development of CHDs, particularly ASDs. Therefore, both extensive recombination of *Isl1-Cre* in non-cardiac tissues and/or reduced *Isl1* expression could be contributing factors to the development of CHDs observed in *Nipbl*$^{Flox/+}$*; Isl1*$^{Cre/+}$ hearts.

Isl1-haploinsufficiency enhances the development of CHDs in _Nipbl_$^{+/-}$ mice

Since *Isl1*-haploinsufficiency alone produced CHDs, we investigated the influence of *Isl1*-haploinsufficiency on heart development in the context of global *Nipbl*-haploinsufficiency, by creating mice doubly haploinsufficient for both genes *(Nipbl*$^{+/-}$*; Isl1*$^{+/-}$*)* (Fig. 2.5A). Double haploinsufficiency for *Isl1* and *Nipbl* resulted in more frequent and severe occurrences of CHDs compared to mice haploinsufficient for either *Nipbl* or *Isl1* alone (Fig. 2.6B). The CHDs observed in *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts were similar to those observed in *Nipbl*$^{+/-}$ hearts, with the addition of PTA. Furthermore, a case was documented where a *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ mouse exhibited both an ASD and VSD concurrently. These findings support the idea that *Isl1* haploinsufficiency enhances the frequency and severity of CHDs in mice when the entire body is *Nipbl*-haploinsufficient.

Gene expression changes potentially contributing to the development of CHDs

RNA sequencing of E10.5 WT, *Isl1*$^{+/-}$, *Nipbl*$^{+/-}$, and *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts revealed gene expression changes potentially contributing to the development of CHDs. Firstly, in all three mutant genotypes, there was overexpression of *Hox* genes, such as *Hoxa3*, which play roles in anterior-posterior patterning, outflow tract patterning, and angiogenesis (*149-151*). These genotypes also showed overexpression of genes associated with sarcoplasmic/endoplasmic reticulum Ca2+ transportation (SERCA) (e.g., *Sln*), which play roles in muscle relaxation (*152*).

Conversely, there was de-enrichment for the expression of Myc target genes, which are involved in cell growth/proliferation (*147, 148*).

Secondly, specific to *Nipbl*$^{+/-}$ and *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts, there was an enrichment of genes associated with the IL6 JAK STAT3 signaling pathway, which is known to promote angiogenesis, the formation of blood vessels (*146*). These hearts also exhibited overexpression of genes associated with chromatin and cohesin (*Hist3h2ba, Stag1*) (*153*), genes associated with protease activity (*Plat, Serpinb6b*), which play roles in vascular hemostasis (Fig. 2.7C) (*154*), and cadherin genes (*Cdh13, Cdh10*), which play roles in epithelial mesenchymal transition, heart tube formation, and chamber formation (*155, 156*). There was also underexpression of genes associated with muscle contraction (*Myoz1, Tnni2*) (*157, 158*) and protocadherin beta genes (*Pcdhb18, Pcdhb19, Pcdhb16, Pcdhb20, Pcdhb22*), which play roles in establishing neural circuits and synapses (*159*). Together, these gene expression changes illuminate key biological pathways and mechanisms potentially underlying the development of CHDs.

Upregulation of *Isl1* prevents *Nipbl*$^{+/-}$ mice from developing more frequent and severe CHDs

Fig. 2.8 showed that RNA sequencing of E10.5 hearts revealed that *Isl1* expression was downregulated in *Isl1*$^{+/-}$ hearts and upregulated in *Nipbl*$^{+/-}$ hearts. However, in *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts, *Isl1* was expressed at WT levels, even though these hearts are hemizygous for *Isl1* and *Nipbl* (Fig. 2.8, D & G). This pattern of expression indicates that *Isl1* expression is upregulated in both *Nipbl*$^{+/-}$ and to *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts: in *Nipbl*$^{+/-}$ hearts, above WT levels, and in *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts, to WT levels. These results suggest two key points: 1) upregulation of *Isl1* expression in the heart is a compensatory response to *Nipbl*-haploinsufficiency, and 2) overexpression of *Isl1* may prevent *Nipbl*$^{+/-}$ mice from developing CHDs to the same extent as

Nipbl$^{+/-}$*; Isl1*$^{+/-}$ mice, since *Nipbl*$^{+/-}$ mice exhibit fewer and less severe CHDs than *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ mice.

Homeodomain gene dosages contributing to the development of CHDs

Other gene expression changes that may contribute to enhancement of CHDs in *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts over that of *Nipbl*$^{+/-}$ hearts were observed among *Pitx1*, *Pax6*, and *Hoxc4*. These genes were not upregulated in *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts to the same extent as they were in *Nipbl*$^{+/-}$ hearts. Since *Isl1* is upregulated above WT levels in *Nipbl*$^{+/-}$ hearts and *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ hearts fail to also upregulate *Isl1* to the same levels as in *Nipbl*$^{+/-}$ hearts (Fig. 2.8), failure to upregulate *Pitx1*, *Hoxc4*, and *Pax6* by *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$ hearts is likely due to the failure to upregulate *Isl1*.

Pitx1, *Hoxc4*, and *Pax6* are interesting because they all encode homeodomain transcription factors and they all play roles in the development of other tissues/organs in a dosage-sensitive manner. For instance, knockout of both copies of *Pitx1* in mouse results in anteriorization of the hind limb bud and loss of skeletal elements in such a way that resembles the skeletal structure of the forelimb (*174*). Similarly, variations in *Pax6* dosage influence the development and severity of micro-opthalmia in *Pax6* mutant mice (*173*), and reduced expression of different *Hoxc* cluster genes in mice results in anterior-posterior skeletal transformations (*175*). Such findings suggest that these genes could also affect heart development in a dosage-sensitive manner. To investigate this hypothesis, future studies could investigate the effects of double haploinsufficiency for *Nipbl* and each of these genes on the development of CHDs. One would anticipate that the resulting CHDs would be more severe than those caused by *Nipbl*-haploinsufficiency alone. Insights gained from such studies could enhance our understanding of the roles these genes play in the development of CHDs in both Cornelia de Lange Syndrome (CdLS) and nonsyndromic cases.

Materials and Methods

Animal Ethics Statement

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

Mouse Lines and Genotyping

All mouse lines were genotyped by polymerase chain reaction (PCR) using genomic DNA obtained from tail biopsies.

Nipbl$^{Flox/Flox}$ Line

The Nipbl$^{Flox/Flox}$ line (*Nipbl*$^{Gt(EUCE313f02)1.1Hmgu}$*/Clf*) has an inverted gene trap cassette encoding *b-geo* that is flanked by *Cre* recombinase target sites in intron 1 of both *Nipbl* alleles (*18*). In this inverted orientation, referred herein as *Flox*, there is no trapping of the *Nipbl* gene and *Nipbl* is expressed normally. As a result, the *Nipbl*$^{Flox/Flox}$ line is phenotypically WT. However, when this cassette is exposed to Cre recombinase, the gene trap cassette gets inverted into a non-inverted orientation which we call FIN. In this non-inverted orientation, trapping of the *Nipbl* gene occurs and *b-geo* is expressed as a reporter of successful gene trapping, making the *Nipbl FIN* allele a null allele. The *Nipbl*$^{Flox/Flox}$ line was maintained on a C57BL/6J background for more than 15 generations and was genotyped as described in (*18*).

Recombinase Mouse Lines

<u>*Nanog-Cre Line:*</u> *The Nanog-Cre* line (Tg(Nanog-cre)#Vlcg, MGI: 5545911) was a gift from Dr. A. Economides of Regeneron Pharmaceuticals (*145*). The *Nanog-Cre* line carries a

transgene encoding a *Cre* recombinase downstream of the promoter of the *Nanog* gene and was maintained on C57BL/6J background. The *Nanog-Cre* line was genotyped using the following primers: Forward: 5'- GCA CTG ATT TCG ACC AGG TT-3' and Reverse: 5'- GCT AAC CAG CGT TTT CGT TC-3'. *Nanog-Cre* amplicon: 200 bp.

Mef2c-Cre Line: The *Mef2c-Cre* line (Tg(Mef2c-cre)2Blk, MGI: 3639735) carries a transgene encoding a *Cre* recombinase downstream of the promoter of the *Mef2c* gene (*143*) and was maintained on a C57BL/6J background. The *Mef2c-Cre* line was genotyped using the following primers: Forward: 5'- GGC TGG TTG GGA TAG AGA AAC- 3' and Reverse: 5'- TTT TCC ATG AGT GAA CGA ACC-3'. *Mef2c-Cre* amplicon: 612 base pairs (bp).

$Isl1^{Cre/+}$ Line: The *$Isl1^{Cre/+}$* line (*$Isl1^{tm1(cre)Sev}$*, MGI: 3623159) is a knock-in transgenic that targeted the start codon of the *Isl1* locus (*144*). The *$Isl1^{Cre/+}$* line was maintained on a C57BL/6J background. The *$Isl1^{Cre/+}$* line was genotyped using the following primers: Forward 1: 5'- GCC ACT ATT TGC CAC CTA GC-3', Reverse 1: 5'- CAA ATC CAA AGA GCC CTG TC-3', Reverse 2: 5'- AGG CAA ATT TTG GTG TAC GG-3'. *Isl1-WT* amplicon: 434 bp. *Isl1-Cre* amplicon: 300 bp.

Stra8-iCre Line: The Stra8-iCre line (Tg(Stra8-icre)1Reb, MGI:3779079) carries a transgene encoding an iCre recombinase downstream of the promoter of the *Stra8* gene (*176*) and was maintained on a C57BL/6J background. The *Stra8-iCre* line was genotyped using the following primers: Forward: 5'-AGG TCA TCT TGC TCC TTC CA-3' and Reverse: 5'-TCC TGT TGT TCA GCT TGC AC-3'. *Stra8-iCre* amplicon: 700 bp.

Stra8-iCre; $Nipbl^{Flox/+}$ Line: The *Stra8-iCre; $Nipbl^{Flox/+}$* line was generated by crossing the *Stra8-iCre* line with the *$Nipbl^{Flox/Flox}$* line. The *Stra8-iCre; $Nipbl^{Flox/+}$* line produces sperm that are either *$Nipbl^{Flox}$* (which is a WT allele) or *$Nipbl^{FIN}$* (which is a null allele). The *Stra8-*

iCre; Nipbl$^{Flox/+}$ line was maintained on a C57BL/6J background and genotyped for *Stra8-Cre* (see *Stra8-iCre* Line) and *Nipbl*$^{Flox/+}$ as described above (see *Nipbl*$^{Flox/Flox}$ Line).

Reporter Mouse Lines

Td-tomato-EGFP Line: *The reporter line used in this study was Td-tomato-EGFP* (*Gt(ROSA)26Sor*$^{tm4(ACTB-tdTomato,-EGFP)Luo}$, MGI:3716464) (*177*). Homozygous *Td-tomato-EGFP* mice on a C57BL/6J background were purchased from The Jackson Laboratory and maintained on a C57BL/6J background. The *Td-tomato-EGFP line was genotyped using the following primers: Forward 1: 5'-CTC TGC TGC CTC CTG GCT TCT-3', Reverse 1: 5'-CGA GGC GGA TCA CAA GCA ATA-3', and Reverse 2: 5'-TCA ATG GGC GGG GGT CGT T-3'. Td-tomato-WT amplicon: 330 bp. Td-tomato-EGFP amplicon: 250 bp.*

Nipbl$^{Flox/Flox}$*; Td-tomato-EGFP Line:* The *Nipbl*$^{Flox/Flox}$*; Td-tomato-EGFP* line was generated by crossing the *Nipbl*$^{Flox/Flox}$ line with the *Td-tomato-EGFP* line. The *Nipbl*$^{Flox/Flox}$*; Td-tomato-EGFP* line was maintained on a C57BL/6J background and genotyped for *Nipbl*$^{Flox/Flox}$ (see *Nipbl*$^{Flox/Flox}$ Line) and *Td-tomato-EGFP (see Td-tomato-EGFP Line)* as described above.

Generation of Mouse Embryos

For Visualization of Cre-Mediated Recombination Domains and Tissue-Specific Depletion of Nipbl Expression

To generate embryos for the visualization of Cre-mediated recombination domains and tissue-specific depletion of *Nipbl* expression, *Nipbl*$^{Flox/Flox}$*; Td-tomato-EGFP doubly homozygous* mice were mated with either *Mef2c-Cre* or *Isl1*$^{Cre/+}$ mice. When *Nipbl* $^{Flox/Flox}$*; Td-tomato-EGFP doubly homozygous* mice are mated with *Mef2c-Cre* mice, the resulting progeny are either *Nipbl*$^{Flox/+}$*; Td-tomato-EGFP* (which are phenotypically WT) or *Mef2c-Cre; Nipbl*$^{Flox/+}$*; Td-tomato-EGFP* (which are depleted for *Nipbl* expression in the *Mef2c-Cre* expression domain).

When *Nipbl* $^{Flox/Flox}$*; Td-tomato-EGFP doubly homozygous* mice are mated with *Isl1*$^{Cre/+}$ mice, the resulting progeny are either *Nipbl*$^{Flox/+}$*; Td-tomato-EGFP* (which are phenotypically WT) or *Isl1*$^{Cre/+}$*; Nipbl*$^{Flox/+}$*; Td-tomato-EGFP* (which are depleted for *Nipbl* expression in the *Isl1-Cre* expression domain).

For Tissue-Specific Depletion of Nipbl Expression Only

To generate embryos for the tissue-specific depletion of *Nipbl* expression, *Nipbl*$^{Flox/Flox}$ mice were mated with either *Mef2c-Cre* or *Isl1*$^{Cre/+}$ mice. When *Nipbl* $^{Flox/Flox}$ mice are mated with *Mef2c-Cre* mice, the resulting progeny are either *Nipbl*$^{Flox/+}$ (which are phenotypically WT) or *Mef2c-Cre; Nipbl*$^{Flox/+}$ (which are depleted for *Nipbl* expression in the *Mef2c-Cre* expression domain). When *Nipbl* $^{Flox/Flox}$ mice are mated with *Isl1*$^{Cre/+}$ mice, the resulting progeny are either *Isl1*$^{+/+}$; *Nipbl*$^{Flox/+}$ (which are phenotypically WT) or *Isl1*$^{Cre/+}$*; Nipbl*$^{Flox/+}$ (which are depleted for *Nipbl* expression in the *Isl1-Cre* expression domain).

For Visualization of Cre-mediated Recombination Domains Only

To generate embryos for the visualization of Cre-mediated recombination domains, *Td-tomato-EGFP homozygous* mice were mated with either *Mef2c-Cre* or *Isl1*$^{Cre/+}$ mice. When *Td-tomato-EGFP doubly homozygous* mice are mated with *Mef2c-Cre* mice, the resulting progeny are either *Td-tomato-EGFP* or *Mef2c-Cre; Td-tomato-EGFP*. When *Td-tomato-EGFP homozygous* mice are mated with *Isl1*$^{Cre/+}$ mice, the resulting progeny are either *Isl1*$^{+/+}$*; Td-tomato-EGFP* or *Isl1*$^{Cre/+}$*; Td-tomato-EGFP*.

For Generation of Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ *Mouse Embryos

To generate embryos that are either WT, haploinsufficient for *Isl1* (*Isl1*$^{+/-}$), haploinsufficient for *Nipbl* (*Nipbl*$^{+/-}$); or haploinsufficient for both *Nipbl* and *Isl1 (Nipbl*$^{+/-}$*; Isl1*$^{+/-}$*)*, we mated *Stra8-iCre*; *Nipbl*$^{Flox/+}$ male mice with *Isl1*$^{Cre/+}$ female mice. *Isl1*$^{Cre/+}$ mice

encode a *Cre*-recombinase in one allele of *Isl1* and are haploinsufficient for *Isl1*, as shown in Fig. 2.4D. *Stra8-iCre* mice encode a *Cre*-recombinase downstream of the promoter for the *Stra8* gene and *Stra8-iCre* recombines in the male germ line. *Nipbl*$^{Flox/+}$ mice have an inverted gene trap cassette encoding *Bgeo* that is flanked by target sites for *Cre*-recombinase in intron 1 of 1 allele of the *Nipbl* gene. In this inverted orientation, which we call *Flox*, 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 into a non-inverted orientation which we call *FIN*. In this non-inverted orientation, trapping of the *Nipbl* gene occurs and *Bgeo* is expressed as a reporter of successful gene trapping. Thus, *Stra8-iCre; Nipbl*$^{Flox/+}$ male mice make sperm that are either *Nipbl*^{Flox} (which is a wildtype allele and referred to here as *Nipbl*$^{+}$) or *Nipbl*^{FIN} (which is a null allele and referred to here as *Nipbl*$^{-}$) and when sperm from these mice fertilize eggs from *Isl1*$^{Cre/+}$ female mice, it results in littermates that are either WT, *Isl1*$^{+/-}$, *Nipbl*$^{+/-}$, or *Nipbl*$^{+/-}$; *Isl1*$^{+/-}$).

Collection of E17.5, E15.5, and E10.5 Mouse Hearts

At either 17.5-,15.5-, or 10.5-days post copulation, pregnant dams were euthanized by CO_2 inhalation followed by cervical dislocation. The uterine horns were dissected and placed in 1X DEPC PBS on ice, and different tissues were dissected with the aid of a dissection microscope. Tissues were collected for genotyping and stored at -20°C. Hearts for LFSM imaging were dissected out of the body and fixed in 4% Paraformaldehyde (PFA) for 2-4 hours at 4°C before they were transferred into clear, unobstructed brain/body imaging cocktails and computational analysis (CUBIC) solution. Hearts for RNA sequencing were dissected out of the body, frozen in liquid nitrogen, and stored at -80°C. Bodies were fixed in 4% PFA at 4°C overnight, before they were transferred into 1X PBS, and stored at 4°C for somite counting.

Zeiss Z.1 Light Sheeting Imaging

CUBIC solution was prepared with the following reagents: 25% w/v urea (Fisher Scientific: AC327380010), 25% w/v N,N,N,N-tetrakis(2-hydroxypropyl)ethylenediamine (Fisher Scientific: T0781500G), 15% w/v TritonX-100 (Fisher Scientific: AC327372500), and 35% w/v nanopure water (*178*). Hearts were cleared in 1 ml of CUBIC solution in a 2 ml tube at 37°C with a rocking motion at 20 rpm for 5-7 days. CUBIC solution was changed every 3-4 days.

1 ml embedding syringes with a diameter of 4.5mm (Fisher Scientific: 14823434) were prepared by cutting off the tip of syringe with a single-edge No. 9 razor blade (VWR: 55411-050). Embedding syringes were checked for jagged edges by drawing up agarose, letting the agarose cool, plunging the solidified agarose up and down, and checking under the microscope for cuts to the agarose. Embedding syringes with jagged edges were discarded. Hearts were removed from CUBIC solution and placed in a 1% low melting point agarose (Invitrogen: 16520050) solution (in water). 1% low melting point agarose was aspirated into the embedding syringe and an individual heart was embedded near the aperture. Then the syringe was placed on a flat surface on ice for at least 10-15 minutes. Embedded hearts were then equilibrated by being suspended and completely submerged in CUBIC solution at 4°C for 3-5 days.

Once equilibrated, embedded hearts were suspended in CUBIC solution and imaged using a Zeiss Z.1 Light Sheet microscope and ZEN software (using a 1.45x chamber and 5x illumination and detection objectives) at the Optical Biology Core Facility, at the University of California, Irvine. The distance settings for the illumination objectives were set to -0.4700 and 0.4700 and the configuration for the clearing chamber was selected. Each sample was rotated to image from either a dorsal or ventral view, and the laser focus and exposure time was adjusted to obtain the best quality image. All samples were imaged as a Z-stack and excited at a 100% laser

intensity with laser 561 nm for tdTomato excitation and 488nm for EGFP excitation. The Z-stack for each sample was set to capture the whole sample within the Z-stack. Z-stack settings: 746 slices, 6.71µm slice thickness. Z-stacks (CZI files) were processed and converted to Imaris files using the Imaris software converter.

Histology

Fixed heart samples were embedded in a 24x24x5mm mold (Fisher Scientific: 22363552) with 1.5% LMP agarose. For orientation purposes, a thin black strip of 1.5% agarose made with India Ink (Pro Art PRO-4100) was added. All hearts were also oriented differently for identification purposes and noted on a map. Once all hearts were embedded, the agarose block was carefully removed from the mold, placed in a Omnisette Tissue Casette (Fisher Scientific: 15197700A), and submerged in a container with 1X PBS. Samples were taken to the Department of Pathology at the University of California Irvine to be paraffin-embedded, sectioned serially at 7 µm with 2 sections per slide, and stained with hematoxylin and Eosin Y (H&E). Paraffin-sectioned heart images were obtained with a Discovery V8 stereomicroscope equipped with Axiovision software (Zeiss).

Heart Analysis

All heart samples from paraffin sectioning and LSFM imaging were positioned or oriented to a canonical angle for heart analysis where the atria are superior to the ventricles, the apex of the heart is inferior, the left atrium and ventricle is on the right, and the right atrium and ventricle is on the left. All samples were assessed from ventral to dorsal. Heart files from LFSM imaging were evaluated for the presence or absence of ASDs or VSDs using the volume viewer plug-in of ImageJ (v. 1.52t) and using a similar criterion outlined in (*18*). Paraffin-sectioned heart images were also evaluated with the same criteria outlined in (*18*). Imaris files of light

sheet imaged hearts were evaluated in ImageJ (v. 1.52t) for the presence or absence of ASD or VSD using the volume viewer plug-in and the criterion outlined in (*18*).

Hearts were scored positive for outflow tract defects as follows: Hearts were scored positive for persistent truncus arteriosus (PTA) when the aorta and pulmonary trunk were observed as fused. Hearts were scored positive for double outlet right ventricle (DORV), in which the aorta and pulmonary trunks are observed as both exiting the right ventricle. Hearts were scored positive for overriding aorta (OA) when the aorta was positioned directly over a ventricular septal defect (VSD), instead of over the left ventricle. Fisher's exact tests were used to calculate *P*-values for the frequency of heart defects (ASD, VSD, and/or outflow tract defects) at E17.5 and E15.5 for all crosses. Ventricular volumes were calculated using the surfaces function on Imaris by manually contouring various sections throughout the ventricles and creating a 3D surface. T-Tests were used to calculate *P*-values for ventricular volume measurements at E17.5 and E15.5.

X-gal Staining

Whole mount x-gal staining was performed as described (*179*) and imaged with a Zeiss Discovery V8 stereomicroscope equipped with Axiovision software.

Reverse Transcription-Quantitative Polymerase Chain Reaction

RNA was extracted from E10.5 WT and *Isl1*$^{Cre/+}$ hearts using the Monarch Total RNA Miniprep Kit (New England Biolabs). cDNA was made using iSCRIPT reverse transcriptase (BioRad). qRT-PCR for *Isl1* was performed using iTaq SYBR green (BioRad) as per manufacturer's instructions, *Rpl4* was used as the housekeeping gene. *Rpl4* Primer 1: 5'-ATC TGG ACG GAG AGT GCT TT-3'. *Rpl4* Primer 2: 5'-GGT CGG TGT TCA TCA TCT TG-3'.

Isl1 Primer 1: 5'- ACC AAT TGT CCA ACC ACC AT-3'. *Isl1* Primer 2: 5'- TCC CAT CCC TAA CAA AGC AC-3'.

RNA Purification, Library Construction, and RNA sequencing

To isolate RNA from 10.5 dpc mouse hearts for library construction, we used the New England BioLabs Monarch Nucleic Acid Purification kit (Cat No: T2010S). Libraries for RNA sequencing were constructed using the Ultra II RNA Library Prep Kit for Illumina (NEB #E7765). RNA libraries were submitted to the UCI Genomics High-Throughput Facility (GHTF) for sequencing on the Illumina Nova Seq6000. RNA libraries were multiplexed and sequenced on a single lane to a depth of approximately 20 million reads each. Sequences were demultiplexed by GHTF.

Read Mapping and Depth Normalization

Read mapping was performed by GHTF using the GRCm38 mm10 mouse reference genome. Genes with high quality alignments were assigned a transcript count of 1 for each read. GHTF subsequently delivered a matrix of transcript counts for each gene across all libraries. To normalize transcript counts for sequencing depth across all libraries, we calculated the sum of transcript counts in each library and divided these sums by the lowest sum, resulting in a scaling factor for each library. Transcript counts were then multiplied by their library's respective scaling factors to arrive at depth normalized counts.

Differential Gene Expression Analysis

To identify differentially expressed genes (DEGs) between pairs of genotypes, we performed an ANOVA test for all genes using normalized transcript counts. Resulting P-values were adjusted for false-discovery using the Benjamini-Hochberg method. Genes with Q-values less than 0.05 indicated that expression of the given gene differed between at least two or more

four genotypes. To identify which genotypes differed in its expression of that gene, we performed a Tukey-Kramer test using an alpha of 0.05.

Principal Component Analysis

Transcript counts from hearts of the same genotype were aggregated together to generate pseudobulked libraries representing hearts of all 4 genotypes: WT, *Isl1*$^{+/-}$*, Nipbl*$^{+/-}$*, and Nipbl*$^{+/-}$*; Isl1*$^{+/-}$. Sequencing depth across all the pseudobulked libraries were normalized by calculating a sum of transcript counts in each library and dividing these sums by the lowest sum, producing a scaling factor for each library. Transcript counts were then multiplied by their library's respective scaling factors to arrive at depth normalized counts.

The mean expressions of detected genes were calculated across all pseudobulked libraries, from which coefficients of variation were also calculated. The coefficients of variation for all genes were plotted against their mean expression. A LOESS fit was performed across all points in R using the *loess* function of the *stats* package. The command syntax used was *loess(dependent_variable ~ independent_variable)*, where *dependent_variable* represents a gene's coefficient of variation and *indepdent_variable* represents a gene's mean expression. All genes were binned into 32 bins along their mean expression and for each bin, the standard deviation of the coefficient of variation from the LOESS fit was calculated. A Z-score was calculated for gene by dividing its coefficient of variation by the standard deviation of its bin. Genes with Z-score greater than 2 were used as input into principal component analysis (PCA). PCA was performed in R using the *prcomp* function from the *stats* package. The command syntax used was *prcomp(data)*, where *data* represents the expression of genes selected as input into PCA across all hearts of all four genotypes.

Acknowledgements

I thank Rosaysela Santos (RS) for suggesting the use of light sheet fluorescent imaging for the identification of defects in mouse hearts. Appreciation extends to Louise Villagomez (LV) and RS for their efforts in adapting the CUBIC clearing reagent and developing protocols for imaging mouse hearts using the Zeiss Z.1 Light Sheet microscope. I acknowledge Brian Bui (BB) and LV for their diligent management of mouse lines, crosses, dissection, PCR genotyping, clearing, and imaging of mouse hearts. Special thanks to RS and Anne L. Calof (ALC) for proposing the examination of ventricular volumes. Recognition goes to BB, LV, and RS for the identification of heart defects and precise measurement of ventricular volumes. I am grateful to BB, LV, and Martha E. Lopez-Burks (MEL) for their X-gal staining of mouse hearts. I also acknowledge MEL for her work in conducting RT-qPCR for Isl1 in mouse hearts, identifying non-viable embryos, measuring crown-rump lengths, and selecting E10.5 hearts for RNA sequencing. I thank Arthur D. Lander (ADL) for our discussions on statistical analysis, which have been invaluable, and we thank him for his guidance. I also thank ALC for recognizing *Pax6, Pitx1*, and *Hoxc4* as homeodomain transcription factors and her hypothesis about the relationship between gene dosage and the development of heart defects. I also thank LV and MEL for their meticulous preparation of figures. Gratitude is due to ALC for her contributions to the visualization of data and figures. I am profoundly grateful to MEL, RS, and ALC for their thorough review and editing of this manuscript. I extend gratitude to the National Institute of Neurological Disorders and Stroke of the National Institutes of Health (NIH) for supporting me with a training grant (NINDS-NIH T32NS082174) and the National Heart, Lung, and Blood Institute of the NIH for supporting this study with a research grant (NHLBI-NIH R01HL138659).

CHAPTER 3

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

This chapter was published in Science Advances (*180*).

Abstract

In animal models, *Nipbl*-deficiency phenocopies gene expression changes and birth defects seen in Cornelia de Lange Syndrome (CdLS), the most common cause of which is *Nipbl*-haploinsufficiency. Previous studies in *Nipbl*$^{+/-}$ mice suggested that heart development is abnormal as soon as cardiogenic tissue is formed. To investigate this, we performed single-cell 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.

Introduction

Each year, 1 of every 33 babies in the United States is born with a birth defect (*181*), the most prevalent of which are congenital heart defects (CHDs), neural tube defects, and cleft lip/palate (*13*). Given the major impact that birth defects have on infant mortality and morbidity (*182, 183*), 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 (*36*), and is characterized by craniofacial anomalies, delays in growth and maturation, intellectual disability, neurological impairments, abnormalities of limbs, especially arms and hands, coupled with issues in the visual, auditory, gastrointestinal, genitourinary, and cardiopulmonary systems (*184*).

The majority of CdLS cases – more than 55% – are caused by heterozygous mutations in the gene *Nipped-B-like* (*NIPBL*) (*185*), named for its homology to the Drosophila gene *Nipped-B*. These mutations often produce a non-functional protein, suggesting CdLS arises from haploinsufficiency (*95*). Remarkably, even a subtle 15% reduction in *NIPBL* gene expression can produce a mild yet recognizable CdLS phenotype (*186*). Such observations highlight the importance of precise *NIPBL* gene dosage in human development.

The *NIPBL* gene encodes a universally conserved protein that plays a role in loading cohesin onto chromosomes (*139*). Cohesin, similarly conserved and ubiquitous, is a four-subunit protein complex (Smc1, Smc3, Rad21, and either SA1 or SA2) essential for chromosome organization and genome stability (*7*). Mutations in the cohesin subunits *SMC1* and *SMC3* account for a small proportion of clinically mild CdLS (~5% and <1%, respectively) (*4, 57, 187*). Additionally, mutations in *HDAC8*, which catalyzes cohesin release from chromatin during mitosis, are found in a distinct subset of CdLS patients (*188*). Mutations in *RAD21* have also surfaced in individuals exhibiting a CdLS-like phenotype but with substantially milder cognitive

impairment (*97*). Collectively, these findings reinforce the idea that impairment of cohesin function contributes to CdLS.

Cohesin was initially identified for its role in sister chromatid cohesion during mitosis (*7*). However, pronounced defects in sister chromatid cohesion or irregularities in mitosis have not been observed in either CdLS patients or *Nipbl*-haploinsufficient (*Nipbl*$^{+/-}$) mice (*10, 112*), suggesting that cohesin has additional functions. Studies in CdLS patient cells and *Nipbl*-haploinsufficiency animal models indicate that cohesin is involved in transcriptional regulation (*189*). Specifically, *Nipbl*-haploinsufficiency leads to alterations in the expression of many hundreds to thousands of genes (*10*). Interestingly, many of the affected genes are regulated by long-distance enhancers (*117*), aligning with the emerging concept of NIPBL and cohesin as critical determinants in DNA-looping (*119*).

Interestingly, most gene expression changes in *Nipbl*-deficient animals are small, usually less than 2-fold. Though likely to be inconsequential individually, such small changes can act collectively to produce structural and functional defects. For example, in zebrafish, joint depletion of two developmental genes downregulated by *nipbl*-deficiency produced a CdLS-like phenotype (*11*), suggesting that altered gene expression is the ultimate cause of developmental and physiological abnormalities. Thus, CdLS exemplifies a class of genetic disorders known as "transcriptomopathies" (*121*), in which the cumulative or synergistic effects of minor disturbances to gene expression lead to developmental abnormalities.

Previously, we reported that *Nipbl*-haploinsufficient mice displayed birth defects phenocopying those in CdLS (*10*). These included CHDs (primarily atrial septal defects), in about 30% of *Nipbl*$^{+/-}$ mice. Subsequently, we used a conditional *Nipbl* allelic series to investigate the role of *Nipbl* expression in the production of CHDs (*18*). That study showed that

Nipbl$^{+/-}$ mice exhibit heart abnormalities early in development: At embryonic day 13.5 (E13.5), 70% of *Nipbl*$^{+/-}$ mice displayed delays in ventricular septal fusion, whereas three days earlier, at E10.5, 100% showed right ventricle hypoplasia (*18*). In situ hybridization experiments showed reduced expression of two transcription factors crucial for early heart progenitor cell growth and differentiation: *Nkx2-5* and *Mesp1* (*18*). That structural abnormalities may begin at the earliest stages of heart development was suggested by results in in *nipbl*-morphant zebrafish, in which defects in the initial migration of cardiogenic mesoderm were detected as early as 18 hours post-fertilization (*11*). These results suggested that, at least for CHDs, causal events may occur as early as gastrulation, when the three primary germ layers form and the earliest progenitor cells of major tissues and organs begin to differentiate (*136*).

In this study, we sought to identify developmental alterations during gastrulation that might account for birth defects in *Nipbl*$^{+/-}$ mice. We used single-cell RNA sequencing (scRNAseq) to compare the cellular compositions, lineage trajectories, and transcriptional landscapes of *Nipbl*$^{+/-}$ mouse embryos to their WT counterparts at two stages spanning the end of gastrulation: the late bud- (LB) stage (E7.5) and cardiac crescent- (CC) stage (E7.75). Our findings reveal that *Nipbl*$^{+/-}$ embryos have the same cell populations as WT embryos but display subtle misallocation of specific mesodermal cell populations. Our evidence suggests this occurs as a result of alterations in specific cell fate decisions, including the choice by mesoderm cells to progress toward a non-cardiac versus cardiac fate. Our observations strongly suggest these events cannot be attributed to changes in apoptosis, cell proliferation, global developmental delay, or the structure of cell lineages. As in earlier research, we observed that most gene expression changes in *Nipbl*-deficient tissues at these stages were small (*10*), but we also identified several key developmental genes that are more markedly misexpressed in *Nipbl*$^{+/-}$

embryos, the most notable of which was *Nanog*. *Nipbl*$^{+/-}$ embryos failed to downregulate *Nanog* at the end of gastrulation, which normally occurs in all but germ cells. As a result, the misexpression of many *Nanog* target genes was observed in all germ layers. We also saw substantial underexpression of *Hox* genes and overexpression of Nodal signaling genes, which play roles in anterior-posterior and left-right patterning, respectively. As a result, we propose a model in which birth defects in CdLS arise from the prolonged overexpression of *Nanog* and dysregulation of developmental pathways governing axial patterning, resulting in the misdirection of cell fate decisions, and misallocation of specific progenitor cell populations.

Results

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

To investigate early factors influencing birth defects in CdLS, we generated WT and *Nipbl*$^{+/-}$ littermate embryos by crossing *Nanog*$^{Cre/+}$ mice (*145*) with *Nipbl*$^{Flox/Flox}$ mice (*18*) as described in Methods and shown in Fig. 3.1A. To ensure an accurate accounting of cell populations and proportions at well-defined stages of early embryonic development, we generated a large excess of embryos and sorted them into groups of narrowly defined stage based on morphological criteria (*190*). Using the 10X Genomics Chromium Single Cell Expression platform (*191*), we performed scRNAseq on 5 WT and 6 *Nipbl*$^{+/-}$ samples of late bud stage (LB) embryos (E7.5), and 8 WT and 8 *Nipbl*$^{+/-}$ samples of cardiac crescent stage (CC) embryos (E7.75) (Fig. 3.1B and Fig. S3.1). Given the low cell numbers in individual LB-stage embryos, and the 10X Genomics Chromium Single Cell Expression system's capture efficiency, samples at the LB stage consisted of pairs of embryos, whereas at the CC-stage they consisted of single embryos. Pairing LB-stage embryos facilitated the capture of rare cell groups, such as primordial germ cells, which typically comprise about 50 cells per embryo at this stage (*192*).

At LB-stage, we captured a median of 2,537 cells per sample, with a median of 17,301 RNA transcripts per cell (Fig. S3.2 and Table S3.3.1). At CC-stage, we captured a median of 4,116 cells per sample, with a median of 14,997 RNA transcripts per cell (Table S3.3.1). As expected, all *Nipbl*$^{+/-}$ embryos expressed both *Cre* and *Betageo*, which report an inactivated *Nipbl* allele (*18*), while WT embryos did not (fig. S3, A and B). *Nipbl*$^{+/-}$ embryos across both stages expressed *Nipbl* at levels that were approximately 50% lower than WT counterparts (fig. S3C).

We corrected for batch effects among embryos of identical stage and genotype using Seurat's integration protocol (*193*) (Fig. 3.1C and Fig. S3.4), which identifies anchors (cells of similar gene expression across samples of the same biological condition) and uses them to align cells into a space shared by all samples. Recognizing that all cells in *Nipbl*$^{+/-}$ tissues exhibit substantial gene expression changes (*10*), and considering that clustering algorithms rely on differences in gene expression among cells (*193*), we implemented measures to prevent our cell clustering from being skewed by gene expression alterations attributable to *Nipbl*-haploinsufficiency. We did this by first clustering cells from WT samples (Fig. 3.1C), using a robust iterative clustering method that considered both intra-cluster stability and inter-cluster variation (Fig. S3.5) to determine the optimal number of clusters for cells at each developmental stage. WT cells were clustered into 18 cell populations at LB-stage (Fig. 3.1D) and 26 cell populations at CC-stage (Fig. 3.1E). We then projected *Nipbl*$^{+/-}$ cells onto these WT cell populations at corresponding stages (Fig. 3.1C). We observed that all *Nipbl*$^{+/-}$ cells could be projected to WT cell populations, with each WT population receiving some *Nipbl*$^{+/-}$ cells (Fig. 3.1, D and E). This pattern held true across individual *Nipbl*$^{+/-}$ samples, as well as the total *Nipbl*$^{+/-}$ cells in aggregate at each stage (Fig. 3.1, F and G). To ensure that any differences

observed in cell populations were not the result of technical artifacts introduced by the projection algorithm, we performed additional studies in which WT cells were projected onto *Nipbl*$^{+/-}$ clusters (Fig. S3.6 and Fig. S3.14). In both cases, *Nipbl*$^{+/-}$ cells contributed to all clusters, but as we describe later, in varying proportions (Fig. 3.3).

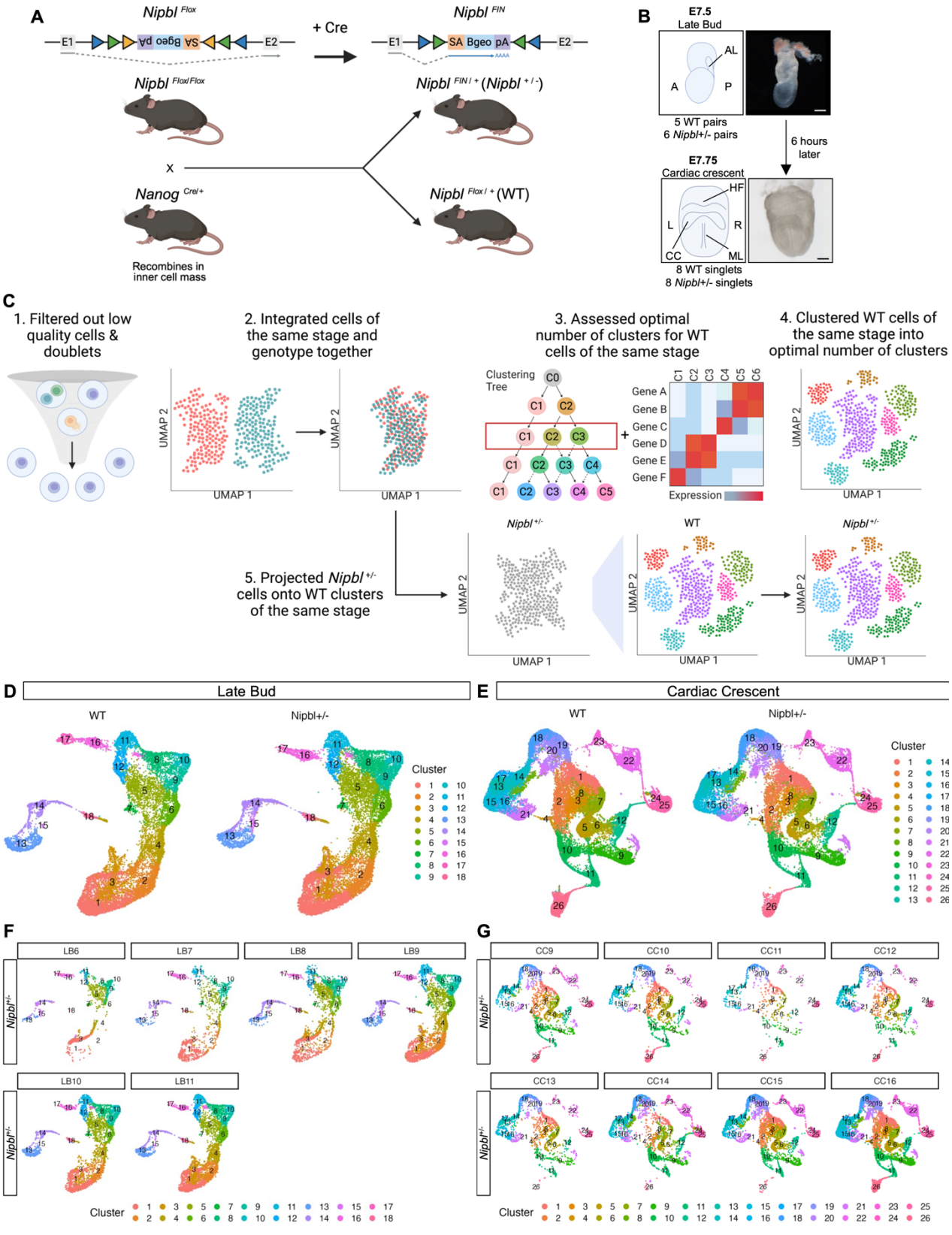


Fig. 3.1. *Nipbl*$^{+/-}$ mice do not lack any cell populations found in WT mice.

(**A**) *Nipbl* alleles used in this study. *Nipbl*^{Flox} contains an inverted gene trap cassette encoding *b-geo* flanked by *Cre* recombinase target sites in intron 1 of *Nipbl* gene alleles (*18*). 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 *b-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 (*145*). (**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.

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

To assign biological identities to clustered cell populations, we used WT cells as a reference. For each WT cell population, we performed differential gene expression analysis (DGEA) by comparing it to all other cells within embryos of identical stage (Mann-Whitney U test). We detected markers indicative of germ layer identity in every cluster (Table S3.3.2), enabling us to associate each cluster with a specific germ layer at each developmental stage (Fig. 3.2A). Figure 2B displays all cells from all samples, categorizing them into germ layers at both LB- and CC-stages of development. Figures 2C and 2D further depict the assignment of cells based on the expression of germ layer markers, at LB- and CC-stages, respectively. These markers include those for ectoderm, such as *Utf1, Sox2,* and *Pou3f1* (*194-196*), mesoderm, including *T, Hand1, Vim, Twist1,* and *Prrx2 (197-201)*, and endoderm, *Ttr (202)*.

To identify cell types within germ layers, we re-applied DGEA after clustering individual germ layers. For the endoderm, this enabled us to identify definitive endoderm progenitors (DEP) (*Sox4* and *Foxa2*) (*203, 204*), as well as extraembryonic endoderm (XEnd) (*Hnf4a* and *Tbx3) (205, 206)* at LB-stage (Fig. 3.2E and Table S3.3). In CC-stage embryos, we captured identical cell types in the endoderm, and additionally, visceral endoderm (VE) (*Hnf4a* and *Esx1*) (*205, 207*) (Fig. 3.2F and Table S3.3). DEPs and XEnd give rise to the gut and placenta, respectively.

Within the ectoderm of LB-stage embryos, we identified early progenitor populations including intraembryonic ectoderm progenitors (IEPs) (*Pou3f1* and *Otx2*) (*196, 208*), neural ectoderm progenitors (NEPs) (*Cdx1* and *Nkx1-2) (209, 210)*, and surface ectoderm (SE) (*Id1, Msx2,* and *Msx1*) (*211*) (Fig. 3.2, E and G, and Table S3.4). Additionally, we detected a small portion of extraembryonic ectoderm (XEct) (*Elf5* and *Tead4*) (*212, 213*), which was expected, as

a small quantity of XEct cells was included in our dissection. By CC-stage, IEPs, NEPs, and XEct were no longer detectable, a phenomenon previously noted (*214*). Rather, we continued to detect surface ectoderm (SE). We also detected six distinct neural populations (Fig. 3.2, F and H, and Table S3.4). These neural populations included: prosencephalic neural plate (*Six3* and *Hesx1*) (*215, 216*), mesencephalic neural plate (*En1* and *Pax2*) (*217, 218*), rhombencephalic neural plate (*Hoxa1* and *Hes3*) (*219, 220*), midline neural plate (*Foxa2* and *Nkx2-9*) (*221, 222*), spinal neural plate (*Sox2* and *Ezr*) (*223, 224*), and neural crest (*Pax3* and *Foxd3*) (*225, 226*). Collectively, these populations constitute the nascent progenitors of the brain, spinal cord, and peripheral nervous system.

Within the mesoderm of LB-stage embryos, we identified six distinct groups of cells, detailed in Fig. 3.2, E and G, and Table S3.5. These consisted of: 1) neuromesodermal progenitors (NMPs): expressing markers *T,* and *Nkx1-2 (197, 210)*; 2) derivatives of NMPs: including *Mesp1*-expressing mesoderm progenitors (MMPs) (*Mesp1*) (*227*), two types of first heart field cells (FHF1 and FHF2) marked by *Hand1 (198)*; cells of the second heart field (SHF) with *Id2* and *Mef2c (228, 229)*; and three kinds of paraxial mesoderm cells, further categorized into paraxial mesoderm progenitors (PMPs) (*Tbx6* and *Meis2) (230, 231)*, trunk paraxial mesoderm (TPM) (*Foxc2* and *Tcf15*) (*232, 233*), and head paraxial mesoderm (HPM) (*Alx1* and *Tcf15*) (*233, 234*); 3) hematopoietic cells: including hemangioblasts (HMBs) (*Tal1* and *Etv2*) (*235, 236*) and primitive erythrocytes (PEs) (*Gata1* and *Tal1*) (*235, 237*); and 4) two rare populations: comprising cells of the primitive node (PN) (*Noto* and *Foxj1*) (*238, 239*), and primordial germ cells (PGCs) (*Tfap2c* and *Msx1*) (*240, 241*). At CC-stage, we detected all mesodermal populations found earlier, along with several that emerge later in development. These comprised lateral plate mesoderm progenitors (LPMP) (*Tlx2* and *T*) (*197, 242*), additional

populations of *Mesp1*-expressing mesoderm progenitors (MMP1 and MMP2) (*Mesp1*) (*227*), cardiomyocytes (CDMs) (*Mef2c* and *Nkx2-5*) (*243, 244*), and presomitic mesoderm (PSM) (*Tcf15* and *Meox1*) (*233, 245*) (Fig. 3.2, F and H, and Table S3.5). Collectively, these progenitors are responsible for the formation of structures including skeletal muscle, bone, blood, and heart.

