INTEGRATING CHROMATIN AND DEVELOPMENT: LDB1 REGULATES GENOME ARCHITECTURE AND GENE EXPRESSION IN MOTOR NEURON DIFFERENTIATION

Justin Tate Demmerle, B.A.

Linacre College

Nuffield Department of Medicine

University of Oxford

NIH-Oxford-Cambridge Scholars Program



A thesis submitted for the degree of

Doctor of Philosophy

Biomedical Sciences

Hilary Term, 2019

Supervised by:

Dr. Lothar Schermelleh, Department of Biochemistry, University of Oxford, UK

Dr. Todd Macfarlan, NICHD, National Institutes of Health, USA

DECLARATION

I declare that this thesis is entirely my own work, except where otherwise stated has not been previously published, and except where otherwise stated describes my own research. Please see Appendix D for details.

Justin Demmerle
Linacre College

DEDICATION

This thesis is dedicated to my late father, Alan Demmerle, M.S.E.E.

1933-2019

And to the memory of my father's mother, Ida P. Rolf, Ph.D.

1896-1979

ABSTRACT

Dynamic changes in chromatin architecture underpin gene expression in development. The chromatin integrator protein LDB1 is a key regulatory node in the organization of developmentally-specific gene expression programs, particularly in motor neurons (MNs) of the vertebrate spinal cord. By coordinating transcription factor binding, chromatin remodeling, and 3D genome organization, LDB1 is an essential chromatin integrator protein in the development of MNs. In this thesis I characterise the deletion of LDB1 in an *in vitro* MN differentiation system in the mouse. By measuring transcription factor (TF) binding, protein association, gene expression, and 3D-genome organization, I specify the genetic elements that LDB1 regulates. I find that LDB1 regulates binding of the promiscuous TFs Isl1 and Lhx3 at a subset of critical MN genes, specified particularly in cooperation with the LDB1 binding partner SSDP1, where it drives open chromatin, H3K27ac deposition, gene expression, and likely enhancer function. Thus, LDB1 integrates multiple levels of specificity to regulate dozens of genes and enhancers that are critical to the expression of MN genes. I propose that this is a key feature of spatial genome regulation and is mediated by LDB1 and the interaction with its binding partner SSDP1. I identify a minimal set of MN-specific genes regulated by LDB1, as well as a set of elements that regulate subtle features of 3D genome organization and transcription. Together, I establish LDB1 as a critical regulator of 3D genome organization, transcription factor assembly, and gene expression, making LDB1 and genes downstream of its effects into promising targets for improved cellular reprogramming strategies.

ACKNOWLEDGEMENTS

I would like to sincerely thank:

My supervisors, Dr. Todd Macfarlan and Dr. Lothar Schermelleh, for giving me the freedom, means, and support to pursue such an ambitious and sometimes tenuous project. They both have my deepest appreciation and loyalty for taking a chance on me and guiding us to create great science together.

My examiners for my Transfer, Prof. Veronica Buckle and Prof. Rob Klose, and for my Confirmation, Prof. Klose and Dr. Jim Hughes, for being a voice of reason, clarifying my goals, and recommending a move achievable scientific approach at each juncture.

The admissions committee, directors, and supporters of the NIH-OxCam Program, including but not limited to Jim Sellers, Ted Pierson, Katie Soucy, and Randi Balletta of the Biomedical Research Alliance for their continued support.

The continued support of my scholarship and the OxCam program by NICHD Scientific Director Constantine Stratakis will always be appreciated, as well as the NICHD travel team, in particular Nancy Richman, who made my frequent trips possible and always worked things out when I needed it.

My thesis examiners, Dr. Elaine Ostrander and Dr. Erika Mancini, for agreeing to work with me on this project.

Particular thanks to Dr. Ezequiel Miron who was once my student and who I then learned from immensely. His character, perseverance, and good faith were an inspiration both during and after our time together. He preceded me at every stage of my academic journey and it was his example that I looked to when I needed one.

None of this study would be possible without the careful dedication of Carson Miller, PhD, in generating LDB1 knockout mice and cell lines on which the following work was performed.

Ming-an Sun and Jinpu Jin both provided essential help with analysis workflows, and the Macfarlan Lab members Sherry Ralls and Gernot Wolf were invaluable in my training and with experimental support at the NIH. Also thanks to all the members of the PGD for their good spirits and thoughtful feedback through the years.