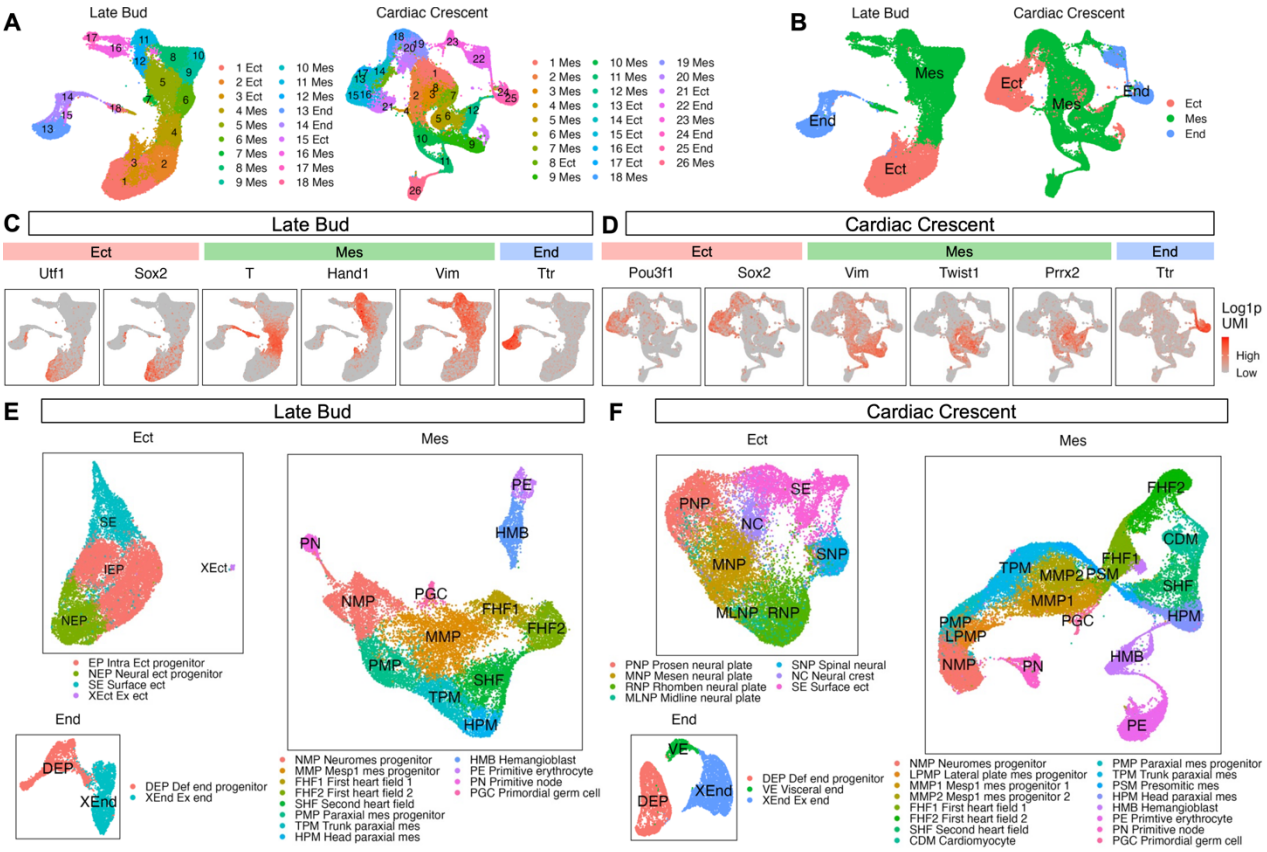

A
Late Bud
Cardiac Crescent
B
Late Bud
Cardiac Crescent
Ect
Mes
End
C
Late Bud
Ect
Mes
End
Utf1
Sox2
T
Hand1
Vim
Ttr
D
Cardiac Crescent
Ect
Mes
End
Pou3f1
Sox2
Vim
Twist1
Prrx2
Ttr
Log1p UMI
High
Low
E
Late Bud
Ect
Mes
SE
PE
HMB
PN
NEP
XEct
NMP
PGC
FHF1
FHF2
MMP
PMP
SHF
TPM
HPM
EP Intra Ect progenitor
NEP Neural ect progenitor
SE Surface ect
XEct Ex ect
NMP Neuromes progenitor
MMP Mesp1 mes progenitor
FHF1 First heart field 1
FHF2 First heart field 2
SHF Second heart field
PMP Paraxial mes progenitor
TPM Trunk paraxial mes
HPM Head paraxial mes
HMB Hemangioblast
PE Primitive erythrocyte
PN Primitive node
PGC Primordial germ cell
End
DEP
XEnd
DEP Def end progenitor
XEnd Ex end
F
Cardiac Crescent
Ect
Mes
PNP
SE
NC
SNP
MNP
MLNP
RNP
FHF2
CDM
FHF1
TPM
MMP2
PSM
SHF
MMP1
PMP
PGC
HPM
LPMP
NMP
PN
HMB
PE
PNP Prosen neural plate
MNP Mesen neural plate
RNP Rhomben neural plate
MLNP Midline neural plate
SNP Spinal neural
NC Neural crest
SE Surface ect
End
VE
DEP
XEnd
DEP Def end progenitor
VE Visceral end
XEnd Ex end
NMP Neuromes progenitor
LPMP Lateral plate mes progenitor
MMP1 Mesp1 mes progenitor 1
MMP2 Mesp1 mes progenitor 2
FHF1 First heart field 1
FHF2 First heart field 2
SHF Second heart field
CDM Cardiomyocyte
PMP Paraxial mes progenitor
TPM Trunk paraxial mes
PSM Presomitic mes
HPM Head paraxial mes
HMB Hemangioblast
PE Primitive erythrocyte
PN Primitive node
PGC Primordial germ cell
G
Late Bud
Ect
Mes
IEP
NEP
SE
XEct
NMP
MMP
FHF1
FHF2
SHF
PMP
TPM
HPM
HMB
PE
PN
PGC
Cdx1
Cdx2
Nkx1-2
Sp5
Gbx2
Hoxa1
Pou5f1
Hes3
Hoxb1
Etv4
T
Id1
Msx2
Msx1
Tfap2a
Hand1
Gata3
Id3
Id4
Dlx5
Id2
Rhox6
Elf5
Sox7
Epas1
Tead4
Rhox9
Esrrb
Gata6
Zfp42
Sox17
Pou3f1
Otx2
Irx3
Hesx1
Sox2
Irx5
Tcf7l1
Tal2
Sox9
Zic5
Pou5f1
T
Nkx1-2
Cdx1
Sp5
Hes7
Mixl1
Mesp1
Cdx4
Hand1
Foxf1
Snai1
Isl1
Cdx2
Hoxb6
Pitx1
Hoxd9
Msx1
Hoxa10
Msx2
Prdm6
Klf6
Lrrfip1
Snai2
Id2
Hmga2
Mef2c
Id3
Gata6
Meis1
Id1
Tbx6
Hoxb1
Meis2
Lef1
Foxb1
Hoxa1
Foxc2
Tcf15
Irx5
Irx1
Twist1
Irx3
Tcf7l1
Alx1
Otx2
Pitx2
Tal1
Etv2
Hhex
Gata2
Fli1
Sox7
Lyl1
Gata1
Runx1
Gfi1b
Klf1
Ikzf1
Noto
Foxj1
Foxa2
Foxd4
Tfap2c
Elf3
Irf1
End
DEP
XEnd
Tead2
Sox4
Hmga2
Tgif1
Hmgb3
Hmgb1
Sp5
Foxa2
Sox11
Foxa1
Foxo4
Id2
Hnf4a
Maf
Msx1
Creb3l3
Runx1
Tulp4
Tbx3
H
Cardiac Crescent
Ect
Mes
PNP
MNP
RNP
MLNP
SNP
NC
SE
NMP
LPMP
MMP1
MMP2
FHF1
FHF2
SHF
CDM
PMP
TPM
PSM
HPM
HMB
PE
PN
PGC
Six3
Hesx1
Pou3f1
Otx2
Foxd4
Lhx5
Fezf1
En1
Sox21
Irx5
Pax2
Foxb1
Irx3
Hoxa1
Hes3
Meis2
Hoxb2
Cdx1
Hnf1b
Tshz1
Foxa2
Nkx2-9
Nkx6-1
Sox2
Gli1
Nkx6-2
Twist1
Foxc2
Prrx2
Tcf15
Meox1
Hmga2
Lef1
Pax3
Foxd3
Sox9
Id1
Nr2f1
Tfap2a
Id3
Peg3
Gata3
Msx1
Cdx1
Nkx1-2
Hes3
Hoxb8
Cdx2
Cdx4
Hand1
Osr1
Hoxd9
Foxf1
Lhx1
Hmga2
Mesp1
Meis1
Pitx1
Tbx4
Hoxa10
Hoxc9
Peg3
Snai2
Lrrfip1
Klf6
Id2
Prrx2
Gata6
Mef2c
Nkx2-5
Gata5
T
Etv4
Hes7
Sp5
Hoxb9
Tbx6
Lef1
Hoxb1
Tcf15
Meox1
Foxc2
Meis2
Pax3
Twist1
Irx5
Irx3
Tbx1
Etv2
Fli1
Hhex
Tal1
Elk3
Klf1
Gata1
Zfpm1
Runx1
Foxa2
Sox9
Noto
Sox4
Irf1
Parp12
Tfap2c
Klf2
Elf3
Msx1
Tlx2
Std Log1p UMI
High
Low
DEP
VE
XEnd
End
Sox4
Hmga2
Hmgb1
Peg3
Hmgb3
Hes1
Zeb2
Pou5f1
Hoxb1
Prrx2
Cdx1
Hes3
Maf
Creb3l3
Glmp
Hnf4a
Hmg20b

Fig. 3.2. 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 layers in LB- and CC-stage embryos. Expression of genes marking germ layers of (**C**) LB- and (**D**) CC-stage embryos in UMAP. UMAP of cell populations in germ layers of (**E**) LB- and (**F**) CC-stage embryos. Heatmap of fold change in expression of the most differentially expressed transcription factor genes (lowest *Q*-values from Mann-Whitney U Test) (*246*) of cell populations from all other cell populations in germ layers of (**G**) LB- and (**H**) CC-stage embryos.

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

We analyzed the cellular composition of embryos at both LB- and CC-stage, quantifying the proportion of cells within each germ layer that contributed to the total cell count. LB-stage mutants exhibited 13% fewer mesoderm cells (Fig. 3.3, A and B), while concurrently displaying a greater proportion of endoderm and ectoderm cells. We sought to ascertain if this decrease in mesoderm was generalized or restricted to specific subpopulations. Focusing on LB-stage embryos, we merged mesodermal cell subpopulations of similar biological identity together (FHF1 + FHF2 into FHF, and PSM + CPM into PM) (Fig. 3.3C) and calculated the percentage of cells across these subpopulations, relative to the overall mesoderm. The findings (Fig. 3.3D) revealed a 77% reduction in PEs *Nipbl*$^{+/-}$ embryos (P=0.06, T-Test). Additionally, differences were observed in two mesodermal derivatives of the neuromesoderm: 1) Mutants had 24% fewer first heart field cells (FHF) (Fig. 3.3D); and 2) *Nipbl*$^{+/-}$ embryos showed a contrasting pattern in paraxial mesoderm (PM), a 33% increase in PM cells compared to WT (Fig. 3.3D).

Since *Nipbl*$^{+/-}$ cells were projected onto WT cell populations, we wondered whether any observed differences between *Nipbl*$^{+/-}$ and WT embryos might have occurred as a result of misprojecting cells, i.e. assigning cells to incorrect clusters. To investigate this, we independently clustered *Nipbl*$^{+/-}$ cells and projected WT cells onto those clusters; we refer to this as reverse projection (Fig. S3.6). If observed differences in the allocation of *Nipbl*$^{+/-}$ cells to particular clusters had been a result of misclassification due to technical limitations of projecting cells, then the same differences should not appear when reverse projection was done. Indeed, if misclassified cells were simply distributed at random one should expect reverse projection to produce differences opposite to those produced by the original, forward projection. This was not

found to be the case, however. Cluster assignments resulting from reverse projection quantitatively resembled those of forward projection (Fig. S3.6, A to D). Germ layer identities, which were based on the expression of specific markers (Table S3.6), were generally the same in the reverse projection analysis as well (Fig. S3.6, E to H).

Finally, we quantified the proportion of cells within each germ layer and mesodermal population following reverse projection. Consistent with the findings from forward projection (Fig. 3.3B), *Nipbl*$^{+/-}$ embryos showed 13% fewer mesoderm cells than WT embryos, while concurrently showing more endoderm and ectoderm cells (Fig. S3.6M). Likewise, the mesoderm of *Nipbl*$^{+/-}$ embryos showed fewer FHF cells (24% fewer) and more PM cells (22% more) than WT embryos (Fig. 3.3D, Fig. S3.6N). (Because very few PEs are present in *Nipbl*$^{+/-}$ embryos (Fig. 3.3D), it was not possible to cluster these cells into a distinct cell population during reverse projection (Fig. S3.6, K and L).) Altogether, these data indicate that differences in cell population sizes observed in *Nipbl*$^{+/-}$ embryos represent true biological changes and are not a consequence of technical issues, such as misprojection.

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

Cellular composition changes very quickly during and after gastrulation; one possible explanation for differences in proportions of cell types could be a change in the overall pace of development, such that embryos of one genotype were slightly delayed or accelerated relative to the other. To investigate whether this was the case for *Nipbl*$^{+/-}$ embryos, we employed pseudotemporal ordering. To ensure that pseudotime accurately mirrored real developmental time, we augmented our analysis with scRNAseq data from an additional two pairs of WT embryos at the early bud (EB) stage and two pairs of WT embryos at the early head fold (EHF) stage (Fig. 3.3E and Fig. S3.7). EB-stage embryos are two hours younger and EHF-stage

embryos are two hours older than LB-stage embryos, so these supplementary samples provided the temporal resolution necessary to discern even minor changes (less than 2 hours) in developmental timing (*190*). We eliminated low-quality cells and doublets (Fig. 3.3F, Fig. S3.8 and Table S3.8) before integrating cells of the same stage (Fig. S3.9). We then clustered EB-stage cells into 8 distinct populations and EHF-stage cells into 13 (Fig. S3.10, A and B). Each population was subsequently annotated as ectoderm (Ect), mesoderm (Mes), or endoderm (End) (Fig. 3.3G, and Fig. S3.10, C and D) based on the expression of specific germ layer markers (Table S3.9).

We ordered cells from each embryo at every developmental stage (EB, LB, and EHF) using URD (*247*), which constructs a diffusion map of transition probabilities and, starting with an assigned group of root cells, performs a probabilistic breadth-first graph search using the transition probabilities. When we visualized the arrangement of cells from each stage based on their density in pseudotime, we found that cells from WT embryos ordered in accordance to their developmental stage: EB came first, followed by LB, and finally EHF (Fig. 3.3H), while showing partial overlap between cells of different stages. This confirmed that the pseudotime values we acquired were an accurate representation of actual developmental timing. When the pseudotime orderings of WT and *Nipbl*$^{+/-}$ LB-stage embryos were compared with each other, we found no statistically significant deviations (Kolmogorov-Smirnov test) (Fig. 3.3I). 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.

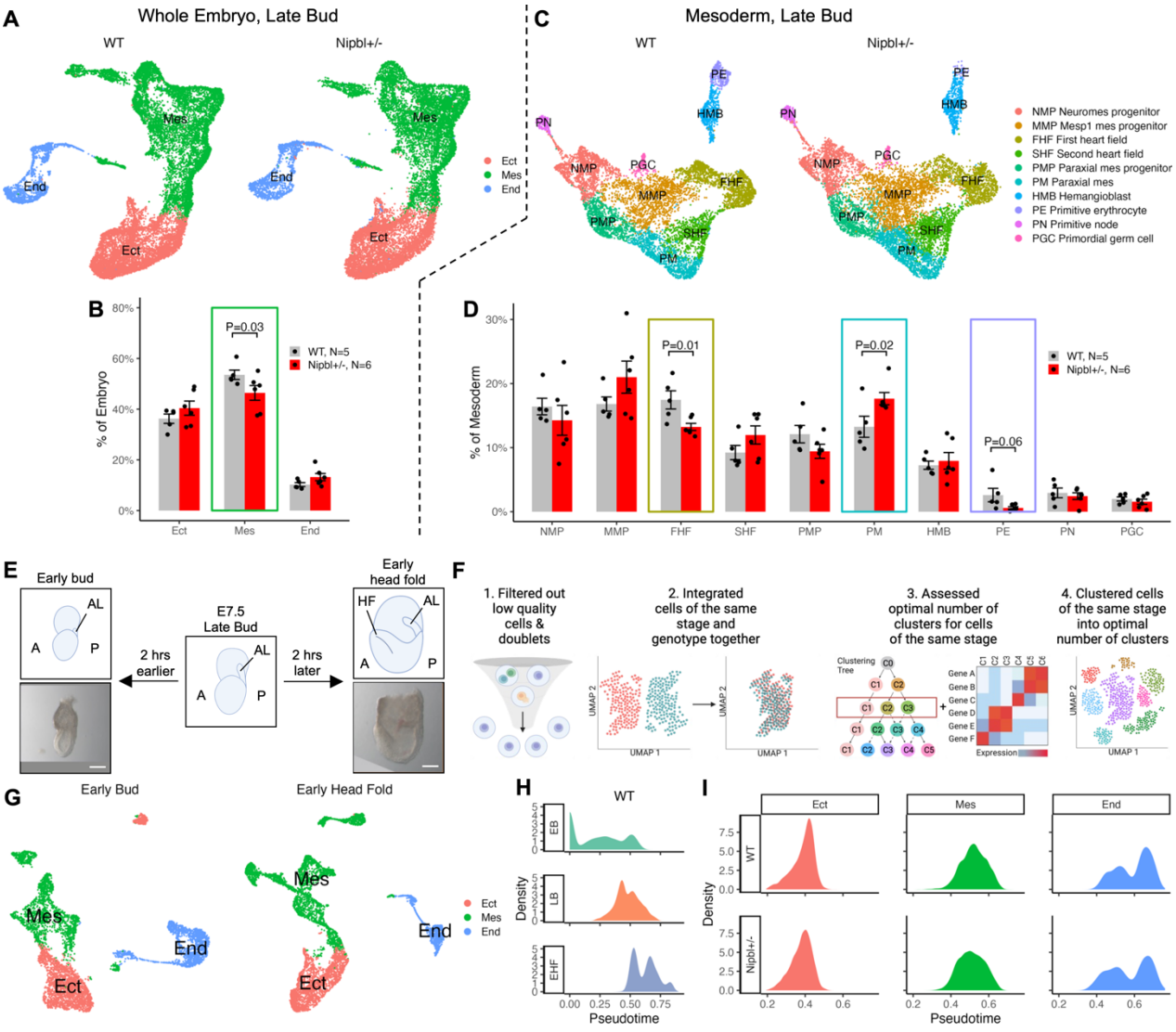


Fig. 3. *Nipbl*$^{+/-}$ embryos exhibit changes in the sizes of mesodermal subpopulations that are not accompanied by changes in developmental timing.

(**A**) UMAP of germ layers in LB-stage WT and *Nipbl*$^{+/-}$ embryos. (**B**) Percentage of cells in germ layers from all cells in LB-stage WT and *Nipbl*$^{+/-}$ embryos. Error bars show standard error of the mean. *P*-values from T-Test. (**C**) UMAP of cell populations in mesoderm of LB-stage WT and *Nipbl*$^{+/-}$ embryos. (**D**) Percentage of cells in mesodermal cell populations from all cells in LB-stage WT and *Nipbl*$^{+/-}$ embryos (bottom). Error bars show SEM. *P*-values from T-Test. (**E**) Lateral view of WT EB- and EHF-stage embryos subjected to scRNAseq. A = anterior, P =

posterior, AL = allantois, L = left, R = right, HF = head fold. Dashed line represents where embryonic tissue was separated from extraembryonic tissue. Scale bar = 100 um. (**F**) Workflow used to filter out low quality cells and doublets and cluster cells at EB- and EHF-stage into optimal number of clusters. (**G**) UMAP of clusters assigned to germ layers in EB- and EHF-stage embryos. (**H**) Density of cells from WT EB-, LB-, and EHF-stage embryos along pseudotime (calculated using URD). (**I**) Density of cells in germ layers of LB-stage WT and *Nipbl*$^{+/-}$ embryos along pseudotime.

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 (*248*) 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. 3.4A). In WT mesoderm, five distinct lineages were discernible (Fig. 3.4B), 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 (*249*), 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. 3.4A). 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. 3.3, B and D).

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

To explore whether variations in cell fate decisions might be responsible for the observed differences in the numbers of FHF and PM cells in *Nipbl*$^{+/-}$ embryos, we employed CellRank (*250*) to compute the likelihood of cells from the mesoderm (excluding HMBs, PEs, PGCs, and PN cells) transitioning terminally into cells of SHF, FHF, or PM fate. HMBs, PEs, PGCs, and PN were omitted from the calculation because our trajectories derived from RNA velocity indicated that these populations do not ultimately transition into SHF, FHF, or PM. In CellRank, the absorption probability defines the likelihood that a specific cell type will make a transition to a terminal state, derived using RNA velocity-directed, random walks from initial to terminal cell states (*250*). We represented these probabilities in UMAP space, with shades of blue to signify low probabilities and red for high probabilities. Fig. 3.4, C and D illustrates that the probabilities of NMPs, MMPs, and PMPs making a terminal transition solely into SHF do not vary significantly between *Nipbl*$^{+/-}$ and WT embryos, as quantified in the adjacent violin plots. However, the transition probabilities into FHF are 27-48% lower in *Nipbl*$^{+/-}$ embryos (Fig. 3.4, E and F), while the probabilities for transitions into PM are 9-11% higher (Fig. 3.4, G and H). Taken together, these results support the hypothesis that disparities in cell numbers in *Nipbl*$^{+/-}$ embryos stem from a misdirection of cells from FHF pathway towards a PM fate (Fig. 3.4I).

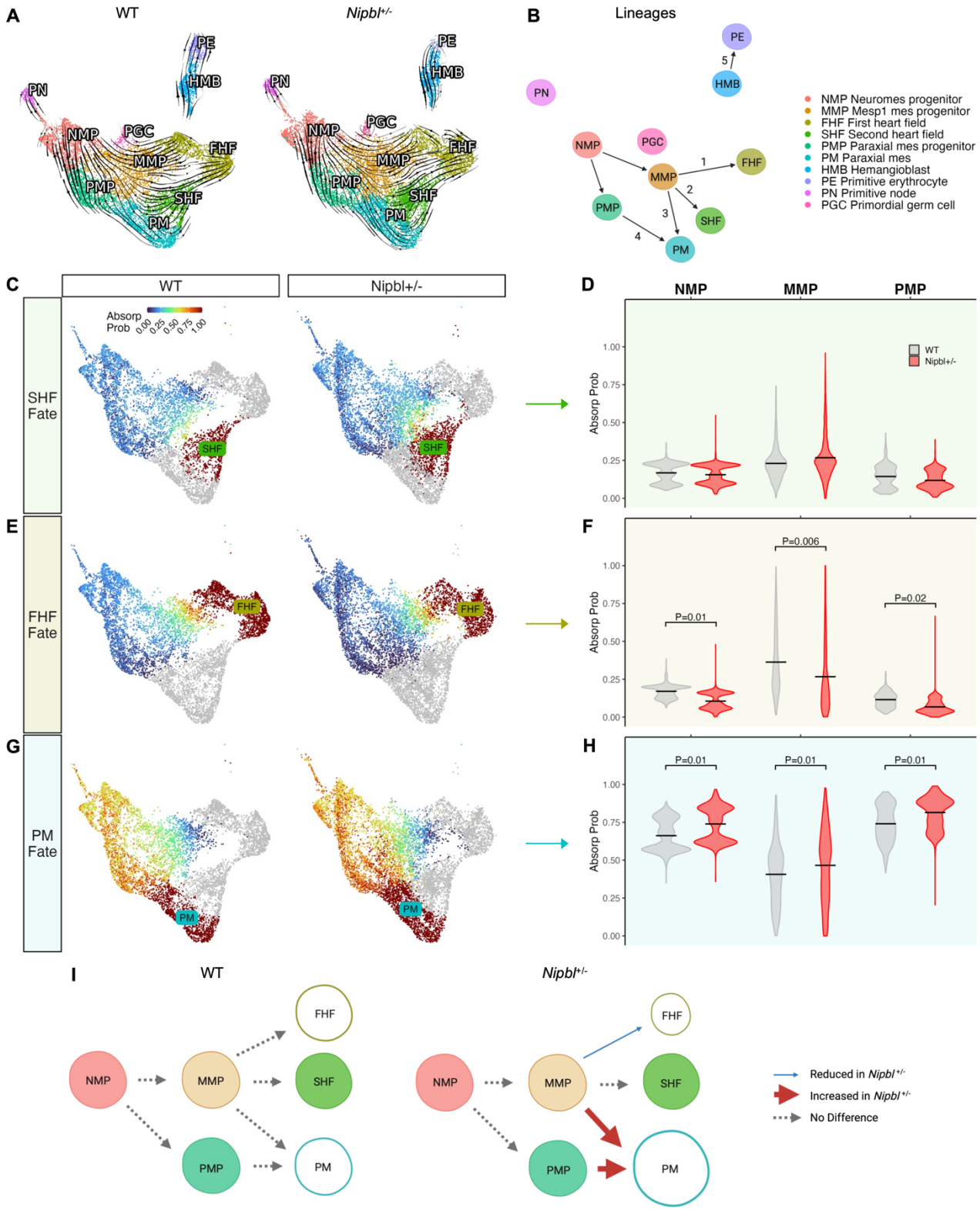

A
WT
Nipbl+/-
PE
HMB
PN
NMP
PGC
FHF
MMP
PMP
SHF
PM
B
Lineages
5
1
2
3
4
NMP Neuromes progenitor
MMP Mesp1 mes progenitor
FHF First heart field
SHF Second heart field
PMP Paraxial mes progenitor
PM Paraxial mes
HMB Hemangioblast
PE Primitive erythrocyte
PN Primitive node
PGC Primordial germ cell
C
WT
Nipbl+/-
Absorp Prob
0.00
0.25
0.50
0.75
1.00
SHF Fate
D
NMP
MMP
PMP
Absorp Prob
WT
Nipbl+/-
E
FHF Fate
F
P=0.01
P=0.006
P=0.02
G
PM Fate
H
P=0.01
P=0.01
P=0.01
I
WT
Nipbl+/-
Reduced in Nipbl+/-
Increased in Nipbl+/-
No Difference

Fig. 3.4. *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 *Nipbl*$^{+/-}$ embryos. Probability of mesoderm cells from LB-stage WT and *Nipbl*$^{+/-}$ embryos terminally transitioning (absorption probabilities calculated using CellRank) into (**C**) SHF, (**E**) FHF, or (**G**) PM fates in UMAP. Violin plot of probability of NMPs, MMPs, and PMPs from LB-stage WT and *Nipbl*$^{+/-}$ embryos terminally transitioning (absorption probabilities) into (**D**) SHF, (**F**) FHF, or (**H**) PM fates. Lines show means. *P*-values from T-Test. (**I**) Schematic illustrating how mesoderm cells in *Nipbl*$^{+/-}$ embryos are misdirected into PM fate at the expense of FHF fate.

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

Altered cell numbers in different mesodermal subpopulations of *Nipbl*$^{+/-}$ embryos could conceivably also stem from modifications in rates of cell death or proliferation. Focusing on apoptotic activity, we carried out differential gene expression analysis (DGEA) on WT and *Nipbl*$^{+/-}$ embryos at LB-stage, using genes associated with apoptosis from two well-known gene sets: Reactome Apoptosis and Hallmark Apoptosis (*251, 252*). We ranked each gene by minimum Q-value, arranging them from lowest to highest, and visualized their expression across all germ layers in *Nipbl*$^{+/-}$ embryos as heatmaps of the fold change relative to their WT counterparts (Fig. S3.11, A and B). For both gene sets, we saw no visible pattern of differences in the expression of genes with even the lowest Q-values between any of the germ layers in both WT and *Nipbl*$^{+/-}$ embryos. Although genes from these gene sets tend to mark cells with the capacity for apoptosis, their expression levels are not necessarily indicative of cells actively undergoing apoptosis. We therefore also turned to a gene set identified in a study that compared 180 apoptosis-associated genes in hematopoietic cells that were either healthy or undergoing apoptosis (*253*), and identified 93 apoptosis-associated genes that were differentially expressed. When we examined these genes in the germ layers of *Nipbl*$^{+/-}$ embryos, we also saw no visible difference with WT embryos (Fig. S3.11C). These data suggest that changes in apoptotic activity are unlikely to be a major contributor to differences in cell type proportions.

To examine cell proliferation, we again employed DGEA across all germ layers of both WT and *Nipbl*$^{+/-}$ embryos at LB-stage, focusing on the meta-PCNA gene set, an ensemble of genes that exhibit the strongest positive correlation with PCNA expression, a recognized biomarker of cell proliferation (*254*). When we visualized the expression of these genes in the same manner as the gene sets above, we found no discernible variations in the expression levels

of the meta-PCNA genes between the germ layers of WT and *Nipbl*$^{+/-}$ embryos (Fig. S3.12A). As another measure of cell proliferation, we used Seurat to assign cells from LB-stage WT and *Nipbl*$^{+/-}$ embryos into phases (G1, S, and G2/M) of the cell cycle (*255*). Seurat does this by calculating a cell cycle phase score based on the expression of canonical S phase and G2/M phase markers. Seurat considers these marker sets to be anticorrelated in their expression, so when cells express neither, they are considered to be in G1 phase. As expected, in all germ layers, cells from LB-stage WT and *Nipbl*$^{+/-}$ embryos assigned into all phases of the cell cycle (Fig. S3.12B). In all germ layers, there was no statistically significant difference in the proportions of cells in each cell cycle between WT and *Nipbl*$^{+/-}$ embryos (Fig. S3.12C). We therefore concluded that extensive alterations in cell proliferation are not likely responsible for disparities in cell subpopulations.

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

Next, we sought to investigate whether the misdirection of mesoderm cells in *Nipbl*$^{+/-}$ embryos at the LB-stage could be attributed to the misexpression of genes that drive the transition of mesoderm cells into either FHF or PM. To answer this question, we also employed CellRank to identify potential driver genes of the FHF and PM fates, using those mesoderm cells that specifically contribute to these developmental pathways (Fig. 3.5, A and D). CellRank achieves this by calculating a correlation coefficient between the likelihood (absorption probabilities) of cells progressing towards a particular lineage fate and the expression level of individual genes (*250*). In this context, genes with positive correlation coefficients are considered drivers, as their expression elevates alongside an increase in absorption probabilities. Conversely, genes with negative correlation coefficients are deemed anti-drivers, as their

expression diminishes with increasing absorption probabilities. To minimize the risk of false discoveries, we set a threshold, accepting genes with correlation coefficients greater than 0.25 as drivers and those less than -0.25 as anti-drivers.

CellRank identified 383 genes as drivers of FHF fate (Fig. 3.5B). 11% of them were transcription factors. Among these were *Hand1* and *Tbx20*, transcription factors previously recognized for their roles in promoting FHF development (*198, 256*). Several transcription factors that have not been previously linked to FHF development were also predicted to be FHF drivers, including *Msx1, Msx2, Foxf1, Hoxb6,* and *Pitx1*. Of these, *Hoxb6* and *Pitx1* are known to play important roles in segmentation and patterning. Gene set over-representation analysis using MSigDB's Hallmark gene sets (*251*), which characterize well-defined biological states and processes, revealed that FHF drivers were significantly enriched in genes associated with TGF-beta signaling (Fig. 3.5C), which has been previously described as important in cardiomyocyte proliferation (*257*). Genes associated with epithelial to mesenchymal transition (EMT) were also significantly enriched (see below). Conversely, CellRank predicted 225 genes to be anti-drivers of FHF fate, with 16% being transcription factors. Among these were genes known to be involved in pluripotency (*Pou5f1*), mesoderm specification/expansion (*T, Cdx1, Sp5,* and *Zic3*), anterior-posterior patterning (*Hoxb1* and *Hoxa1*), and paraxial mesoderm development (*Hes7, Meis2, Tbx6, Gbx2,* and *Foxb1*).

CellRank predicted 66 genes to be drivers and 115 genes to be anti-drivers of PM fate (Fig. 3.5E). A substantial proportion (22%) of PM drivers were transcription factors. These included *Tcf15* and *Foxc2*, transcription factors that have been previously recognized for their role in regulating PM development (Fig. 3.5E) (*232, 258*). The *Irx* transcription factors *Irx5, Irx1,* and *Irx3* – known for their roles in segmentation during development (*259*) – were also

identified as PM drivers. PM drivers did not exhibit significant over-representation of any Hallmark gene sets (Fig. 3.5F). Of all PM anti-drivers, 20% were identified as transcription factors, including *Hand1, Cdx2, Msx2, Msx1,* and *Cdx4*. Intriguingly, *Hand1, Msx2, and Msx1* had also been predicted as drivers for FHF. This overlap prompted us to conduct a comparative analysis between PM anti-drivers and FHF drivers. Our examination revealed that 40% of all PM anti-drivers were concurrently FHF drivers (Fig. 3.5, G and H).

The confluence of expressed genes between FHF drivers and PM anti-drivers seems to indicate that during normal development, increased expression of these shared genes by mesoderm cells predisposes them to transition into FHF, while decreased expression of these same genes steers them towards PM (Fig. 3.5I). This set of shared genes may thus be considered to be a "lineage branching signature".

How does *Nipbl*-haploinsufficiency impact the expression of driver and anti-driver genes? Examination of *Nipbl*$^{+/-}$ embryos revealed significant misexpression in NMP, MMP, and PMP cells of numerous driver and anti-driver genes associated with both FHF (first heart field) and PM (paraxial mesoderm) (Fig. 3.5J, Fig. S3.13). Interestingly, most affected driver and anti-driver genes were downregulated. To ascertain whether this was more than coincidental, we employed a Chi-square analysis, which revealed a striking pattern: *Nipbl*$^{+/-}$ embryos underexpressed FHF drivers at a higher frequency than they overexpressed them (166 out of 174 differentially expressed genes) (Fig. 3.5J). Similarly, they commonly underexpressed PM anti-drivers (40 out of 51) (Fig. S3.13), aligning with expectations since many PM anti-drivers serve dual roles as FHF drivers.

Strikingly, genes associated with EMT were highly enriched among FHF drivers that were underexpressed in *Nipbl*$^{+/-}$ embryos (Fig. 3.5K). A total of 15 EMT genes was found to be

underexpressed, including the EMT transcription factors *Msx1* and *Snai2*, and signaling genes such as *Igfbp4*, *Pdgfrb*, and *Tgfb1* (Fig. 3.5J). These results suggest that, in the context of *Nipbl*-haploinsufficiency, skewed differentiation of mesoderm cells towards the PM lineage at the expense of FHF lineage may be attributable to the underexpression of genes driving FHF fate. These results are consistent with the idea that genes associated with EMT play a role in this lineage misdirection.

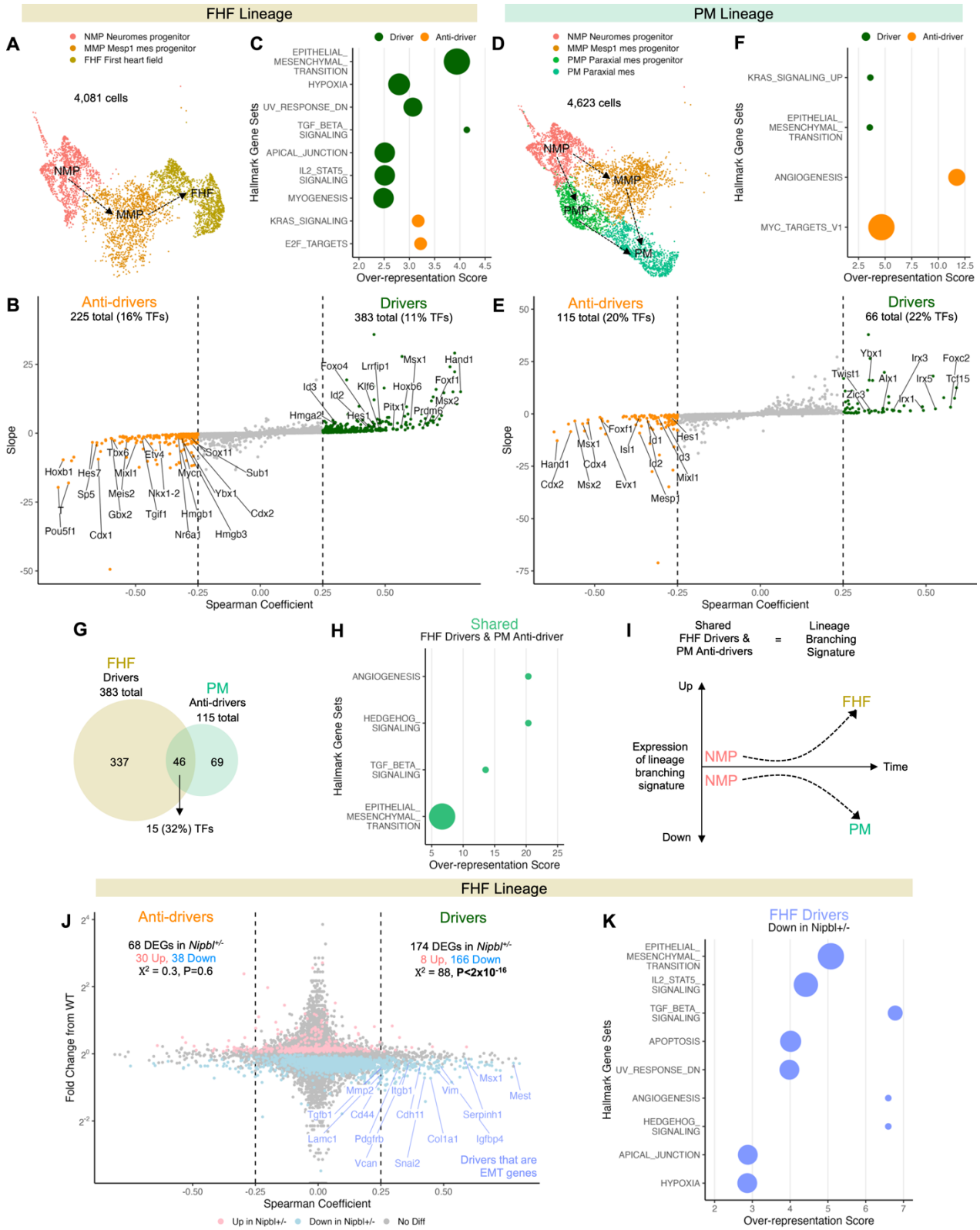

FHF Lineage
PM Lineage
A
NMP Neuromes progenitor
MMP Mesp1 mes progenitor
FHF First heart field
4,081 cells
NMP
MMP
FHF
C
Driver
Anti-driver
Hallmark Gene Sets
EPITHELIAL_MESENCHYMAL_TRANSITION
HYPOXIA
UV_RESPONSE_DN
TGF_BETA_SIGNALING
APICAL_JUNCTION
IL2_STAT5_SIGNALING
MYOGENESIS
KRAS_SIGNALING
E2F_TARGETS
Over-representation Score
D
NMP Neuromes progenitor
MMP Mesp1 mes progenitor
PMP Paraxial mes progenitor
PM Paraxial mes
4,623 cells
NMP
MMP
PMP
PM
F
Driver
Anti-driver
KRAS_SIGNALING_UP
EPITHELIAL_MESENCHYMAL_TRANSITION
ANGIOGENESIS
MYC_TARGETS_V1
Over-representation Score
B
Anti-drivers
225 total (16% TFs)
Drivers
383 total (11% TFs)
Slope
Spearman Coefficient
Foxo4
Lrrfip1
Msx1
Hand1
Id3
Klf6
Hoxb6
Foxf1
Id2
Msx2
Hmga2
Hes1
Pitx1
Prdm6
Tbx6
Etv4
Sox11
Hoxb1
Hes7
Mixl1
Myc
Sub1
Sp5
Meis2
Nkx1-2
Ybx1
Gbx2
Tgif1
Hmgb1
Cdx2
Pou5f1
Cdx1
Nr6a1
Hmgb3
E
Anti-drivers
115 total (20% TFs)
Drivers
66 total (22% TFs)
Slope
Spearman Coefficient
Ybx1
Irx3
Foxc2
Twist1
Alx1
Irx5
Tcf15
Zic3
Irx1
Foxf1
Id1
Hes1
Msx1
Isl1
Id3
Hand1
Cdx4
Id2
Mixl1
Cdx2
Msx2
Evx1
Mesp1
G
FHF
Drivers
383 total
PM
Anti-drivers
115 total
337
46
69
15 (32%) TFs
H
Shared
FHF Drivers & PM Anti-driver
Hallmark Gene Sets
ANGIOGENESIS
HEDGEHOG_SIGNALING
TGF_BETA_SIGNALING
EPITHELIAL_MESENCHYMAL_TRANSITION
Over-representation Score
I
Shared FHF Drivers & PM Anti-drivers
=
Lineage Branching Signature
Up
Down
Expression of lineage branching signature
NMP
NMP
FHF
PM
Time
FHF Lineage
J
Anti-drivers
68 DEGs in Nipbl+/-
30 Up, 38 Down
X2 = 0.3, P=0.6
Drivers
174 DEGs in Nipbl+/-
8 Up, 166 Down
X2 = 88, P<2x10-16
Fold Change from WT
Spearman Coefficient
Msx1
Mmp2
Itgb1
Vim
Mest
Tgfb1
Cd44
Cdh11
Serpinh1
Lamc1
Pdgfrb
Col1a1
Igfbp4
Vcan
Snai2
Drivers that are EMT genes
Up in Nipbl+/-
Down in Nipbl+/-
No Diff
K
FHF Drivers
Down in Nipbl+/-
Hallmark Gene Sets
EPITHELIAL_MESENCHYMAL_TRANSITION
IL2_STAT5_SIGNALING
TGF_BETA_SIGNALING
APOPTOSIS
UV_RESPONSE_DN
ANGIOGENESIS
HEDGEHOG_SIGNALING
APICAL_JUNCTION
HYPOXIA
Over-representation Score

Fig. 3.5. *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 correlated (anti-drivers) with the transition (absorption probabilities from Fig. 3.3J) of mesoderm cells from WT LB-stage embryos into (**B**) FHF or (**E**) PM fates. Correlation coefficients calculated using Spearman's Rank Correlation. Genes with correlation coefficient greater than 0.25 or less than -0.25 were considered drivers and anti-drivers, respectively. Slope refers to change in gene expression along absorption probability. Over-representation score of MSigDB's Hallmark gene sets among drivers and anti-drivers of (**C**) FHF and (**F**) PM fate. (**G**) Venn diagram of genes shared between PM anti-drivers and FHF drivers. (**H**) Over-representation score of MSigDB's Hallmark gene sets among shared FHF drivers and PM anti-drivers. (**I**) Expression of lineage branching signature as NMPs transition into FHF or PM fates. (**J**) Fold change in expression of genes in mesoderm cells of FHF lineage of LB-stage *Nipbl*$^{+/-}$ embryos differentially expressed ($Q < 0.05$, Mann-Whitney U Test) from that of WT embryos along Spearman's Rank Correlation coefficient from Fig. 3.5B. Genes associated with EMT are labeled. (**K**) Over-representation score of MSigDB's Hallmark gene sets among shared FHF drivers downregulated in FHF lineage of *Nipbl*$^{+/-}$ mice.

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

Using DGEA to compare the germ layers of LB-stage WT and $Nipbl^{+/-}$ embryos, we found that $Nipbl^{+/-}$ embryos misexpressed hundreds of genes across all germ layers. Underexpression was more common than overexpression in all cases (Fig. 3.6A), supporting a broad role for *Nipbl* in enhancing gene expression. The majority of gene expression changes were subtle, i.e., less than 2-fold (Fig. 3.6A). This pattern of gene expression changes agrees with earlier studies (*10*). Gene set enrichment analysis (GSEA) for MSigDB's Hallmark gene sets found that $Nipbl^{+/-}$ embryos showed enrichment, and no de-enrichment, for 4 out of 40 gene sets: oxidative phosphorylation in the mesoderm; Mtorc1 signaling in the ectoderm; and G2M checkpoint, Myc targets, and E2F targets in the endoderm (Fig. 3.6B). In these cases, however, enrichment was driven by a relatively small number of the genes in these sets; the vast majority were either not differentially expressed in $Nipbl^{+/-}$ embryos or showed changes much smaller than 1.5-fold (Fig. 3.6C).

Genes exhibiting substantial changes in expression – greater than 2-fold upregulated or downregulated – were nevertheless observed across every germ layer of $Nipbl^{+/-}$ embryos (Fig. 3.6A). Individual curation of these genes unveiled associations with developmental processes (Fig. 3.6D). Within the mesoderm, two- to six-fold upregulation of genes involved in Nodal signaling, including *Tdgf1, Eomes, Lefty2, Lhx1*, was observed (*260-263*) . At the same time, three clusters of genes were robustly downregulated: those associated with primitive erythropoiesis, including *Gata1* and *Klf1* (2-3 fold) (*237, 264*); *Hox* genes such as *Hoxb8, Hoxb5,* and *Hoxa5* (2-fold); and genes associated with Notch signaling, including *Hey2* and *Nr2f2* (2-fold) (*246, 265*) (Fig. 3.6D).

In the ectoderm, five transcription factor genes tied to the hedgehog signaling pathway were downregulated more than 2-fold, including *Pax2, Sox9, Nkx6-1, Sox1,* and *Irx2* (*266-270*) (Fig. 3.6D). In the endoderm, there was a distinct 2-3 fold upregulation of several adult hemoglobin genes such as *Hbb-bt, Hba-a2, Hba-a1,* and *Hbb-bs* (Fig. 3.6D). Intriguingly, consistent patterns were evident in both the ectoderm and endoderm, mirroring the mesoderm's trends. Genes related to Nodal signaling were upregulated, and *Hox* genes were noticeably downregulated, across all germ layers. The magnitude of changes in Nodal signaling genes and *Hox* genes, being greater than 2-fold in both directions, and their critical roles in various germ layers, strongly suggest that disruptions in these pathways could be major contributors to phenotypes observed in *Nipbl*$^{+/-}$ mice.

When we compared the sets of differentially expressed genes (DEGs) obtained from forward projection (Fig. 3.6) with those obtained from reverse projection (Fig. S3.14), we found greater than 94% of DEGs in mesoderm and ectoderm were identical (Fig. S3.14, A and B). For the endoderm, reverse projection showed more than 75% of the same DEGs (Fig. S3.14C). As was the case for changes in the relative sizes of cell populations, we conclude that DEGs observed in *Nipbl*$^{+/-}$ embryos represent true biological differences and are not a consequence of technical issues in cell classification.

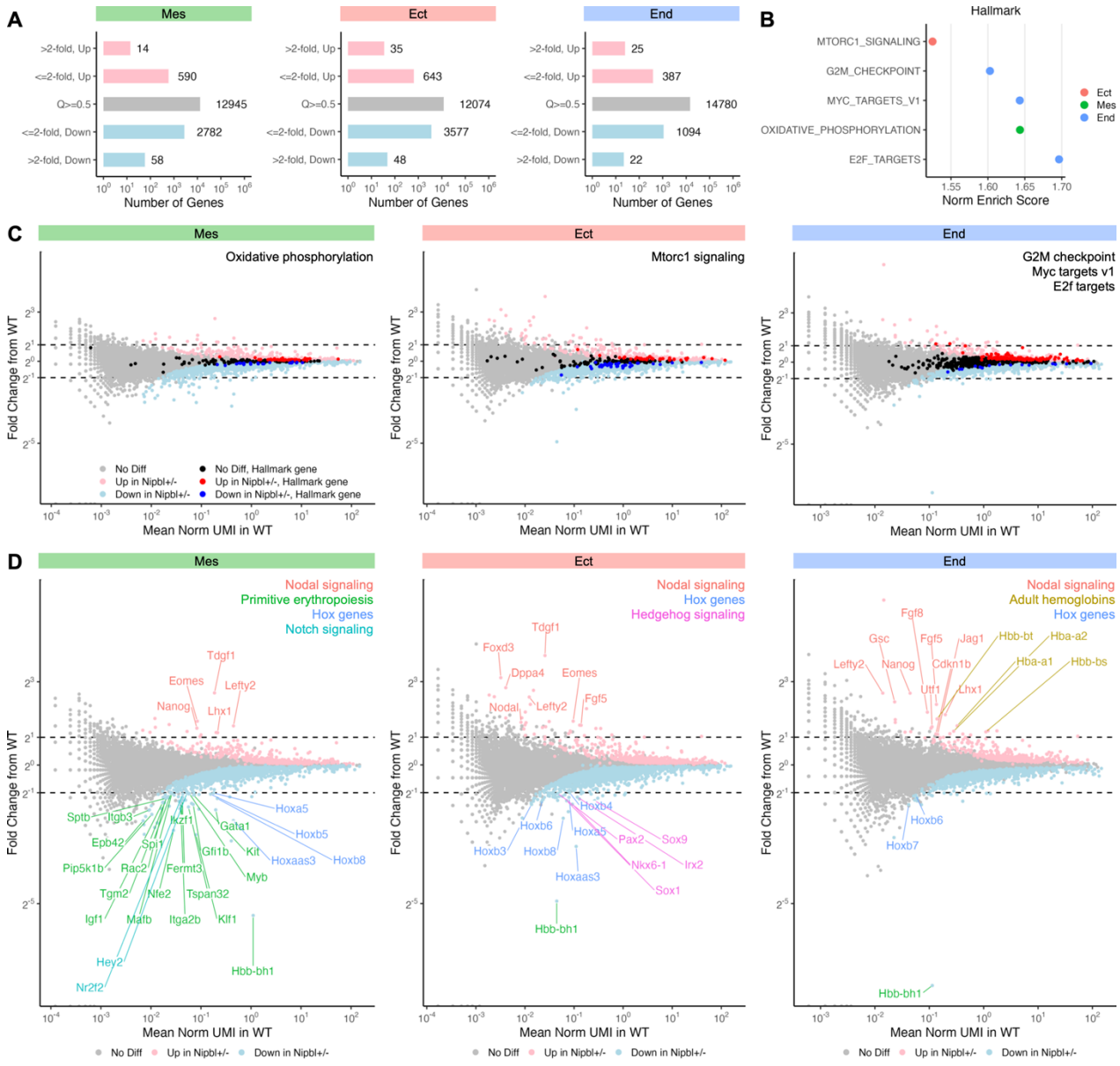


Fig. 3.6. *Nipbl*$^{+/-}$ mice show large changes in the expression of major developmental regulators in all germ layers.

(**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. 3.5B 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 ($Q < 0.05$, Mann-Whitney U Test) in germ layers of LB-stage *Nipbl*$^{+/-}$ embryos showing large changes in expression (> 2-fold) from that of WT embryos along their average expression in WT embryos.

Nipbl$^{+/-}$ mice overexpress *Nanog* during and after gastrulation

To uncover developmental processes that might be the most broadly impacted in *Nipbl*$^{+/-}$ embryos, we used the large, significant, changes in gene expression (greater than 2-fold up or down) as input data for STRING (*271*), a database and algorithm that constructs a network of potential gene interactions, in this case limiting predictions to those informed by experiments demonstrating co-expression or protein-protein interactions. The outcome of this analysis (Figs. 7, A and B) revealed that many genes with substantial expression changes in *Nipbl*$^{+/-}$ embryos are predicted to interact either directly or indirectly with *Nanog*. Indeed, *Nanog* itself was overexpressed across all three germ layers in *Nipbl*$^{+/-}$ embryos, with a more pronounced overexpression, exceeding 2-fold, in the mesoderm and endoderm (Fig. 3.7B).

Nanog, which encodes a transcription factor, is notably expressed at two junctures during mouse embryonic development. Initially, it appears at the blastocyst stage, where it plays an essential role in sustaining the pluripotency of the inner cell mass cells (*272*). Subsequently, during gastrulation, *Nanog* expression temporarily surges, only to be silenced as cells transition out of the primitive streak (*273*). This expression pattern could be observed directly in our data: In Fig. 3.7C, we ordered mesodermal cell populations from LB-stage embryos by their stage in gastrulation, as inferred from their RNA velocity positions (Fig. 3.4A). In WT NMPs, *Nanog* is expressed at a relatively high level. As cells move through successive stages of gastrulation within WT embryos, *Nanog's* expression exhibits a consistent decline (Fig. 3.7C), reaching near zero in cells of the first and second heart field (FHF and SHF). An exception to this trend occurs in primordial germ cells, which are known to maintain elevated *Nanog* expression (*274*). In contrast, in *Nipbl*$^{+/-}$ embryos (Fig. 3.7C), *Nanog* expression remained elevated across all cell populations. Although expression still falls after the NMP stage in *Nipbl*$^{+/-}$ cells, *Nanog* never

falls to WT levels. This is evident even in primordial germ cells. Moreover, the overexpression of *Nanog* in *Nipbl*$^{+/-}$ embryos is sustained up to the CC-stage (Fig. 3.7D and Table S3.3.10), where it reaches a 10-fold increase over WT levels. These observations indicate that *Nipbl*-haploinsufficiency leads to a marked failure of *Nanog* downregulation.

Another transcription factor that plays a role in maintaining pluripotency is *Pou5f1* (Oct4). From studies in embryonic stem cells, it is known that *Pou5f1* promotes the expression of *Nanog*, and conversely, *Nanog* promotes the expression of *Pou5f1* (*275*). Given this, we were curious about whether *Nipbl*$^{+/-}$ embryos might also overexpress of *Pou5f1*. Despite the fact that *Pou5f1* is more highly expressed than *Nanog*, we did not observe any statistically significant difference in *Pou5f1* expression across mesodermal cell types or throughout LB-stage *Nipbl*$^{+/-}$ embryos (Fig. 3.7E, Fig. 3.7F, and Table S3.3.11). However, a more than 2-fold elevation in *Pou5f1* expression was detected in *Nipbl*$^{+/-}$ embryos at CC-stage (Fig. 3.7F).

We sought to determine whether elevated *Nanog* expression in *Nipbl*$^{+/-}$ embryos during gastrulation might reflect some sort of non-specific overactivity of the *Nanog* gene. To investigate this hypothesis, we looked at *Nanog* expression in WT and *Nipbl*$^{+/-}$ embryonic stem cells (ESCs), as ESCs are known to express *Nanog* (as do to the blastocyst inner cell mass cells from which ESCs are derived). We generated WT and *Nipbl*$^{+/-}$ ESCs by treating *Nipbl*$^{Flex/+}$ (*Nipbl*$^{+/-}$) ESCs with Flp recombinase (which inverts a gene trap in *Nipbl*-Flex ESCs in such a way that reverses gene trapping) to produce *Nipbl*$^{Flrt/+}$ (WT) ESCs (Fig. 3.7G) (*18*). We assessed *Nanog* expression in nine independent clones from each genotype, using quantitative reverse transcription polymerase chain reaction (RT-qPCR). We observed no discernible difference in *Nanog* expression between *Nipbl*$^{+/-}$ and WT ESCs, a pattern that was also observed for *Pou5f1* (Fig. 3.7H). As a control, we showed that *Nipbl*$^{+/-}$ ESCs exhibit the expected reduced expression

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

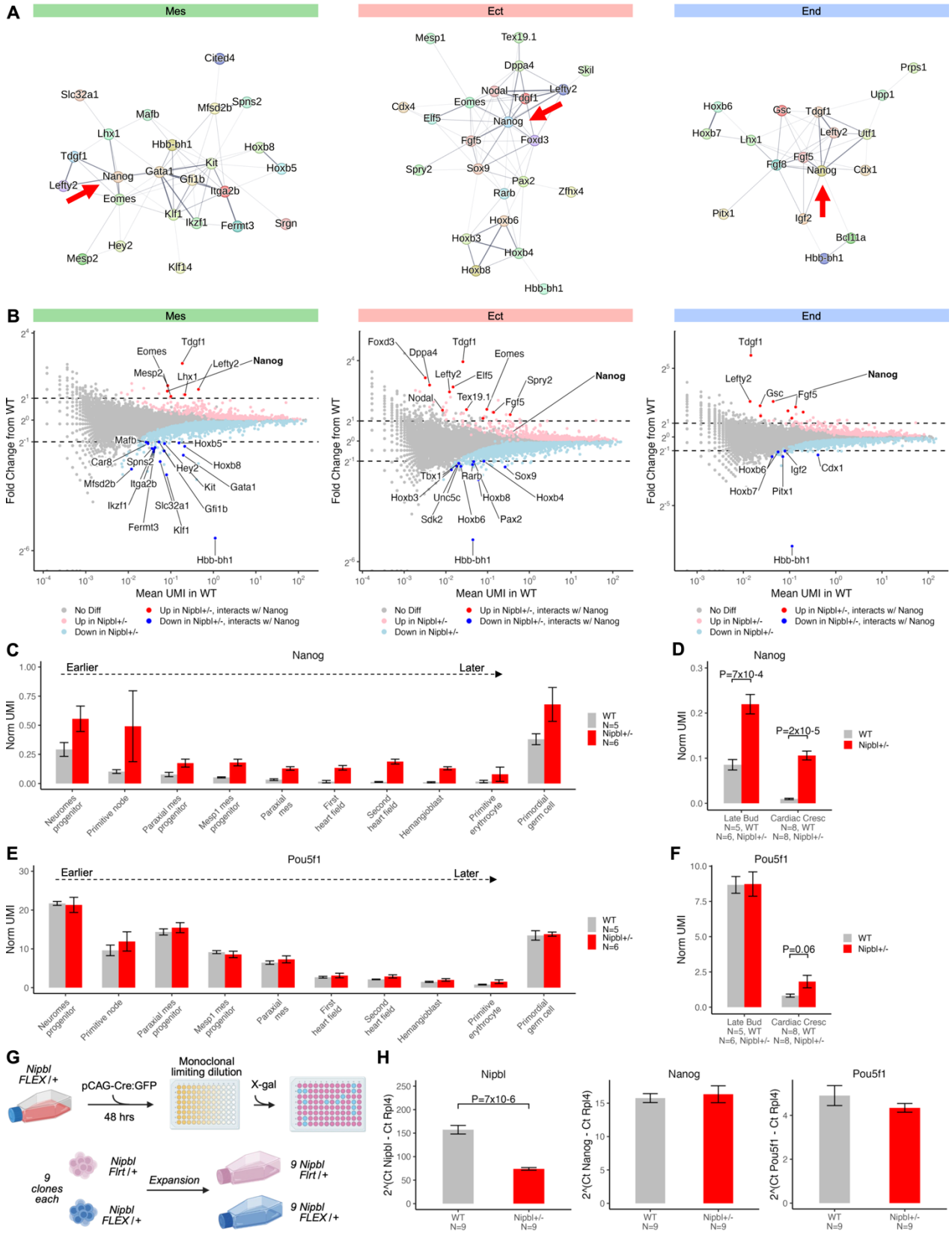

A
Mes
Ect
End
B
Mes
Ect
End
Fold Change from WT
Mean UMI in WT
No Diff
Up in Nipbl+/-
Down in Nipbl+/-
Up in Nipbl+/-, interacts w/ Nanog
Down in Nipbl+/-, interacts w/ Nanog
C
Nanog
Earlier
Later
Norm UMI
WT N=5
Nipbl+/- N=6
Neuromes progenitor
Primitive node
Paraxial mes progenitor
Mesp1 mes progenitor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

Fig. 3.7. *Nipbl*$^{+/-}$ mice overexpress *Nanog* during and after gastrulation.

(**A**) Differentially expressed genes (greater than 2-fold upregulated or downregulated) between the germ layers of WT and *Nipbl*$^{+/-}$ embryos predicted by STRING to interact with each other. (**B**) Fold change in expression of differentially expressed genes ($Q < 0.05$, Mann-Whitney U Test) in germ layers of LB-stage *Nipbl*$^{+/-}$ embryos showing large changes in expression (> 2-fold) from that of WT embryos along their average expression in WT embryos that are predicted by STRING to interact with *Nanog*. (**C**) Expression of *Nanog* in mesodermal cell populations of LB-stage WT and *Nipbl*$^{+/-}$ embryos ordered from left (earlier) to right (later) by their RNA velocity positions in Fig. 3.4A. Error bars show SEM. (**D**) Expression of *Nanog* in WT and *Nipbl*$^{+/-}$ embryos from LB- to CC-stage. Error bars show SEM. *P*-values from T-Test. € Expression of *Pou5f1* in mesodermal cell populations of LB-stage WT and *Nipbl*$^{+/-}$ embryos ordered from left (earlier) to right (later) by their RNA velocity positions in Fig. 3.4A. Error bars show SEM. (**F**) Expression of *Pou5f1* in WT and *Nipbl*$^{+/-}$ embryos from LB- to CC-stage. Error bars show SEM. *P*-values from T-Test. (**G**) Monoclonal generation of *Nipbl*$^{Flrt/+}$ (WT) and *Nipbl*$^{Flex/+}$ (*Nipbl*$^{+/-}$) ESCs. (**H**) Expression of *Nipbl*, *Nanog*, and *Pou5f1* in WT and *Nipbl*$^{+/-}$ embryonic stem cells as measured by RT-qPCR and normalized to the housekeeping gene, *Rpl4*. Error bars show SEM. *P*-values from T-Test.

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

As a critical regulator of pluripotency in early embryonic development, *Nanog* directly influences the expression of a multitude of other genes. This relationship prompted us to ask how many of the gene expression differences in *Nipbl*$^{+/-}$ embryos might be attributable to the overexpression of *Nanog*. To address this question, we took advantage of data from a recent study, by Tiana et al., 2022 (*276*), in which mice were engineered to express *Nanog* under doxycycline-inducible (Dox) control. In that study, bulk RNA sequencing was performed on untreated (*Nanog* Dox-) and Dox-treated (*Nanog* Dox+) embryos at stages E7.5 and E9.5, with E7.5 corresponding closely to the LB-stage embryos analyzed here. The E7.5 *Nanog* Dox+ embryos that were sequenced had been treated with doxycycline from E4.5 to E7.5, while E9.5 *Nanog* Dox+ embryos were treated with doxycycline from E6.5 to E9.5. Differential gene expression analysis, performed in that study, found hundreds of gene expression changes between *Nanog* Dox- and *Nanog* Dox+ embryos at both stages (*276*). We subsequently analyzed the results of their analysis in four ways.

First, we plotted the fold change in expression of genes in *Nipbl*$^{+/-}$ embryos, as a whole, from that of WT embryos, concentrating on LB-stage (Fig. 3.8A). (CC-stage comparisons are shown in Fig. S3.15A, and Fig. S3.15B.) In Fig. 3.8A, genes significantly overexpressed in *Nipbl*$^{+/-}$ embryos are colored pink and those significantly underexpressed, light blue. DEGs in *Nipbl*$^{+/-}$ embryos that are also significantly overexpressed or underexpressed in E7.5 *Nanog* Dox+ embryos, are colored red and blue, respectively. The results are summarized in Fig. 3.8B. At LB-stage, 68.8% of overexpressed and 61.5% of underexpressed genes in *Nipbl*$^{+/-}$ embryos were also overexpressed or underexpressed, respectively, in E7.5 *Nanog* Dox+ embryos. A

binomial test confirmed that *Nanog* Dox+ DEGs are highly over-represented in *Nipbl*$^{+/-}$ embryos (Fig. 3.8B).

Second, we again plotted the fold change in expression of genes in LB-stage *Nipbl*$^{+/-}$ embryos from that of WT embryos, but this time separately by germ layer, rather than across whole embryos (Fig. 3.8C). Genes colored red and blue are DEGs in *Nipbl*$^{+/-}$ embryos that are also overexpressed or underexpressed in *Nanog* Dox+ embryos. When we manually curated genes that were expressed with more than 2-fold change (Fig. 3.8C), we found that all germ layers showed upregulation of genes involved with Nodal signaling (*Tdgf1, Lefty2, Eomes, Lhx1,* and *Ffg5*) (*260-263, 277*) and downregulation of *Hox* genes (*Hoxb8 and Hoxb6*). Additionally, the mesoderm of *Nipbl*$^{+/-}$ embryos underexpressed genes associated with primitive erythropoiesis (*Hbb-bh1*, *Klf1*, and *Gata1*) (*237, 264*) and hedgehog signaling (*Pax2* and *Sox9*) (*266, 267*). (CC-stage gene expression changes are discussed in Fig. S3.15C.) These findings are consistent with the idea that many of the largest changes in gene expression that occur in *Nipbl*$^{+/-}$ embryos, as well as the developmental pathways they regulate, could be attributable to overexpression of *Nanog*.

Third, we analyzed how closely *Nipbl*$^{+/-}$ embryos mirrored the gene expression alterations of *Nanog* Dox+ embryos in terms of magnitude and direction. To do this, we generated a plot illustrating the fold-changes in expression of individual genes in *Nipbl*$^{+/-}$ embryos compared to their fold change in *Nanog* Dox+ embryos (Fig. 3.8D). We colored genes that were significantly overexpressed in *Nanog* Dox+ embryos pink and those significantly underexpressed, light blue. Additionally, we calculated the regression line depicting the relationship between fold changes in gene expression in *Nipbl*$^{+/-}$ *vs. Nanog* Dox+ embryos. At LB-stage, the slope of the regression line between *Nipbl*$^{+/-}$ and E7.5 *Nanog* Dox+ embryos was

0.8, implying that, on average, the fold changes in gene expression in *Nipbl*$^{+/-}$ embryos are quantitatively similar to those in *Nanog* Dox+ embryos (Fig. 3.8D). Fig. 3.8D also depicts that, at LB-stage, *Nipbl*$^{+/-}$ embryos upregulate 46% of the same genes that were overexpressed in E7.5 *Nanog* Dox+ embryos and downregulate 93% of the same genes that were underexpressed in E7.5 *Nanog* Dox+ embryos. (Similar analysis was performed at CC-stage, and is described in Fig. S3.15D.) These findings highlight a remarkable similarity in gene expression changes between *Nipbl*$^{+/-}$ and *Nanog* Dox+ embryos.

Fourth, we examined the magnitudes of fold changes in individual gene expression in *Nipbl*$^{+/-}$ embryos and *Nanog* Dox+ embryos, focusing on the 50 most highly up- and down-regulated genes in *Nanog* Dox+ embryos. As shown in Fig. 3.8E, we categorized genes into two groups at each stage: those overexpressed in *Nanog* Dox+ embryos (ranked from top to middle) and those underexpressed in *Nanog* Dox+ embryos (ranked from bottom to middle), ordering them by ascending Q-value. Next, we constructed a heatmap that visually captures the fold changes in gene expression within *Nanog* Dox+ whole embryos, contrasting them with the corresponding fold changes in the individual germ layers of *Nipbl*$^{+/-}$ embryos (Fig. 3.8E). E7.5 *Nanog* Dox+ embryos overexpressed genes linked with the Nodal signaling pathway (*Tdgf1, Eomes, Fgf5, Lhx1, Foxd3, Lefty2,* and *Nodal*) (*260-263, 277-279*) and those integral to maintaining pluripotency (*Nanog* and *Pou5f1*) (Fig. 3.8E) (*275*). Concurrently, they underexpressed genes associated with primitive erythropoiesis (*Hbb-bh1, Tal1,* and *Lmo2*) (*280, 281*) and specific *Hox* genes (*Hoxa1, Hoxb3os,* and *Hoxb1*). Strikingly, this pattern of fold changes was mirrored, to a remarkable degree, in the germ layers of LB-stage *Nipbl*$^{+/-}$ embryos (Fig. 3.8E). (The corresponding analysis for CC-stage embryos is given in Fig. S3.15E.)

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

Given the large overlap between the gene expression changes observed in LB-stage *Nipbl*$^{+/-}$ embryos and E7.5 *Nanog* Dox+ embryos, we wondered how many differentially expressed genes in *Nipbl*$^{+/-}$ embryos might be direct targets of Nanog. Although genome-wide patterns of Nanog binding have not been characterized for embryos at this stage, many groups have used chromatin immunoprecipitation sequencing (ChIP-seq) to define Nanog binding sites in mESCs. We turned to a recent study (*282*), in which Avsec et al. captured 52,456 Nanog binding sites in the genome of mESCs, and reported that 3,645 genes showed one or more Nanog binding sites located within 250 nt on either side of their transcription start sites (Fig. 3.8F). On comparing these genes with those up- and down-regulated in LB-stage *Nipbl*$^{+/-}$ embryos, we found that 25% of the overexpressed genes and 22% of the underexpressed genes match those genes with Nanog binding sites in mESCs (Fig. 3.8, G and H). This represented significant enrichment for Nanog binding sites in differentially vs. non-differentially expressed genes ($P<1x10^{-12}$ and $P<2x10^{-16}$, Fisher's Exact Tests). These data suggest that a substantial portion of differentially expressed genes in LB-stage *Nipbl*$^{+/-}$ embryos could indeed be direct targets of Nanog.

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

As shown above in Fig. 3.3K and Fig. S3.13, FHF drivers and PM antidrivers are downregulated in LB-stage *Nipbl*$^{+/-}$ mice. Given the large overlap between downregulated genes in LB-stage *Nipbl*$^{+/-}$ embryos and downregulated genes in E7.5 *Nanog* Dox+ embryos (Fig. 3.8, B and D), we wondered how many of the downregulated FHF drivers and PM antidrivers in *Nipbl*$^{+/-}$ embryos could have their changes in expression attributed to the overexpression of *Nanog*. To answer this question, we compared the FHF drivers and PM antidrivers

downregulated in LB-stage *Nipbl*$^{+/-}$ embryos to those genes exhibiting downregulation in E7.5 *Nanog* Dox+ embryos. We found that approximately 40% of the FHF drivers and 38% of the PM antidrivers that were downregulated in *Nipbl*$^{+/-}$ embryos were also downregulated in E7.5 *Nanog* Dox+ embryos (Fig. 3.8, I and J). This raises the possibility that overexpression of *Nanog* leads to the downregulation of FHF drivers and PM antidrivers, and that this results in the misallocation of mesoderm cells to a PM fate at the expense of FHF (Fig. 3.3 and Fig. 3.8K).

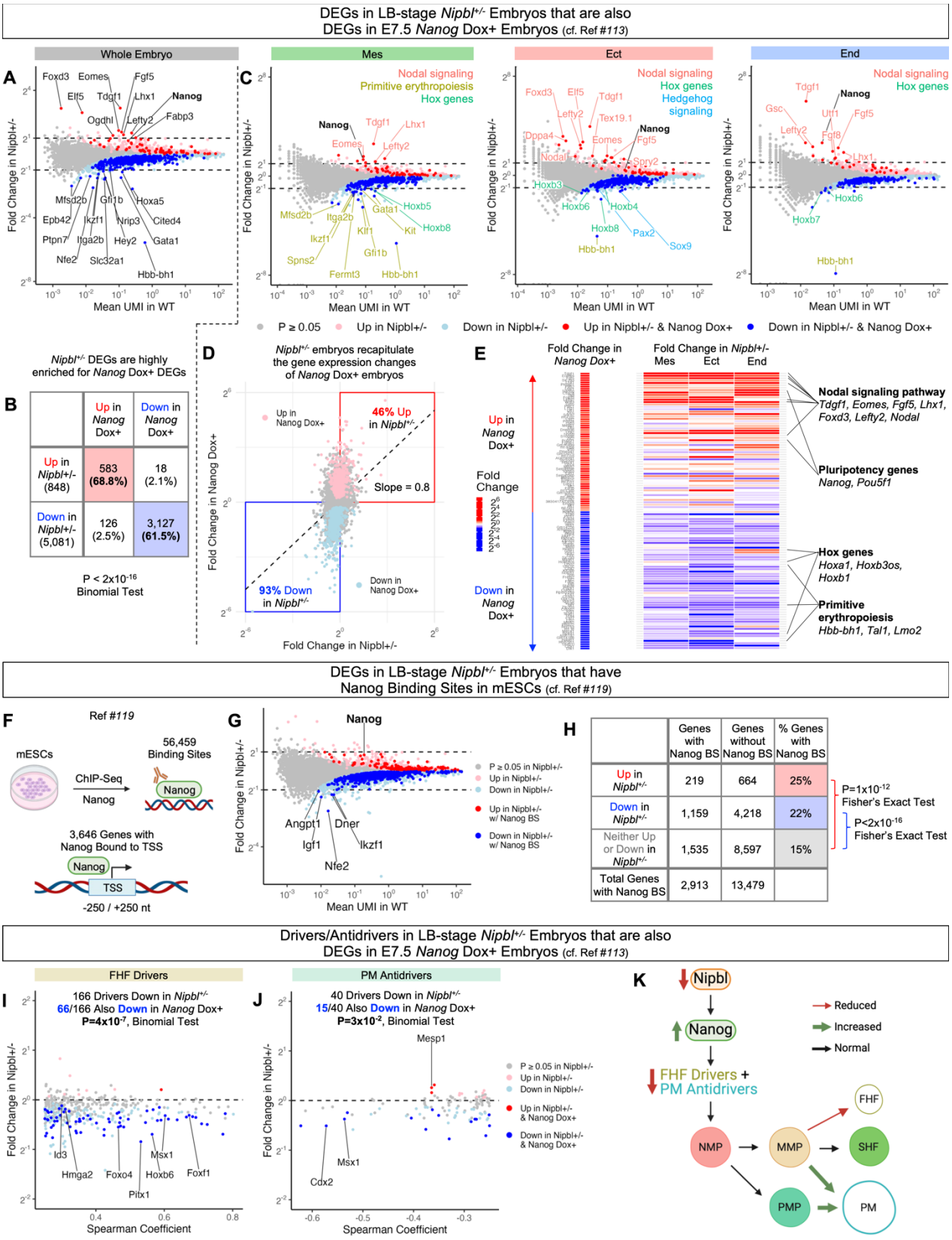

DEGs in LB-stage Nipbl+/- Embryos that are also DEGs in E7.5 Nanog Dox+ Embryos (cf. Ref #113)
Whole Embryo
Mes
Ect
End
Nodal signaling
Primitive erythropoiesis
Hox genes
Hedgehog signaling
Fold Change in Nipbl+/-
Mean UMI in WT
P ≥ 0.05
Up in Nipbl+/-
Down in Nipbl+/-
Up in Nipbl+/- & Nanog Dox+
Down in Nipbl+/- & Nanog Dox+
Nipbl+/- DEGs are highly enriched for Nanog Dox+ DEGs
Up in Nanog Dox+
Down in Nanog Dox+
Up in Nipbl+/- (848)
583 (68.8%)
18 (2.1%)
Down in Nipbl+/- (5,081)
126 (2.5%)
3,127 (61.5%)
P < 2x10-16
Binomial Test
Nipbl+/- embryos recapitulate the gene expression changes of Nanog Dox+ embryos
46% Up in Nipbl+/-
Slope = 0.8
93% Down in Nipbl+/-
Fold Change in Nanog Dox+
Fold Change in Nipbl+/-
Fold Change in Nanog Dox+
Fold Change in Nipbl+/-
Mes Ect End
Nodal signaling pathway
Tdgf1, Eomes, Fgf5, Lhx1, Foxd3, Lefty2, Nodal
Pluripotency genes
Nanog, Pou5f1
Hox genes
Hoxa1, Hoxb3os, Hoxb1
Primitive erythropoiesis
Hbb-bh1, Tal1, Lmo2
DEGs in LB-stage Nipbl+/- Embryos that have Nanog Binding Sites in mESCs (cf. Ref #119)
Ref #119
mESCs
ChIP-Seq
Nanog
56,459 Binding Sites
3,646 Genes with Nanog Bound to TSS
TSS
-250 / +250 nt
Mean UMI in WT
Genes with Nanog BS
Genes without Nanog BS
% Genes with Nanog BS
Up in Nipbl+/- 219 664 25%
Down in Nipbl+/- 1,159 4,218 22%
Neither Up or Down in Nipbl+/- 1,535 8,597 15%
Total Genes with Nanog BS 2,913 13,479
P=1x10-12 Fisher's Exact Test
P<2x10-16 Fisher's Exact Test
Drivers/Antidrivers in LB-stage Nipbl+/- Embryos that are also DEGs in E7.5 Nanog Dox+ Embryos (cf. Ref #113)
FHF Drivers
166 Drivers Down in Nipbl+/-
66/166 Also Down in Nanog Dox+
P=4x10-7, Binomial Test
PM Antidrivers
40 Drivers Down in Nipbl+/-
15/40 Also Down in Nanog Dox+
P=3x10-2, Binomial Test
Spearman Coefficient
Nipbl
Nanog
FHF Drivers + PM Antidrivers
Reduced
Increased
Normal
NMP
MMP
PMP
FHF
SHF
PM

Fig. 3.8. LB-stage *Nipbl*$^{+/-}$ mice replicate the gene expression changes of *Nanog* overexpression.

(**A**) Fold change in expression of DEGs in LB-stage *Nipbl*$^{+/-}$ embryos ($Q < 0.05$, Mann-Whitney U Test) that are also DEGs in E7.5 *Nanog* Dox+ embryos ($Q < 0.05$, T-Test) (*276*). (**B**) Percentage of DEGs in LB-stage *Nipbl*$^{+/-}$ embryos that are also DEGs in E7.5 *Nanog* Dox+ embryos (*276*). (**C**) Fold change in expression of DEGs in the germ layers of LB-stage *Nipbl*$^{+/-}$ embryos that are also DEGs in E7.5 *Nanog* Dox+ embryos (*276*). (**D**) Fold change in expression of genes in LB-stage *Nipbl*$^{+/-}$ embryos versus their fold change in E7.5 *Nanog* Dox+ embryos. DEGs in E7.5 *Nanog* Dox+ embryos (*276*) are colored pink and light blue. (**E**) Heatmap of the fold change in expression of DEGs from E7.5 *Nanog* Dox+ embryos (lowest *Q*-values from T-Test) (*276*) in the germ layers of LB-stage *Nipbl*$^{+/-}$ embryos. (**F**) ChIP sequencing for Nanog in mESCs from (*282*) identified 3,646 genes with one or more Nanog binding sites within +/- 250 nt of their TSSs. (**G**) Fold change in expression of DEGs in LB-stage *Nipbl*$^{+/-}$ embryos with one or more Nanog binding sites from (*282*). (**H**) Percentage of DEGs in LB-stage *Nipbl*$^{+/-}$ embryos with one or more Nanog binding sites from (*282*). Fold change in expression of DEGs in mesoderm cells of (**I**) FHF and (**J**) PM lineages of LB-stage *Nipbl*$^{+/-}$ embryos that are also DEGs in E7.5 *Nanog* Dox+ embryos (*276*). Transcription factor genes are labeled. (**K**) Reduction of *Nipbl* levels leads to the upregulation of *Nanog* in LB-stage *Nipbl*$^{+/-}$ embryos and downregulation of FHF drivers and PM antidrivers, resulting in the misallocation of mesoderm cells to PM at the expense of FHF.

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

We noticed that LB-stage *Nipbl*$^{+/-}$ embryos exhibited a significant underexpression, exceeding 2-fold, in *Hox* genes across all germ layers, especially genes within the *Hoxb* cluster (Fig. 3.6D and Fig. 3.8C). As *Hox* genes are integral to spatial patterning in development, we compared the expression of all *Hox* genes in *Nipbl*$^{+/-}$ embryos with their WT counterparts. The results, shown in Fig. 3.9, A and B, revealed extensive misexpression at both LB- and CC-stage. At LB-stage, any *Hox* gene that was significantly differentially expressed was underexpressed in *Nipbl*$^{+/-}$ embryos in every germ layer where expression was detectable (Fig. 3.9A). At CC-stage, we observed a trend of underexpression across all germ layers for the majority of misexpressed *Hox* genes (Fig. 3.9B), but exceptions were found in the mesoderm and ectoderm, where a select few *Hox* genes were overexpressed. These included *Hoxb1* and *Hoxb2* in the mesoderm and *Hoxa1* in the ectoderm, all of which are characterized as anterior *Hox* genes.

Hox genes are organized into chromosomal clusters that turn on expression in a wave-like temporal pattern, producing highly structured and sequential activation during development. Specifically, genes at the 5' end of the cluster, known as anterior *Hox* genes, are activated early in development, and progressively silenced as genes further forward the 3' end (posterior *Hox* genes) are turned on (*149*). To investigate whether our data reflected this known pattern, we examined the *Hoxb* genes, as they were the most prominently expressed *Hox* genes at both LB- and CC-stage. After ordering them according to their chromosomal positions, we plotted their expression within the mesoderm and ectoderm of WT embryos at the corresponding developmental stages. In the mesoderm of LB-stage embryos anterior *Hoxb* genes were actively expressed, while posterior *Hoxb* genes were only faintly detectable, if at all (Fig. 3.9C). At CC-stage, the expression of anterior *Hoxb* genes decreased from their LB-stage levels, accompanied

by the emergence of posterior *Hoxb* gene expression (Fig. 3.9C). This dynamic shift in expression aligns closely with the established understanding of *Hox* gene regulation during development. In contrast, the ectoderm exhibited a less pronounced pattern, with both anterior and posterior *Hoxb* genes expressing at minimal levels (Fig. 3.9D). No discernible difference in the expression of anterior *Hoxb* genes was detected in the ectoderm between LB- and CC-stage embryos (Fig. 3.9D).

Next we looked at the expression of *Hoxb* genes in the mesoderm of *Nipbl*$^{+/-}$ embryos. As in WT embryos, anterior *Hoxb* genes were actively expressed at LB-stage, while posterior *Hoxb* genes were silent (Fig. 3.9E). However, *Hoxb1* and *Hoxb2* were expressed at levels that were conspicuously lower than in WT (Fig. 3.9E). At CC-stage, *Nipbl*$^{+/-}$ embryos reflected the general trend found in WT embryos, with anterior *Hoxb* genes waning in expression as the posterior *Hoxb* genes began their ascent (Fig. 3.9F, note axis scale). Interestingly, the two most anterior Hox genes, *Hoxb1* and *Hoxb2* did not decline as much, proportionally, as they did in WT embryos, with the outcome being that, by CC-stage, expression of *Hoxb1* and *Hoxb2* in *Nipbl*$^{+/-}$ embryos was actually higher than in WT (Fig. 3.9F). These results collectively provide evidence for a temporal disruption in the *Hox* gene expression program in *Nipbl*$^{+/-}$ embryos. Specifically, *Hox* genes appear to be delayed in initiating expression at the LB-stage and similarly tardy in becoming suppressed at the CC-stage.

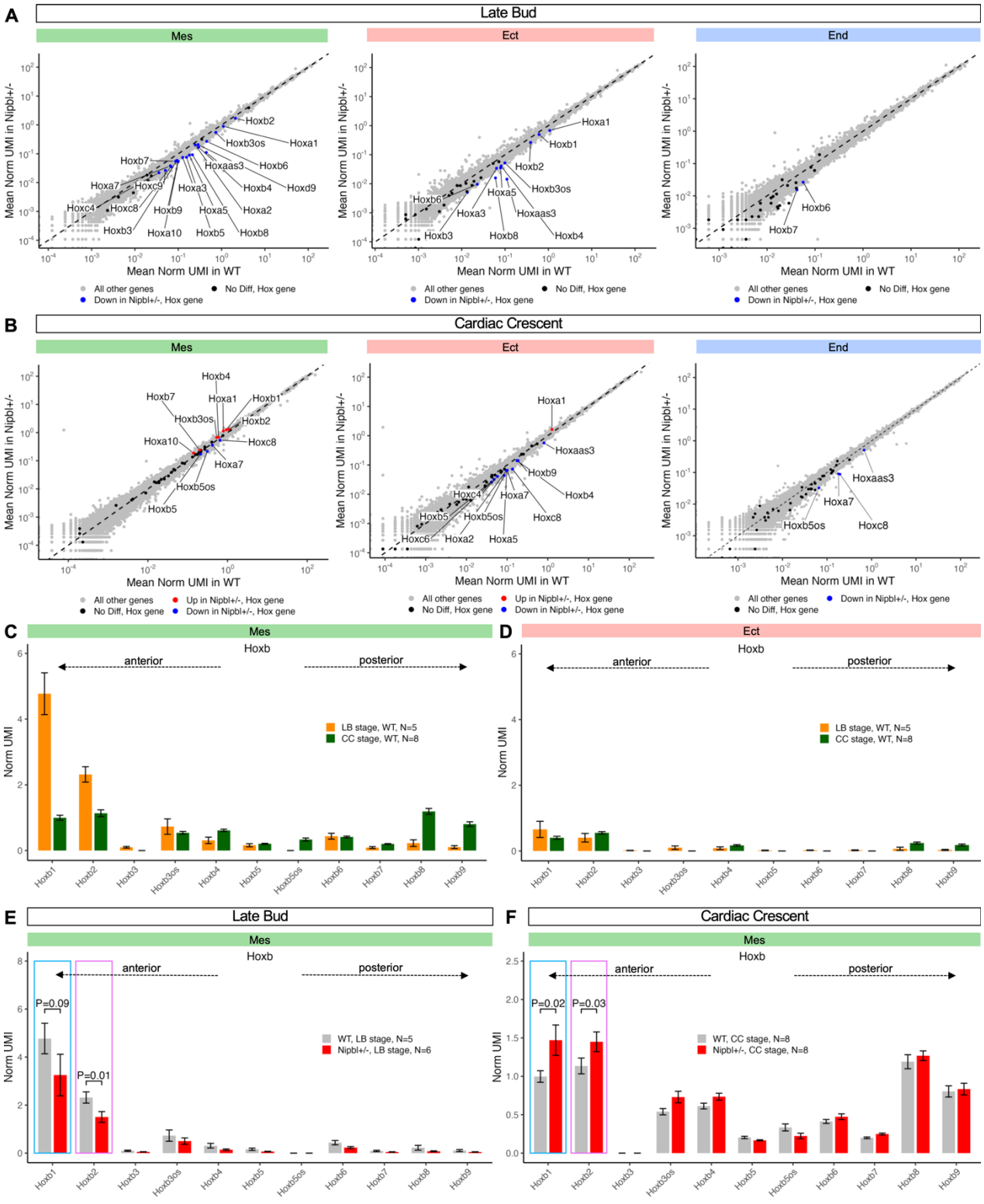
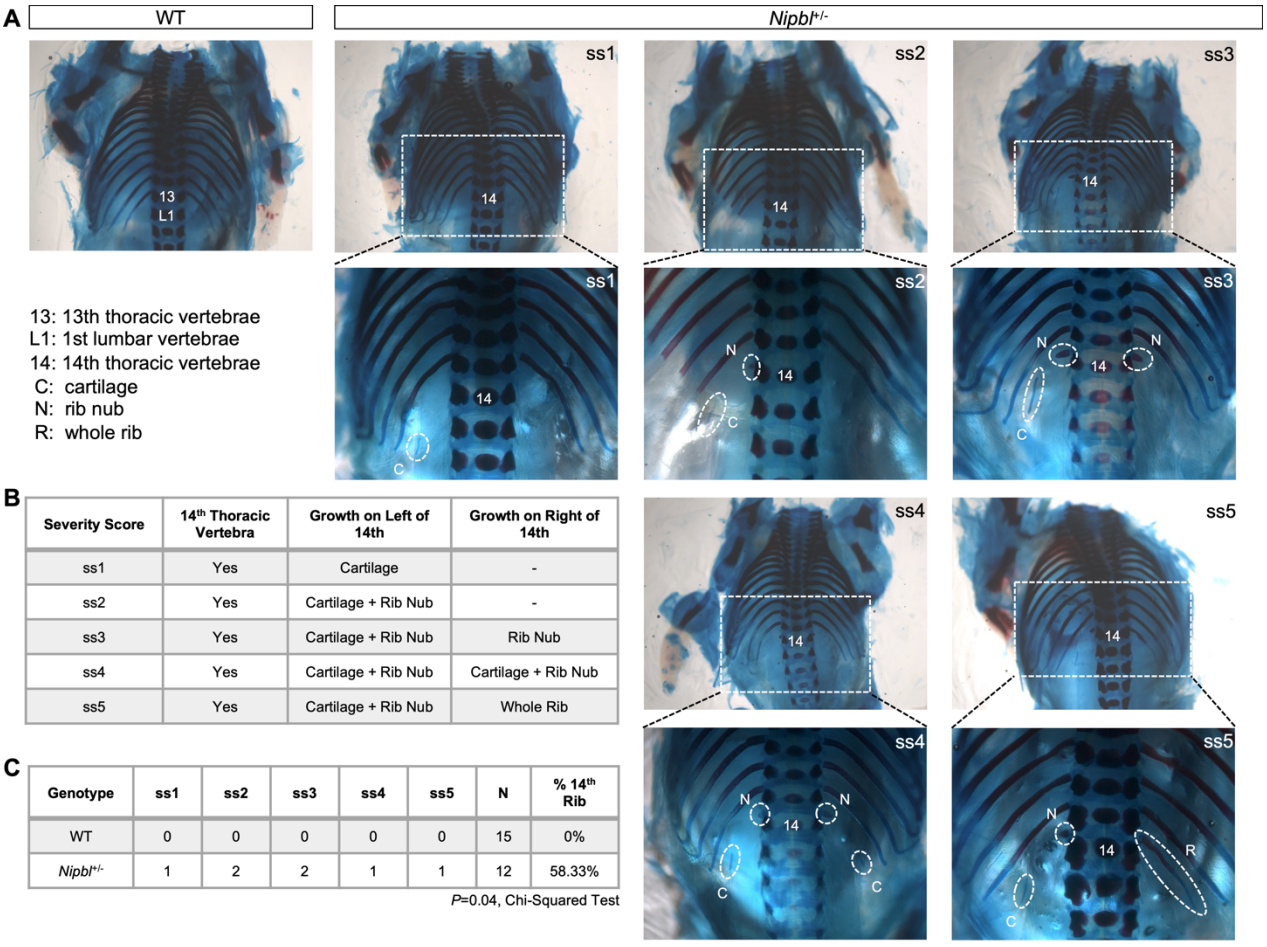


Fig. 3.9. *Nipbl*$^{+/-}$ mice show a delay in the expression of anterior *Hox* genes.

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

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

A previous study (*283*) showed that a knockout of *Hoxb1-Hoxb9* in mice led to the anteriorization of the axial skeleton, characterized by an increased number of thoracic vertebrae with ribs. Since *Nipbl*$^{+/-}$ mice underexpress *Hox* genes and show delayed regulation of anterior *Hox* genes, we were curious if similar anteriorizations could be detected in *Nipbl*$^{+/-}$ mice (Fig. 3.9). To explore this, we analyzed the vertebrae and ribs of 15 WT and 12 littermate *Nipbl*$^{+/-}$ embryos at E18.5, using alizarin red and alcian blue stains. In WT embryos, we confirmed the presence of the normal count of 13 ribs (Fig. 3.10A). In contrast, many *Nipbl*$^{+/-}$ embryos displayed 14 thoracic vertebrae, the 14th of which bore a range of rib growths, often displaying left-right asymmetry (Fig. 3.10A). We classified the growths of these 14th vertebrae based on severity, ranking them from low (ss1) to high (ss5), and documented left-right differences (Fig. 3.7B). Three out of 7 *Nipbl*$^{+/-}$ embryos displayed partial growth of a 14th rib on the left side only (S1-S2). These growths were either cartilage only or disconnected formation of rib nub with cartilage, but never any whole ribs. Four out of 7 *Nipbl*$^{+/-}$ embryos showed bilateral growth (S3-S5). Partial growth occurred on both the left and right sides, but complete formation a whole rib only ever occurred, in one instance, on the right side (S5). In total, 58% of E18.5 *Nipbl*$^{+/-}$ embryos demonstrated some form of rib growth from a 14th thoracic vertebra (Fig. 3.10C). These findings suggest that anteriorization of the axial vertebrae does occur in *Nipbl*$^{+/-}$ embryos, possibly as a result of delayed activation (and inactivation) of anterior *Hox* gene expression.