David Pinto, who wrote the 3D-SIM analysis code, is an excellent man and a great thinker in the imaging space. He is accompanied in influence by Ian Dobbie, from whom I learned so much at Micron, as well as the rest of the Oxford contingent who made it the most productive scientific home of my life so far.

Credit is also due to the individuals, too numerous to be named here, who helped me so many times with issues large and small, from printing to software to hardware to making time in the lab more hospitable. I appreciate all the small kindnesses you've shown me over the years and I hope to return them throughout my career.

I also wish to acknowledge all the work from the many scientists who developed the 3D genome field to what it is today, and whom I was not able to cite here for space constraints. Drs. James Shapiro and James Holaska were also major scientific influences and I appreciate the opportunity to learn from them.

I want to give credit to my beloved wife Christina, who was has supported me over the last four years and whose mind and heart are a constant source of joy in my life.

And, most importantly, I want to thank my forebearers, grandparents, and my parents Joy Belluzzi and Alan Demmerle for providing me with the privilege, love, and good fortune to undertake such a luxury as trying to understand the universe.

Thanks again, to all of you.

TABLE OF CONTENTS

DECLARATION **2**
DEDICATION **3**
ABSTRACT **4**
ACKNOWLEDGEMENTS **5**
TABLE OF CONTENTS **7**
TABLE OF FIGURES **9**
TABLE OF TABLES **12**
LIST OF APPENDICES **13**
LIST OF ABBREVIATIONS **14**
CHAPTER 1. INTRODUCTION **15**
1.1 - MULTI-DIMENSIONAL GENOMIC ORGANIZATION 15
Structure-Function Relationships in the Biological Information Space *15*
Genomic Structure as a Driver of Cellular Identity in Development *20*
1.2 - LDB1 IS A CHROMATIN INTEGRATOR PROTEIN 28
1.3 – LIM-HD TFS DRIVE MOTOR NEURON DIFFERENTIATION 41
1.4 - MODES AND MOTIVATIONS OF UNDERSTANDING 3D GENOME ORGANIZATION 47
1.5 - GAPS IN KNOWLEDGE OF LDB1 FUNCTION AND DEVELOPMENTAL CHROMATIN REMODELING 50
1.6 - AIMS OF THIS STUDY 55
CHAPTER 2. METHODS **57**
2.1 - BIOCHEMICAL, CELL, AND MOLECULAR BIOLOGY METHODS 57
Cell Culture *57*
Western Blots *58*
FACS *59*
Immunofluorescence *59*
2.2 - SEQUENCING METHODS 60
ChIP-Seq *60*
Nuclear RNA-Seq and RT-qPCR *61*
ATAC-Seq *62*
Hi-C *62*
Hi-ChIP *63*
Capture-C *65*
2.3 - SUPER-RESOLUTION MICROSCOPY METHODS 66
Structured Illumination Microscopy *67*
Structured Illumination Microscopy Analysis *67*
2.4 – QUANTIFICATION AND STATISTICAL ANALYSIS METHODS 68
RNA-Seq analysis *69*
ATAC-Seq analysis *69*
Hi-C analysis *69*
Hi-ChIP analysis *70*
CHAPTER 3. LDB1 REGULATES LIM-HD TF OCCUPANCY AND H3K27AC DEPOSITION IN COOPERATION WITH SSDP1 ..... **71**
3.1 – LDB1 AND INDUCTION OF ISL1-FLAG AND LHX3-HA REGULATE MNX1 EXPRESSION IN AN IN-VITRO MODEL OF MOTOR NEURON DIFFERENTIATION 71

3.2 – LDB1 DELETION OR MUTATION OF ITS SSDP1-INTERACTING DOMAIN REDUCES LIM-HD TF OCCUPANCY AT KEY MOTOR NEURON GENES AND GENOME-WIDE....78
3.3 – LIM-HD TF AND SSDP1 OCCUPANCY, AND CHROMATIN ACCESSIBILITY AT LDB1 PEAKS DEPENDS ON LDB1 AND ITS INTERACTION WITH SSDP1....90
3.4 - H3K27AC OCCUPANCY IS REDUCED AT KEY MN GENES IN THE ABSENCE OF LDB1....95
3.5 – LDB1 MARKS SITES OF OPEN CHROMATIN VIA ATAC-SEQ....99
3.6 – SSDP1 MARKS A SUBSET OF LDB1 AND LIM-HD TF BOUND SITES THAT ARE SPECIFICALLY ACCESSIBLE AND H3K27AC-OCCUPIED....104