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. 3.10. *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. 3.10B. 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.

Discussion

The *Nipbl*$^{+/-}$ mouse, a model of the multisystem birth defects syndrome, Cornelia de Lange Syndrome (CdLS) (*64*), offers a distinct lens through which to explore the genetic origins of birth defects. *Nipbl*$^{+/-}$ mice display many of the same birth defects observed in human CdLS and a key feature of *Nipbl*-deficiency across various organisms is quantitative alterations in gene expression, including both upregulation and downregulation, affecting at least hundreds of genes in every tissue (*10, 18, 284*). Previous research into animal models of *Nipbl*-deficiency suggested that the root causes of birth defects in CdLS likely manifest during a period when progenitor cell populations are being formed for all major tissue and organ lineages (*11, 18*). Therefore, to elucidate how early, cell-type-specific changes in gene expression might contribute to the onset of birth defects, we used scRNAseq to conduct a comparative analysis between *Nipbl*$^{+/-}$ embryos and their WT littermates at key developmental stages, focusing on the conclusion of gastrulation, LB-stage (approximately E7.5) to early CC-stage (approximately E7.75) (*136*).

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

In Fig. 3.1, we showed that *Nipbl*$^{+/-}$ embryos do not lack any cell populations found in WT embryos. However, LB-stage *Nipbl*$^{+/-}$ embryos possess fewer total mesoderm cells, and within the mesoderm, have fewer PEs and FHF cells, and more PM cells (Fig. 3.2). These changes, while not always large, are likely to be of physiological importance. For example, the observation that PEs are reduced in number is consistent with defects in blood formation and circulation observed in *Nipbl*-morphant zebrafish (*11*), and thrombocytopenia in CdLS (*89*).

In addition, the reduction in FHF cell number could contribute to the presence of CHDs in *Nipbl*$^{+/-}$ mice and individuals with CdLS (*18, 79*). In a prior study in *Nipbl*$^{+/-}$ mice, the

majority of CHDs observed were in heart regions (right ventricle, atrial/ventricular septa) considered to be derivatives of the second heart field (SHF) (*18*). Although FHF cells are not thought to give rise directly to "SHF structures", they likely exert indirect effects on the development of SHF derivatives. This echoes the conclusions of Santos et al., 2016 (*18*), who found that complex interactions between different cell types (even between non-cardiogenic and cardiogenic cell types), influence the incidence of CHDs in *Nipbl*$^{+/-}$ mice.

The increased number of PM cells at LB-stage is intriguing, as it correlates with observed defects in PM derivatives, including the axial skeleton (Fig. 3.10) and limb digits in *Nipbl*$^{+/-}$ mice (*285*). Notably, both such deficits are features of CdLS (*60*). How an overabundance of PM cells could contribute to limb defects is not obvious. Nevertheless, these findings suggest that birth defects in *Nipbl*-deficient organisms arise not only from the absence of specific progenitor cell populations, but also from progenitor cell misallocation events that alter the relative proportions of cell types.

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

Nipbl$^{+/-}$ embryos at LB-stage did not exhibit global changes in apoptotic activity, cell proliferation, developmental timing, or overall lineage trajectories (Fig. S3.11, Fig. S3.12, Fig. 3.3, and Fig. 3.4). However, we produced evidence that nascent mesoderm cells in *Nipbl*$^{+/-}$ embryos differentiated more frequently into PM than FHF, compared to WT (Fig. 3.4, C to I). This led us to conclude that misallocation of mesoderm cells in *Nipbl*$^{+/-}$ embryos is driven by alterations in early cell fate choices that occur along the main pathways of otherwise unaltered lineage trajectories. These findings suggest that some types of structural birth defects, such as

heart defects, arise from quantitative alterations to cell fate determination, rather than a complete disruption of an entire lineage pathway.

In Fig. 3.5, we presented evidence that reduced differentiation of mesoderm cells into FHF in *Nipbl*$^{+/-}$ embryos could be attributed to the underexpression of genes predicted to drive FHF differentiation. Interestingly, underexpressed FHF driver genes show strong enrichment for genes associated with EMT (Fig. 3.5, J and K). EMT not only occurs during the emergence of mesenchymal mesoderm and FHF cells during gastrulation; it also plays a role in the contribution of later cell lineages (SHF cells, endocardial cells, and neural crest cells) to various cardiac structures. SHF cells undergo EMT as they participate in the elongation of the heart tube and outflow tract (*286*), and endocardial cells undergo EMT as they form structures such as the cardiac cushion (*160*). Later in development, migratory neural crest cells, which arise from neural epithelium by EMT, contribute to the aorticopulmonary septum and parts of the outflow tract (*287*). Taken together, these observations suggest that dysregulation of EMT genes may play an important role in the development of heart defects in *Nipbl*$^{+/-}$ mice.

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

In contrast to our prior studies of *Nipbl*$^{+/-}$ mice, in which differences in gene expression were invariably found to be small (less than 2-fold) (*10*), the present study demonstrated that, in gastrulation stage embryos, some genes show much larger shifts in expression (as high as eight-fold) (Figs. 6D and 8H). One of the largest of such changes was in *Nanog,* which displayed overexpression in LB-stage *Nipbl*$^{+/-}$ embryos, and which persisted to CC-stage (Figs. 7 and 8). Using data from Tiana et al., 2022 (*276*), we demonstrated that many of the gene expression changes in *Nipbl*$^{+/-}$ embryos mirror those seen in embryos with induced *Nanog* overexpression (Fig. 3.8, A to E, and Fig. S3.15, A to E). Moreover, many *Nipbl*$^{+/-}$ DEGs are likely to be direct

Nanog targets, as the corresponding genes display enrichment for Nanog binding sites (Fig. 3.8, F and H). Comparisons of our data with those of Tiana et al., 2022 (*276*) also suggest that misallocation of mesoderm cells to a PM fate at the expense of FHF, resulting from downregulation of FHF drivers / PM antidrivers, may be a consequence of *Nanog* overexpression (Fig. 3.3 and Fig. 3.8K).

A question that arises from these observations is whether all gene expression alterations observed in *Nipbl*$^{+/-}$ embryos might be due to *Nanog* overexpression. For a number of reasons, we do not think that this is the case. First, for the largest gene expression shifts (greater than two-fold over- or underexpressed), only 39% at LB-stage and 18% at CC-stage were misexpressed in the same direction as DEGs in embryos in which *Nanog* was overexpressed (Fig. 3.8 and Fig. S3.15) (*276*). Second, although *Nanog* continues to be overexpressed in *Nipbl*$^{+/-}$ embryos from LB- to CC-stage, the magnitude of its expression decreases substantially by CC-stage (Fig. 3.7). Third, *Nipbl*-haploinsufficiency itself causes significant, large expression changes in other genes not implicated as consequences of *Nanog* overexpression. For instance, we observed reduced expression of genes linked to the Notch signaling pathway in the mesoderm, and increased expression of adult hemoglobin genes in the endoderm (Fig. 3.6). Therefore, we conclude that the extensive gene expression changes in *Nipbl*$^{+/-}$ embryos are unlikely to be due to *Nanog* overexpression alone.

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

In Figs. 4D and 8G, we showed that large decreases in *Hox* gene expression occur across all germ layers of LB-stage *Nipbl*$^{+/-}$ embryos. While only a subset of *Hox* genes is underexpressed by more than two-fold, many *Hox* genes across all germ layers show small (less than two-fold) but significant decreases in expression (Fig. 3.9A). We also provide evidence for

a temporal disturbance in anterior *Hox* gene expression in *Nipbl*$^{+/-}$ embryos, whereby *Nipbl*-haploinsufficiency postpones the onset of anterior *Hox* gene expression at LB-stage and delays their subsequent deactivation at CC-stage (Fig. 3.9, C to F). These shifts in *Hox* gene expression dynamics mirror those in a previous study of *nipbl*-morphant zebrafish, in which genome-location-specific misexpression of *hox* genes was also observed (*288*). In Fig. 3.10, we showed that anteriorization of thoracic vertebrae occurs in E18.5 *Nipbl*$^{+/-}$ mice (Fig. 3.9). Such alterations in axial skeletal development parallel skeletal anomalies that have been identified in individuals with CdLS, including fused, absent, or misshapen ribs (*62*). Since *Nipbl*-deficiency is the dominant form of CdLS, it is interesting to speculate that these skeletal anomalies might originate from underexpression of and delayed anterior *Hox* gene activity during gastrulation.

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

In Fig. 3.4D, we showed that *Nipbl*$^{+/-}$ embryos at LB-stage show large overexpression of genes associated with the Nodal signaling pathway across all germ layers. Nodal signaling, pivotal for left-right patterning (*289*), ensures the appropriate positioning and morphogenesis of the musculoskeletal system and internal organs (*290*). In Fig. 3.10, we showed that growth of a 14th vertebra in E18.5 *Nipbl*$^{+/-}$ embryos displayed left-right asymmetry, with the growth of a whole 14th rib only occurring on the right side of the embryo. Such patterning alterations are consistent with the heightened incidence of right-sided skeletal anomalies in CdLS (*60*). Changes in left-right patterning are also consistently associated with intestinal malrotation and some types of cardiac anomalies observed in *Nipbl*$^{+/-}$ mice, *nipbl*-morphant zebrafish, and CdLS (*10, 17, 18*). These data suggest that left-right patterning defects in CdLS might be due to overexpression of Nodal signaling genes as early as gastrulation.

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

The results of this study, together with prior studies of mouse and zebrafish models, reinforce the notion that the chromosomal location of genes is an important factor in their sensitivity to variations in *Nipbl* levels. For example, our prior studies of *Nipbl*$^{+/-}$ embryonic brain demonstrated that the largest expression changes among genes in the 22-gene protocadherin beta (*Pcdhb*) cluster were found in those genes situated at the 5' and 3' ends of the cluster (*10*), where CTCF sites are located (*291*). In developing pectoral fins of *nipbl*-morphant zebrafish, *hox* genes located near the 3' end of three separate *hox* clusters (*hoxa*, *hoxc*, and *hoxd*) show a position-dependent pattern of overexpression (*288*). Similarly, in the present study, we showed that anterior *Hoxb* genes – located at the 3' end of the *Hoxb* cluster – are preferentially underexpressed in the mesoderm of *Nipbl*$^{+/-}$ embryos at gastrulation (Fig. 3.8E). Thus, we consistently observe a strong influence of chromosomal location on gene sensitivity to *Nipbl* levels, particularly within gene clusters, and this influence is already apparent at gastrulation.

Recent studies suggest that changes in gene expression due to *Nipbl*-deficiency may be the result of global alterations to chromosomal structure and organization. For instance, drastic reduction of *Nipbl* expression in mouse hepatocytes resulted in genome-wide depletion of topologically associated domains (TADs) and Hi-C interaction peaks (*292*), both of which require chromatin looping (*293*). In fact, recent studies have demonstrated that both the formation and rate of chromatin loop extrusion were reduced *in vitro* when a CdLS pathogenic mutation was introduced into *NIPBL* (*294*).

Chromatin looping within TADs is thought to bring distant cis-regulatory elements, such as enhancers, into proximity with target promoters. Interestingly, Nipbl has been reported to

preferentially bind to enhancers and promoters (*295*), and ChiP-seq studies show reduced enhancer-promoter interactions in *Nipbl*$^{+/-}$ mouse embryonic fibroblasts (*296*). Since chromatin looping is essential for gene activation, impaired loop extrusion as a consequence of reduced Nipbl levels may provide an explanation for why *Nipbl*-haploinsufficiency causes more genes to be downregulated than upregulated in gastrula-stage mouse embryos (e.g., >80% of mesodermal genes, Fig. 3.6A), a trend that has been observed in studies of other tissues (*296*).

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

Materials and Methods

Statement on Care and Use of Animals

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

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

WT and *Nipbl*$^{+/-}$ mouse littermates were generated by mating *Nanog*$^{Cre/+}$ mice (*145*) and *Nipbl*$^{Flox/Flox}$ mice (*18*). *Nipbl*$^{Flox/Flox}$ mice have an inverted gene trap cassette encoding *b-geo* that is flanked by *Cre* recombinase target sites in intron 1 of *Nipbl* alleles (Fig. 3.1A) (*18*). In this inverted orientation, referred herein as Flox, 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 into a non-inverted orientation which we call FIN (Fig. 3.1A). In this non-inverted orientation, trapping of the *Nipbl* gene occurs and *b-geo* is expressed as a reporter of successful gene trapping. Therefore, the *Nipbl* FIN allele is a null allele. *Nanog*$^{Cre/+}$ mice carry a transgene encoding a *Cre* recombinase downstream of a promoter of the *Nanog* gene and initiate recombination in the earliest cells of the embryo (*145*). Consequently, mating *Nanog*$^{Cre/+}$ mice with *Nipbl*$^{Flox/Flox}$ mice results in littermates that are either *Nipbl*$^{Flox/+}$ or *Nipbl*$^{FIN/+}$, entirely. *Nipbl*$^{Flox/+}$ mice express *Nipbl* at wildtype levels (Fig. 3.1A) and show no defects, making them essentially WT (*18*). *Nipbl*$^{FIN/+}$ mice express *Nipbl* at levels ~50% lower than WT (fig. S3), and show defects similar to those observed in CdLS, making them essentially *Nipbl*$^{+/-}$ (*18*).

Timing of Mouse Pregnancies

To generate female mice that were pregnant on the same day, male mice were singly housed and female mice were group-housed in groups of five for a minimum of one week to synchronize their estrous cycles, thus taking advantage of the Lee-Boot effect (*297*). At the beginning of the night cycle, the bedding from female cages were discarded and the bedding from at least two male cages were transferred into each of them. At the beginning of the third night cycle after which the females were exposed to male bedding, all females were transferred into male cages, resulting in two females per male, thus taking advantage of the Whitten effect

(*298*). At the end of the third night cycle, the females were inspected for vaginal plugs. Those that had vaginal plugs were considered potentially pregnant. To time the dissection of embryos from potentially pregnant females, we considered the end of the 12-hour night cycle after which the vaginal plug was discovered as E0 and dissected the embryos at the following times post E0: E7.5 (7 days + 12 hours) for late bud stage, E7.75 (7 days + 18 hours) for cardiac crescent stage, E7.41 (7 days + 10 hours) for early bud stage, E7.58 (7 days + 14 hours) for early head fold stage embryos, and at E18.5 (18 days + 12 hours).

Dissection of Mouse Embryos

Pregnant female mice were euthanized by CO_2 inhalation followed by cervical dislocation. The uterine horns were dissected out of their abdomens with dissection forceps and scissors and placed in a petri dish with 1X DEPC PBS on ice. Individual deciduae were separated from one another and transferred into their own petri dishes with 1X DEPC PBS on ice. Embryos were dissected out of each deciduae under a dissection microscope as described in (*299*). The Reichert's membrane was removed from each embryo. The ectoplacental cone was separated from embryos collected for scRNAseq, and transferred by forcep into microcentrifuge tubes, where they were kept on ice or stored at 20°C as tissue for polymerase chain reaction (PCR) genotyping. The exocoelom was separated from embryos collected for scRNAseq and transferred into a welled plate by wide bore pipette tip with fixative on ice as tissue for genotyping by X-gal stain. Embryos for scRNAseq were transferred by wide bore pipette tip with 1X DEPC PBS into microcentrifuge tubes and kept on ice until dissociation. Tails were separated from E18.5 embryos collected for Alcian Blue – Alizarin Red staining. Embryos for Alcian Blue – Alizarin Red staining were transferred by forcep into 10% neutral buffered formalin in scintillation vials and kept at 4°C for 24 hours.

Genotyping of Mouse Embryos and ES Cells

X-gal Stain

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

Polymerase Chain Reaction

Ectoplacental cones were treated with 50 ul 60 ug/ml proteinase K in PBND (50 mM KCl, 10 mM Tris-HCl, 2.5 mM MgCl2, 0.1 mg/ml gelatin, 0.45% Nonident P40, and 0.45% Tween 20), for 1 hour at 55℃ to extract DNA for PCR. To deactivate proteinase K so that it does not interfere with DNA amplification, they were then incubated at 95℃ for 10 min. To genotype mouse embryos as either *Nipbl*$^{Flox/+}$ (WT) or *Nipbl*$^{FIN/+}$ (*Nipbl*$^{+/-}$) and ES cells as either *Nipbl*$^{Flrt/+}$ (WT) or *Nipbl*$^{FLEX/+}$ (*Nipbl*$^{+/-}$), standard Taq PCR was performed on the extracted DNA using the primers and thermocyling protocol previously described in (*18*). Primer 1: 5'-CTCCGC CTCCTCTTCCTCCATC-3'. Primer 2: 5'-CCTCCCCCGTGCCTTCCTTGAC-3'. Primer 3: 5'-

TTTGAGGGGACGACGACAGTCT-3'. Thermocycling conditions: 1 cycle a 95°C for 30 sec, 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 hold at 4°C. PCR products were treated with loading dye and electrophoresed in agarose gel stained with SYBER Safe DNA Gel Stain (Invitrogen: S33102) without cleanup and visualized under UV light. Flox conformation: 782 base pairs (bp). FIN conformation 518 bp. Flrt conformation: 735 bp. FLEX conformation: 652 bp.

Single Cell RNA Sequencing

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

20 ul single cell suspensions were submitted to the Genomics High Throughput Facility (GHTF) at the University of California, Irvine (*300*) for scRNAseq using 10X Genomics' Chromium Next GEM Single Cell 3' Kit v3.1 (10x Genomics: 1000268), Chromium Next GEM Chip G Single Cell Kit (10x Genomics: 1000120), and Chromium Controller. Passage through the Chromium Controller resulted in sample, cell, and transcript barcoded cDNA, which GHTF amplified by PCR. GHTF assessed the quality and quantity of the amplified cDNA by electrophoresis using Agilent's Agilent High Sensitivity DNA Kit (Agilent: 5067-4626) and Bioanalyzer before library construction. Constructed libraries, representing embryonic samples, were multiplexed and sequenced by GHTF on the Illumina HiSeq 4000 to a minimum depth of 20 million read pairs per cell. GHTF demultiplexed Illumina's raw base call (BCL) files and returned FASTQ files as deliverables.

Read Mapping and Cell Calling

Cell Ranger v3.0 was used to map reads onto the GRCm38/mm10 C57BL/6J Mus musculus genome/transcriptome assembly and call cells. Cell Ranger does this by using a read mapper called STAR (*301*), which performs splicing-aware mapping of reads to the genome. Reads are considered confidentially mapped to the genome with a MAPQ of 255. Exonic reads are further mapped to annotated transcripts. A read that is compatible with the exons of an annotated transcript, and aligned to the same strand, is considered mapped to the transcriptome. CellRanger called cells using the EmptyDrops method described in (*302*).

Normalization of Library Depth

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

Removal of Low-Quality Cells and Doublets

Cells exceeding three median absolute deviations in any one of the following criteria among cells of the same stage were considered either low-quality cells or doublets (*303*) and removed: 1) percentage of mitochondrial genes expressed, 2) number of genes expressed, and/or 3) number of transcripts detected (Fig. S3.2 and Fig. S3.8).

Normalization of Cell Depth

Seurat v3.0 was used to normalize the read depth between cells of the same stage using the SCtransform method, which is described in (*304*). SCtransform does this by modeling the read counts in a regularized negative binomial model to determine the variation due to read depth and then adjusting that variance according to genes of similar abundances. SCtransform was also used to normalize the read depth between cells when cells were subset from the whole embryo into germ layers and clusters.

Batch Effect Correction

Seurat v3.0 was used to correct for batch effects among libraries of the same stage and genotype. Seurat does this by identifying a set of shared variable genes among the libraries being considered, and using these genes, identifies pairs of cells between any two libraries whose expression of these genes are similar to other. These pairs of cells act as anchors between libraries for batch effect correction and integration (*193*).

Clustering of WT Cells

Seurat v3.0 was used to cluster WT cells. Seurat does this by first performing principal component analysis on the shared variable genes identified during batch effect correction and integration. Principal components whose explained variances exceeded 2 median absolute deviations were used to calculate k-nearest neighbors and construct a shared nearest-neighbor

graph. Clusters were determined by optimization of the modularity function using the Louvain algorithm. The number of clusters was controlled by modulating the resolution function.

The optimal number of clusters was determined by clustering cells at increasing consecutive numbers of clusters and generating a clustering tree (Fig. S3.5) visualizing how cell cluster identities change as the number of clusters consecutively increase. Clusters are stable when a large proportion of cells are derived from a single preceding cluster rather than multiple preceding clusters. We adopted Shannon Entropy as a measure of these proportions as a measure of intra-cluster stability. A low Shannon Entropy represents high intra-cluster stability. We visually inspected the clustering tree to determine which number of clusters maximized the number of clusters while at the same time minimizing the total Shannon Entropy across all clusters at that number of clusters.

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

Differential Gene Expression Analysis Between Clusters

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

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

Seurat v3.0 was used to project *Nipbl*$^{+/-}$ cells onto WT clusters. Seurat does this by identifying a set of shared variable genes among WT samples, and using these genes, identifies pairs of cells between WT cells and *Nipbl*$^{+/-}$ cells whose expression of these genes are similar to other. These pairs of cells act as anchors between WT cells and *Nipbl*$^{+/-}$ cells for projecting and sorting *Nipbl*$^{+/-}$ cells into WT clusters (*193*).

Reverse Projection

Using the same method that was used to cluster WT cells (see Clustering of WT Cells), cells from LB-stage *Nipbl*$^{+/-}$ embryos were first clustered and annotated (Fig. S3.6A). Cells from LB-stage WT embryos were then projected onto the *Nipbl*$^{+/-}$ clusters (Fig. S3.6, B to D) (see Projection of *Nipbl*$^{+/-}$ Cells onto WT Clusters).

Differential Gene Expression Analysis Between Genotypes

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

Pseudotime

URD was used to calculate the pseudotime of cells from EB-, LB-, and EHF-stage embryos (*247*). URD does this by constructing a diffusion map of transition probabilities and starting with an assigned group of root cells, performs a probabilistic breadth-first graph search using the transition probabilities. This moves step-wise outward from the root cells, until the

entire graph is visited. Several simulations are run, and then pseudotime is calculated as the average iteration that visited each cell.

Construction of Lineage Trajectories

Velocyto (*305*) was used to count the numbers of spliced and unspliced transcripts per gene, using default parameters. Reads aligning to exonic regions were counted as spliced. Reads aligning to intronic regions were counted as unspliced. Reads aligning to exon-intron boundaries were considered ambiguous and excluded from downstream analyses. scVelo (*248*) was used estimate RNA velocities. Counts were normalized using the pp.filter_and_normalize() function, moments of unspliced vs spliced abundances were computed using the pp.moments() function, and velocities were computed using the tl.velocity() function, all using default parameters. Lineage trajectories were visually inferred from stream plots of computed RNA velocities.

Calculating Fate Probabilities

CellRank was used to calculate the fate probabilities of mesoderm cells (*250*). CellRank does this by performing RNA velocity-directed random walks from initial cell states to terminal cell states. Fate probabilities correspond to the fraction of walks in which a cell was a part of that that terminated in a particular terminal cell state. Terminal states were set with the *set_terminal_states()* function, absorption probabilities were computed with the *compute_absorption_probabilities()* function, and driver genes were computed with the *compute_lineage_drivers()* function, all with default parameters.

Cell Cycle Phase Assignment

Seurat was used to assign cells into G1, S, or G2/M phase based on the expression of markers of S phase and G2/M phase provided by Seurat. Using the CellCycleScoring() function, Seurat calculated scores for the expression of S phase and G2/M phase markers, while

considering the expression of these marker genes to be anticorrelated to one another. When cells express neither, they are considered to be in G1 phase.

Identification of Drivers and Anti-drivers

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

Gene Set Over-Representation Analysis

clusterProfiler (*306*) was used to perform gene set over-representation analysis. It does this by performing a Fisher's Exact Test on a contingency table of the genes in a gene set that match or do not match the genes of interest. P-values were corrected for false discovery using the Bonferroni correction method. Gene sets with Q-values less than 0.05 were considered statistically significant and over represented by the genes of interest.

Gene Set Enrichment Analysis

FGSEA (*307*) was used to perform gene set enrichment analysis.

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

ES cells were grown in Glasgow's MEM, 15% heat-inactivated FBS (ES cell qualified, Hyclone SH30071.03E), 1X Glutamine, 1X Pen-Strep, 5mM mercaptoethanol, and 1000 U/ml leukemia inhibitory factor (LIF: ESGRO Millipore). Following limited dilution single cell

cloning, a clone of EUC313f02 *Nipbl*$^{FLEX/+}$ ES cells (European Conditional Mouse Mutagenesis Program) was transfected with pCAG-Cre:GFP (Addgene #13776) using Lipofectamine 2000 (Invitrogen) to convert the *Nipbl*^{FLEX} allele to the *Nipbl*^{Flrt} conformation in vitro. 48 hours post-transfection, cells were plated at 1 cell/well into several 96 well plates. Clonal colonies were isolated, and the clones were stained for β-galactosidase (lacZ) activity using X-gal. For LacZ staining, the ES cells were fixed for 5 minutes in 2 mM MgCl2, 0.5% glutaraldehyde in 1X PBS, followed by 3 washes with 1xPBS at room temperature. X-gal staining (5 mM K3Fe(CN)6, 5 mM K4Fe(CN)6, 2 mM MgCl2, 1 mg/mL X-gal in 1X PBS) was performed at 37°C until the blue precipitate was detected. Colonies positive for X-gal staining, *Nipbl*$^{FLEX/+}$ ES cells, and negative for X-gal staining, *Nipbl*$^{Flrt/+}$, were verified by PCR genotyping, described above.

Reverse Transcription-Quantitative Polymerase Chain Reaction

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

Identification of Genes with Nanog Binding Sites

In Avsec et al., 2021 (*282*), ChIP-seq was performed for Nanog in mESCs and 56,459 peaks were called (Fig. 3.8F). We considered the called peaks from that study as representing Nanog binding sites. Using the narrowPeak file from that study (which contains the called peaks), the GRCm38/mm10 C57BL/6J Mus musculus genome annotation, and ChIPseeker v3, we identified which genes had a peak +/- 250 nt of their TSS. We identified a total of 3,646 genes with a Nanog binding site at their TSS (Fig. 3.8F).

Alcian Blue – Alizarin Red Staining

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

Acknowledgements

I am deeply grateful to Henry Chang, Louise Villagomez, and Brian Bui for their meticulous efforts in maintaining mouse lines, conducting crosses, dissecting and imaging mouse

embryos, performing X-gal staining, assisting with the preparation of single cell suspensions, and PCR genotyping. Thanks also to Melanie Oakes of the University of California Irvine's (UCI) Genomics and Research Technology Hub for preparing the single cell RNA sequencing libraries and sequencing them. I appreciate Qing Nie and Arthur D. Lander (ADL) for their invaluable discussions on cluster analysis and ADL's innovative approach for ordering embryos in pseudotime in such a way that reflects real time. I appreciate Jesse Kreger (JK) and Adam L. MacLean (ALM) for recommending the use of CellRank and their assistance in predicting the cell fate probabilities of mesoderm cells. My gratitude extends to ADL and Anne L. Calof (ALC) for their insightful annotations of gene expression changes into biologically informative groups. I am grateful to Martha E. Lopez-Burks for performing RT-qPCR for *Nipbl*, *Nanog*, and *Pou5f1* in mouse embryonic stem cells, providing the protocol for the alizarin red and alcian blue staining of bone and cartilage, and assisting with the identification of defects in thoracic vertebrae. Thanks also to Christopher Wu for his analysis and visualization of *Hox* gene expression. I also thank ADL for finding the study by Tiana et al., 2022, in which *Nanog* was overexpressed in mouse embryos. I further acknowledge ADL and ALC for their helpful discussions on data analysis and figure visualization. To JK, ALM, ADL, and ALC for reviewing and editing this work. Gratitude is also due to UCI's Center for Complex Biological Systems and the Cornelia de Lange Syndrome Foundation for supporting the initial sequencing of embryos. I am also thankful to the National Science Foundation-Simons Center for Multiscale Cell Fate Research for supporting our collaboration with JK and ALM. To the National Institute of Neurological Disorders and Stroke of the National Institutes of Health (NIH) for supporting me with a training grant (NINDS-NIH T32NS082174), and the National Heart, Lung, and Blood Institute of the NIH for supporting this study with a research grant (NHLBI-NIH R01HL138659).

CHAPTER 4

Conclusions

A need for understanding the origins of birth defects

Approximately 3% of newborns worldwide are affected by birth defects (*1*). While these conditions are relatively uncommon, they represent the primary cause of infant mortality in developed nations (*21*) and can lead to enduring health challenges for those who survive beyond infancy. This underscores the urgent need to develop interventions that not only alleviate the clinical severity of birth defects but also prevent their onset. Various factors, including environmental exposures, infections during pregnancy, nutritional deficiencies, and genetic influences – as well as their interactions – have been shown to contribute to the development of birth defects. Despite identifying multiple causes, these factors explain only about 20-30% of birth defects, leaving the origins of the majority (70-80%) unknown (*2*).

Birth defect syndromes often present multiple, concurrent birth defects that resemble the characteristics of more common, isolated birth defects. These syndromes usually stem from a single genetic origin, which facilitates their replication in animal models (*31, 32*). Creating an animal model of a birth defect syndrome would enable investigation of the causal mechanisms behind several types of birth defects, even in cases where the causes of many isolated defects remain unknown. This method offers a more efficient approach than creating animal models for studying each defect individually and could yield significant insights into the common underlying mechanisms of various birth defects.

Using the *Nipbl*-haploinsufficient mouse model to study birth defects

Cornelia de Lange Syndrome (CdLS) is a well-studied birth defect syndrome for which an animal model has been developed. This syndrome arises from mutations in proteins that are part of or associated with the cohesin complex. Key proteins involved include *NIPBL* (*5*), *SMC1A* (*4*), *SMC3* (*4*), *HDAC8 (96),* and *RAD21 (97)*. *SMC1A, SMC3*, and *RAD21* are components of the cohesin complex itself (*7, 97*), while *NIPBL* and *HDAC8* interact with the complex (*6, 98*). The majority and the severest CdLS cases are caused by mutations in *NIPBL* (*5*), which encodes a protein that plays a role in loading the cohesin complex onto DNA (*6*). This process allows the cohesin complex to encircle chromatin, through which it plays a pivotal role in regulating gene expression.

Nipbl$^{+/-}$ mice have demonstrated numerous similarities with human CdLS, displaying a range of symptoms that include prenatal and postnatal growth retardation, cardiac septal defects, delayed bone development, a lean body phenotype, and microbrachycephaly with distinct craniofacial alterations. These mice also exhibit behavioral anomalies, ophthalmological irregularities, cerebellar hypoplasia, and hearing impairments. In certain aspects, the *Nipbl*$^{+/-}$ mouse shows remarkable parallels with human CdLS, such as variable growth rates in early postnatal life (*122*) and a characteristic upturned nose (*123*). However, some typical CdLS features like upper limb reduction, fusion abnormalities (*57, 60*), and ventricular septal defects are not present in the *Nipbl*$^{+/-}$ mouse (*81*). Despite these differences, the *Nipbl*$^{+/-}$ mouse serves as a strong animal model that effectively represents many facets of CdLS.

Research using cell lines from individuals with CdLS and animal models lacking cohesin complex proteins has provided strong evidence of their crucial role in gene regulation. In cases of *Nipbl*-haploinsufficiency, the most prevalent form of CdLS, there are changes in the

expression of hundreds to thousands of genes (*9-12*). Although the alterations in gene expression in *Nipbl*-deficient animals are typically subtle, often less than 2-fold, these minor changes collectively contribute to significant structural and functional defects. Thus, studies of CdLS and corresponding animal models have shed light on a critical aspect regarding the origins of birth defects: the cumulative or synergistic effects of slight disturbances in gene expression can lead to developmental abnormalities.

The role of *Isl1*-haploinsufficiency in the development of CHDs

Approximately 30% of *Nipbl*$^{+/-}$ mice exhibit congenital heart defects (CHDs) (*10, 18*), a prevalence similar to that observed in individuals with CdLS; however, these defects in mice are primarily atrial septal defects. Additionally, *Nipbl*$^{+/-}$ mice demonstrate other cardiac abnormalities at earlier developmental stages. For instance, at embryonic day 13.5 (E13.5), 70% of *Nipbl*$^{+/-}$ mice show delays in ventricular septal fusion. Even earlier, at embryonic day 10.5 (E10.5), a striking 100% of *Nipbl*$^{+/-}$ mice display right ventricle hypoplasia.

A previous study investigated which specific cell lineages are necessary and sufficient for the development of CHDs in *Nipbl*-haploinsufficient mice, similar to those seen in mice completely deficient in *Nipbl* (*18*). The study utilized a *Nipbl* allelic series featuring a conditional/invertible gene trap and mouse lines expressing Cre recombinase, to create *Nipbl*-haploinsufficiency in neural crest cells (via *Wnt-Cre*) and in the cells of the cardiac crescent (via *Nkx2-5-Cre*). The results indicated that *Nipbl*-haploinsufficiency in neural crest cells alone did not lead to CHDs. However, inducing *Nipbl*-haploinsufficiency in the cardiac crescent cells resulted in CHDs. These cardiac crescent cells comprise both first heart field and second heart field (SHF) cells. Interestingly, the heart abnormalities observed in *Nipbl*$^{+/-}$ mice, which include atrial septal defects (ASDs), incomplete fusion of the ventricular septum, and right ventricle

hypoplasia, occur in tissues derived from the SHF. These findings collectively suggest that the SHF may be a source of CHDs in *Nipbl*$^{+/-}$ mice.

Since prior research suggested the SHF as a potential source of CHDs in *Nipbl*$^{+/-}$ mice, we used light sheet fluorescence microscopy (LSFM) to explore this hypothesis. We created *Nipbl*-haploinsufficiency in the SHF using two Cre recombinase-expressing mouse lines: *Mef2c-Cre* and *Isl1*$^{Cre/+}$. Interestingly, while *Mef2c-Cre*-mediated *Nipbl*-haploinsufficiency in the SHF did not lead to CHDs (Fig. 2.2, A and B), *Isl1*$^{Cre/+}$ did, resulting in a range of CHDs remarkably similar to those observed in *Nipbl*$^{+/-}$ hearts (Fig. 2.2, D and E). We attributed this difference to a significant reduction in *Isl1* expression in *Isl1*$^{Cre/+}$ hearts, which was over 60% lower compared to wild type (Fig. 2.4D). Furthermore, our results demonstrate that reduced *Isl1* expression alone can lead to the development of CHDs, particularly atrial septal defects (ASDs) (Fig. 2.4, A and B).

Given that *Isl1*-haploinsufficiency alone resulted in CHDs, we explored its impact on heart development within the context of global *Nipbl*-haploinsufficiency. To this end, we engineered mice doubly haploinsufficient for both genes *(Nipbl*$^{+/-}$*; Isl1*$^{+/-}$*)* (Fig. 2.5A). Double haploinsufficiency for *Isl1* and *Nipbl* led to more frequent and severe CHDs compared to mice haploinsufficient for either gene alone (Fig. 2.6B). These results underscore the notion that *Isl1*-haploinsufficiency exacerbates the frequency and severity of CHDs in the presence of global *Nipbl*-haploinsufficiency, highlighting a synergistic effect between these genetic factors in heart development.

In Figure 2.8, RNA sequencing of E10.5 hearts indicated that *Isl1* expression is downregulated in *Isl1*$^{+/-}$ hearts and upregulated in *Nipbl*$^{+/-}$ hearts. Intriguingly, in *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts, where both genes are hemizygous, *Isl1* expression reverted to WT levels (Fig. 2.8, D &

G). This suggests a regulatory adjustment in *Isl1* expression across different genetic backgrounds: it is increased in *Nipbl*$^{+/-}$ hearts above WT levels and adjusted to WT levels in *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts. These findings imply that the upregulation of *Isl1* in the heart may serve as a compensatory mechanism in response to *Nipbl*-haploinsufficiency. Additionally, the relative overexpression of *Isl1* in *Nipbl*$^{+/-}$ mice might be protective, preventing the development of CHDs to the extent seen in *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ mice, in which the incidence and severity of CHDs are greater.

In *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts, the expression levels of the genes *Pitx1*, *Pax6*, and *Hoxc4* were not upregulated to the same extent as in *Nipbl*$^{+/-}$ hearts. This difference in gene expression may contribute to the enhanced severity and frequency of CHDs observed in *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ mice compared to *Nipbl*$^{+/-}$ mice. Since *Isl1* is upregulated above WT levels in *Nipbl*$^{+/-}$ hearts but not in *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts, the inability of *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts to similarly upregulate *Isl1* (Fig. 2.8) may impede the upregulation of *Pitx1, Hoxc4*, and *Pax6*. This suggests a link between the level of *Isl1* expression and the regulatory control of these genes.

Isl1, Pitx1, Hoxc4, and *Pax6* are intriguing because they all encode homeodomain transcription factors that play integral roles in a variety of developmental processes. *Isl1* specifically plays a role in heart development (*308*). It is expressed in SHF cells that contribute to forming several cardiac structures, including parts of the right ventricle, outflow tract, and atria. Although *Pitx1, Hoxc4*, and *Pax6* are not typically associated with heart development, their roles in other tissues and organs are developmentally significant and similarly dosage-sensitive. *Pitx1* distinguishes hindlimb features from forelimbs (*309*) and plays a role in defining the left-right asymmetry of internal organs like the liver, lungs, and gut (*310*). *Hoxc4*, part of the *Hox* gene family, may define vertebral segments during embryonic development and influence the

development of the spinal cord and thoracic structures (*149*). *Pax6* is required for the development of various eye structures such as the lens, retina, and cornea (*162*), and it also helps define the regional identity of specific brain structures (*163*). These roles suggest that alterations in the expression of these genes could potentially influence heart development in a dosage-sensitive manner.

Progenitor cell misallocation in *Nipbl*$^{+/-}$ embryos foreshadows birth defects in CdLS

As stated earlier, 30% of *Nipbl*$^{+/-}$ mice display CHDs. These mice also demonstrate heart abnormalities at early stages of development: At embryonic day 13.5 (E13.5), 70% of *Nipbl*$^{+/-}$ mice exhibit delays in ventricular septal fusion, and even earlier, at E10.5, 100% showed right ventricle hypoplasia (*18*). In situ hybridization experiments revealed reduced expression of two transcription factors, *Nkx2-5* and *Mesp1,* which are essential for the growth and differentiation of early heart progenitor cells (*18*). These findings indicate that the causal events for CHDs in these mice might occur as early as gastrulation, the stage when the three primary germ layers form and the earliest progenitor cells of major tissues and organs begin to differentiate (*136*).

To elucidate developmental changes during gastrulation that could lead to birth defects in *Nipbl*$^{+/-}$ mice, we employed single-cell RNA sequencing (scRNAseq). This technique allowed us to analyze and compare the cellular compositions, lineage trajectories, and transcriptional landscapes of *Nipbl*$^{+/-}$ mouse embryos with their WT counterparts. We focused our analysis on two critical stages spanning the end of gastrulation: the late bud (LB) stage at embryonic day 7.5 (E7.5) and the cardiac crescent (CC) stage at embryonic day 7.75 (E7.75). This approach enabled a detailed examination of the early developmental differences potentially responsible for observed congenital heart defects.

In Figure 3.1, we demonstrated that *Nipbl*$^{+/-}$ embryos contain all the same cell populations as WT embryos. However, at LB-stage, *Nipbl*$^{+/-}$ embryos have a reduced total number of mesoderm cells. Specifically, within the mesoderm, there is a decrease in the number of primitive erythrocytes (PE) and first heart field (FHF) cells, coupled with an increase in paraxial mesoderm (PM) cells, as shown in Figure 3.2. Although these changes are not always substantial, they are likely significant in terms of physiological impact.

The reduction in the number of FHF cells may contribute to the occurrence of CHDs in *Nipbl*$^{+/-}$ mice and individuals with CdLS (*18, 79*). The majority of CHDs in *Nipbl*$^{+/-}$ mice, particularly those affecting the right ventricle and atrial/ventricular septa, are associated with regions typically derived from the SHF (*18*). While FHF cells do not directly give rise to SHF structures, they likely have indirect effects on the development of these areas. This idea supports the findings of Santos et al., 2016, who noted that complex interactions between various cell types – including those not typically involved in cardiac development – can significantly influence the incidence of CHDs in *Nipbl*$^{+/-}$ mice (*18*). This suggests a broader, systemic interaction in heart development, where changes in one cell population can impact the formation of critical cardiac structures.

The observed increase in PM cells at LB-stage is particularly noteworthy, as it corresponds with defects in PM derivatives, such as the axial skeleton and limb digits in *Nipbl*$^{+/-}$ mice, as illustrated in Figure 3.10 (*285*). These specific abnormalities are also characteristic features of CdLS (*60*). The mechanism by which an excess of PM cells contributes to limb defects is not immediately clear. However, these findings indicate that birth defects in *Nipbl*-deficient organisms may result not only from the absence of specific progenitor cell populations but also from the misallocation of progenitor cells, in such a way that disrupts the normal

balance of cell types. This alteration in cellular proportions could lead to developmental anomalies in structures derived from these cells.

Nipbl$^{+/-}$ embryos at the LB-stage did not exhibit significant changes in apoptotic activity, cell proliferation, developmental timing, or overall lineage trajectories, as shown in Figures S3.11, S3.12, 3.3, and 3.4. This observation suggests that these factors are not responsible for the observed disparities in the numbers of FHF and PM cells. Instead, our data indicate that in *Nipbl*$^{+/-}$ embryos, nascent mesoderm cells differentiated more frequently into PM rather than FHF cells, compared to WT counterparts (Figure 3.4, C to I). This shift suggests that misallocation of mesoderm cells in *Nipbl*$^{+/-}$ embryos results from changes in early cell fate decisions within the main pathways of lineage trajectories that are otherwise normal. These findings imply that certain types of structural birth defects, such as heart defects, may stem from subtle quantitative shifts in cell fate determination, rather than drastic disruptions of entire lineage pathways.

Nanog overexpression precedes birth defects in _Nipbl_$^{+/-}$ mice

In contrast to earlier studies on *Nipbl*$^{+/-}$ mice, where we typically observed modest changes in gene expression (less than 2-fold), our current research reveals that some genes in gastrulation stage embryos exhibit much larger shifts in expression, with increases as high as eight-fold, as shown in Figures 3.6D and 3.8H. Specifically, the gene *Nanog* was significantly overexpressed in LB-stage *Nipbl*$^{+/-}$ embryos, a trend that continued into CC-stage (Figs. 3.7 and 3.8). Utilizing data from Tiana et al., 2022, we found that many gene expression changes in *Nipbl*$^{+/-}$ embryos are similar to those observed in embryos with induced overexpression of *Nanog* (Fig. 3.8, A to E, and Fig. S3.15, A to E). Furthermore, many differentially expressed genes (DEGs) in *Nipbl*$^{+/-}$ embryos are likely direct targets of Nanog, given the enrichment of Nanog

binding sites in these genes (Figures 3.8, F and H). This suggests a significant regulatory influence of *Nanog* overexpression on developmental processes in *Nipbl*$^{+/-}$ embryos.

Impact of *Nanog* overexpression on developmental pathways in *Nipbl*$^{+/-}$ mice

Comparisons of our data with those of Tiana et al., 2022, indicate that *Nanog* overexpression leads to the downregulation of FHF drivers and upregulation of PM antidrivers (Figs. 3.3 and 3.8K). This dysregulation likely contributes to the misallocation of mesoderm cells in such a way that favors a PM fate at the expense of FHF. Furthermore, *Nanog* overexpression results in the underexpression of *Hox* genes, which are crucial for anterior-posterior patterning and axial skeleton development. In Figure 3.10, we documented the anteriorization of thoracic vertebrae in E18.5 *Nipbl*$^{+/-}$ mice, mirroring skeletal anomalies observed in individuals with CdLS, such as fused, absent, or misshapen ribs (*62*). Additionally, *Nanog* overexpression leads to a significant upregulation of genes in the Nodal signaling pathway across all germ layers. Nodal signaling is essential for left-right patterning and the proper positioning and morphogenesis of the musculoskeletal system and internal organs (*289*). In Figure 3.10, we reported abnormal growth of a 14th vertebra exhibiting left-right asymmetry in E18.5 *Nipbl*$^{+/-}$ embryos, with a complete 14th rib forming only on the right side. This pattern aligns with the increased occurrence of right-sided skeletal anomalies in CdLS (*60*). Taken together, these findings suggest that overexpression of *Nanog* and its downstream effects – downregulation of FHF drivers and PM antidrivers, underexpression of *Hox* genes, and upregulation of Nodal signaling genes – are precursor events leading to the birth defects observed in *Nipbl*$^{+/-}$ mice.

Can reduction of *Nanog* rescue the phenotypes observed in *Nipbl*$^{+/-}$ mice?

Since gene expression changes in LB-stage *Nipbl*$^{+/-}$ mice are associated with Nanog binding sites at their promoters and linked to *Nanog* overexpression, this prompted us to explore

whether reducing *Nanog* could mitigate the transcriptional and structural abnormalities observed in these mice. To this end, we utilized a mouse model where one *Nanog* allele was replaced with a GFP reporter (*Nanog* $^{GFP/+}$) (Fig. 4.1A), rendering these mice haploinsufficient for *Nanog* with reduced expression levels compared to WT mice. We additionally employed *Stra8-iCre;* *Nipbl*$^{Flox/+}$ male mice, which produce sperm that are either *Nipbl* $^{+}$ or *Nipbl* $^{-}$. Crossing these mice generates offspring that are either: WT (*Nipbl*$^{+/+}$*; Nanog*$^{+/+}$), *Nanog*-haploinsufficient (*Nanog*$^{+/-}$*)*, *Nipbl*-haploinsufficient (*Nipbl*$^{+/-}$), or doubly- haploinsufficiency for both genes (*Nipbl*$^{+/-}$*; Nanog*$^{+/-}$*)*. The latter genotype should not exhibit elevated *Nanog* levels as seen in mice only haploinsufficient for *Nipbl*.

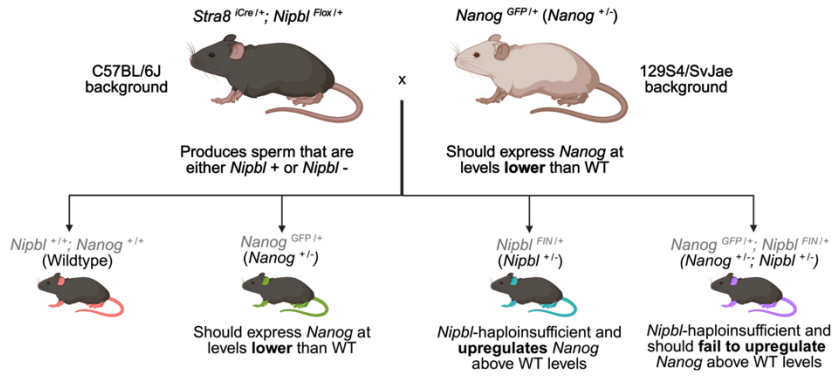
We crossed *Stra8-iCre; Nipbl*$^{Flox/+}$ males with *Nanog* $^{GFP/+}$ females, and analyzed 29 offspring across 8 litters and three embryonic stages (E18.5, E15.5, and E8.5) (Fig. 4.1B). Contrary to expectations, the resulting genotypic distribution did not follow Mendelian ratios: approximately 41% were WT, *Nanog*$^{+/-}$, 7% *Nipbl*$^{+/-}$, and only 3% were doubly haploinsufficient (*Nipbl*$^{+/-}$*; Nanog*$^{+/-}$). This presents a challenge in generating sufficient numbers of *Nipbl*$^{+/-}$ and *Nipbl*$^{+/-}$*; Nanog*$^{+/-}$ mice for a thorough rescue study, as only two *Nipbl*$^{+/-}$ and one *Nipbl*$^{+/-}$*;* *Nanog*$^{+/-}$ littermates were obtained out of 29 littermates.

The rarity of progeny carrying the mutant *Nipbl* alelle, despite successful generation in other crosses like *Stra8-iCre; Nipbl*$^{Flox/+}$ by *Isl1*$^{Cre/+}$ or *Nanog-Cre* by *Nipbl*$^{Flox/Flox}$ raises questions. One possibility is the impact of genetic background. Crosses like *Stra8-iCre;* *Nipbl*$^{Flox/+}$ by *Isl1*$^{Cre/+}$ or *Nanog-Cre* by *Nipbl*$^{Flox/Flox}$ only involved mice on a C57BL/6J background, whereas the *Stra8-iCre; Nipbl*$^{Flox/+}$ by *Nanog* $^{GFP/+}$ cross involves a *Nanog* $^{GFP/+}$ mouse that is on a 129S4/SvJae background. Therefore, the mutant Nipbl allele may have trouble surviving on a mixed C57BL/6J 129S4/SvJae background. To address this, transitioning the

Nanog $^{GFP/+}$ mice onto a C57BL/6J background could potentially improve the outcomes of this cross.

Another hypothesis for the low yield of *Nipbl*$^{+/-}$ and *Nipbl*$^{+/-}$*; Nanog*$^{+/-}$ mice might involve the interaction between sperm and eggs. If sperm carrying the mutant *Nipbl* alelle struggle to fertilize eggs carrying the mutant Nanog allele, an alternative approach could involve using male *Nanog* $^{GFP/+}$ mice to produce sperm carrying the mutant *Nanog* allele and female mice to produce eggs carrying the mutant *Nipbl* allele. This arrangement might enhance the likelihood of successful fertilization. Maybe sperm carrying the mutant *Nipbl* allele can fertilize eggs carrying the *Nipbl* mutant allele, but not the other way around.

A



B

	WT	*Nanog*$^{+/-}$	*Nipbl*$^{+/-}$	*Nipbl*$^{+/-}$; *Nanog*$^{+/-}$	Total
E18.5 (2 litters)	4	2	0	0	6
E15.5 (3 litters)	6	4	1	1	12
E8.5 (3 litters)	2	8	1	0	11
Total	12	14	2	1	29

Fig. 4.1. Mating *Stra8-iCre*; *Nipbl*$^{Flox/+}$ male mice with *Nanog* $^{GFP/+}$ female mice does not produce Mendelian genotypic frequencies. (A) *Stra8-iCre*; *Nipbl*$^{Flox/+}$ male mice make sperm that are either *Nipbl* $^{+}$ or *Nipbl* $^{-}$. Mating *Stra8-iCre*; *Nipbl*$^{Flox/+}$ male mice with *Nanog*$^{GFP/+}$ female mice produces littermates that are either WT (*Nipbl*$^{+/+}$; *Nanog*$^{+/+}$), *Nanog*-haploinsufficient (*Nanog*$^{+/-}$), *Nipbl*-haploinsufficient (*Nipbl*$^{+/-}$), or doubly- haploinsufficiency for both genes (*Nipbl*$^{+/-}$; *Nanog*$^{+/-}$). (B) Table of frequencies of genotypes obtained across different embryonic stages after crossing *Stra8-iCre*; *Nipbl*$^{Flox/+}$ male mice with *Nanog*$^{GFP/+}$ female mice. We did not detect *Nipbl*$^{+/-}$ or *Nipbl*$^{+/-}$; *Nanog*$^{+/-}$ embryos at Mendelian frequencies.

Nipbl's effects on chromatin topology and gene expression in *Nipbl*$^{+/-}$ embryos

The findings reported in Chapter 3, alongside previous research using mouse and zebrafish models, underscore the significance of chromosomal gene location in determining their sensitivity to fluctuations in *Nipbl* levels. For instance, in earlier work on *Nipbl*$^{+/-}$ embryonic brains, it was noted that the most pronounced gene expression changes within the 22-gene protocadherin beta (Pcdhb) cluster occurred at the genes located at the 5' and 3' ends (*10*), which are sites of CTCF binding (*291*). Similarly, in zebrafish with reduced *nipbl* expression, *hox* genes situated near the 3' end of the *hoxa*, *hoxc*, and *hoxd* clusters exhibited a position-dependent overexpression pattern in developing pectoral fins (*288*). In the current study, we observed that the anterior *Hoxb* genes, located at the 3' end of the *Hoxb* cluster, are consistently underexpressed in the mesoderm of *Nipbl*$^{+/-}$ embryos during gastrulation (Fig. 3.8E). These observations reinforce the strong impact of chromosomal positioning on gene sensitivity to *Nipbl* levels, an effect that is particularly pronounced within gene clusters and manifests as early as gastrulation.

Recent research indicates that the gene expression changes associated with *Nipbl* deficiency might stem from broad alterations in chromosomal structure and organization. For example, a significant reduction in *Nipbl* expression in mouse hepatocytes leads to a genome-wide decrease in topologically associated domains (TADs) and Hi-C interaction peaks (*292*), which are both dependent on chromatin looping mechanisms (*293*). In vitro experiments have shown that the formation and pace of chromatin loop extrusion are diminished when a CdLS pathogenic mutation is introduced into *NIPBL* (*294*). Chromatin looping within TADs is critical for bringing distant cis-regulatory elements like enhancers close to their target promoters. Intriguingly, *Nipbl* is known to bind preferentially to enhancers and promoters (*295*), and ChiP-

seq studies have revealed diminished enhancer-promoter interactions in *Nipbl*$^{+/-}$ mouse embryonic fibroblasts (*296*). Given that chromatin looping is crucial for gene activation, impaired loop extrusion due to reduced *Nipbl* levels could explain the predominance of gene expression downregulation observed in gastrula-stage mouse embryos (Fig. 3.6A).

Nipbl-haploinsufficiency results in reduced chromatin accessibility at enhancers and CTCF sites

To investigate the influence of *Nipbl* on chromatin topology, we performed single-nucleus RNA and ATAC sequencing (10X Genomics Single Cell Multiome ATAC + Gene Expression) on two WT and two *Nipbl*$^{+/-}$ mouse embryos at the late bud stage. ATAC sequencing assesses chromatin accessibility throughout the genome, facilitating the identification of DNA regions more open to transcription factors and regulatory proteins. By conducting ATAC sequencing alongside RNA sequencing, we can use the RNA libraries to categorize nuclei into biological groups corresponding to different germ layers. This would allow us to associate specific chromatin accessibility measures with particular germ layers. Although we initially performed multiome sequencing on two WT embryos, only one met our stringent quality control criteria for both their RNA and ATAC libraries. Therefore, we included only this one WT sample in the analyses discussed below.

From the single WT sample that met our quality control standards, we isolated 1,957 nuclei. In contrast, we captured a total of 7,280 nuclei from the two *Nipbl*$^{+/-}$ embryos. Across both WT and *Nipbl*$^{+/-}$ samples, the RNA libraries showed a median of 4,006 unique molecular identifiers (UMIs) per nucleus, which corresponded to a median of 2,367 genes per nucleus. For the ATAC libraries, the median capture was 863 fragments per nucleus, equating to 476 peaks per nucleus. We verified that nuclei from the *Nipbl*$^{+/-}$ embryos expressed Cre recombinase and

Bgeo, although these data are not shown here. Additionally, our analysis confirmed that the *Nipbl*$^{+/-}$ embryos exhibited a 50% reduction in *Nipbl* expression compared to the WT embryo (data not shown).

Initially, we clustered nuclei from the WT embryo using Signac (*311*), using genes with a residual variance in UMI count greater than 1 and peaks with a fragment count above the 75th percentile. This approach resulted in the WT nuclei forming only eight clusters (data not shown), which is significantly fewer than the number previously observed in LB-stage embryos necessary for accurate annotation into germ layers (Fig. 3.1 and Fig. 3.2). We speculate that Signac's limited ability to segregate nuclei into more defined clusters, as achieved using Seurat, may stem from the inclusion of peaks that lack high variability. To determine the germ layer identities of each nucleus, we projected the RNA libraries of the nuclei onto the single-cell RNA libraries of LB-stage WT embryos previously obtained in Fig. 3.1. Projection revealed that nuclei from both WT and *Nipbl*$^{+/-}$ embryos were represented in all germ layers (Fig. 4.2). Consequently, our multiome sequencing of LB-stage WT and *Nipbl*$^{+/-}$ embryos effectively captured nuclei from every germ layer, mirroring the results from single-cell RNA sequencing of embryos of the same stage and genotypes.

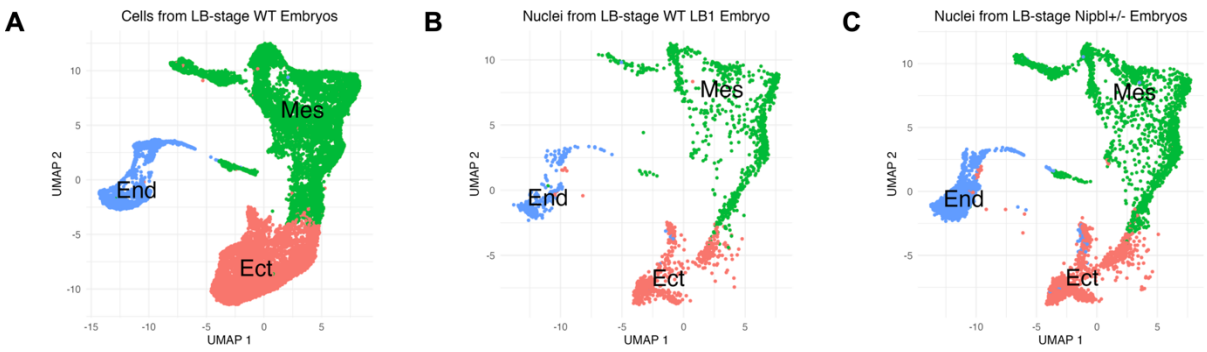


Fig. 4.2. Single cell multiome (RNA + ATAC) sequencing of LB-stage WT and *Nipbl*$^{+/-}$ embryos captured nuclei from all germ layers. (A) UMAP of germ layers detected in single cell RNA sequencing of LB-stage WT embryos. UMAP of germ layers detected in (B) one LB-stage WT embryo and (C) two LB-stage *Nipbl*$^{+/-}$ embryos.

In our analysis, we detected 69,553 chromatin accessibility peaks in the mesoderm of both WT and *Nipbl*$^{+/-}$ embryos, 60,486 in the ectoderm, and 47,056 in the endoderm. The smaller number of peaks in the endoderm likely reflects its smaller size in LB-stage embryos, as depicted in Fig. 4.1. Using ENCODE's Registry of Candidate cis-Regulatory Elements (cCREs) (*312*), we categorized these peaks as either intragenic or intergenic, and further annotated them as enhancers or CTCF sites. Across all germ layers, intragenic peaks comprised approximately 67-70% of the total, while intergenic peaks accounted for 30-33% (Table 4.1). Among these, 44-52% of all peaks were identified as intragenic enhancers, and 16% as intergenic enhancers. Intragenic CTCF sites represented 5-6% of all peaks, with intergenic CTCF sites making up 4-5%. These data suggest that the majority of accessible chromatin in LB-stage embryos corresponds to regions encoding genes and enhancers, with a relatively small proportion representing CTCF sites.

Table 4.1. Genes and enhancers make up the majority of accessible chromatin in LB-stage WT and *Nipbl*$^{+/-}$ embryos, with CTCF sites making up the minority. Frequency and percentage of peaks detected in germ layers of LB-stage WT and *Nipbl*$^{+/-}$ embryos following annotation using ENCODE's Registry of Candidate cis-Regulatory Elements (cCREs).

Germ Layer	Intragenic	Intergenic	Intragenic Enhancer	Intergenic Enhancer	Intragenic CTCF Site	Intergenic CTCF Site
Mesoderm 69,553	46,290 (67%)	23,263 (33%)	30,939 (44%)	10,909 (16%)	3,960 (6%)	3,376 (5%)
Ectoderm 60,486	40,787 (67%)	19,699 (33%)	28,396 (47%)	9,565 (16%)	3,398 (6%)	2,934 (5%)
Endoderm 47,056	33,149 (70%)	13,907 (30%)	24,660 (52%)	7,319 (16%)	2,150 (5%)	1,830 (4%)

To assess the differential accessibility of chromatin between WT and *Nipbl*$^{+/-}$ embryos, we plotted the average accessibility of each annotated peak across pseudobulked embryos (*Nipbl*$^{+/-}$ on the Y-axis and WT on the X-axis) and calculated the slope of the line of best fit. For peaks overlapping genes, the slope was close to 1 across all germ layers (Fig. 4.3A), indicating nearly identical accessibility between the two genotypes. Conversely, for peaks not overlapping genes, the slopes were lower: 0.61 in the mesoderm and 0.66 in the ectoderm (Fig. 4.3B), with a slope of 0.81 in the endoderm, where the y-intercept was below zero. This suggests that the average accessibility of non-gene-overlapping peaks is reduced in *Nipbl*$^{+/-}$ embryos.

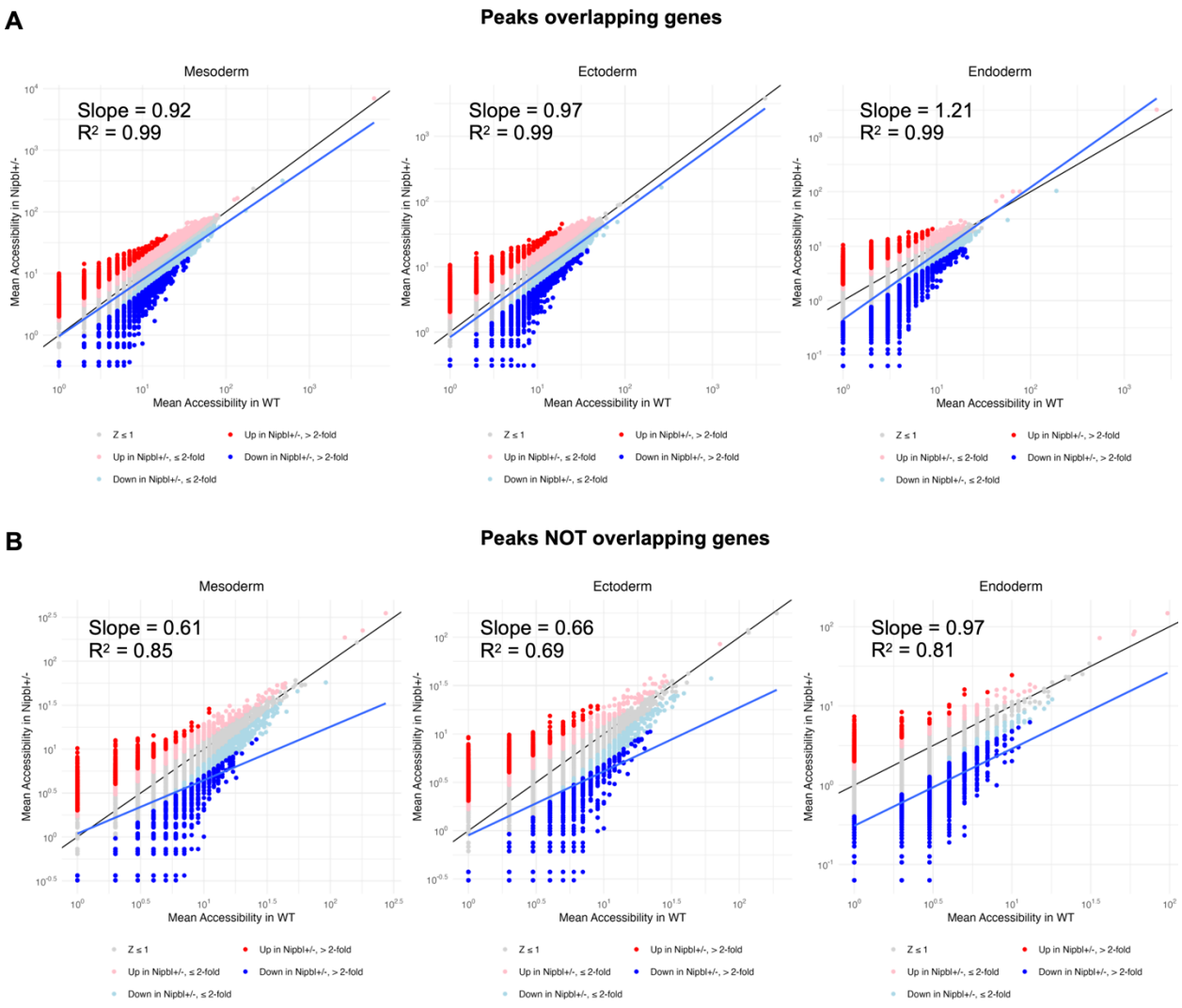


Fig. 4.3. LB-stage *Nipbl*$^{+/-}$ embryos show reduced accessibility of intergenic chromatin. (A) Average accessibility of peaks overlapping in germ layers of LB-stage *Nipbl*$^{+/-}$ embryos vs. one WT embryo. Peaks with a Z score less than 1 are colored grey and were considered to not be accessibly different between WT and *Nipbl*$^{+/-}$ embryos. Those colored shades of red or blue had a Z score greater than 1 and were considered to be accessibly different. Peaks colored pink or blue showed a less than 2-fold difference in increased or decreased accessibility among *Nipbl*$^{+/-}$ embryos. Peaks colored red or blue showed a more than 2-fold difference in increased or decreased accessibility among *Nipbl*$^{+/-}$ embryos. The black line represents a line with a slope of

1 and a y-intercept of zero. The blue line represents the line of best fit and the slope and R^2 values correspond to that line. (B) Average accessibility of peaks NOT overlapping in germ layers of LB-stage *Nipbl*$^{+/-}$ embryos vs. one WT embryo.

In the case of enhancers overlapping genes, no significant differences in accessibility were observed between WT and *Nipbl*$^{+/-}$ embryos (Fig. 4.4A). However, enhancers not overlapping genes showed reduced accessibility with slopes ranging from 0.5 to 0.76 across all germ layers (Fig. 4.4B). For CTCF sites, reduction in accessibility was the most pronounced: slopes ranged from 0.13 to 0.18 in the mesoderm and 0.24 in the ectoderm (Fig. 4.5), indicating the lowest accessibility among the categories discussed so far. In the endoderm, slopes ranged from 0.82 to 0.88, but with a y-intercept also less than zero. These findings highlight a substantial decrease in accessibility at intergenic enhancers and CTCF sites (whether intragenic or intergenic) in *Nipbl*$^{+/-}$ embryos, compared to other genomic regions.

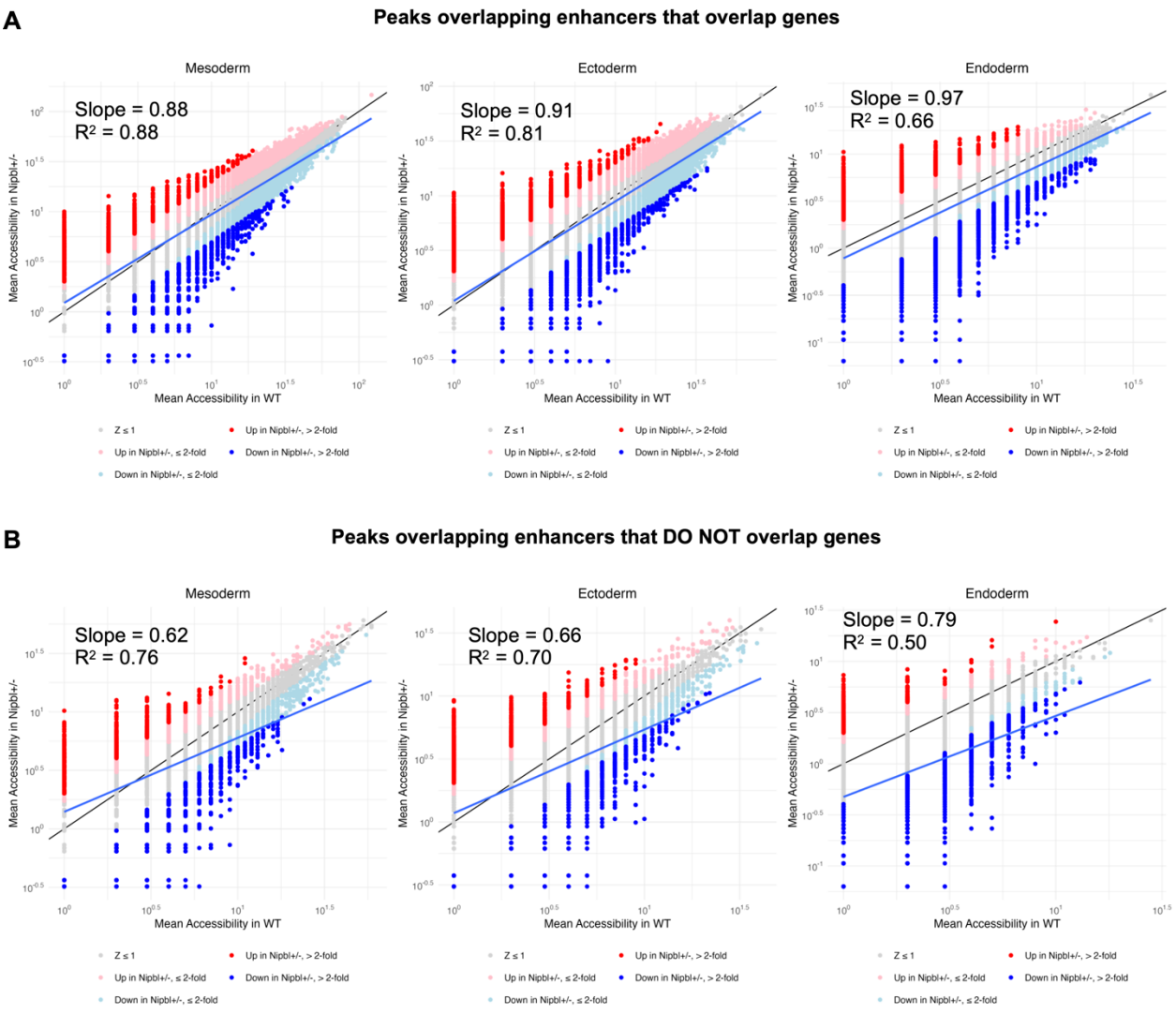


Fig. 4.4. LB-stage *Nipbl*$^{+/-}$ embryos show reduced accessibility of intergenic enhancers. (A) Average accessibility of peaks overlapping enhancers that overlap genes in germ layers of LB-stage *Nipbl*$^{+/-}$ embryos vs. one WT embryo. Peaks with a Z score less than 1 are colored grey and were considered to not be accessibly different between WT and *Nipbl*$^{+/-}$ embryos. Those colored shades of red or blue had a Z score greater than 1 and were considered to be accessibly different. Peaks colored pink or blue showed a less than 2-fold difference in increased or decreased accessibility among *Nipbl*$^{+/-}$ embryos. Peaks colored red or blue showed a more than 2-fold difference in increased or decreased accessibility among *Nipbl*$^{+/-}$ embryos. The black line

represents a line with a slope of 1 and a y-intercept of zero. The blue line represents the line of best fit and the slope and R^2 values correspond to that line. (B) Average accessibility of peaks overlapping enhancers that do NOT overlap genes in germ layers of LB-stage *Nipbl*$^{+/-}$ embryos vs. one WT embryo.

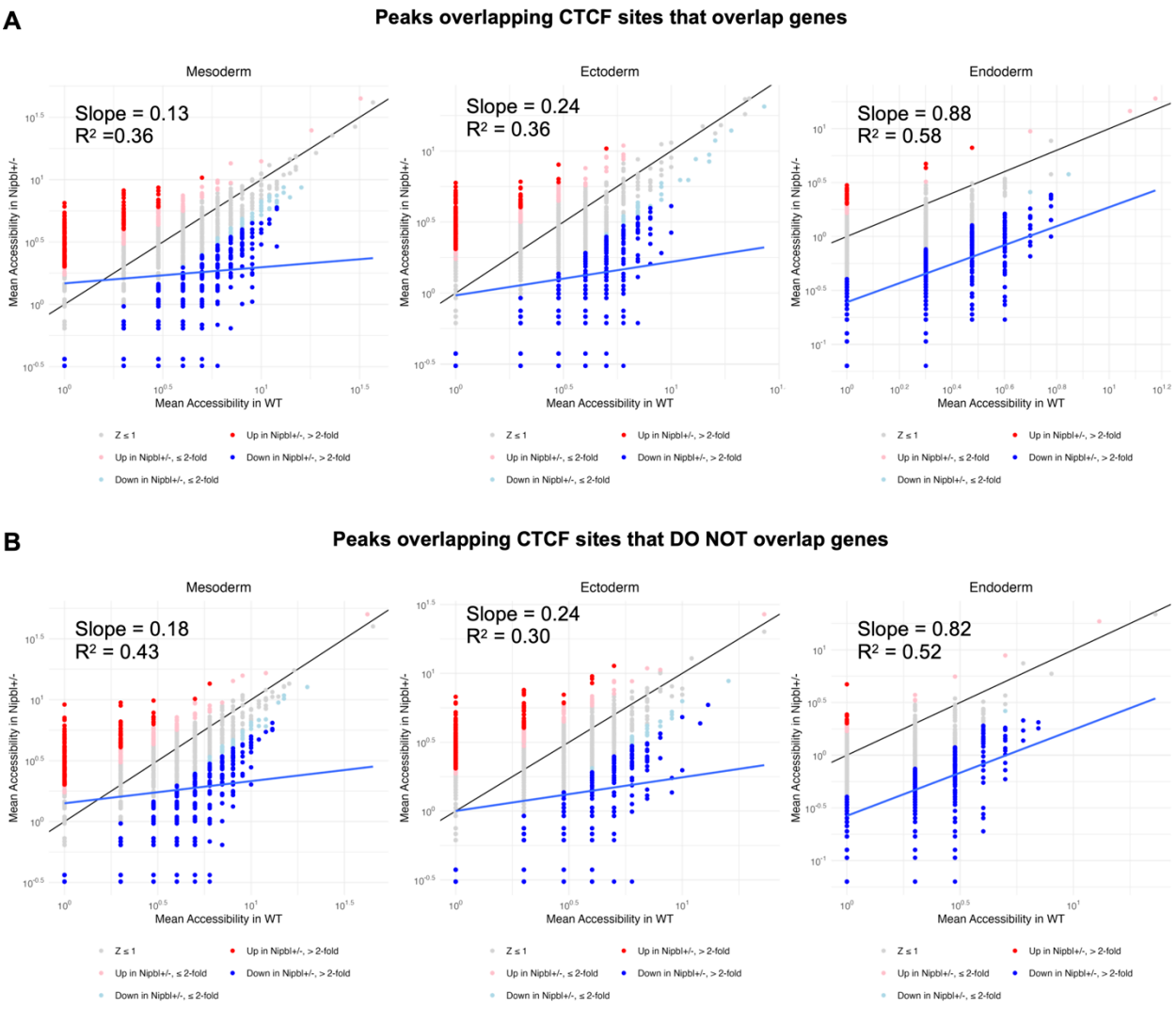


Fig. 4.5. LB-stage *Nipbl*$^{+/-}$ embryos show reduced accessibility of intragenic and intergenic CTCF sites. (A) Average accessibility of peaks overlapping CTCF sites that overlap genes in germ layers of LB-stage *Nipbl*$^{+/-}$ embryos vs. one WT embryo. Peaks with a Z score less than 1 are colored grey and were considered to not be accessibly different between WT and *Nipbl*$^{+/-}$ embryos. Those colored shades of red or blue had a Z score greater than 1 and were considered to be accessibly different. Peaks colored pink or blue showed a less than 2-fold difference in increased or decreased accessibility among *Nipbl*$^{+/-}$ embryos. Peaks colored red or blue showed a more than 2-fold difference in increased or decreased accessibility among *Nipbl*$^{+/-}$ embryos. The

black line represents a line with a slope of 1 and a y-intercept of zero. The blue line represents the line of best fit and the slope and R^2 values correspond to that line. (B) Average accessibility of peaks overlapping CTCF sites that do NOT overlap genes in germ layers of LB-stage *Nipbl*$^{+/-}$ embryos vs. one WT embryo.

Unchanged chromatin accessibility at gene loci in *Nipbl*$^{+/-}$ embryos suggests that the previously observed downregulation of gene expression in *Nipbl*$^{+/-}$ tissues is not due to decreased accessibility at the genes themselves. Instead, the diminished accessibility at intergenic enhancers and CTCF sites, both intragenic and intergenic, implies that the dysregulation of gene expression in *Nipbl*$^{+/-}$ mice likely stems from altered accessibility at these regulatory elements. The reduction in accessibility at enhancers is particularly noteworthy because *Nipbl* is known to bind preferentially to enhancers (*291*). Similarly, reduced accessibility at CTCF sites is significant because *Nipbl* haploinsufficiency has been shown to disproportionately affect the expression of genes proximal to CTCF sites compared to those more distal (*295*). Since CTCF sites generally function as insulators, restricting interactions between chromosomal elements within their bounds and external elements, decreased accessibility at these sites could lead to a breakdown in this insulation. This disruption may allow cis-regulatory elements, such as enhancers, to influence promoters outside their usual range of effect. Additionally, observed reduction in chromatin accessibility at enhancers and CTCF sites due to *Nipbl*-haploinsufficiency suggests that *Nipbl* plays a critical role in maintaining chromatin in an open state at these key regulatory sites.

Interestingly, the *Nanog* gene is situated between two CTCF sites (Fig. 4.6), one 12 kb downstream of Nanog and the other 47kb upstream of Nanog. Interestingly, *Nanog* is the only gene between these two CTCF sites and it is closer to the downstream CTCF site than the upstream site. Additionally, there is a distant upstream enhancer located near the upstream CTCF site of *Nanog*, which exhibits greater accessibility in *Nipbl*$^{+/-}$ embryos compared to the WT embryo. These findings suggest that the overexpression of *Nanog* observed in LB-stage *Nipbl*$^{+/-}$ embryos might be due to increased accessibility of this distal upstream enhancer and/or

decreased accessibility at the CTCF sites flanking the *Nanog* locus. The former scenario is straightforward: enhanced accessibility at an enhancer typically leads to increased gene expression. The latter scenario, however, is less straightfoward. Reduced accessibility at *Nanog's* CTCF sites could potentially disrupt their insulatory function, allowing cis-regulatory elements outside these boundaries to influence *Nanog* expression, thereby contributing to its overexpression.

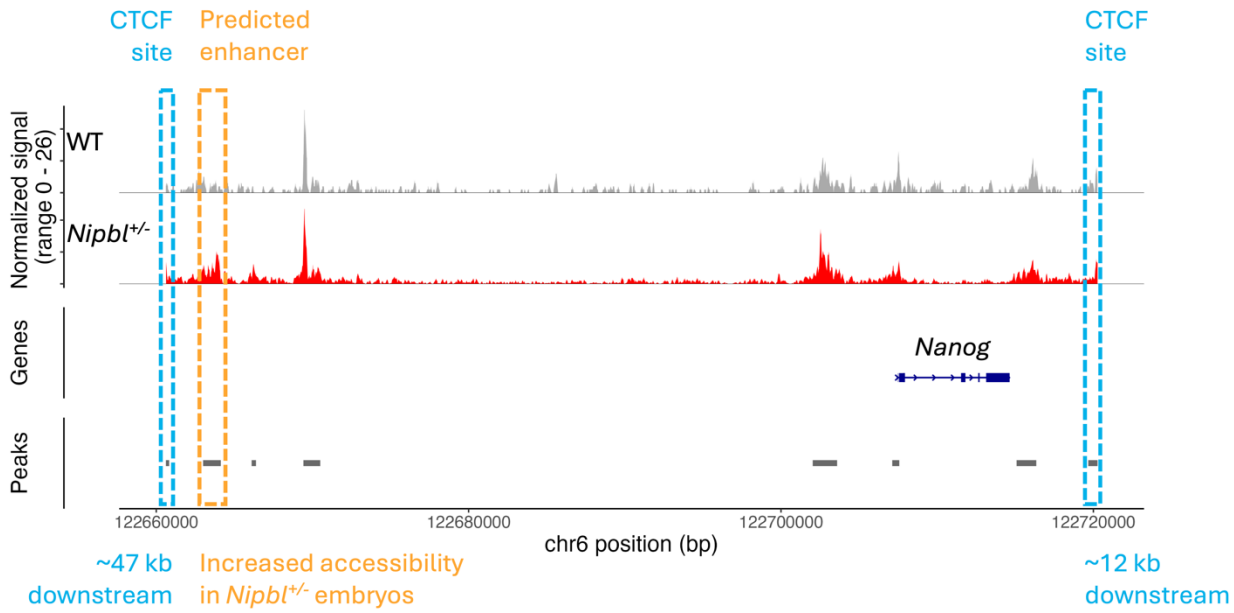


Fig. 4.6. The *Nanog* gene is located between two CTCF sites, with a predicted distal enhancer element that shows increased accessibility in LB-stage *Nipbl*$^{+/-}$ embryos. From top to bottom, this figure shows three tracks: at the top, DNA accessibility in the one WT embryo and the two *Nipbl*$^{+/-}$ embryos used in single nucleus multiome sequencing; in the middle, the location of the Nanog gene; and at the bottom, the locations chromatin accessibility peaks. CTCF sites and predicted enhancers as annotated by ENCODE's Registry of Candidate cis-Regulatory Elements (cCREs).

Given that our analysis included only one WT sample and two *Nipbl*$^{+/-}$ samples, determining the statistical significance of the observed reduced accessibility at enhancers and CTCF sites in *Nipbl*$^{+/-}$ embryos remains challenging. Nonetheless, the consistent reduction across all germ layers (Fig. 4.3 and Fig. 4.4) suggests that we may not need to conduct further multiome sequencing of WT and *Nipbl*$^{+/-}$ embryos to establish these differences, which is fortunate as multiome sequencing is costly. Instead, a more economical approach could involve conducting ATAC sequencing on whole embryos. The differences noted across the germ layers are likely to be detectable even when averaged across the entire embryo. To this end, we have recently carried out ATAC sequencing on 7 WT embryos and 7 *Nipbl*$^{+/-}$ embryos at LB-stage, and we are currently awaiting the return of the FASTQ files. The increased number of biological replicates for each genotype will help confirm whether there is a global reduction in chromatin accessibility at enhancers and CTCF sites in *Nipbl*$^{+/-}$ embryos and how these phenomena relate to the overexpression of *Nanog*.

Why do birth defects develop?

Globally, around 3% of newborns are affected by birth defects (*1*). These anomalies can impact various body parts and systems and are the leading cause of infant mortality in developed countries (*21*). Beyond the immediate health implications, birth defects pose enduring challenges for those who survive past infancy, often necessitating continuous medical care to address physical, developmental, or cognitive disabilities. The repercussions of birth defects extend beyond individual health, exerting substantial emotional and financial strains on families and placing significant demands on healthcare systems. This underscores the urgent need to develop therapies not only to alleviate the clinical severity of birth defects but also to prevent them from occurring.

Birth defects can arise from environmental exposures, infections during pregnancy, nutritional deficiencies, genetic factors, and combinations thereof. Although these identified causes account for only about 20-30% of all birth defects, leaving the origins of the majority (70-80%) unknown (*2*), ongoing research into syndromic birth defects has provided valuable insights. For instance, individuals with Cornelia de Lange Syndrome (CdLS) show multiple concurrent defects, often reflecting those seen in more common, isolated birth defects. Individuals with CdLS typically exhibit noticeable abnormalities in various body parts, especially the face and limbs, accompanied by intellectual disability. Most cases of CdLS are linked to mutations in genes that affect the cohesin complex, essential for the maintenance of chromatin structure. More specifically, however, most cases of CdLS are caused by haploinsufficiency of the *NIPBL* gene, which plays a role in loading the cohesin complex onto chromatin.

Birth defects can be caused by small changes in gene expression

Initially recognized for its role in sister chromatid cohesion, the cohesin complex is now also known to play a crucial role in chromatin looping, facilitating the interaction between distant enhancers and promoters to regulate global gene expression. Studies of cell lines from patients with the *Nipbl*-haploinsufficient form of CdLS have revealed that *Nipbl*-haploinsufficiency results in alterations in the expression of hundreds of genes. Similar patterns have been observed in animal models of *Nipbl*-haploinsufficiency, including *nipbl*-morphant zebrafish and $Nipbl^{+/-}$ mice. Although these changes in gene expression are statistically significant, they are generally small in magnitude. Consequently, it is unlikely that the alteration in any single gene's expression is responsible for the development of defects. Instead, it appears more probable that the cumulative effect of these minor changes contributes to the birth defects

seen in these models. These findings highlight a critical insight from studies of *Nipbl*-haploinsufficiency in CdLS and its animal models: birth defects can arise from the collective effect of small, yet statistically significant, alterations in gene expression.

Birth defects can be caused by deficiencies in tissues outside of the tissues of interest

The second insight from studies of animal models with *Nipbl*-haploinsufficiency in CdLS concerns the contribution of non-target tissues to birth defects. As shown in Fig. 2.2, using the $Isl1^{Cre/+}$ mouse line (but not the *Mef2c-Cre* line) to create *Nipbl*-haploinsufficiency in the second heart field (SHF) led to heart defects similar to those observed in mice with global *Nipbl*-haploinsufficiency. Further analysis, shown in Fig. 2.4, suggests that these heart defects could partly be due to reduced *Isl1* expression in $Isl1^{Cre/+}$ mice, which also exhibit heart defects without concurrent *Nipbl*-haploinsufficiency. However, we showed in Fig. 2.3 that *Isl1-Cre* l recombine extensively in non-cardiac tissues such as the hindlimbs, dorsal root ganglia, pharyngeal arches, trigeminal ganglia, and diencephalon. This widespread recombination suggests that the heart defects observed in $Isl1^{Cre/+}$ mice may not solely be due to changes within the SHF but could also result from deficiencies in other tissues affected by the *Isl1-Cre*-driven *Nipbl*-haploinsufficiency. These findings indicate that the origins of heart and other birth defects may involve multiple tissues that do not directly form the affected organ systems, highlighting a complex interdependence between different body tissues during development.

The causes of birth defects may occur earlier in development than expected

The third insight from studies using animal models of *Nipbl*-haploinsufficient CdLS is that birth defects may originate very early in development, possibly as early as during gastrulation. This crucial phase marks the establishment of germ layers and the emergence of the earliest progenitor cell populations, which eventually form the various structures of the body.

Supporting evidence from studies of congenital heart defects in *Nipbl*$^{+/-}$ mice shows that 30% exhibit congenital heart defects (CHDs), primarily atrial septal defects (*18*). Notably, at embryonic day 13.5 (E13.5), 70% of *Nipbl*$^{+/-}$ mice demonstrate delays in ventricular septal fusion, and by embryonic day 10.5 (E10.5), all *Nipbl*$^{+/-}$ mice exhibit right ventricle hypoplasia (*18*). Earlier still, *Nipbl*$^{+/-}$ mice show reduced levels of crucial transcription factors, namely *Nkx2-5* in the second heart field by embryonic day 7.75 (E7.75), and *Mesp1* in mesoderm progenitors by E7.5 (*18*). Similar early developmental disruptions are evident in *nipbl*-morphant zebrafish, where defects in cardiogenic mesoderm migration are detectable just 18 hours post-fertilization (*11*). These findings underscore that birth defects can arise during the formative stages of tissue lineage establishment, highlighting the importance of early developmental processes in congenital anomalies.

Developmental delay may not significantly contribute to the development of birth defects

Because the anomalies observed in CdLS and animal models of CdLS often resemble premature forms of the final bodily structures, it has been hypothesized that overall developmental delay might significantly contribute to birth defects in such genetic syndromes. However, recent studies of *Nipbl*$^{+/-}$ mice offer a new perspective, suggesting that developmental delay may not be a primary factor in birth defects. Single cell RNA sequencing and pseudotemporal ordering of early *Nipbl*$^{+/-}$ mouse embryos during the early bud, late bud, and early head fold stages, as shown in Fig. 3.3, reveal that the developmental timing of *Nipbl*$^{+/-}$ germ layers aligns with that of their wildtype counterparts. This consistency suggests that developmental delays are unlikely to significantly influence the onset of birth defects, even during gastrulation, a critical period in embryonic development. Consequently, these findings

prompt a reevaluation of other potential mechanisms that could explain the origins of birth defects.

Birth defects can be caused by misallocation of progenitor cell populations

The fifth insight from studies on *Nipbl*$^{+/-}$ mice underscores that birth defects may arise from the misallocation of early progenitor cell populations. Single-cell RNA sequencing of the mesoderm in gastrulation-stage *Nipbl*$^{+/-}$ mice has provided compelling evidence for this. The mesoderm predominantly forms the developing heart in mice. Two of the earliest cell populations to give rise to the heart include the first heart field and second heart field. As depicted in Fig. 3.3, LB-stage *Nipbl*$^{+/-}$ embryos show a reduced total number of mesoderm cells, specifically fewer primitive erythrocytes and first heart field cells, with a concomitant increase in paraxial mesoderm cells.

Reduction in primitive erythrocytes aligns with the blood formation and circulation defects observed in *nipbl*-morphant zebrafish (*11*) and the thrombocytopenia seen in CdLS (*89*). Similarly, the decreased number of first heart field cells may also underlie heart defects in these mice, given their role in early heart development. The increase in paraxial mesoderm cells at the LB-stage is also particularly noteworthy, as it correlates with defects in derivatives of this cell type, such as the axial skeleton and limb digits, as shown in Fig. 3.10. The mechanisms, however, behind how an excess of paraxial mesoderm cells leads to limb defects remain unclear.

Further analysis revealed that these misallocations are not attributable to changes in apoptotic activity, cell proliferation, developmental timing, or overall lineage trajectories. Instead, as shown in Fig. 3.4, the misallocation is driven by shifts in early cell fate decisions, with nascent mesoderm cells in *Nipbl*$^{+/-}$ embryos more frequently differentiating into paraxial mesoderm rather than first heart field, compared to wild-type. This alteration is further supported

by the underexpression of genes essential for first heart field differentiation, detailed in Fig. 3.5. These findings suggest that birth defects in *Nipbl*$^{+/-}$ mice might stem from the dysregulation of key genes governing cell fate, leading to an abnormal distribution of cells that ultimately impacts the development of specific tissues and organs.

Birth defects can be caused by large changes in gene expression

Sixth, our recent findings from studies of *Nipbl*$^{+/-}$ mice at the gastrulation stage, as shown in Figures 6 and 8, contrast with previous studies where gene expression differences were typically minimal (less than 2-fold) (*10*). Here, we observed large shifts in gene expression, with changes reaching up to eight-fold. Notably, these changes included the overexpression of Nodal signaling genes and a pluripotency transcription factor, *Nanog*, alongside the underexpression of *Hox* genes. The overexpression of Nodal signaling is particularly significant due to its critical role in left-right body patterning (*289*), which influences the proper positioning and morphogenesis of the musculoskeletal system and internal organs (*290*). This overexpression correlates with the right ventricle hypoplasia seen in *Nipbl*$^{+/-}$ mice, intestinal malrotation in nipbl-morphant zebrafish, and CdLS (*10, 17, 18*), as well as with a higher incidence of more severe right-sided skeletal anomalies in CdLS (*60*). Conversely, the underexpression of Hox genes, which are essential for anterior-posterior patterning, is linked to the anteriorization of the thoracic vertebrae in these embryos, observed at a frequency of about 58% by E18.5. Meanwhile, the overexpression of *Nanog*, which typically ceases after gastrulation, further complicates the gene expression profile because, as we demonstrated in Fig. 8, overexpression of Nodal signaling and underexpression of *Hox* genes may be due to the overexpression of *Nanog*. Collectively, these substantial changes in the expression of key developmental genes suggest that birth defects

in *Nipbl*$^{+/-}$ mice can be precipitated by significant alterations in the expression of genes that govern crucial developmental processes.