CHAPTER 4. LDB1 REGULATES LIM-HD TF-INDUCED GENE EXPRESSION PROGRAMS....105
4.1 – LDB1 IS RELIABLY INDUCED AND IS NOT COMPENSATED FOR BY LDB2 IN THE MOTOR NEURON INDUCTION SYSTEM....106
4.2 – LDB1 REGULATES GENE EXPRESSION OF SELECT MOTOR NEURON GENES....109
4.3 – LDB1 DELETION AFFECTS GENERAL DIFFERENTIATION AND LIM-HD TF-INDUCED MOTOR NEURON PROGRAMS....110
4.4 – IDENTIFICATION OF A MINIMAL MOTOR NEURON GENE EXPRESSION PROGRAM....114
4.5 - DISCUSSION....116

CHAPTER 5. 3D-SIM IMAGING OF LDB1 AND LIM-HD TFS....118
5.1 – 3D STRUCTURED ILLUMINATION MICROSCOPY OF MNC COMPONENTS....118
5.2 – ANALYSIS OF LIM-HD TF FOCI POSITIONS....121
5.3 – ASSOCIATION OF ISL1-FLAG AND MNX1 LOCI....126
5.4– LDB1 ASSOCIATES WITH LHX3-HA IN CLEAR, POSSIBLY MULTI-MERIC FOCI....128
5.5 – DISCUSSION....130

CHAPTER 6. INTERACTIONS BETWEEN MNC COMPLEX COMPONENTS....132
6.1 – LIM-HD TFS DO NOT REQUIRE LDB1 TO INTERACT....132
6.2 – LDB1-SSDP1 INTERACTION IS NECESSARY FOR ISL1-FLAG RECRUITMENT TO THE COMPLEX....133

CHAPTER 7. LDB1 REGULATES GENOMIC ORGANIZATION AT A SELECT NUMBER OF REGIONS....137
7.1 – LDB1 DELETION REGULATES 3D GENOMIC ORGANIZATION WITH LIM-HD TF AND H3K27AC OCCUPANCY AND GENE EXPRESSION AT SOME MN-INDUCED GENES....137
7.2 – LDB1 DELETION REDUCES LOOP FORMATION AT A SUBSET OF LDB1- AND H3K27AC MEDIATED GENOMIC INTERACTION SITES....149
7.3 – DESIGN AND INITIAL RESULTS OF CAPTURE-C TARGETING MOTOR NEURON AND LDB1-RELATED ELEMENTS....155
7.4 - DISCUSSION....159

CHAPTER 8. DISCUSSION....161
8.1 – FUNCTION OF LDB1....161
8.2 - MECHANISMS OF LDB1 FUNCTION IN MOTOR NEURON SPECIFICATION....167
8.3 - CHROMATIN INTEGRATOR PROTEINS AS CRITICAL HUBS FOR DEVELOPMENT AND DISEASE INTERVENTION....170
8.4 - INTEGRATING GENOMIC AND SPATIAL MAPS....174
8.5 - CONCLUSIONS....175
APPENDIX A: EQUIPMENT AND MATERIALS....179
APPENDIX B: STATISTICS AND CONTROLS....183
APPENDIX C: CODE....183
APPENDIX D: PUBLICATIONS RELATED TO OR INCLUDING PARTS OF THIS THESIS....183