Birth defects can be caused by changes in chromatin topology

Finally, the seventh insight from studies of *Nipbl*$^{+/-}$ mice highlights that birth defects may result from global changes in chromatin topology. Multiome sequencing of late bud stage WT and *Nipbl*$^{+/-}$ embryos has brought some illumination to this phenomenon. In Figures 4.3 and 4.4, we demonstrate that *Nipbl*$^{+/-}$ embryos exhibit reduced chromatin accessibility, particularly at intergenic chromatin, including intergenic enhancer sites. Furthermore, in Figure 4.5, we show that *Nipbl*$^{+/-}$ embryos exhibit diminished accessibility at CTCF binding sites, regardless of whether they are intragenic or intergenic, which play roles in organizing chromatin into topologically associating domains (TADs). These TADs facilitate interactions between enhancers and promoters within the same domain while preventing interactions with elements outside the domain, primarily through chromatin looping mediated by the cohesin complex and *Nipbl*.

The most significant reductions in accessibility were observed at CTCF sites, suggesting a disruption of TAD structures in *Nipbl*$^{+/-}$ embryos. Such disruption has been correlated with global reductions in TADs, as evidenced by Hi-C studies in mouse hepatocytes exhibiting *Nipbl*-haploinsufficiency (*292*). This disruption likely leads to aberrant interactions between promoters and distant cis-regulatory elements, which could explain the extensive gene expression changes observed in these models, particularly affecting genes associated with Nodal signaling, *Hox* genes, and *Nanog*. These genes are typically tightly regulated by enhancers and silencers confined within specific TADs, and their misregulation may directly contribute to the developmental anomalies seen in *Nipbl*$^{+/-}$ mice.

Why do birth defects develop? Drawing on evidence from studies of *Nipbl*-deficient animal models of Cornelia de Lange Syndrome (CdLS), particularly the *Nipbl*$^{+/-}$ mouse, we have identified several key insights into the causal origins of birth defects. Firstly, birth defects can result from small but statistically significant changes in the expression of multiple genes. Secondly, they can arise from deficiencies in tissues not directly involved in the affected structures. Thirdly, the underlying causes of birth defects may manifest very early in development, sooner than previously recognized. Fourthly, developmental delay may not significantly contribute to the development of birth defects. Fifthly, misallocation of progenitor cell populations can lead to structural anomalies. Sixthly, large shifts in gene expression (greater than two-fold) precede and likely cause birth defects. Seventhly, and perhaps most critically, changes in chromatin topology may underpin these gene expression differences, suggesting a fundamental genomic disorganization in *Nipbl*$^{+/-}$ tissues.

This last point is particularly important as it potentially explains the mechanism behind the diverse gene expression anomalies observed in these models. Future research should therefore concentrate on the impact of *Nipbl* haploinsufficiency on chromatin topology. While preliminary data from chromatin accessibility studies reported here have been informative, advancing our understanding will require employing techniques such as chromatin conformation capture to more precisely delineate the differences in topologically associating domains and the interactions between enhancers and promoters. This will further clarify how chromatin structure contributes the gene expression changes observed *Nipbl*$^{+/-}$ mice and the development of birth defects in CdLS.

REFERENCES

1. S. E. Parker *et al.*, Updated National Birth Prevalence estimates for selected birth defects in the United States, 2004-2006. *Birth Defects Res A Clin Mol Teratol* **88**, 1008-1016 (2010).

2. M. L. Feldkamp, J. C. Carey, J. L. B. Byrne, S. Krikov, L. D. Botto, Etiology and clinical presentation of birth defects: population based study. *BMJ* **357**, j2249 (2017).

3. A. de Knecht-van Eekelen, R. C. Hennekam, Historical study: Cornelia C. de Lange (1871-1950)--a pioneer in clinical genetics. *Am J Med Genet* **52**, 257-266 (1994).

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

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

6. I. Parenti *et al.*, MAU2 and NIPBL Variants Impair the Heterodimerization of the Cohesin Loader Subunits and Cause Cornelia de Lange Syndrome. *Cell Rep* **31**, 107647 (2020).

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

8. D. Donze, C. R. Adams, J. Rine, R. T. Kamakaka, The boundaries of the silenced HMR domain in Saccharomyces cerevisiae. *Genes Dev* **13**, 698-708 (1999).

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

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

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

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

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

14. J. I. Hoffman, S. Kaplan, The incidence of congenital heart disease. *J Am Coll Cardiol* **39**, 1890-1900 (2002).

15. M. E. Oster *et al.*, Temporal trends in survival among infants with critical congenital heart defects. *Pediatrics* **131**, e1502-1508 (2013).

16. A. A. Richards, V. Garg, Genetics of congenital heart disease. *Curr Cardiol Rev* **6**, 91-97 (2010).

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

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

19. L. D. Botto, A. Correa, J. D. Erickson, Racial and temporal variations in the prevalence of heart defects. *Pediatrics* **107**, E32 (2001).

20. R. G. Kelly, M. E. Buckingham, A. F. Moorman, Heart fields and cardiac morphogenesis. *Cold Spring Harb Perspect Med* **4**, (2014).

21. M. B. Forrester, R. D. Merz, First-year mortality rates for selected birth defects, Hawaii, 1986-1999. *Am J Med Genet A* **119A**, 311-318 (2003).

22. Y. Wang, J. Hu, C. M. Druschel, R. S. Kirby, Twenty-five-year survival of children with birth defects in New York State: a population-based study. *Birth Defects Res A Clin Mol Teratol* **91**, 995-1003 (2011).

23. E. A. Miller *et al.*, Maternal exposure to tobacco smoke, alcohol and caffeine, and risk of anorectal atresia: National Birth Defects Prevention Study 1997-2003. *Paediatr Perinat Epidemiol* **23**, 9-17 (2009).

24. S. Broor, A. Kapil, J. Kishore, P. Seth, Prevalence of rubella virus and cytomegalovirus infections in suspected cases of congenital infections. *Indian J Pediatr* **58**, 75-78 (1991).

25. J. D. van Gool, H. Hirche, H. Lax, L. De Schaepdrijver, Folic acid and primary prevention of neural tube defects: A review. *Reprod Toxicol* **80**, 73-84 (2018).

26. J. O. Szot *et al.*, Bi-allelic Mutations in NADSYN1 Cause Multiple Organ Defects and Expand the Genotypic Spectrum of Congenital NAD Deficiency Disorders. *Am J Hum Genet* **106**, 129-136 (2020).

27. C. A. Hobbs *et al.*, Genetic epidemiology and nonsyndromic structural birth defects: from candidate genes to epigenetics. *JAMA Pediatr* **168**, 371-377 (2014).

28. H. Zhu, S. Kartiko, R. H. Finnell, Importance of gene-environment interactions in the etiology of selected birth defects. *Clin Genet* **75**, 409-423 (2009).

29. D. Patterson, Molecular genetic analysis of Down syndrome. *Hum Genet* **126**, 195-214 (2009).

30. C. H. Gravholt, M. H. Viuff, S. Brun, K. Stochholm, N. H. Andersen, Turner syndrome: mechanisms and management. *Nat Rev Endocrinol* **15**, 601-614 (2019).

31. E. Morice, [Relevance of animal models in the study of human pathologies: a mouse model of Down syndrome]. *Biol Aujourdhui* **204**, 3-8 (2010).

32. F. J. Probst, M. L. Cooper, S. W. Cheung, M. J. Justice, Genotype, phenotype, and karyotype correlation in the XO mouse model of Turner Syndrome. *J Hered* **99**, 512-517 (2008).

33. G. J. Kurlander, W. DeMyer, Roentogenology of the Brachmann-de Lange syndrome. *Radiology* **88**, 101-110 (1967).

34. A. Leavitt, N. Dinno, C. Davis, Cornelia de Lange syndrome in a mother and daughter. *Clin Genet* **28**, 157-161 (1985).

35. J. Pie *et al.*, Mutations and variants in the cohesion factor genes NIPBL, SMC1A, and SMC3 in a cohort of 30 unrelated patients with Cornelia de Lange syndrome. *Am J Med Genet A* **152A**, 924-929 (2010).

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

37. T. Kinjo *et al.*, A Case of Cornelia de Lange Syndrome: Difficulty in Prenatal Diagnosis. *Case Rep Obstet Gynecol* **2019**, 4530491 (2019).

38. A. D. Kline, M. Barr, L. G. Jackson, Growth manifestations in the Brachmann-de Lange syndrome. *Am J Med Genet* **47**, 1042-1049 (1993).

39. G. Boog, F. Sagot, N. Winer, A. David, M. F. Nomballais, Brachmann-de Lange syndrome: a cause of early symmetric fetal growth delay. *Eur J Obstet Gynecol Reprod Biol* **85**, 173-177 (1999).

40. S. Dorval *et al.*, A novel RAD21 mutation in a boy with mild Cornelia de Lange presentation: Further delineation of the phenotype. *Eur J Med Genet* **63**, 103620 (2020).

41. M. Cates, D. F. Billmire, M. J. Bull, J. L. Grosfeld, Gastroesophageal dysfunction in Cornelia de Lange syndrome. *J Pediatr Surg* **24**, 248-250 (1989).

42. B. Parma *et al.*, Complex nutritional deficiencies in a large cohort of Italian patients with Cornelia de Lange syndrome spectrum. *Am J Med Genet A* **182**, 2094-2101 (2020).

43. A. Cereda *et al.*, A new prognostic index of severity of intellectual disabilities in Cornelia de Lange syndrome. *Am J Med Genet C Semin Med Genet* **172**, 179-189 (2016).

44. H. M. Saal, C. A. Samango-Sprouse, L. A. Rodnan, K. N. Rosenbaum, D. A. Custer, Brachmann-de Lange syndrome with normal IQ. *Am J Med Genet* **47**, 995-998 (1993).

45. C. Gervasini *et al.*, Molecular characterization of a mosaic NIPBL deletion in a Cornelia de Lange patient with severe phenotype. *Eur J Med Genet* **56**, 138-143 (2013).

46. P. F. Ajmone *et al.*, Communication, cognitive development and behavior in children with Cornelia de Lange Syndrome (CdLS): preliminary results. *Am J Med Genet B Neuropsychiatr Genet* **165B**, 223-229 (2014).

47. E. Pearson *et al.*, Low speech rate but high gesture rate during conversational interaction in people with Cornelia de Lange syndrome. *J Intellect Disabil Res* **65**, 601-607 (2021).

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

49. K. Sarimski, Communication, social-emotional development and parenting stress in Cornelia-de-Lange syndrome. *J Intellect Disabil Res* **41 (Pt 1)**, 70-75 (1997).

50. M. Heald, D. Adams, C. Oliver, Profiles of atypical sensory processing in Angelman, Cornelia de Lange and Fragile X syndromes. *J Intellect Disabil Res* **64**, 117-130 (2020).

51. J. Moss, C. Oliver, L. Nelson, C. Richards, S. Hall, Delineating the profile of autism spectrum disorder characteristics in Cornelia de Lange and Fragile X syndromes. *Am J Intellect Dev Disabil* **118**, 55-73 (2013).

52. T. P. Berney, M. Ireland, J. Burn, Behavioural phenotype of Cornelia de Lange syndrome. *Arch Dis Child* **81**, 333-336 (1999).

53. C. Oliver, J. Sloneem, S. Hall, K. Arron, Self-injurious behaviour in Cornelia de Lange syndrome: 1. Prevalence and phenomenology. *J Intellect Disabil Res* **53**, 575-589 (2009).

54. S. Srivastava *et al.*, Repetitive and Self-injurious Behaviors in Children with Cornelia de Lange Syndrome. *J Autism Dev Disord* **51**, 1748-1758 (2021).

55. A. Verrotti *et al.*, Epilepsy in patients with Cornelia de Lange syndrome: a clinical series. *Seizure* **22**, 356-359 (2013).

56. M. T. Whitehead, U. D. Nagaraj, P. L. Pearl, Neuroimaging features of Cornelia de Lange syndrome. *Pediatr Radiol* **45**, 1198-1205 (2015).

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

58. M. J. Pablo *et al.*, High rate of autonomic neuropathy in Cornelia de Lange Syndrome. *Orphanet J Rare Dis* **16**, 458 (2021).

59. M. A. Grados, M. H. Alvi, S. Srivastava, Behavioral and psychiatric manifestations in Cornelia de Lange syndrome. *Curr Opin Psychiatry* **30**, 92-96 (2017).

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

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

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

63. A. M. Elliott, L. Kibria, M. H. Reed, The developmental spectrum of proximal radioulnar synostosis. *Skeletal Radiol* **39**, 49-54 (2010).

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

65. S. Luzzani, F. Macchini, A. Valade, D. Milani, A. Selicorni, Gastroesophageal reflux and Cornelia de Lange syndrome: typical and atypical symptoms. *Am J Med Genet A* **119A**, 283-287 (2003).

66. K. Husain, P. Fitzgerald, G. Lau, Cecal volvulus in the Cornelia de Lange syndrome. *J Pediatr Surg* **29**, 1245-1247 (1994).

67. I. T. Kim, J. W. Park, W. C. Choi, A Korean case of Cornelia de Lange syndrome. *Korean J Ophthalmol* **19**, 153-155 (2005).

68. A. V. Levin, D. J. Seidman, L. B. Nelson, L. G. Jackson, Ophthalmologic findings in the Cornelia de Lange syndrome. *J Pediatr Ophthalmol Strabismus* **27**, 94-102 (1990).

69. T. Wygnanski-Jaffe *et al.*, Ophthalmologic findings in the Cornelia de Lange Syndrome. *J AAPOS* **9**, 407-415 (2005).

70. H. A. Marres, C. W. Cremers, P. H. Jongbloet, Hearing levels in the Cornelia de Lange syndrome. A report of seven cases. *Int J Pediatr Otorhinolaryngol* **18**, 31-37 (1989).

71. K. C. Janek *et al.*, Improvement in hearing loss over time in Cornelia de Lange syndrome. *Int J Pediatr Otorhinolaryngol* **87**, 203-207 (2016).

72. P. Marchisio *et al.*, Otitis media with effusion and hearing loss in children with Cornelia de Lange syndrome. *Am J Med Genet A* **146A**, 426-432 (2008).

73. J. Hamilton, W. A. Clement, H. Kubba, Otolaryngological presentations of Cornelia de Lange syndrome. *Int J Pediatr Otorhinolaryngol* **78**, 1548-1550 (2014).

74. M. T. Goodban, Survey of speech and language skills with prognostic indicators in 116 patients with Cornelia de Lange syndrome. *Am J Med Genet* **47**, 1059-1063 (1993).

75. A. S. Toker, S. Ay, H. Yeler, I. Sezgin, Dental findings in Cornelia de Lange syndrome. *Yonsei Med J* **50**, 289-292 (2009).

76. T. Hirai *et al.*, [Anesthetic management of an infant with Cornelia de Lange syndrome]. *Masui* **55**, 454-456 (2006).

77. C. Oliver *et al.*, Cornelia de Lange syndrome: extending the physical and psychological phenotype. *Am J Med Genet A* **152A**, 1127-1135 (2010).

78. Y. Shen, D. Zhao, L. Sun, X. Yang, X. Yan, Congenital vaginal obstruction in a female with Cornelia de Lange syndrome: A case report. *Front Endocrinol (Lausanne)* **13**, 886235 (2022).

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

80. A. Nakajima, A. Ohshima, H. Fukayama, T. Kinoshita, Perioperative Management of a Patient With Cornelia de Lange Syndrome and Tetralogy of Fallot. *Anesth Prog* **66**, 159-161 (2019).

81. A. Selicorni *et al.*, Analysis of congenital heart defects in 87 consecutive patients with Brachmann-de Lange syndrome. *Am J Med Genet A* **149A**, 1268-1272 (2009).

82. S. Yurttutan *et al.*, Cornelia de Lange syndrome in a preterm infant presenting with coarctation of aorta and perimembranous ventricular septal defect. *Genet Couns* **22**, 435-438 (2011).

83. G. Y. Oudit, C. M. Chow, W. J. Cantor, Calcific bicuspid aortic valve disease in a patient with Cornelia de Lange syndrome: linking altered Notch signaling to aortic valve disease. *Cardiovasc Pathol* **15**, 165-167 (2006).

84. S. Das, K. Irpachi, R. Kalra, B. Airan, Aortopulmonary window and double outlet right ventricle: a rare combination. *Ann Card Anaesth* **17**, 245-246 (2014).

85. S. Jyonouchi, J. Orange, K. E. Sullivan, I. Krantz, M. Deardorff, Immunologic features of Cornelia de Lange syndrome. *Pediatrics* **132**, e484-489 (2013).

86. I. C. Verma, Cornelia de lange syndrome. Report of a case. *Indian J Pediatr* **37**, 152-154 (1970).

87. B. Ingram, E. A. Frost, Post-operative respiratory complications after palatoplasty in a 19-month-old female with Cornelia de Lange syndrome--a case report. *Middle East J Anaesthesiol* **21**, 419-421 (2011).

88. V. Cavalleri *et al.*, Thrombocytopenia and Cornelia de Lange syndrome: Still an enigma? *Am J Med Genet A* **170A**, 130-134 (2016).

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

90. H. Uzun, D. A. Senses, M. Uluba, K. Kocabay, A newborn with Cornelia de Lange syndrome: a case report. *Cases J* **1**, 329 (2008).

91. D. H. Nicholson, M. F. Goldberg, Ocular abnormalities in the de Lange syndrome. *Arch Ophthalmol* **76**, 214-220 (1966).

92. C. C. Huang, I. Emanuel, S. W. Huang, T. Y. Chen, Two cases of the de Lange syndrome in Chinese infants. *J Pediatr* **71**, 251-254 (1967).

93. B. Beck, Familial occurrence of Cornelia de Lange's syndrome. *Acta Paediatr Scand* **63**, 225-231 (1974).

94. D. Kumar, C. E. Blank, B. L. Griffiths, Cornelia de Lange syndrome in several members of the same family. *J Med Genet* **22**, 296-300 (1985).

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

96. F. J. Kaiser *et al.*, Loss-of-function HDAC8 mutations cause a phenotypic spectrum of Cornelia de Lange syndrome-like features, ocular hypertelorism, large fontanelle and X-linked inheritance. *Hum Mol Genet* **23**, 2888-2900 (2014).

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

98. T. Dasgupta, J. Antony, A. W. Braithwaite, J. A. Horsfield, HDAC8 Inhibition Blocks SMC3 Deacetylation and Delays Cell Cycle Progression without Affecting Cohesin-dependent Transcription in MCF7 Cancer Cells. *J Biol Chem* **291**, 12761-12770 (2016).

99. C. H. Haering, A. M. Farcas, P. Arumugam, J. Metson, K. Nasmyth, The cohesin ring concatenates sister DNA molecules. *Nature* **454**, 297-301 (2008).

100. S. Gruber, C. H. Haering, K. Nasmyth, Chromosomal cohesin forms a ring. *Cell* **112**, 765-777 (2003).

101. A. Losada, T. Yokochi, R. Kobayashi, T. Hirano, Identification and characterization of SA/Scc3p subunits in the Xenopus and human cohesin complexes. *J Cell Biol* **150**, 405-416 (2000).

102. S. Remeseiro, A. Cuadrado, G. Gomez-Lopez, D. G. Pisano, A. Losada, A unique role of cohesin-SA1 in gene regulation and development. *EMBO J* **31**, 2090-2102 (2012).

103. M. Mirkovic, R. A. Oliveira, Centromeric Cohesin: Molecular Glue and Much More. *Prog Mol Subcell Biol* **56**, 485-513 (2017).

104. T. Winters, F. McNicoll, R. Jessberger, Meiotic cohesin STAG3 is required for chromosome axis formation and sister chromatid cohesion. *EMBO J* **33**, 1256-1270 (2014).

105. Y. Murayama, F. Uhlmann, Biochemical reconstitution of topological DNA binding by the cohesin ring. *Nature* **505**, 367-371 (2014).

106. C. H. Haering, J. Lowe, A. Hochwagen, K. Nasmyth, Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol Cell* **9**, 773-788 (2002).

107. S. Gruber *et al.*, Evidence that loading of cohesin onto chromosomes involves opening of its SMC hinge. *Cell* **127**, 523-537 (2006).

108. Y. Murayama, F. Uhlmann, DNA Entry into and Exit out of the Cohesin Ring by an Interlocking Gate Mechanism. *Cell* **163**, 1628-1640 (2015).

109. V. Makrantoni, A. L. Marston, Cohesin and chromosome segregation. *Curr Biol* **28**, R688-R693 (2018).

110. S. Rankin, Complex elaboration: making sense of meiotic cohesin dynamics. *FEBS J* **282**, 2426-2443 (2015).

111. F. Uhlmann, Chromosome cohesion and segregation in mitosis and meiosis. *Curr Opin Cell Biol* **13**, 754-761 (2001).

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

113. R. A. Rollins, P. Morcillo, D. Dorsett, Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. *Genetics* **152**, 577-593 (1999).

114. V. Parelho *et al.*, Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132**, 422-433 (2008).

115. A. Nitzsche *et al.*, RAD21 cooperates with pluripotency transcription factors in the maintenance of embryonic stem cell identity. *PLoS One* **6**, e19470 (2011).

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

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

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

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

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

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

122. I. Barisic *et al.*, Descriptive epidemiology of Cornelia de Lange syndrome in Europe. *Am J Med Genet A* **146A**, 51-59 (2008).

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

124. A. C. Gittenberger-de Groot, M. M. Bartelings, M. C. Deruiter, R. E. Poelmann, Basics of cardiac development for the understanding of congenital heart malformations. *Pediatr Res* **57**, 169-176 (2005).

125. A. Keyte, M. R. Hutson, The neural crest in cardiac congenital anomalies. *Differentiation* **84**, 25-40 (2012).

126. K. A. Lawson, R. A. Pedersen, Cell fate, morphogenetic movement and population kinetics of embryonic endoderm at the time of germ layer formation in the mouse. *Development* **101**, 627-652 (1987).

127. S. J. Kinder *et al.*, The orderly allocation of mesodermal cells to the extraembryonic structures and the anteroposterior axis during gastrulation of the mouse embryo. *Development* **126**, 4691-4701 (1999).

128. C. V. Ragni *et al.*, Amotl1 mediates sequestration of the Hippo effector Yap1 downstream of Fat4 to restrict heart growth. *Nat Commun* **8**, 14582 (2017).

129. S. M. Meilhac, M. Esner, R. G. Kelly, J. F. Nicolas, M. E. Buckingham, The clonal origin of myocardial cells in different regions of the embryonic mouse heart. *Dev Cell* **6**, 685-698 (2004).

130. F. Lescroart *et al.*, Clonal analysis reveals a common origin between nonsomite-derived neck muscles and heart myocardium. *Proc Natl Acad Sci U S A* **112**, 1446-1451 (2015).

131. K. Ivanovitch, S. Temino, M. Torres, Live imaging of heart tube development in mouse reveals alternating phases of cardiac differentiation and morphogenesis. *Elife* **6**, (2017).

132. M. H. Kaufman, V. Navaratnam, Early differentiation of the heart in mouse embryos. *J Anat* **133**, 235-246 (1981).

133. M. Buckingham, S. Meilhac, S. Zaffran, Building the mammalian heart from two sources of myocardial cells. *Nat Rev Genet* **6**, 826-835 (2005).

134. S. D. Vincent, M. E. Buckingham, How to make a heart: the origin and regulation of cardiac progenitor cells. *Curr Top Dev Biol* **90**, 1-41 (2010).

135. F. Rochais, K. Mesbah, R. G. Kelly, Signaling pathways controlling second heart field development. *Circ Res* **104**, 933-942 (2009).

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

137. J. R. Priest *et al.*, De Novo and Rare Variants at Multiple Loci Support the Oligogenic Origins of Atrioventricular Septal Heart Defects. *PLoS Genet* **12**, e1005963 (2016).

138. S. M. Ware, J. L. Jefferies, New Genetic Insights into Congenital Heart Disease. *J Clin Exp Cardiolog* **S8**, (2012).

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

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

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

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

143. M. P. Verzi, D. J. McCulley, S. De Val, E. Dodou, B. L. Black, The right ventricle, outflow tract, and ventricular septum comprise a restricted expression domain within the secondary/anterior heart field. *Dev Biol* **287**, 134-145 (2005).

144. L. Yang *et al.*, Isl1Cre reveals a common Bmp pathway in heart and limb development. *Development* **133**, 1575-1585 (2006).

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

146. L. H. Wei *et al.*, Interleukin-6 promotes cervical tumor growth by VEGF-dependent angiogenesis via a STAT3 pathway. *Oncogene* **22**, 1517-1527 (2003).

147. T. A. Baudino *et al.*, c-Myc is essential for vasculogenesis and angiogenesis during development and tumor progression. *Genes Dev* **16**, 2530-2543 (2002).

148. C. V. Dang, MYC, metabolism, cell growth, and tumorigenesis. *Cold Spring Harb Perspect Med* **3**, (2013).

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

150. M. Roux, B. Laforest, M. Capecchi, N. Bertrand, S. Zaffran, Hoxb1 regulates proliferation and differentiation of second heart field progenitors in pharyngeal mesoderm and genetically interacts with Hoxa1 during cardiac outflow tract development. *Dev Biol* **406**, 247-258 (2015).

151. S. Contarelli, V. Fedele, D. Melisi, HOX Genes Family and Cancer: A Novel Role for Homeobox B9 in the Resistance to Anti-Angiogenic Therapies. *Cancers (Basel)* **12**, (2020).

152. H. Xu, H. Van Remmen, The SarcoEndoplasmic Reticulum Calcium ATPase (SERCA) pump: a potential target for intervention in aging and skeletal muscle pathologies. *Skelet Muscle* **11**, 25 (2021).

153. N. L. Arruda *et al.*, Distinct and overlapping roles of STAG1 and STAG2 in cohesin localization and gene expression in embryonic stem cells. *Epigenetics Chromatin* **13**, 32 (2020).

154. J. C. Rau, L. M. Beaulieu, J. A. Huntington, F. C. Church, Serpins in thrombosis, hemostasis and fibrinolysis. *J Thromb Haemost* **5 Suppl 1**, 102-115 (2007).

155. A. Gheldof, G. Berx, Cadherins and epithelial-to-mesenchymal transition. *Prog Mol Biol Transl Sci* **116**, 317-336 (2013).

156. B. S. Soh *et al.*, N-cadherin prevents the premature differentiation of anterior heart field progenitors in the pharyngeal mesodermal microenvironment. *Cell Res* **24**, 1420-1432 (2014).

157. Y. Yoshimoto, M. Ikemoto-Uezumi, K. Hitachi, S. I. Fukada, A. Uezumi, Methods for Accurate Assessment of Myofiber Maturity During Skeletal Muscle Regeneration. *Front Cell Dev Biol* **8**, 267 (2020).

158. S. S. Sung *et al.*, Mutations in TNNT3 cause multiple congenital contractures: a second locus for distal arthrogryposis type 2B. *Am J Hum Genet* **73**, 212-214 (2003).

159. S. L. Peek, K. M. Mah, J. A. Weiner, Regulation of neural circuit formation by protocadherins. *Cell Mol Life Sci* **74**, 4133-4157 (2017).

160. A. von Gise, W. T. Pu, Endocardial and epicardial epithelial to mesenchymal transitions in heart development and disease. *Circ Res* **110**, 1628-1645 (2012).

161. B. Zhang *et al.*, Forkhead box O6 (FoxO6) promotes cardiac pathological remodeling and dysfunction by activating Kif15-TGF-beta1 under aggravated afterload. *MedComm (2020)* **4**, e383 (2023).

162. D. Lima Cunha, G. Arno, M. Corton, M. Moosajee, The Spectrum of PAX6 Mutations and Genotype-Phenotype Correlations in the Eye. *Genes (Basel)* **10**, (2019).

163. D. Duan, Y. Fu, G. Paxinos, C. Watson, Spatiotemporal expression patterns of Pax6 in the brain of embryonic, newborn, and adult mice. *Brain Struct Funct* **218**, 353-372 (2013).

164. D. G. Johnson, R. Schneider-Broussard, Role of E2F in cell cycle control and cancer. *Front Biosci* **3**, d447-448 (1998).

165. D. C. Fingar *et al.*, mTOR controls cell cycle progression through its cell growth effectors S6K1 and 4E-BP1/eukaryotic translation initiation factor 4E. *Mol Cell Biol* **24**, 200-216 (2004).

166. L. McShane, I. Tabas, G. Lemke, M. Kurowska-Stolarska, P. Maffia, TAM receptors in cardiovascular disease. *Cardiovasc Res* **115**, 1286-1295 (2019).

167. S. Mazzotta *et al.*, Distinctive Roles of Canonical and Noncanonical Wnt Signaling in Human Embryonic Cardiomyocyte Development. *Stem Cell Reports* **7**, 764-776 (2016).

168. Y. Tian *et al.*, Characterization and in vivo pharmacological rescue of a Wnt2-Gata6 pathway required for cardiac inflow tract development. *Dev Cell* **18**, 275-287 (2010).

169. H. Laeremans, S. S. Rensen, H. C. Ottenheijm, J. F. Smits, W. M. Blankesteijn, Wnt/frizzled signalling modulates the migration and differentiation of immortalized cardiac fibroblasts. *Cardiovasc Res* **87**, 514-523 (2010).

170. D. J. Henderson, H. M. Phillips, B. Chaudhry, Vang-like 2 and noncanonical Wnt signaling in outflow tract development. *Trends Cardiovasc Med* **16**, 38-45 (2006).

171. W. Li *et al.*, PLCE1 promotes myocardial ischemia-reperfusion injury in H/R H9c2 cells and I/R rats by promoting inflammation. *Biosci Rep* **39**, (2019).

172. T. N. Durrant, I. Hers, PI3K inhibitors in thrombosis and cardiovascular disease. *Clin Transl Med* **9**, 8 (2020).

173. A. Schedl *et al.*, Influence of PAX6 gene dosage on development: overexpression causes severe eye abnormalities. *Cell* **86**, 71-82 (1996).

174. S. Nemec *et al.*, Pitx1 directly modulates the core limb development program to implement hindlimb identity. *Development* **144**, 3325-3335 (2017).

175. H. Suemori, S. Noguchi, Hox C cluster genes are dispensable for overall body plan of mouse embryonic development. *Dev Biol* **220**, 333-342 (2000).

176. P. I. Sadate-Ngatchou, C. J. Payne, A. T. Dearth, R. E. Braun, Cre recombinase activity specific to postnatal, premeiotic male germ cells in transgenic mice. *Genesis* **46**, 738-742 (2008).

177. M. D. Muzumdar, B. Tasic, K. Miyamichi, L. Li, L. Luo, A global double-fluorescent Cre reporter mouse. *Genesis* **45**, 593-605 (2007).

178. M. V. Gomez-Gaviro *et al.*, Optimized CUBIC protocol for three-dimensional imaging of chicken embryos at single-cell resolution. *Development* **144**, 2092-2097 (2017).

179. C. Heffner, S. Murray. (The Jackson Laboratory, www.jax.org, 2024).

180. S. Chea *et al.*, Gastrulation-stage gene expression in Nipbl(+/-) mouse embryos foreshadows the development of syndromic birth defects. *Sci Adv* **10**, eadl4239 (2024).

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

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

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

184. A. D. Kline *et al.*, Cornelia de Lange syndrome: clinical review, diagnostic and scoring systems, and anticipatory guidance. *Am J Med Genet A* **143A**, 1287-1296 (2007).

185. Z. A. Bhuiyan *et al.*, Genotype-phenotype correlations of 39 patients with Cornelia De Lange syndrome: the Dutch experience. *J Med Genet* **43**, 568-575 (2006).

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

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

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

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

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

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

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

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

194. A. Okuda *et al.*, UTF1, a novel transcriptional coactivator expressed in pluripotent embryonic stem cells and extra-embryonic cells. *EMBO J* **17**, 2019-2032 (1998).

195. S. Zhang, W. Cui, Sox2, a key factor in the regulation of pluripotency and neural differentiation. *World J Stem Cells* **6**, 305-311 (2014).

196. Q. Zhu *et al.*, The transcription factor Pou3f1 promotes neural fate commitment via activation of neural lineage genes and inhibition of external signaling pathways. *Elife* **3**, (2014).

197. V. Wilson, L. Manson, W. C. Skarnes, R. S. Beddington, The T gene is necessary for normal mesodermal morphogenetic cell movements during gastrulation. *Development* **121**, 877-886 (1995).

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

199. B. Saykali *et al.*, Distinct mesoderm migration phenotypes in extra-embryonic and embryonic regions of the early mouse embryo. *Elife* **8**, (2019).

200. H. Bildsoe *et al.*, Transcriptional targets of TWIST1 in the cranial mesoderm regulate cell-matrix interactions and mesenchyme maintenance. *Dev Biol* **418**, 189-203 (2016).

201. B. Leussink *et al.*, Expression patterns of the paired-related homeobox genes MHox/Prx1 and S8/Prx2 suggest roles in development of the heart and the forebrain. *Mech Dev* **52**, 51-64 (1995).

202. G. S. Kwon, A. K. Hadjantonakis, Transthyretin mouse transgenes direct RFP expression or Cre-mediated recombination throughout the visceral endoderm. *Genesis* **47**, 447-455 (2009).

203. L. F. Chu *et al.*, Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. *Genome Biol* **17**, 173 (2016).

204. I. Burtscher, H. Lickert, Foxa2 regulates polarity and epithelialization in the endoderm germ layer of the mouse embryo. *Development* **136**, 1029-1038 (2009).

205. W. S. Chen *et al.*, Disruption of the HNF-4 gene, expressed in visceral endoderm, leads to cell death in embryonic ectoderm and impaired gastrulation of mouse embryos. *Genes Dev* **8**, 2466-2477 (1994).

206. S. F. Khan *et al.*, The roles and regulation of TBX3 in development and disease. *Gene* **726**, 144223 (2020).

207. Y. Li, P. Lemaire, R. R. Behringer, Esx1, a novel X chromosome-linked homeobox gene expressed in mouse extraembryonic tissues and male germ cells. *Dev Biol* **188**, 85-95 (1997).

208. A. Simeone *et al.*, Two vertebrate homeobox genes related to the Drosophila empty spiracles gene are expressed in the embryonic cerebral cortex. *EMBO J* **11**, 2541-2550 (1992).

209. B. I. Meyer, P. Gruss, Mouse Cdx-1 expression during gastrulation. *Development* **117**, 191-203 (1993).

210. A. Rodrigo Albors, P. A. Halley, K. G. Storey, Lineage tracing of axial progenitors using Nkx1-2CreER(T2) mice defines their trunk and tail contributions. *Development* **145**, (2018).

211. Y. Qu *et al.*, Transcriptome and proteome characterization of surface ectoderm cells differentiated from human iPSCs. *Sci Rep* **6**, 32007 (2016).

212. M. Donnison *et al.*, Loss of the extraembryonic ectoderm in Elf5 mutants leads to defects in embryonic patterning. *Development* **132**, 2299-2308 (2005).

213. N. Nishioka *et al.*, Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. *Mech Dev* **125**, 270-283 (2008).

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

215. G. Gestri *et al.*, Six3 functions in anterior neural plate specification by promoting cell proliferation and inhibiting Bmp4 expression. *Development* **132**, 2401-2413 (2005).

216. C. L. Andoniadou *et al.*, Lack of the murine homeobox gene Hesx1 leads to a posterior transformation of the anterior forebrain. *Development* **134**, 1499-1508 (2007).

217. M. T. Alves dos Santos, M. P. Smidt, En1 and Wnt signaling in midbrain dopaminergic neuronal development. *Neural Dev* **6**, 23 (2011).

218. D. H. Rowitch, A. P. McMahon, Pax-2 expression in the murine neural plate precedes and encompasses the expression domains of Wnt-1 and En-1. *Mech Dev* **52**, 3-8 (1995).

219. A. Gavalas, C. Ruhrberg, J. Livet, C. E. Henderson, R. Krumlauf, Neuronal defects in the hindbrain of Hoxa1, Hoxb1 and Hoxb2 mutants reflect regulatory interactions among these Hox genes. *Development* **130**, 5663-5679 (2003).

220. C. S. Hong, J. P. Saint-Jeannet, The b-HLH transcription factor Hes3 participates in neural plate border formation by interfering with Wnt/beta-catenin signaling. *Dev Biol* **442**, 162-172 (2018).

221. G. Cho, Y. Lim, I. T. Cho, J. C. Simonet, J. A. Golden, Arx together with FoxA2, regulates Shh floor plate expression. *Dev Biol* **393**, 137-148 (2014).

222. A. Holz *et al.*, The transcription factors Nkx2.2 and Nkx2.9 play a novel role in floor plate development and commissural axon guidance. *Development* **137**, 4249-4260 (2010).

223. C. Papanayotou *et al.*, A mechanism regulating the onset of Sox2 expression in the embryonic neural plate. *PLoS Biol* **6**, e2 (2008).

224. J. D. Lee, N. F. Silva-Gagliardi, U. Tepass, C. J. McGlade, K. V. Anderson, The FERM protein Epb4.1l5 is required for organization of the neural plate and for the epithelial-mesenchymal transition at the primitive streak of the mouse embryo. *Development* **134**, 2007-2016 (2007).

225. C. Milet, F. Maczkowiak, D. D. Roche, A. H. Monsoro-Burq, Pax3 and Zic1 drive induction and differentiation of multipotent, migratory, and functional neural crest in Xenopus embryos. *Proc Natl Acad Sci U S A* **110**, 5528-5533 (2013).

226. L. Teng, N. A. Mundell, A. Y. Frist, Q. Wang, P. A. Labosky, Requirement for Foxd3 in the maintenance of neural crest progenitors. *Development* **135**, 1615-1624 (2008).

227. S. S. Chan *et al.*, Mesp1 patterns mesoderm into cardiac, hematopoietic, or skeletal myogenic progenitors in a context-dependent manner. *Cell Stem Cell* **12**, 587-601 (2013).

228. M. R. Jongbloed *et al.*, Expression of Id2 in the second heart field and cardiac defects in Id2 knock-out mice. *Dev Dyn* **240**, 2561-2577 (2011).

229. E. Dodou, M. P. Verzi, J. P. Anderson, S. M. Xu, B. L. Black, Mef2c is a direct transcriptional target of ISL1 and GATA factors in the anterior heart field during mouse embryonic development. *Development* **131**, 3931-3942 (2004).

230. T. Sadahiro *et al.*, Tbx6 Induces Nascent Mesoderm from Pluripotent Stem Cells and Temporally Controls Cardiac versus Somite Lineage Diversification. *Cell Stem Cell* **23**, 382-395 e385 (2018).

231. F. Cecconi, G. Proetzel, G. Alvarez-Bolado, D. Jay, P. Gruss, Expression of Meis2, a Knotted-related murine homeobox gene, indicates a role in the differentiation of the forebrain and the somitic mesoderm. *Dev Dyn* **210**, 184-190 (1997).

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

233. O. R. Davies *et al.*, Tcf15 primes pluripotent cells for differentiation. *Cell Rep* **3**, 472-484 (2013).

234. J. M. Khor, C. A. Ettensohn, Transcription Factors of the Alx Family: Evolutionarily Conserved Regulators of Deuterostome Skeletogenesis. *Front Genet* **11**, 569314 (2020).

235. M. Gering, Y. Yamada, T. H. Rabbitts, R. K. Patient, Lmo2 and Scl/Tal1 convert non-axial mesoderm into haemangioblasts which differentiate into endothelial cells in the absence of Gata1. *Development* **130**, 6187-6199 (2003).

236. N. Koyano-Nakagawa, D. J. Garry, Etv2 as an essential regulator of mesodermal lineage development. *Cardiovasc Res* **113**, 1294-1306 (2017).

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

238. A. Beckers, L. Alten, C. Viebahn, P. Andre, A. Gossler, The mouse homeobox gene Noto regulates node morphogenesis, notochordal ciliogenesis, and left right patterning. *Proc Natl Acad Sci U S A* **104**, 15765-15770 (2007).

239. J. L. Stubbs, I. Oishi, J. C. Izpisua Belmonte, C. Kintner, The forkhead protein Foxj1 specifies node-like cilia in Xenopus and zebrafish embryos. *Nat Genet* **40**, 1454-1460 (2008).

240. J. Schemmer *et al.*, Transcription factor TFAP2C regulates major programs required for murine fetal germ cell maintenance and haploinsufficiency predisposes to teratomas in male mice. *PLoS One* **8**, e71113 (2013).

241. J. Sun, M. C. Ting, M. Ishii, R. Maxson, Msx1 and Msx2 function together in the regulation of primordial germ cell migration in the mouse. *Dev Biol* **417**, 11-24 (2016).

242. E. L. Maska *et al.*, A Tlx2-Cre mouse line uncovers essential roles for hand1 in extraembryonic and lateral mesoderm. *Genesis* **48**, 479-484 (2010).

243. C. A. Desjardins, F. J. Naya, The Function of the MEF2 Family of Transcription Factors in Cardiac Development, Cardiogenomics, and Direct Reprogramming. *J Cardiovasc Dev Dis* **3**, (2016).

244. V. Serpooshan *et al.*, Nkx2.5+ Cardiomyoblasts Contribute to Cardiomyogenesis in the Neonatal Heart. *Sci Rep* **7**, 12590 (2017).

245. T. Jukkola *et al.*, Meox1Cre: a mouse line expressing Cre recombinase in somitic mesoderm. *Genesis* **43**, 148-153 (2005).

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

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

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

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

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

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

252. M. Milacic *et al.*, The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Res* **52**, D672-D678 (2024).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

282. Z. Avsec *et al.*, Base-resolution models of transcription-factor binding reveal soft motif syntax. *Nat Genet* **53**, 354-366 (2021).

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

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

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

286. C. Cortes, A. Francou, C. De Bono, R. G. Kelly, Epithelial Properties of the Second Heart Field. *Circ Res* **122**, 142-154 (2018).

287. R. M. George, G. Maldonado-Velez, A. B. Firulli, The heart of the neural crest: cardiac neural crest cells in development and regeneration. *Development* **147**, (2020).

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

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

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

291. T. H. Kim *et al.*, Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**, 1231-1245 (2007).

292. W. Schwarzer *et al.*, Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51-56 (2017).

293. A. S. Hansen, C. Cattoglio, X. Darzacq, R. Tjian, Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus* **9**, 20-32 (2018).

294. M. Panarotto, I. F. Davidson, G. Litos, A. Schleiffer, J. M. Peters, Cornelia de Lange syndrome mutations in NIPBL can impair cohesin-mediated DNA loop extrusion. *Proc Natl Acad Sci U S A* **119**, e2201029119 (2022).

295. D. Hnisz *et al.*, Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947 (2013).

296. D. A. Newkirk *et al.*, The effect of Nipped-B-like (Nipbl) haploinsufficiency on genome-wide cohesin binding and target gene expression: modeling Cornelia de Lange syndrome. *Clin Epigenetics* **9**, 89 (2017).

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

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

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

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

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

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

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

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

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

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

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

308. J. Ren, D. Miao, Y. Li, R. Gao, Spotlight on Isl1: A Key Player in Cardiovascular Development and Diseases. *Front Cell Dev Biol* **9**, 793605 (2021).

309. M. Logan, C. J. Tabin, Role of Pitx1 upstream of Tbx4 in specification of hindlimb identity. *Science* **283**, 1736-1739 (1999).

310. D. P. Szeto *et al.*, Role of the Bicoid-related homeodomain factor Pitx1 in specifying hindlimb morphogenesis and pituitary development. *Genes Dev* **13**, 484-494 (1999).

311. T. Stuart, A. Srivastava, S. Madad, C. A. Lareau, R. Satija, Single-cell chromatin state analysis with Signac. *Nat Methods* **18**, 1333-1341 (2021).

312. E. P. Consortium *et al.*, Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699-710 (2020).

APPENDIX A

Supplementary Materials for Chapter 2

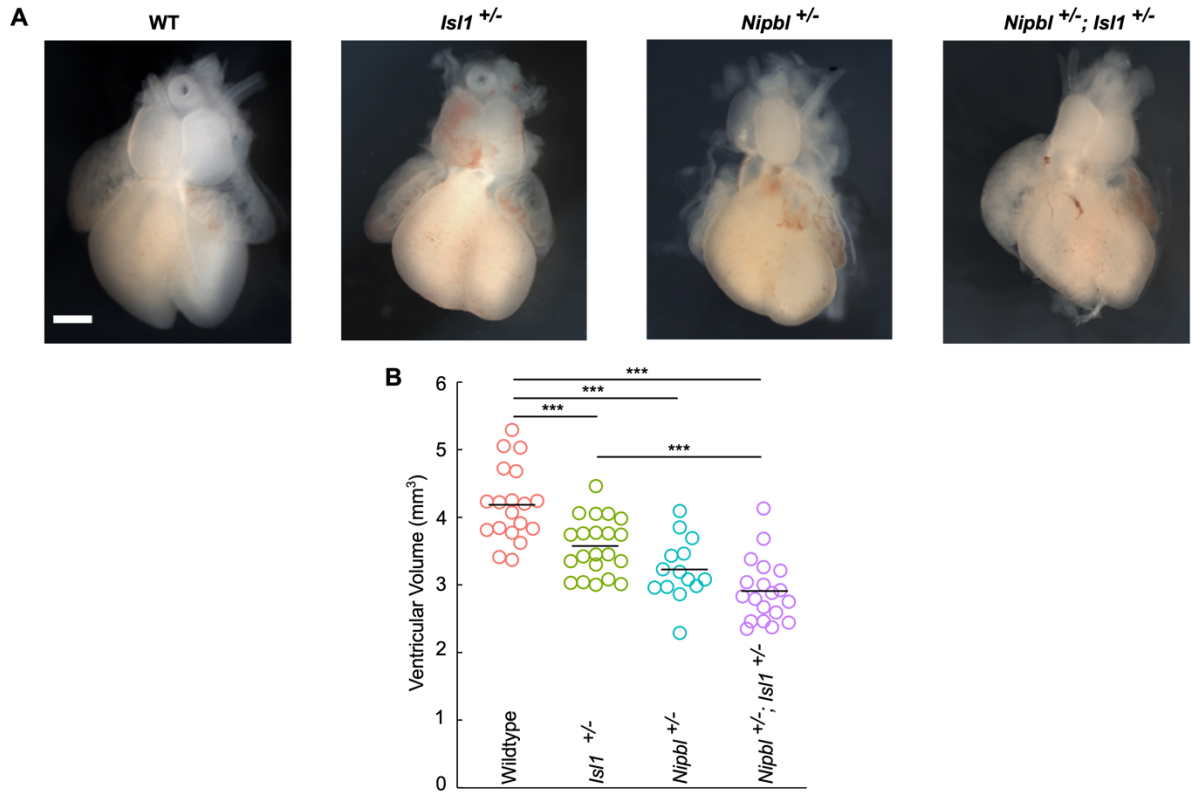


Fig. S2.1. *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts have smaller hearts than WT hearts. (A) Representative darkfield images of E15.5 WT, *Isl1*$^{+/-}$, *Nipbl*$^{+/-}$, and *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts. *Isl1*$^{+/-}$, *Nipbl*$^{+/-}$, and *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ hearts appear smaller than WT. Scale bar = 500 uM. (B) Ventricular volumes of E15.5 WT (*Nipbl*$^{+/+}$*; Isl1*$^{+/+}$) (N=19), *Nipbl*$^{+/-}$ (N=21), *Isl1*$^{+/-}$ (N=14), and *Nipbl*$^{+/-}$*; Isl1*$^{+/-}$ (N=19) hearts. Ventricular volumes were measured in ImageJ using LSFM images of hearts. Bars show means. *** represents $P \leq 0.001$. P-values from Tukey-Kramer Test.

APPENDIX B

Supplementary Materials for Chapter 3

These supplementary materials were published in Science Advances (*180*).

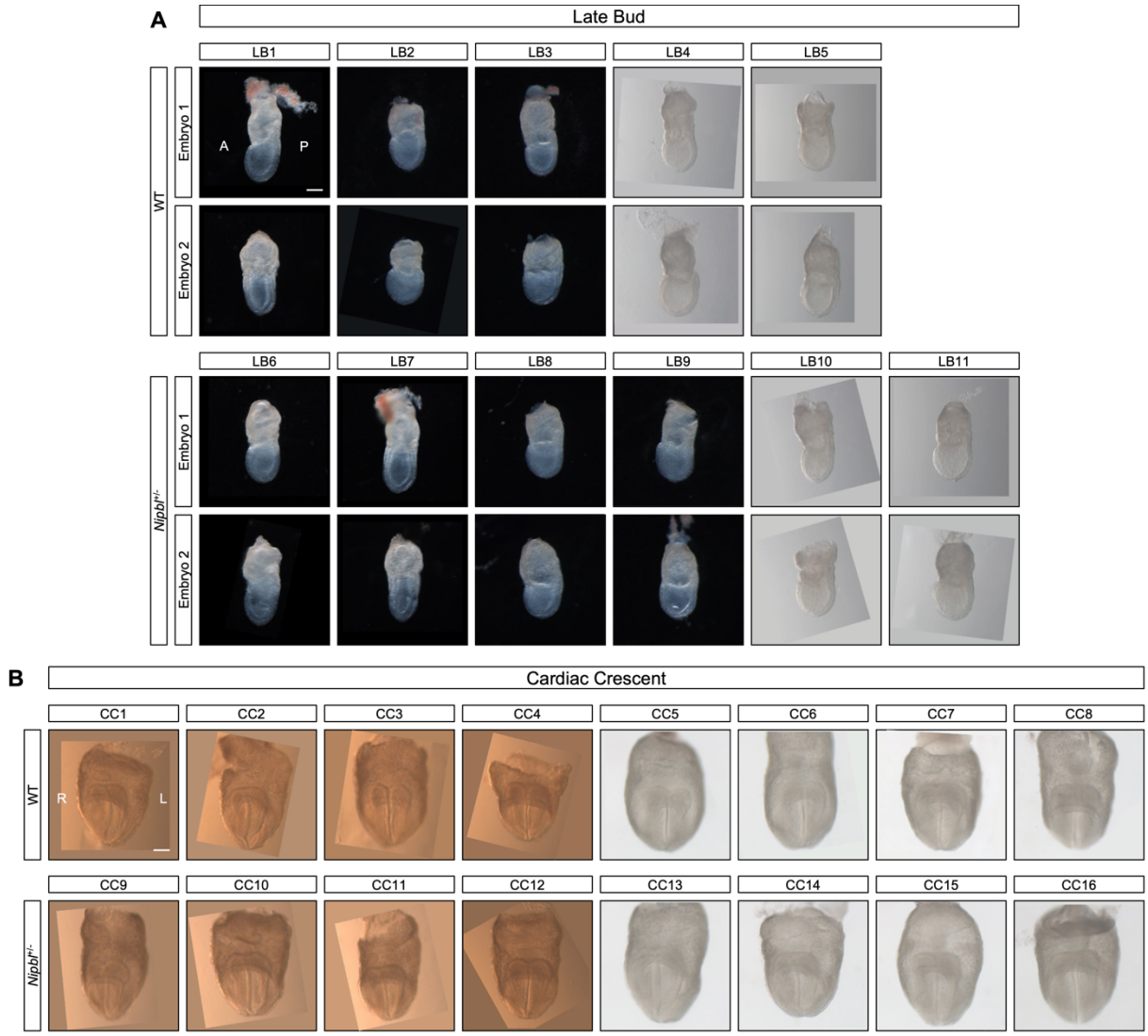


Fig. S3.1. Images of LB- and CC-stage WT and *Nipbl*$^{+/-}$ mouse embryos used for scRNAseq.

Lateral view of (**A**) LB-stage and (**B**) anterior view of CC-stage WT and *Nipbl*$^{+/-}$ mouse embryos. A = anterior, P = posterior, L = left, R = right, and scale bar = 100 um.

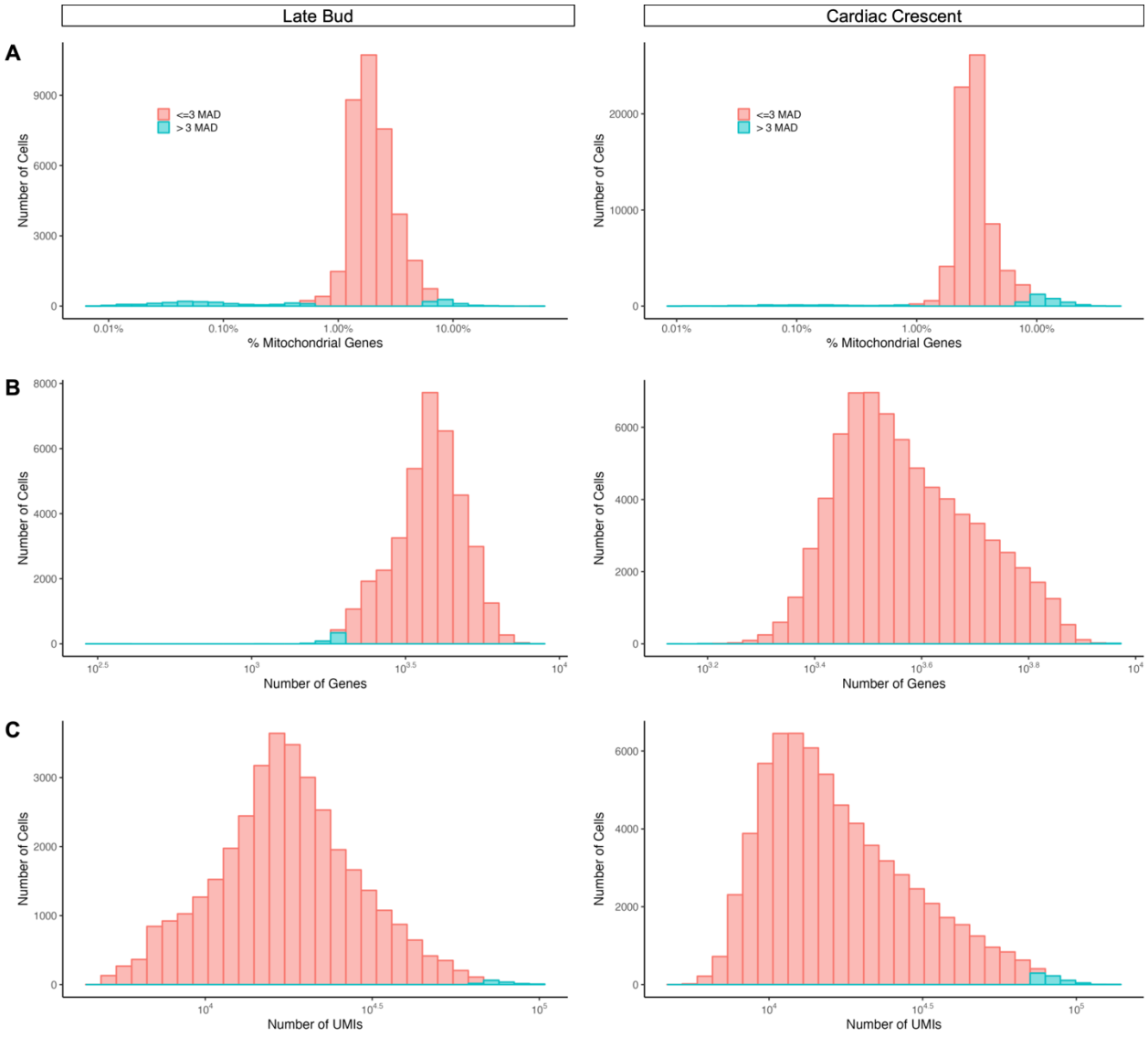


Fig. S3.2. Metrics used to identify low-quality cells and doublets among LB- and CC-stage embryos.

At each stage, we considered cells that exceeded three median absolute deviations from percent mitochondrial genes expressed, number of genes expressed, and number of transcripts detected per cell, low-quality cells and/or doublets and removed them. Histograms of (**A**) percent mitochondrial genes expressed, (**B**) number of genes expressed, and (**C**) number of transcripts

detected per cell at LB- and CC-stage colored by whether cells were less than or equal to 3 median absolute deviations (red) or greater than 3 median absolute deviations (blue).

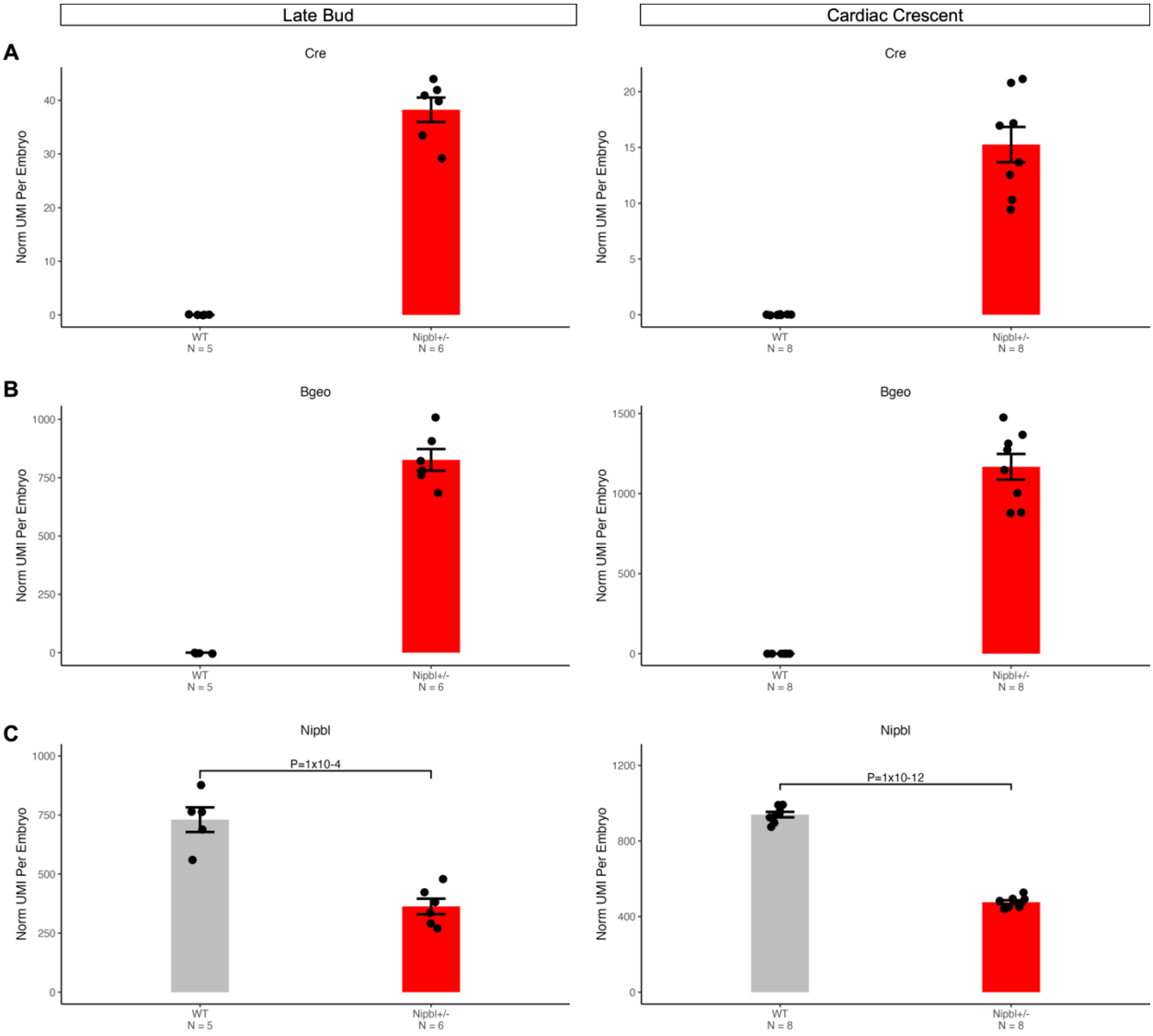


Fig. S3.3. Expression of Cre, Bgeo, and *Nipbl* in LB- and CC-stage WT and *Nipbl*$^{+/-}$ embryos.

Nipbl$^{+/-}$ embryos at LB- and CC-stage express Cre and Bgeo, while WT embryos do not. *Nipbl*$^{+/-}$ embryos also express *Nipbl* at levels lower than that of WT embryos. Normalized transcripts of

(A) *Cre*, (B) *Bgeo*, and (C) *Nipbl* in LB- and CC-stage embryos. Each dot indicates an individual sample. Error bars show standard error of the mean. *P*-value from T-Test.

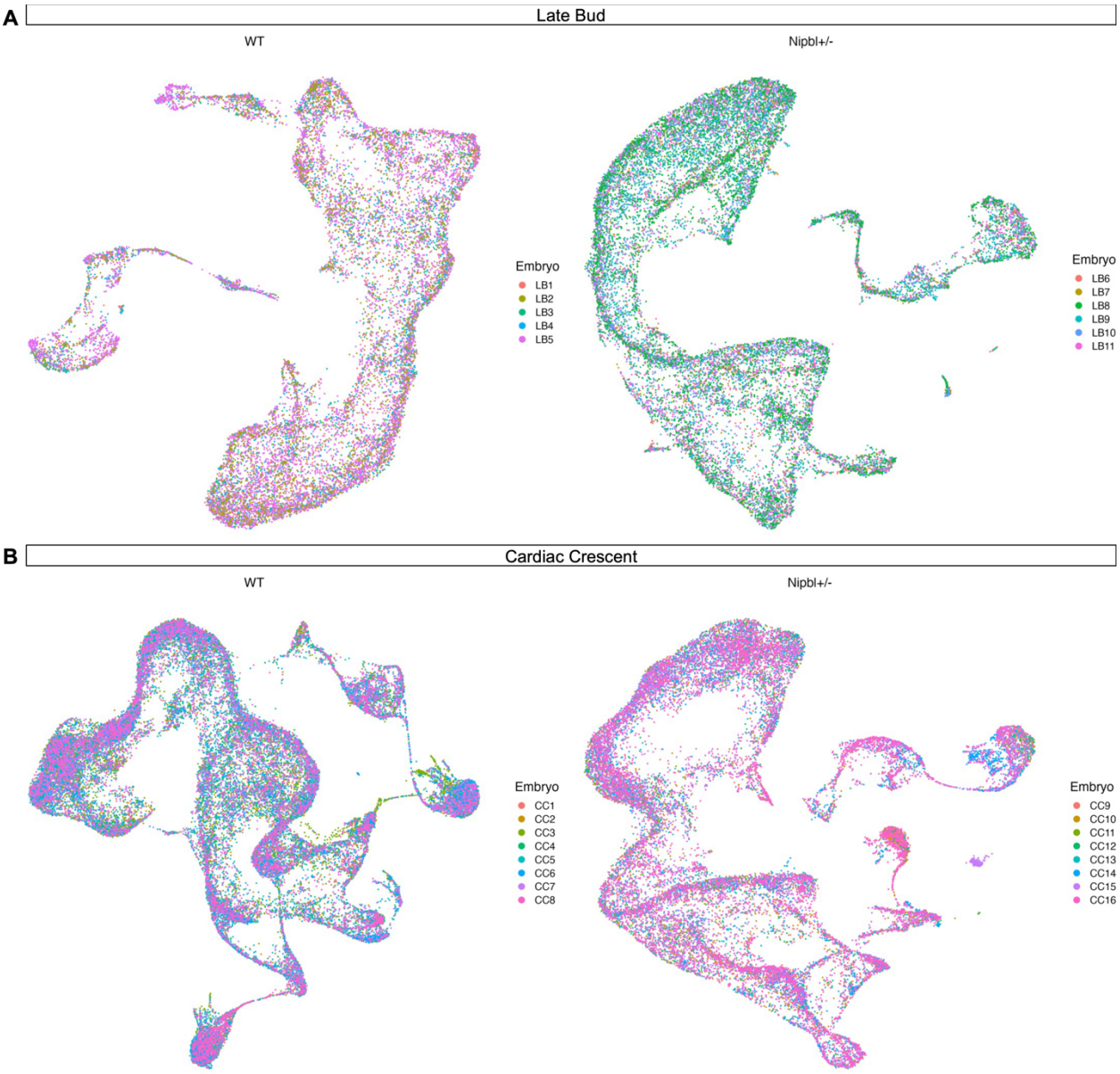


Fig. S3.4. Batch effect correction and integration of cells of the same stage (LB- or CC-stage) and genotype (WT or *Nipbl*$^{+/-}$).

We corrected batch effects among embryos by integrating cells from the same stage and genotype together. UMAP of WT and *Nipbl*$^{+/-}$ cells colored by embryo at (A) LB- and (B) CC-stage.

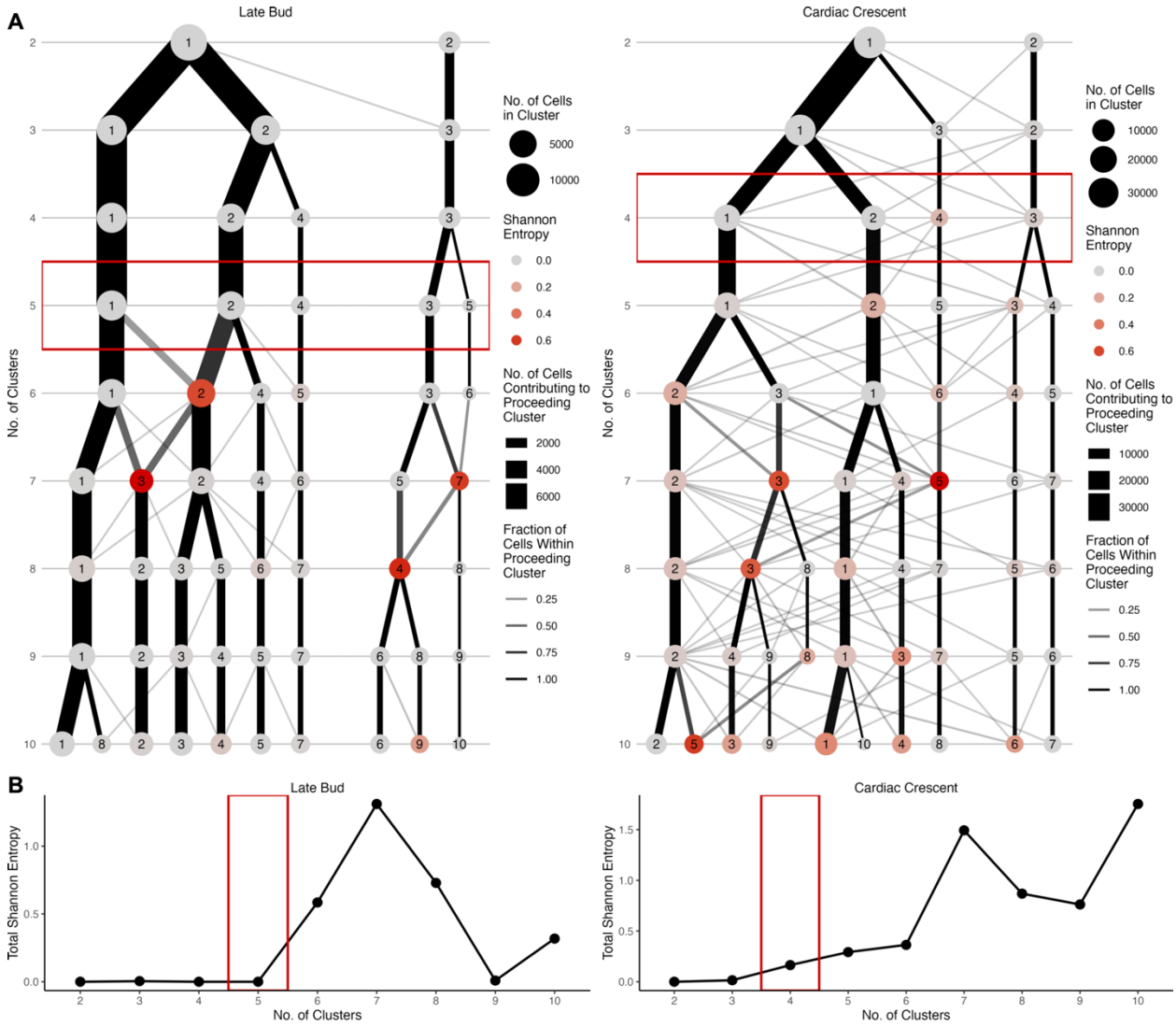


Fig. S3.5. Example of optimized clustering using clustering tree and Shannon Entropy.

(**A**) Following consecutive clustering of WT cells from, we generated a clustering tree to visualize how cell cluster identities change as the number of clusters consecutively increase. Nodes represent clusters and edges represent cells from preceding clusters. Clusters are stable when a large proportion of cells are derived from a single preceding cluster rather than multiple preceding clusters. We adopted Shannon Entropy as a measure of these proportions as a measure of intra-cluster stability. A low Shannon Entropy represents high intra-cluster stability. Here, WT cells from LB-stage embryos can only be clustered as high as 5 clusters before they can be

subclustered into further subclusters. WT cells from CC-stage embryos can only be clustered as high as 4 clusters before they can be subclustered into further clusters. (**B**) Total Shannon Entropy of all clusters among WT cells from LB- or CC-stage embryos at consecutively increasing number of clusters.

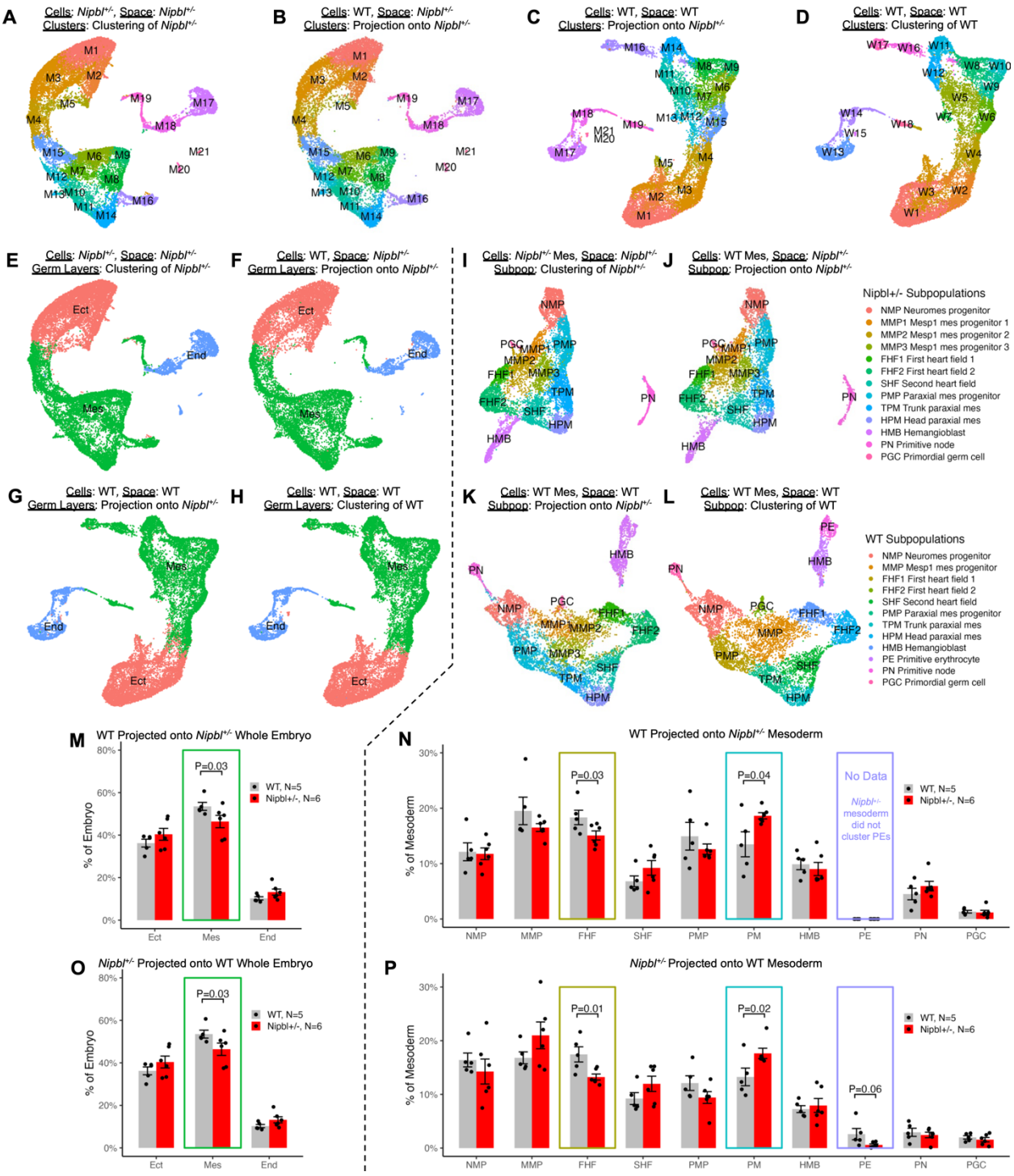


Fig. S3.6. Percentage of cells in germ layers and mesodermal subpopulations of LB-stage embryos following projection of WT cells onto clustered *Nipbl*$^{+/-}$ cells.

(**A**) LB-stage *Nipbl*$^{+/-}$ cells in UMAP of *Nipbl*$^{+/-}$ cells colored by cluster after clustering *Nipbl*$^{+/-}$ cells separately from WT. (**B**) LB-stage WT cells in UMAP of *Nipbl*$^{+/-}$ cells colored by clusters found in *Nipbl*$^{+/-}$ cells after projecting WT cells onto *Nipbl*$^{+/-}$ cells. (**C**) LB-stage WT cells in UMAP of WT cells colored by clusters found in *Nipbl*$^{+/-}$ cells after projecting WT cells onto *Nipbl*$^{+/-}$ cells. (**D**) LB-stage WT cells in UMAP of WT cells colored by clusters after clustering WT cells separately from *Nipbl*$^{+/-}$. (**E**) LB-stage *Nipbl*$^{+/-}$ cells in UMAP of *Nipbl*$^{+/-}$ cells colored by germ layer after clustering *Nipbl*$^{+/-}$ cells separately from WT. (**F**) LB-stage WT cells in UMAP of *Nipbl*$^{+/-}$ cells colored by germ layers found in *Nipbl*$^{+/-}$ cells after projecting WT cells onto *Nipbl*$^{+/-}$ cells. (**G**) LB-stage WT cells in UMAP of WT cells colored by germ layers found in *Nipbl*$^{+/-}$ cells after projecting WT cells onto *Nipbl*$^{+/-}$ cells. (**H**) LB-stage WT cells in UMAP of WT cells colored by germ layer after clustering WT cells separately from *Nipbl*$^{+/-}$. (**I**) LB-stage *Nipbl*$^{+/-}$ mesoderm cells in UMAP of *Nipbl*$^{+/-}$ mesoderm colored by cell population after clustering *Nipbl*$^{+/-}$ mesoderm separately from WT. (**J**) LB-stage WT mesoderm cells in UMAP of *Nipbl*$^{+/-}$ mesoderm colored by cell populations found in *Nipbl*$^{+/-}$ mesoderm after projecting WT mesoderm onto *Nipbl*$^{+/-}$ mesoderm. (**K**) LB-stage WT mesoderm cells in UMAP of WT mesoderm colored by cell populations found in *Nipbl*$^{+/-}$ mesoderm after projecting WT mesoderm onto *Nipbl*$^{+/-}$ mesoderm. (**L**) LB-stage WT mesoderm cells in UMAP of WT mesoderm colored by cell population after clustering WT mesoderm separately from *Nipbl*$^{+/-}$. (**M**) Percentage of cells in germ layers from all cells in LB-stage WT and *Nipbl*$^{+/-}$ embryos after projecting WT cells onto *Nipbl*$^{+/-}$ cells. Error bars show standard error of the mean. *P*-values from T-Test. (**N**) Percentage of cells in mesodermal cell populations from all mesoderm cells in LB-stage WT and *Nipbl*$^{+/-}$ embryos after projecting WT mesoderm onto *Nipbl*$^{+/-}$ mesoderm. Error bars show SEM. *P*-values from T-Test. (**O**) Percentage of cells in germ layers from all

cells in LB-stage WT and *Nipbl*$^{+/-}$ embryos after projecting *Nipbl*$^{+/-}$ cells onto WT cells. Same as This figure is from Fig. 3.3B. Error bars show standard error of the mean. *P*-values from T-Test. (**P**) Percentage of cells in mesodermal cell populations from all mesoderm cells in LB-stage WT and *Nipbl*$^{+/-}$ embryos after projecting *Nipbl*$^{+/-}$ mesoderm onto WT mesoderm. This figure is from Fig. 3.3D. Error bars show SEM. *P*-values from T-Test.

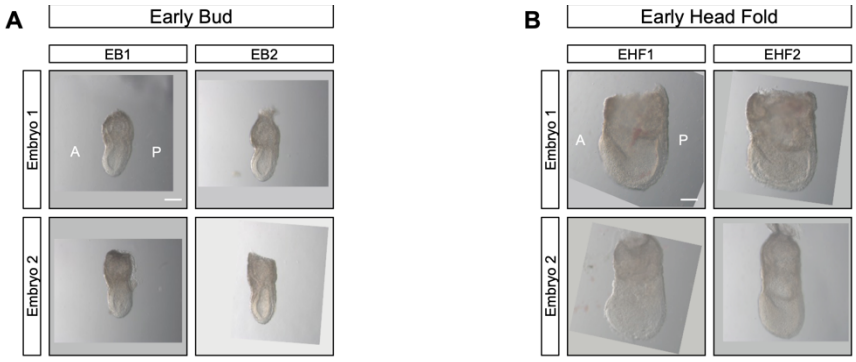


Fig. S3.7. Images of EB- and EHF-stage WT mouse embryos used for scRNAseq.

Lateral view of WT mouse embryos at (**A**) EB- and (**B**) EHF- stage. A = anterior, P = posterior. Dashed line represents where embryonic tissue was separated from extraembryonic tissue. Scale bar = 100 um.

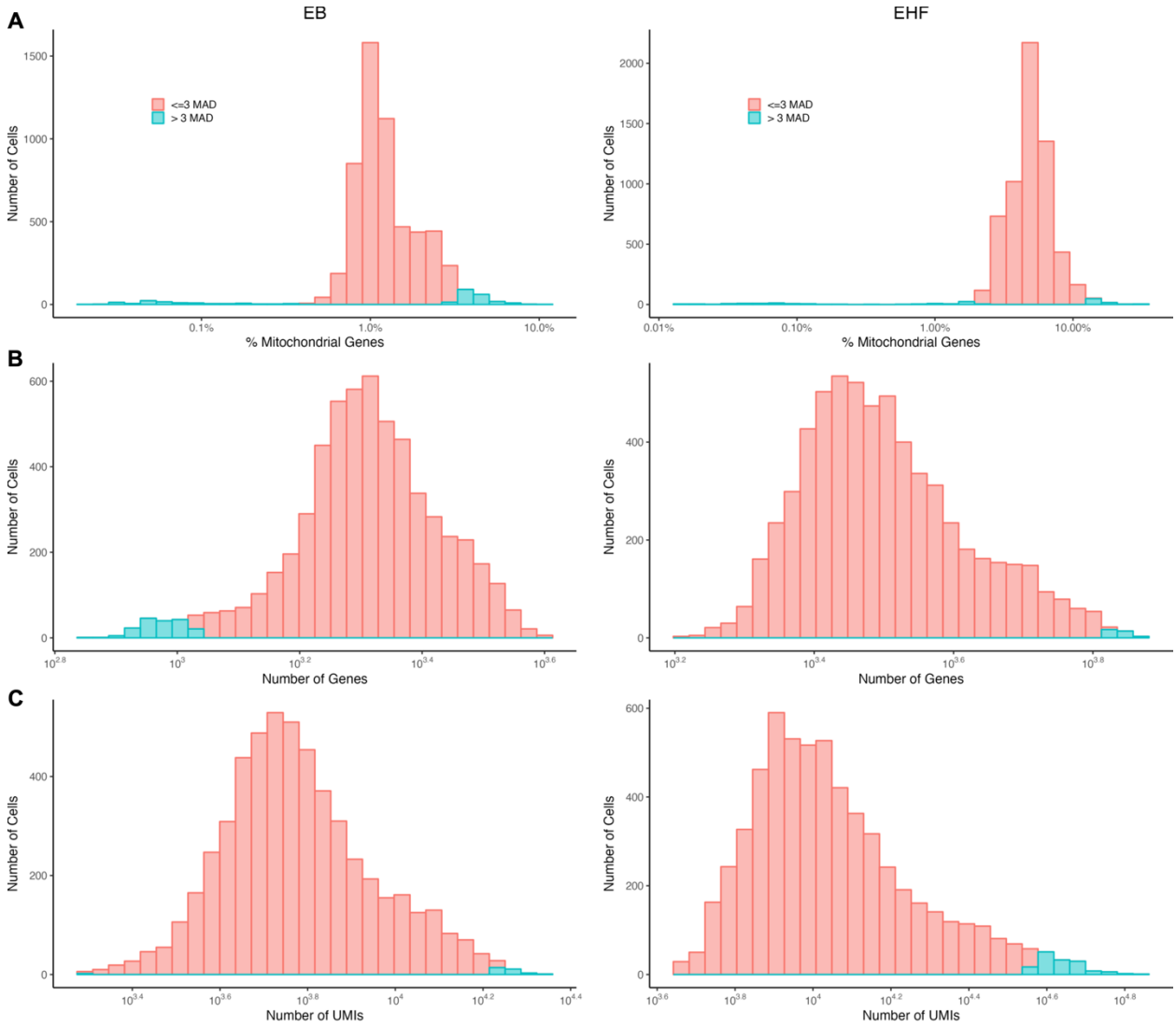


Fig. S3.8. Metrics used to identify low-quality cells and doublets among EB- and EHF-stage embryos.

At each stage, we considered cells that exceeded three median absolute deviations from percent mitochondrial genes expressed, number of genes expressed, and number of transcripts detected per cell low-quality cells and/or doublets and removed them. Histograms of (**A**) percent mitochondrial genes expressed, (**B**) number of genes expressed, and (**C**) number of transcripts detected per cell at EB- and EHF-stage colored by whether cells were less than or equal to 3

median absolute deviations (red) or greater than 3 median absolute deviations (blue). Cells colored in blue were removed.

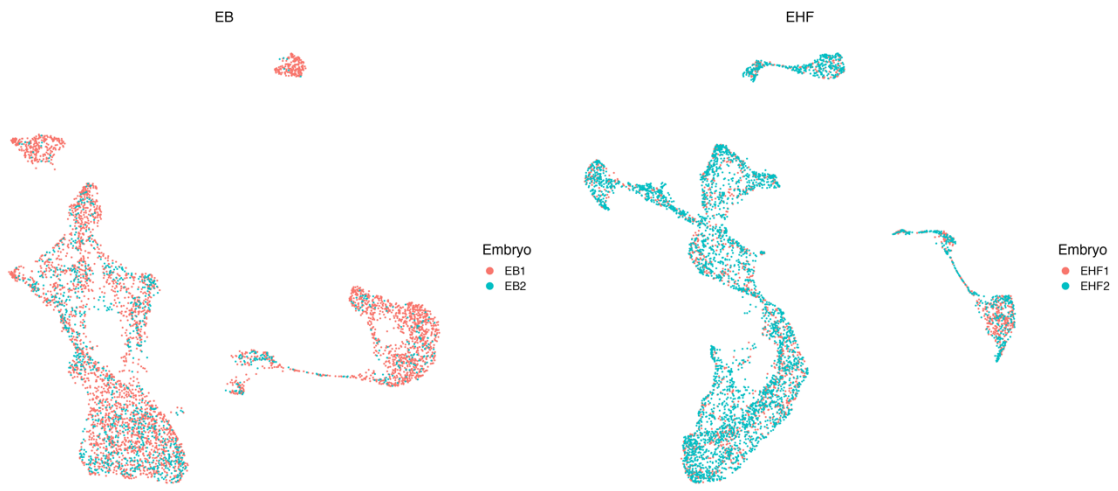


Fig. S3.9. Batch effect correction and integration of cells of the same stage (EB- or EHF-stage).

We removed batch effects among EB- and EHF-stage embryos of the same stage and integrated embryos of the same stage together. UMAP of EB- and EHF-stage cells colored by embryo.

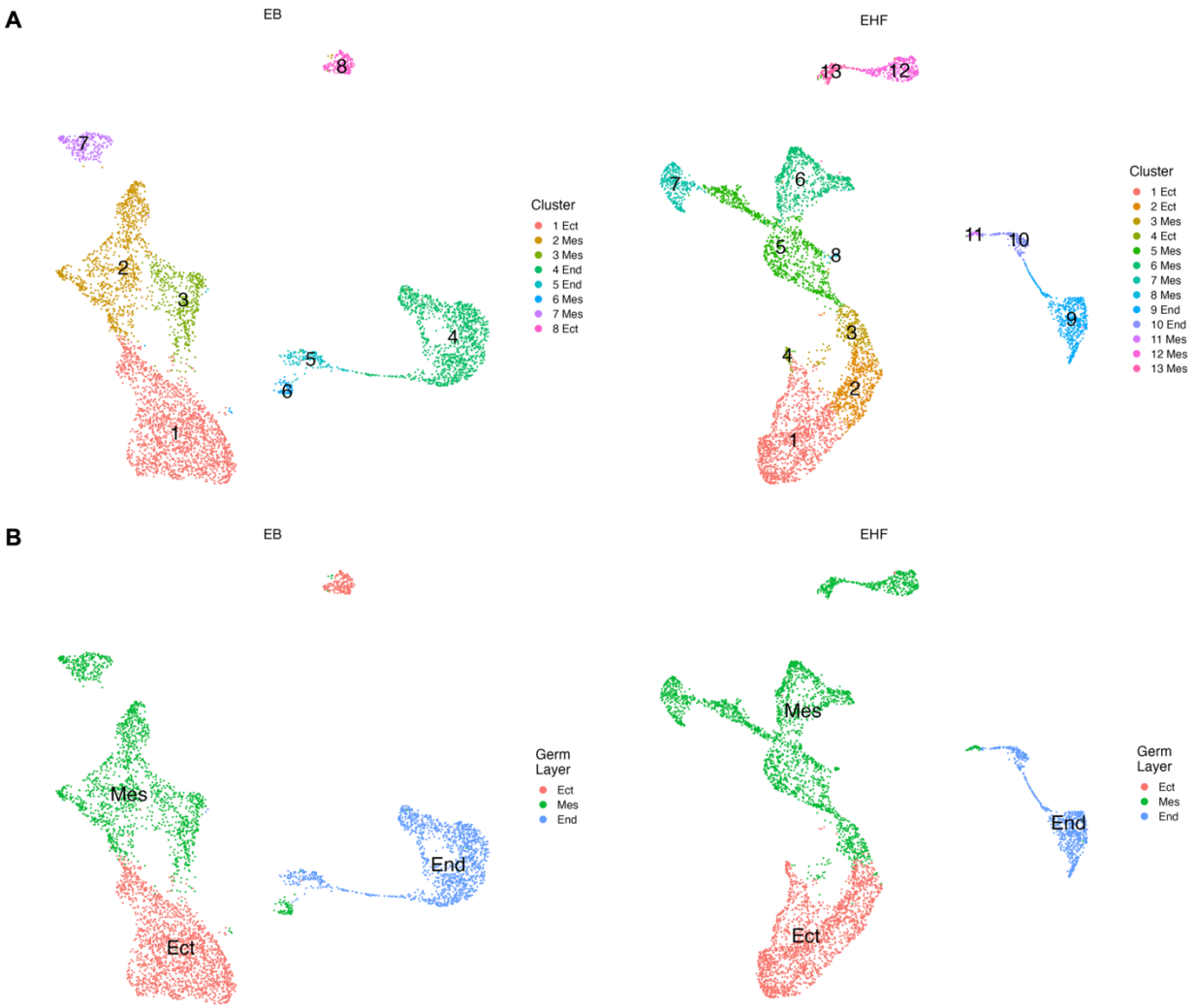


Fig. S3.10. Clustering of EB- and EHF-stage cells from WT embryos and annotation thereof as either ectoderm, mesoderm, or endoderm.

UMAP of EB- and EHF-stage cells colored by (**A**) cell population or (**B**) germ layer.

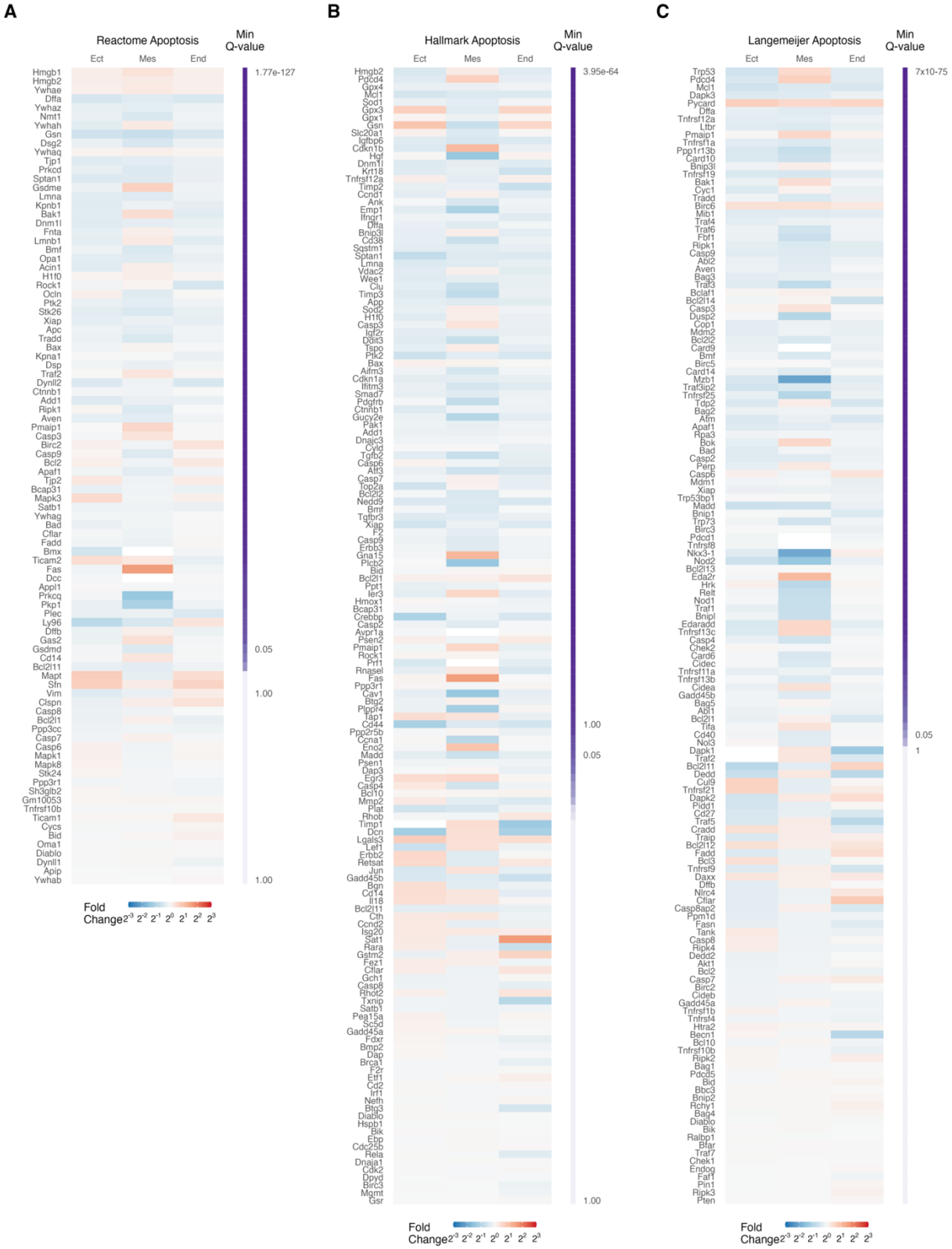

A
Reactome Apoptosis
Min Q-value
Ect
Mes
End
1.77e-127
0.05
1.00
1.00
Fold Change
B
Hallmark Apoptosis
Min Q-value
Ect
Mes
End
3.95e-64
1.00
0.05
1.00
Fold Change
C
Langemeijer Apoptosis
Min Q-value
Ect
Mes
End
7x10-75
0.05
1
Fold Change

Fig. S3.11. Expression of genes associated with apoptosis in *Nipbl*$^{+/-}$ embryos relative to WT.

The germ layers of *Nipbl*$^{+/-}$ embryos do not substantially overexpress or underexpress genes associated with apoptosis. Heatmap of fold change in expression in genes from the (**A**) Reactome Apoptosis, (**B**) Hallmark Apoptosis, and (**C**) Langemeijer Apoptosis gene sets in germ layers of *Nipbl*$^{+/-}$ embryos from that of WT embryos. Genes are ordered from top to bottom by minimum *Q*-value. *Q*-value from Bonferroni corrected *P*-value from Mann-Whitney U test.

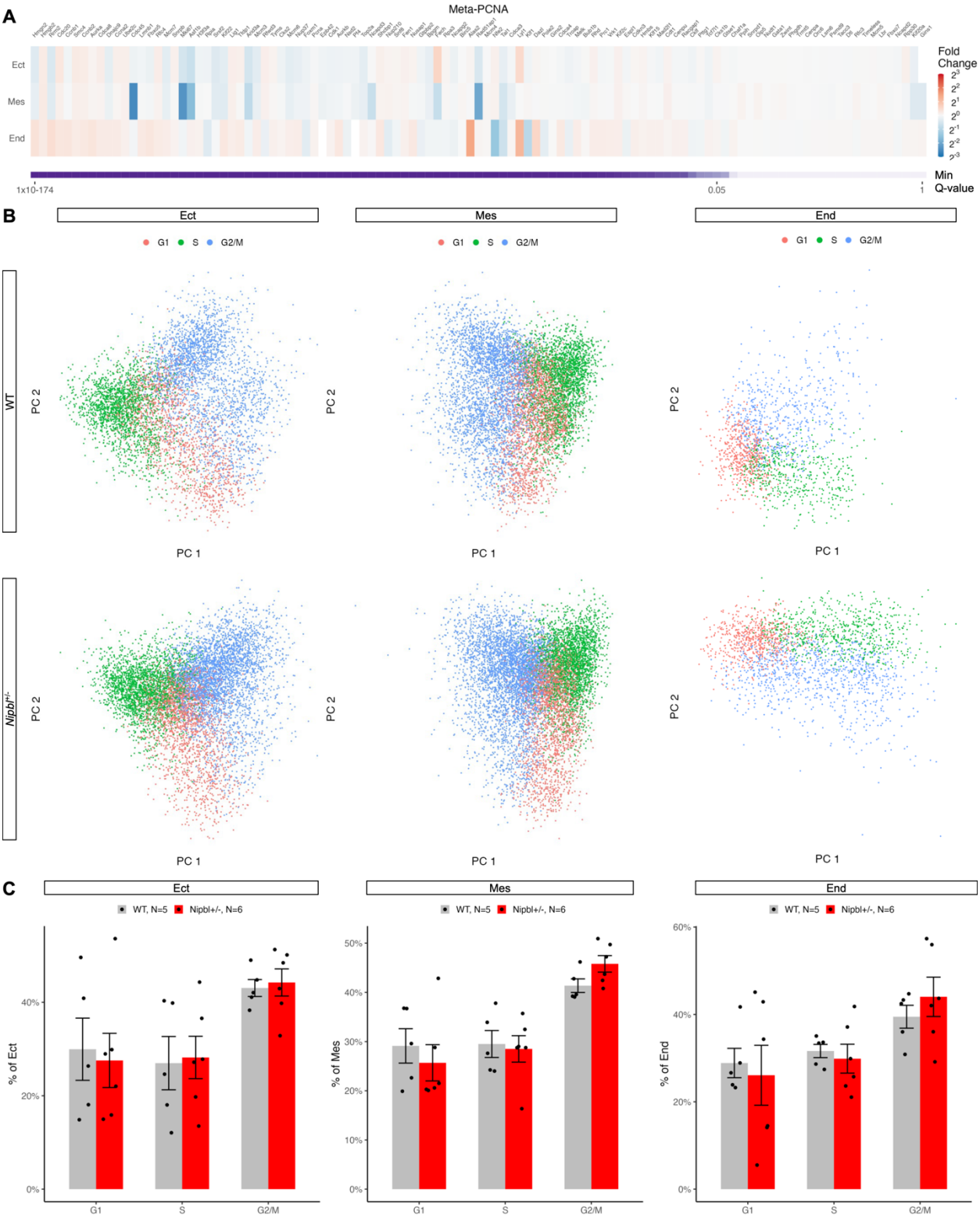


Fig. S3.12. Expression of meta-PCNA genes and cell cycle phase composition of *Nipbl*$^{+/-}$ embryos relative to WT.

The germ layers of *Nipbl*$^{+/-}$ embryos do not substantially overexpress or underexpress genes associated with cell proliferation. (**A**) Heatmap of fold change in expression in genes from the meta-PCNA gene set in germ layers of *Nipbl*$^{+/-}$ embryos from that of WT embryos. Genes are ordered from top to bottom by minimum *Q*-value. *Q*-value from Bonferroni corrected *P*-value from Mann-Whitney U test. (**B**) Principal component analysis was performed cells from each germ layer of WT and *Nipbl*$^{+/-}$ embryos using cell cycle phase genes marking S phase and G2/M phase. Seurat was used to assign cells into G1, S, or G2/M phase based on the expression of these markers. Expression of these marker sets are considered anticorrelated. When cells express neither, they are considered to be in G1 phase. In all germ layers, cells from WT and *Nipbl*$^{+/-}$ embryos assigned into all three phases. (**C**) Percentage of cells in each cell cycle phase across all germ layers of WT and *Nipbl*$^{+/-}$ embryos. There was no statistically significant difference in the percentage of cells in each cell cycle phase between that of WT and *Nipbl*$^{+/-}$ embryos. Error bars show standard error of the mean.

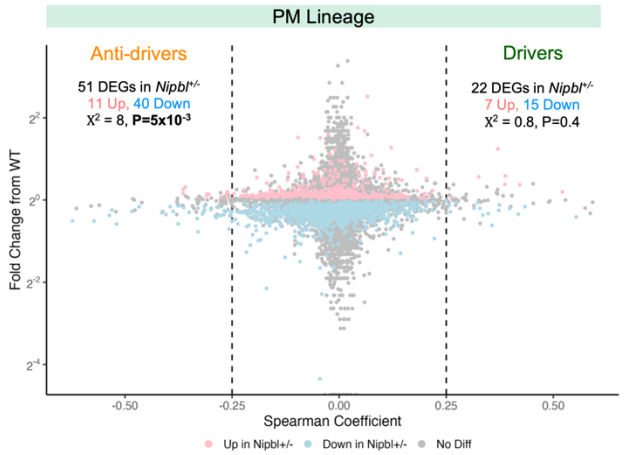


Fig. S3.13. Paraxial mesoderm lineage of LB-stage *Nipbl*$^{+/-}$ embryos, showing underexpression of paraxial mesoderm anti-drivers.

Fold change in expression of genes in mesoderm cells of PM lineage of LB-stage *Nipbl*$^{+/-}$ embryos differentially expressed ($Q < 0.05$, Mann-Whitney U Test) from that of WT embryos along Spearman's Rank Correlation coefficient from Fig. 3.5E.

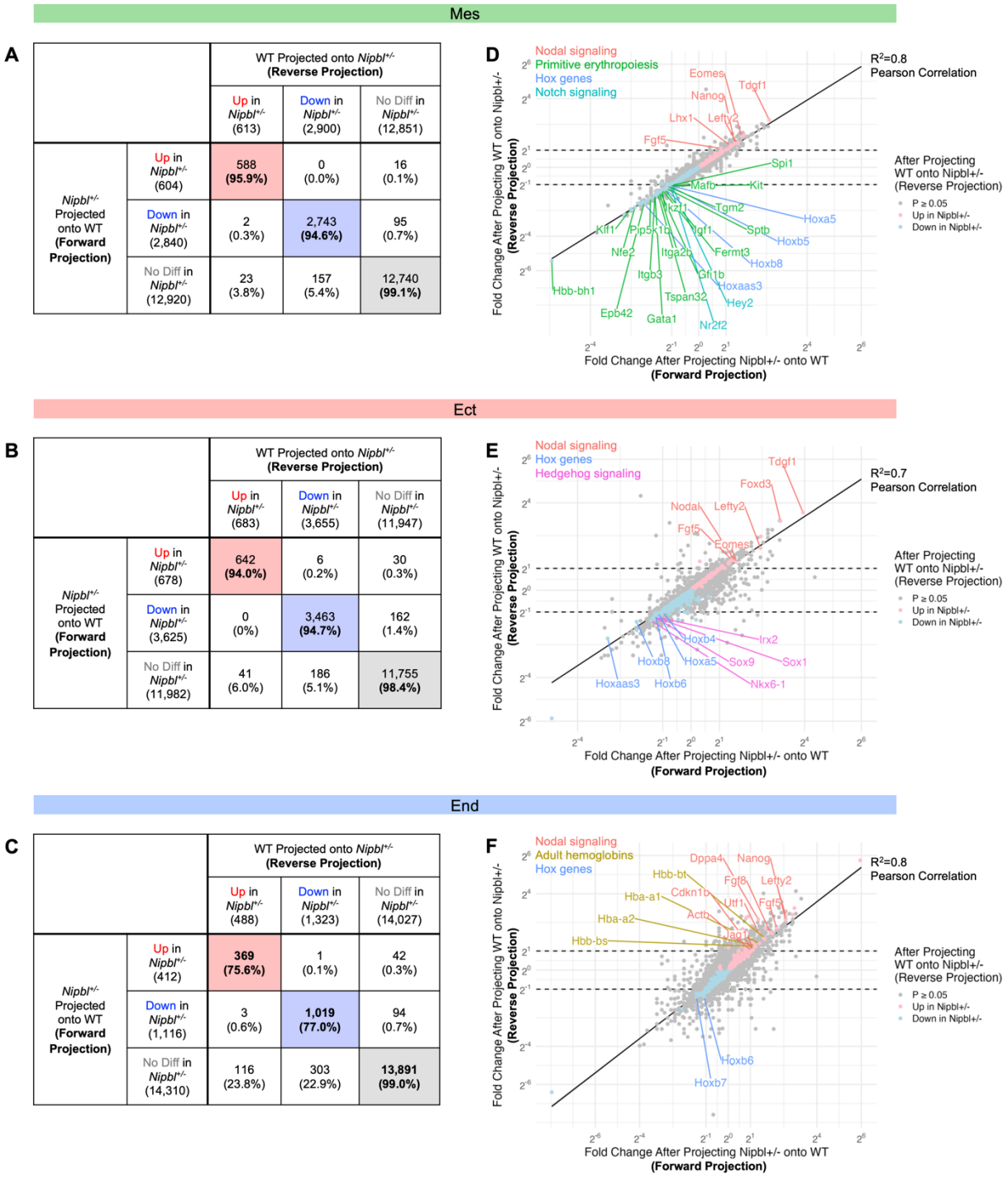


Fig. S3.14. Differentially expressed genes between germ layers of WT and *Nipbl*$^{+/-}$ embryos, determined following projection of WT cells onto *Nipbl*$^{+/-}$ germ layers.

Percentage of differentially expressed genes ($Q < 0.05$, Mann-Whitney U Test) in (**A**) mesoderm, (**B**) ectoderm, and (**C**) endoderm of LB-stage *Nipbl*$^{+/-}$ embryos resulting from the projection of WT cells onto *Nipbl*$^{+/-}$ germ layers (reverse projection) matching those differentially expressed in the germ layers of *Nipbl*$^{+/-}$ embryos resulting from the projection of *Nipbl*$^{+/-}$ cells onto WT germ layers (forward projection). Fold change in expression of differentially expressed genes ($Q < 0.05$, Mann-Whitney U Test) in the (**D**) mesoderm, (**E**) ectoderm, and (**F**) endoderm of LB-stage *Nipbl*$^{+/-}$ embryos following projection of WT cells onto *Nipbl*$^{+/-}$ germ layers (reverse projection) versus the fold change in expression of the same genes in the germ layers of WT embryos following projection of *Nipbl*$^{+/-}$ cells onto WT germ layers (forward projection).

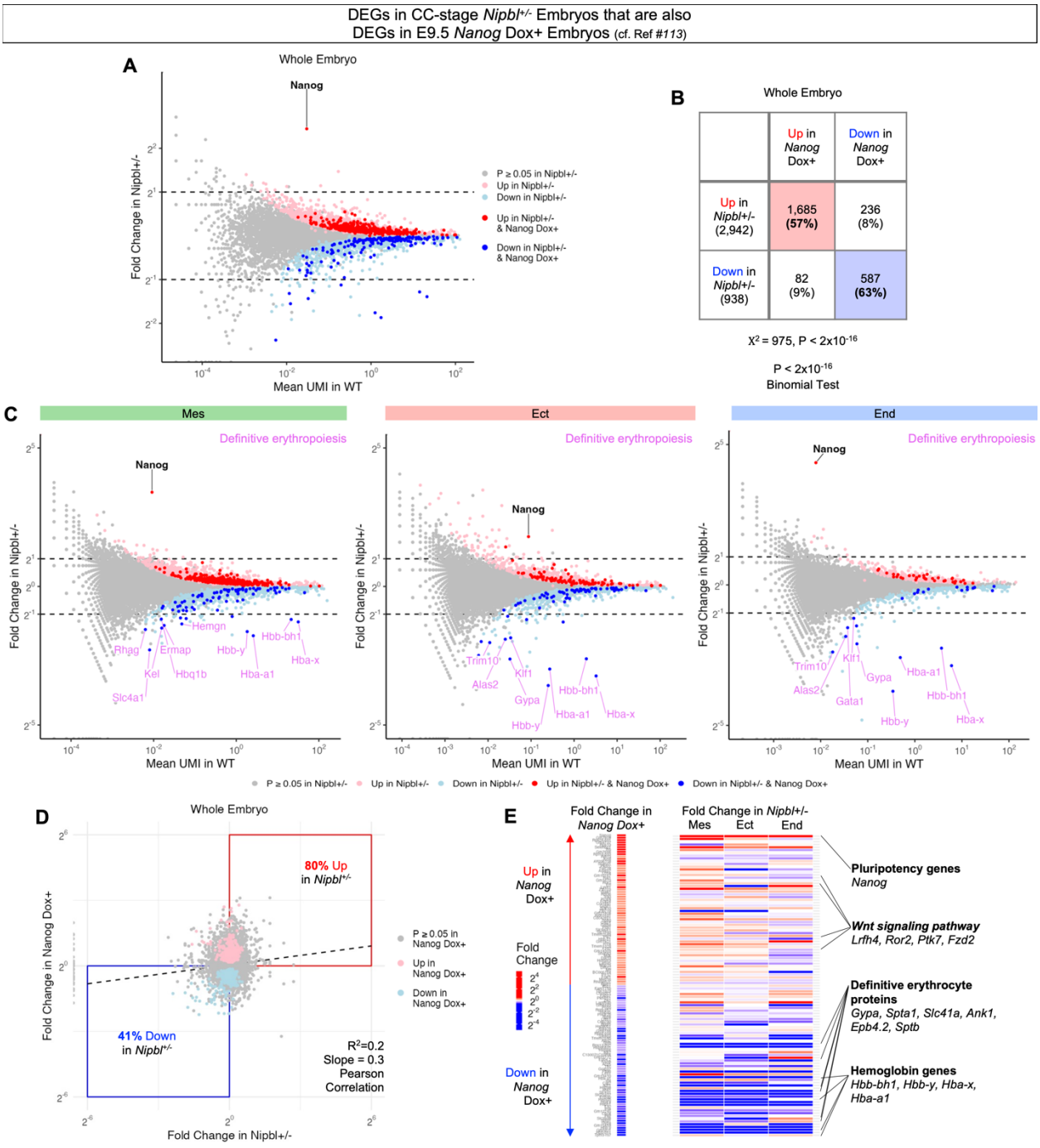


Fig. S3.15. CC-stage *Nipbl*$^{+/-}$ mice replicate the gene expression changes of *Nanog* overexpression.

(**A**) Fold change in expression of genes in CC-stage *Nipbl*$^{+/-}$ embryos that were differentially expressed ($Q < 0.05$, Mann-Whitney U Test) from WT embryos along with their average

expression in WT embryos. Genes that are were also differentially expressed ($Q < 0.05$, T-Test) in E9.5 mouse embryos transgenic for doxycycline inducible expression of *Nanog* (*Nanog* Dox+) from that of WT embryos (*Nanog* Dox-) from *(113)* are colored dark red (upregulated) and dark blue (downregulated). (**B**) Percentage of genes in CC-stage *Nipbl*$^{+/-}$ embryos that were differentially expressed ($Q < 0.05$, Mann-Whitney U Test) from WT embryos, and that were also differentially expressed in the same direction in E9.5 *Nanog* Dox+ embryos. At CC-stage, 57% of overexpressed and 63% of underexpressed genes in *Nipbl*$^{+/-}$ embryos were also overexpressed or underexpressed, respectively, in E9.5 *Nanog* Dox+ embryos. (**C**) Fold change in expression of genes in the germ layers of CC-stage *Nipbl*$^{+/-}$ embryos that were differentially expressed ($Q < 0.05$, Mann-Whitney U Test) from WT embryos along with their average expression in WT embryos. Genes that are were also differentially expressed ($Q < 0.05$, T-Test) in E9.5 *Nanog* Dox+ embryos are colored dark red and dark blue. At CC-stage, all germ layers of *Nipbl*$^{+/-}$ embryos showed downregulation of genes associated with definitive erythropoiesis, encompassing adult hemoglobins including *Hbb-y, Hba-a1,* and *Hba-x*. (**D**) Fold change in expression of genes in CC-stage *Nipbl*$^{+/-}$ embryos versus their fold change in E9.5 *Nanog* Dox+ embryos. Genes differentially expressed ($Q < 0.05$, T-Test) in E9.5 *Nanog* Dox+ embryos are colored dark red and dark blue. At CC-stage, the slope of the regression line comparing misexpression in *Nipbl*$^{+/-}$ and E9.5 *Nanog* Dox+ embryos was found to be 0.3 (panel D). Additionally, at CC-stage, *Nipbl*$^{+/-}$ embryos upregulated 80% of the same genes that were overexpressed in E7.5 *Nanog* Dox+ embryos and downregulated 41% of the same genes underexpressed in E7.5 *Nanog* Dox+ embryos. (**E**) Heatmap of fold change in expression in germ layers of CC-stage *Nipbl*$^{+/-}$ embryos from that of WT embryos of the most differentially expressed genes from E9.5 *Nanog* Dox+ embryos from that of Nanog Dox- embryos (lowest *Q*-

values from T-Test) from *(113)*. Genes are ordered from top to middle (overexpressed in *Nanog* Dox+ embryos) and bottom to middle (underexpressed in *Nanog* Dox+ embryos) by lowest to higher *Q*-value. Underexpressed genes in *Nanog* Dox+ embryos were strongly mirrored in the germ layers of CC-stage *Nipbl*$^{+/-}$ embryos (panel E). In CC-stage *Nipbl*$^{+/-}$ embryos, only a few overexpressed genes observed in *Nanog* Dox+ embryos were replicated, the most prominent of which was *Nanog* itself.

Table S3.3.1. Number of cells, transcripts, and genes captured per LB- and CC-stage WT and *Nipbl*$^{+/-}$ embryos.

scRNAseq of LB- and CC-stage WT and *Nipbl*$^{+/-}$ embryos captured thousands of cells per embryo, thousands of transcripts per cell, and detected thousands of genes per cell. Table shows stage of each embryo, genotype of each embryo, number of cells captured per embryo, median number of UMIs captured per cell, and median number of genes detected per cell.

Stage	Genotype	Sample	Cells	Median UMIs/Cell	Median Genes/Cell
LB	WT	LB1	2,445	17,998	4,242
LB	WT	LB2	1,247	17,011	3,947
LB	WT	LB3	2,537	16,476	3,830
LB	WT	LB4	7,119	18,624	3,940
LB	WT	LB5	2,321	18,413	4,076
LB	WT	LB1-LB5	15,669	18,078	4,001
LB	*Nipbl*$^{+/-}$	LB6	1,700	9,953.5	2,997
LB	*Nipbl*$^{+/-}$	LB7	1,411	18,111	4,110
LB	*Nipbl*$^{+/-}$	LB8	4,218	11,875	3,168
LB	*Nipbl*$^{+/-}$	LB9	3,969	15,752	3,702
LB	*Nipbl*$^{+/-}$	LB10	3,431	17,123	3,847
LB	*Nipbl*$^{+/-}$	LB11	4,840	18,910.5	4,000.5
LB	*Nipbl*$^{+/-}$	LB6-LB11	19,569	16,424	3,780
LB	*WT & Nipbl*$^{+/-}$	LB1-LB11	35,238	17,300	3,894
CC	WT	CC1	1858	15,682	3,796
CC	WT	CC2	3537	14,967	3,711
CC	WT	CC3	4550	13,788	3,609
CC	WT	CC4	5202	14,782	3,549
CC	WT	CC5	7448	14,882	3,453
CC	WT	CC6	7759	14,547	3,438
CC	WT	CC7	4767	15,386	3,716
CC	WT	CC8	6573	15,797	3,652
CC	WT	CC1-CC8	41,694	14,916	3,580
CC	*Nipbl*$^{+/-}$	CC9	2104	15,861	3,837
CC	*Nipbl*$^{+/-}$	CC10	2197	14,968	3,643
CC	*Nipbl*$^{+/-}$	CC11	669	16,370	3,994

CC	*Nipbl*$^{+/-}$	CC12	1614	15,140	3,750
CC	*Nipbl*$^{+/-}$	CC13	2025	15,525	3,797
CC	*Nipbl*$^{+/-}$	CC14	3683	13,302	3,493
CC	*Nipbl*$^{+/-}$	CC15	4919	14,986	3,656
CC	*Nipbl*$^{+/-}$	CC16	8236	15,888	3,699
CC	*Nipbl*$^{+/-}$	CC9-16	25,447	15,122	3,677
CC	*WT & Nipbl*$^{+/-}$	CC1-CC16	67,141	14,997	3,618
LB & CC	*WT & Nipbl*$^{+/-}$	LB1-LB11 & CC1-CC16	102,379	15,429	3,676

Table S3.3.2. Differential overexpression of markers of germ layer identity by cell populations in LB- and CC-stage embryos.

Stage	Cell Population	Markers	Germ Layer
LB	1	*Dnmt3b, Utf1 (31), Pou3f1 (33), Eras, Sox2 (32), Epcam*	Ectoderm
LB	2	*Gbx2, Pou5f1, Cdx2, Cdkn1a*	Ectoderm
LB	3	*Tfap2a, Gata3, Tfap2c*	Ectoderm
LB	4	*T (34), Mixl1, Wnt3a, Nkx1-2, Pou5f1, Nkx2-1*	Mesoderm
LB	5	*Mesp1, Hand1 (35), Foxf1, Fn1, Snai1, Isl1, Bmp4, Vim (36)*	Mesoderm
LB	6	*Hoxb1, Tbx6, Dll3, Dll1, Lef1, Frzb*	Mesoderm
LB	7	*Dppa3, Psmb8, Ifitm3, Tapbp, Irf1*	Mesoderm
LB	8	*Smarcd3, Myl7, Gata6, Cdkn1c, Mef2c, Meis1*	Mesoderm
LB	9	*Foxc2, Twist1 (37), Fn1, Hmga2*	Mesoderm
LB	10	*Foxc2, Twist1 (37), Pitx2, Foxc1, Cd24a, Hmga2, Prrx2 (38)*	Mesoderm
LB	11	*Hand1 (35), Bmp4, Foxf1, Msx2*	Mesoderm
LB	12	*Hand1 (35), Pitx1, Bambi, Foxf1, Bmp4*	Mesoderm
LB	13	*Cubn, Apob, Amn, Apom, Apoa4, Dab2, Pla2g12b*	Endoderm
LB	14	*Cldn6, Krt18, Krt8, Foxa2, Sox17, Cer1, Cldn7*	Endoderm
LB	15	*Tex19.1, Dppa4, Zfp4*	Ectoderm
LB	16	*Tal1, Kdr, Lmo2, Etv2*	Mesoderm
LB	17	*Gata1, Tal1, Lmo2, Hbb-bh1, Gfi1b*	Mesoderm

LB	18	*1110017D15Rik, Fam183b, Tekt1, Dynlrb2*	Mesoderm
CC	1	*Lef1, Tbx6, Dll3, Hes7, Hoxb1, Dll1, Cdx1*	Mesoderm
CC	2	*Hoxd9, Lef1, Msx1, Hoxb6, Hoxb8, Nkd1, Msx2, Hoxb9*	Mesoderm
CC	3	*Mdk, Vim (36), Hand2, Arg1, Fn1, Nrp1, Pdgfra, Ppp3ca*	Mesoderm
CC	4	*Dppa3, Psmb8, Ifitm3, Gdf3, Psme1, Irf1*	Mesoderm
CC	5	*Twist1 (37), Prrx2 (38), Foxc2, Foxd1, Prrx1*	Mesoderm
CC	6	*Meis1, Gata6, Isl1, Arg1, Mef2c, Meis2, Smarcd3, Hand2, Nr2f2, Irx3*	Mesoderm
CC	7	*Foxc2, Foxc1, Prrx2 (38), Twist1 (37)*	Mesoderm
CC	8	*Arg1, Pkm, Tubb5, Trim28, Eef1a1, Sox2 (32), Actb, Igf2bp1, Pou3f1 (33)*	Ectoderm
CC	9	*Acta2, Krt8, Krt18, Tagln*	Mesoderm
CC	10	*Phlda2, Cdkn1c, Peg10, Slc38a4*	Mesoderm
CC	11	*Etv2, Egfl7, Ramp2, Kdr, Fli1, Hhex, Tal1, Lmo2, Flt1, Rasip1, Esam*	Mesoderm
CC	12	*Myl7, Actc1, Tnnt2, Myl4*	Mesoderm
CC	13	*Pou3f1 (33), Sox2 (32), En1, Pax2, Dnmt3b*	Ectoderm
CC	14	*Hoxa1, Pax6, Sox2 (32), Meis2, Sox3*	Ectoderm

CC	15	*Six3, Hesx1, Pou3f1 (33), Otx2, Sox2 (32)*	Ectoderm
CC	16	*Pax3, Sox9, Foxd3*	Ectoderm
CC	17	*Ptch1, Foxa2, Sox2 (32), Nkx6-1, Pou3f1 (33), Sfrp1*	Ectoderm
CC	18	*Cdx2, Nr6a1, Pou5f1, Ccnd1*	Mesoderm
CC	19	*Fst, Fgf8, T (34), Etv4*	Mesoderm
CC	20	*Msx1, Cdx4, Cdx2, Fgf8, Etv4, T (34), Evx1, Klf5, Nkx1-2*	Mesoderm
CC	21	*Epcam, Cldn7, Rab25, Cldn6, Krt18, Krt8*	Ectoderm
CC	22	*Krt18, Epcam, Krt8*	Endoderm
CC	23	*Shh, T (34), Nog, Foxa2*	Mesoderm
CC	24	*Apob, Amn, Afp, Apom, Rbp4, Ttr (39), Apoa4, Apoa1, Cubn*	Endoderm
CC	25	*Ttr (39), Rbp4, Apoa1, Apom*	Endoderm
CC	26	*Hba-x, Hbb-bh1, Hba-a1, Hba-a2*	Mesoderm

Table S3.3. Differential overexpression of markers of biological identity by endodermal cell populations in LB- and CC-stage embryos.

Stage	Cell Population	Markers	Biological Identity
LB	13	*Hnf4a (42), Tbx3 (43), S100a1, Rbp4, Cubn, Mt1, Slc2a2, Apoa2, Ctsb, Aldob, Apom, Lrp2, Cltc, Aprt*	Extraembryonic endoderm
LB	14	*Sox4 (40) Mdk, Tead2, Cer1, Ccnd2, Dkk1, Sfrp1, Foxa2 (41), Bambi, Fzd2, Ccnd1*	Definitive endoderm progenitor
CC	22	*Krt18, Slc2a1, Cd24a, Epcam, Ptma, Hmgn1, Tpt1, Rps2, Bsg, Rps23, Cfl1, Ppia,*	Definitive endoderm progenitor
CC	24	*Esx1 (44), Ldha, Fam162a, Eno1, Slc16a3, Pgk1, Tubb5, Pkm, Aldoa, Hspa8*	Visceral endoderm
CC	25	*Rbp4, Apoa1, Ttr, Apom*	Extraembryonic endoderm

Table S3.4. Differential overexpression of markers of biological identity by ectodermal cell populations in LB- and CC-stage embryos.

Stage	Cell Population	Markers	Biological Identity
LB	1	*Dnmt3b, Pou3f1 (33), Otx2 (45), Utf1, Eras, Irx3, Hesx1, Sox2*	Ectoderm progenitor
LB	2	*Cdx1 (46), Nkx1-2 (47), Cdx2, Fst, Gbx2, Pou5f1, Cdkn1a, Fgf8*	Neural ectoderm progenitor
LB	3	*Id1 (45) Msx2 (45), Msx1 (45), Tfap2a*	Surface ectoderm
LB	15	*Elf5 (49), Dppa4, Tex19.1, Grem2, Bmp8b, Sox7, Tead4 (50), Esrrb, Dppa2, Gata6, Nog, Zfp42, Sox17*	Extraembryonic ectoderm
CC	8	*Sox2 (60), Ezr (61), Arg1, Fn1, Twist1, Foxc2, Prrx2, Hmga2*	Spinal neural plate
CC	13	*En1 (54), Pax2 (55), Pou3f1, Zic3, Zeb2, Wnt1, Sox2, Pax5, Sox3*	Mesencephalic neural plate
CC	14	*Hoxa1 (56), Hes3 (57), Mdk, Meis2, Tshz1, Pax6, Pou5f1, Ptn*	Rhombencephalic neural plate
CC	15	*Six3 (52), Hesx1 (53), Pou3f1, Otx2, Lhx5, Fezf1, Rax*	Prosencephalic neural plate
CC	16	*Pax3 (62), Foxd3 (63), Sox9*	Neural crest
CC	17	*Ptch1, Foxa2 (58), Nkx2-9 (59), Nkx6-1, Cfc1, Sox2, Gli1, Lefty1, Ccnd1, Sfrp1*	Midline neural plate
CC	21	*Krt8, Krt18, Epcam*	Surface ectoderm

Table S3.5. Differential overexpression of markers of biological identity by mesodermal cell populations in LB- and CC-stage embryos.

Stage	Cell Population	Markers	Biological Identity
LB	4	*Pou5f1, T (34), Nkx1-2 (47), Epcam, Cdh1, Wnt3a, Mixl1, Cdx2, Mgst1, Dnmt3b, Cdkn1a*	Neuromesodermal progenitor
LB	5	*Mesp1 (64), Hand1, Isl1, Cdx2, Evx1*	*Mesp1*-expressing mesoderm progenitor
LB	6	*Tbx6 (67), Meis2 (68), Hoxb1, Dll1, Dll3, Lef1, Frzb*	Paraxial mesoderm progenitor
LB	7	*Dppa3, Ifitm3, Psmb8, Tfap2c (77), Msx1 (78), Psme2, Psme1, Irf1, Tapbp*	Primordial germ cell
LB	8	*Id2 (65), Smarcd3, Myl7, Cdkn1c, Mef2c (66), Gata6*	Second heart field
LB	9	*Foxc2 (69), Tcf15 (70), Twist1, Cd24a, Tcf7l1, Lef1*	Trunk paraxial mesoderm
LB	10	*Alx1 (70), Tcf15 (70), Cd24a, Pitx2, Twist1, Foxc2, Foxc1, Prrx2, Six1, Bmp7, Jag1*	Head paraxial mesoderm
LB	11	*Hand1 (35), Bmp4, Msx2, Foxf1*	First heart field 2
LB	12	*Pitx1, Hand1 (35), Bambi, Krt8, Msx1, Krt18, Foxf1, Bmp4, Col1a1, Msx2, Cdx2*	First heart field 1
LB	16	*Kdr, Tal1 (72), Lmo2, Etv2 (73)*	Hemangioblast
LB	17	*Gata1 (72), Hbb-bh1, Tal1 (72), Runx1*	Primitive erythrocyte
LB	18	*Noto (75), Foxj1 (76), 1110017D15Rik, Tekt1, Pifo, Enkur, Fam183b*	Primitive node

CC	1	*Tbx6, Dll3, Lef1, Hes7, Dll1, Hoxb1, Cdx1*	Trunk paraxial mesoderm
CC	2	*Mesp1 (64), Hand1, Msx1, Lef1, Msx2, Cdx2, Nkd1, Sall4, Tbx3*	*Mesp1*-expressing mesoderm progenitor 1
CC	3	*Mesp1 (64), Meis1, Meis2, Hoxa1, Tshz1, Hoxa5, Foxp1, Pbx3*	*Mesp1*-expressing mesoderm progenitor 2
CC	4	*Dppa3, Psmb8, Ifitm3, Irf1, Dnd1, Gdf3, Psme1, Parp12, Psme2, Nanos3, Ifitm2, Tapbp*	Primordial germ cell
CC	5	*Twist1, Prrx2, Foxc2, Foxd1, Prrx1*	Head paraxial mesoderm
CC	6	*Meis1, Gata6, Isl1, Irx3, Meis2, Irx5*	Second heart field
CC	7	*Tcf15 (70), Meox1 (82), Foxc2, Ptn, Pax3, Foxc1, Cadm1, Mdk, H3f3a, Epb41l3, Foxd1, Fbn2, Six1, Ncam1, Notch1, Prrx2, Efnb2, Pdgfra, Twist1, Zic2*	Presomitic mesoderm
CC	9	*Krt18, Krt8, Tagln, Hand1, Acta2, Sparc, Cnn2, Vim, Bmp4, Csrp2, Nrp1, Serpinh1, Myl6*	First heart field 2
CC	10	*Hand1, Bambi, Bmp4, Twist1*	First heart field 1
CC	11	*Etv2, Ramp2, Egfl7, Kdr, Fli1, Hhex, Rasip1, Flt1, Tal1, Esam, Lmo2*	Hemangioblast
CC	12	*Mef2c (80), Nkx2-5 (81), Myl7, Actc1, Tnnt2, Myl4*	Cardiomyocyte
CC	18	*Cdx2, Sox2, Nr6a1, Pou5f1*	Neuromesodermal progenitor

CC	19	*T, Fgf8, Fst, Etv4, Fgf17*	Paraxial mesoderm progenitor
CC	20	*Tlx2 (79), Msx1, Cdx4, Cdx2, Evx1os, Etv4, Fgf8, T (34)*	Lateral plate mesoderm progenitor
CC	23	*Shh, Foxa2, Nog, Krt18*	Primitive node
CC	26	*Hba-x, Hbb-bh1, Hba-a1, Hba-a2*	Primitive erythrocyte

Table S3.6. Differential overexpression of markers of germ layer identity by cell populations in LB-stage *Nipbl*$^{+/-}$ embryos.

Stage	Cell Population	Markers	Germ Layer
LB	1	*Dnmt3b, Utf1 (31)*	Ectoderm
LB	2	*Dlx5, Tfap2a*	Ectoderm
LB	3	*Pou5f1, Dnmt3b, Hes3*	Ectoderm
LB	4	*Cdx1, Mixl1, Pou5f1*	Mesoderm
LB	5	*Msx1, Tfap2a*	Ectoderm
LB	6	*Foxc2, Lefty2*	Mesoderm
LB	7	*Mesp1, Rspo3, Isl1, Wnt2*	Mesoderm
LB	8	*Myl7, Smarc3d, Cdkn1c, Gata6, Phlda2, Mef2c*	Mesoderm
LB	9	*Tcf15, Foxc2*	Mesoderm
LB	10	*Hand1 (35), Bmp4*	Mesoderm
LB	11	*Hoxb6, Hand1 (35), Hoxd9, Pitx1, Foxf1*	Mesoderm
LB	12	*Mesp1, Mixl1*	Mesoderm
LB	13	*Dppa3, Dnd1*	Mesoderm
LB	14	*Hand1 (35), Krt18, Krt8, Foxf1, Bmp4,*	Mesoderm
LB	15	*Tbx6, Hoxb1*	Mesoderm
LB	16	*Tal1, Lmo2*	Mesoderm
LB	17	*Ttr, Apoa1*	Endoderm
LB	18	*Slc116a1, Emb*	Endoderm
LB	19	*Pifo, Cfap126*	Mesoderm
LB	20	*Tex19.1, Dppa4*	Endoderm
LB	21	*Hs3st1, Srgn*	Endoderm

Table S3.7. Differential overexpression of markers of biological identity by mesodermal cell populations in LB-stage *Nipbl*$^{+/-}$ embryos.

Stage	Cell Population	Markers	Biological Identity
LB	4	*Pou5f1, T (34)*	Neuromesodermal progenitor
LB	6	*Foxc2 (69), Tcf15 (70)*	Trunk paraxial mesoderm
LB	7	*Mesp1 (64), Rspo3, Hoxb1, Wnt2, Isl1, Meis2 (68)*	Mesp1-expressing mesoderm progenitor 3
LB	8	*Myl7, Smarcd3, Gata6, Cdkn1c, Mef2c (66)*	Second heart field
LB	9	*Otx2, Cer1*	Head paraxial mesoderm
LB	10	*Bmp4, Hand1 (35)*	Mesp1-expressing mesoderm progenitor 2
LB	11	*Hoxb6, Hoxd9*	First heart field 2
LB	12	*Mixl1, Mesp1 (64)*	Mesp1-expressing mesoderm progenitor 1
LB	13	*Dppa3, Psmb8, Dnd1, Nanos3, Ifitm3, Prdm14*	Primordial germ cell
LB	14	*Krt18, Krt8*	First heart field 2
LB	15	*Tbx6 (67), Rbp1, Hoxb1, Aldh1a2*	Paraxial mesoderm progenitor
LB	16	*Tal1 (72), Lmo2*	Hemangioblast
LB	19	*Pifo, Cfap126, Foxj1 (76)*	Primitive node

Table S3.8. Number of cells, transcripts, and genes captured by scRNAseq of EB- and EHF-stage WT embryos.

scRNA-seq of EB- and EHF-stage of WT mouse embryos captured thousands of cells per embryo, thousands of transcripts per cell, and detected thousands of genes per cell. Table shows genotype of each embryo, number of cells captured per embryo, median number of UMIs captured per cell, and median number of genes detected per cell.

Embryo(s)	Genotype	Cells	Median UMIs/Cell	Median Genes/Cell
EB1	WT	4109	5693	2064
EB2	WT	1185	5943	2120
EB1-EB2	WT	5294	5740	2073.5
EHF1	WT	1969	11208	3271
EHF2	WT	3839	9587	2891
EHF1-EHF2	WT	5808	10162	3030

Table S3.9. Differential overexpression of markers of germ layer identity by cell populations in EB- and EHF-stage embryos.

Stage	Cell Population	Markers	Germ Layer
EB	1	*Pou5f1, Dnmt3b, Epcam, Utf1, Pkm*	Ectoderm
EB	2	*Vim, Hand1, Hmga2, Fn1, Actb, Fgf3, Tpm4, Myl7, Cald1*	Mesoderm
EB	3	*Lhx1, Eomes, Fgf3, Snai1, Fn1, Mixl1*	Mesoderm
EB	4	*Apom, Apoa1, Dab2, Ttr, Apob, Amn*	Endoderm
EB	5	*Cer1, Gsc, Lhx1, Sox17, Fgf5, Foxa2*	Endoderm
EB	6	*Foxj1, 1110017D15Rik, Tekt1, Pifo, Cfap126*	Mesoderm
EB	7	*Lmo2, Fgf3, Tal1, Runx1, Fli1, Kdr*	Mesoderm
EB	8	*Elf5, S100a6, Nr0b1, Anxa5, Rhox5, Trap1a, Tex19.1, Hspb1, Tfap2c, Dnmt3l, Dppa4, Cd9, Lgals1, Zfp42*	Ectoderm
EHF	1	*Dnmt3b, Gapdh, Epcam, Sox2, Pou3f1*	Ectoderm
EHF	2	*Pou5f1, Gbx2, Sox3, Cdx2*	Ectoderm
EHF	3	*T, Fst, Nkx1-2, Cdx2, Pou5f1, Wnt3a, Fgf8*	Mesoderm
EHF	4	*Tacstd2, Wnt6, Wnt4, Lrp2, Lrpap1, Krt18, Krt8, Wt1*	Ectoderm
EHF	5	*Lef1, Dll1, Pcdh19, Dll3, Tbx6, Cdh2*	Mesoderm
EHF	6	*Hmga2, Twist1, Prrx2, Foxc2*	Mesoderm
EHF	7	*Hand1, Krt18, Krt8, Foxf1, Tagln, Vim, Serpinh1, Ppic,*	Mesoderm

		Cnn2, Tpm1, Nrp1, Acta2, Bmp4, Myl6	
EHF	8	*Dppa3, Ifitm3, Gdf3, Dnd1, Parp9, Tfap2c, Klf2, Parp12, Prdm1*	Mesoderm
EHF	9	*Amn, Apom, Apoa1, Cubn, Apob, Ttr*	Endoderm
EHF	10	*Krt8, Krt18, Sox17, Foxa1, Foxa2*	Endoderm
EHF	11	*Foxj1, 1110017D15Rik, Dynlrb2, Cfap52, Fam183b, Cfap126*	Mesoderm
EHF	12	*Lmo2, Gata1, Hbb-bh1, Runx1*	Mesoderm
EHF	13	*Kdr, Fli1, Etv2, Lmo2*	Mesoderm

Table S3.3.10. Expression of *Nanog* in LB- and CC-stage embryos.

Genotype = genotype of embryo, Embryo = embryo identifier, *Nanog* = normalized transcript counts for *Nanog* in cell type.

Stage	Genotype	Embryo	*Nanog*
Late bud	Wildtype	LB1	0.05051302
Late bud	Wildtype	LB2	0.11359867
Late bud	Wildtype	LB3	0.08
Late bud	Wildtype	LB4	0.10866373
Late bud	Wildtype	LB5	0.07430341
Late bud	*Nipbl*$^{+/-}$	LB6	0.3121175
Late bud	*Nipbl*$^{+/-}$	LB7	0.20864662
Late bud	*Nipbl*$^{+/-}$	LB8	0.24953096
Late bud	*Nipbl*$^{+/-}$	LB9	0.18756219
Late bud	*Nipbl*$^{+/-}$	LB10	0.16984127
Late bud	*Nipbl*$^{+/-}$	LB11	0.19027661
Cardiac crescent	Wildtype	CC1	0.00668338
Cardiac crescent	Wildtype	CC2	0.01766304
Cardiac crescent	Wildtype	CC3	0.01175446
Cardiac crescent	Wildtype	CC4	0.00583771
Cardiac crescent	Wildtype	CC5	0.01085305
Cardiac crescent	Wildtype	CC6	0.00340408
Cardiac crescent	Wildtype	CC7	0.010921
Cardiac crescent	Wildtype	CC8	0.01019946
Cardiac crescent	*Nipbl*$^{+/-}$	CC9	0.15524194
Cardiac crescent	*Nipbl*$^{+/-}$	CC10	0.12248629
Cardiac crescent	*Nipbl*$^{+/-}$	CC11	0.08695652
Cardiac crescent	*Nipbl*$^{+/-}$	CC12	0.12486993
Cardiac crescent	*Nipbl*$^{+/-}$	CC13	0.11223629
Cardiac crescent	*Nipbl*$^{+/-}$	CC14	0.09711432
Cardiac crescent	*Nipbl*$^{+/-}$	CC15	0.07511408
Cardiac crescent	*Nipbl*$^{+/-}$	CC16	0.07357057

Table S3.3.11. Expression of *Pou5f1* in LB- and CC-stage embryos.

Genotype = genotype of embryo, Embryo = embryo identifier, *Pou5f1* = normalized transcript counts for *Pou5f1* in cell type.

Stage	Genotype	Embryo	*Pou5f1*
Late bud	Wildtype	LB1	8.3875296
Late bud	Wildtype	LB2	8.28772803
Late bud	Wildtype	LB3	8.98193548
Late bud	Wildtype	LB4	10.6762115
Late bud	Wildtype	LB5	7.02786378
Late bud	*Nipbl*$^{+/-}$	LB6	7.82741738
Late bud	*Nipbl*$^{+/-}$	LB7	12.1409774
Late bud	*Nipbl*$^{+/-}$	LB8	8.35897436
Late bud	*Nipbl*$^{+/-}$	LB9	10.3502488
Late bud	*Nipbl*$^{+/-}$	LB10	6.68201058
Late bud	*Nipbl*$^{+/-}$	LB11	7.03855826
Cardiac crescent	Wildtype	CC1	0.66165414
Cardiac crescent	Wildtype	CC2	1.05661232
Cardiac crescent	Wildtype	CC3	0.48193296
Cardiac crescent	Wildtype	CC4	0.83712785
Cardiac crescent	Wildtype	CC5	0.76383764
Cardiac crescent	Wildtype	CC6	0.43211854
Cardiac crescent	Wildtype	CC7	1.3815071
Cardiac crescent	Wildtype	CC8	0.96033545
Cardiac crescent	*Nipbl*$^{+/-}$	CC9	0.92271505
Cardiac crescent	*Nipbl*$^{+/-}$	CC10	0.74710542
Cardiac crescent	*Nipbl*$^{+/-}$	CC11	4.35652174
Cardiac crescent	*Nipbl*$^{+/-}$	CC12	1.9760666
Cardiac crescent	*Nipbl*$^{+/-}$	CC13	2.9164557
Cardiac crescent	*Nipbl*$^{+/-}$	CC14	0.91287458
Cardiac crescent	*Nipbl*$^{+/-}$	CC15	1.2965953
Cardiac crescent	*Nipbl*$^{+/-}$	CC16	1.39984006