REFERENCES....185

TABLE OF FIGURES

FIGURE 1.1 - THE EUKARYOTIC GENOME IS INTRINSICALLY STRUCTURED IN THREE DIMENSIONS 21
FIGURE 1.2 - GENOME ORGANIZATION AT THE SUB-NUCLEAR SCALE IS COMPLEX AND HIGHLY STRUCTURED 22
FIGURE 1.3 - THE RESEARCH SPACE IN WHICH CHROMATIN ORGANIZATION IS UNDERSTOOD 25
FIGURE 1.4 - LDB1 AS A CHROMATIN INTEGRATOR PROTEIN 26
FIGURE 1.5 – LDB1 DOMAIN STRUCTURE AND EVOLUTIONARY CONSERVATION 29
FIGURE 1.6 - EFFECTS OF LDB1 DELETION ON EMBRYOS AND EMBRYONIC STEM CELLS 32
FIGURE 1.7 – KNOWN ROLES OF LDB1 IN MEDIATING TRANSCRIPTION FACTOR COMPLEXES 39
FIGURE 1.8 – HYPOTHESIS: LDB1 REGULATES GENOMIC ARCHITECTURE AND ACTIVITY AT MOTOR NEURON GENES 55
FIGURE 3.1 – MOUSE CROSSING SCHEME FOR LDB1 -/- INDUCIBLE ESCS 72
FIGURE 3.2 – INDUCIBLE MOTOR NEURON SYSTEM DESIGN AND DIFFERENTIATION PROTOCOL 73
FIGURE 3.3 – EXPERIMENTAL SETUP TO DETERMINE EFFECTS OF LDB1 KNOCKOUT 74
FIGURE 3.4 – MNX1 IS INDUCED AT THE PROTEIN LEVEL BY ISL1-FLAG AND LHX3-HA IN A LDB1-DEPENDENT MANNER 75
FIGURE 3.5 – FIGURE 3.5 - MNX1 EXPRESSION DEPENDS ON THE LDB1-SSDP1 INTERACTION 76
FIGURE 3.6 – LDB1- AND DOXYCYCLINE- DEPENDENT INDUCTION OF MNX1 AND TARGET LIM-HD TFS. 77
FIGURE 3.7 – LDB1 REGULATES OCCUPANCY OF LIM-HD TFS, SSDP1, H3K27AC, AND OPEN CHROMATIN AT THE *MNX1* LOCUS 79
FIGURE 3.8 – LDB1 REGULATES OCCUPANCY OF LIM-HD TFS, SSDP1, H3K27AC, AND OPEN CHROMATIN AT THE *NRP1* LOCUS 80
FIGURE 3.9 – LDB1 REGULATES OCCUPANCY OF LIM-HD TFS, SSDP1, H3K27AC, AND OPEN CHROMATIN AT THE *SLIT3* LOCUS 81
FIGURE 3.10 – LDB1 REGULATES OCCUPANCY OF LIM-HD TFS, SSDP1, H3K27AC, AND OPEN CHROMATIN AT THE ARHGAP36 LOCUS 82
FIGURE 3.11 – LDB1 ASSOCIATES WITH LIM-HD TFS, OPEN CHROMATIN, H3K27AC, AND SSDP1 84
FIGURE 3.12 – MNC COMPONENTS, H3K27AC, AND OPEN CHROMATIN ARE REDUCED AT LDB1 PEAKS IN THE ABSENCE OF LDB1 85
FIGURE 3.13 – LDB1 BINDING MOTIFS ARE SIMILAR TO PAIRS OF LIM-HD MOTIFS, WHICH ARE ABLATED IN THE ABSENCE OF LDB1 86
FIGURE 3.14 - PROPOSED MNC TARGET SITES ARE UNPAIRED IN THE GENOME, AND NOT INDICATIVE OF LDB1 BINDING 88
FIGURE 3.15 – LDB1 AND ITS SSDP1 INTERACTION REGULATES OCCUPANCY OF ISL1-FLAG, LHX3-HA, H3K27AC, AND SSDP1 91
FIGURE 3.16 – ISL1-FLAG CO-OCCUPANCY WITH LHX3-HA, H3K27AC, AND SSDP1 IS DEPENDENT ON LDB1 AND ITS SSDP1-INTERACTING DOMAIN 93
FIGURE 3.17 – H3K27AC IS DIFFERENTIALLY REGULATED ON A SUBSET OF LDB1 PEAKS 96
FIGURE 3.18 – H3K27AC RETAINS OCCUPANCY AT ISL1-FLAG SITES INDEPENDENTLY OF LDB1 FUNCTION 97
FIGURE 3.19 - LDB1 PEAKS ARE MADE ACCESSIBLE VIA-ATAC-SEQ IN AN LDB1- AND SSDP1-DEPENDENT MANNER 100
FIGURE 3.20 – ISL1-FLAG PEAKS ARE MADE ACCESSIBLE VIA-ATAC-SEQ IN AN LDB1- AND SSDP1-DEPENDENT MANNER 102

FIGURE 3.21 - LDB1 ASSOCIATES WITH A SELECT GROUP OF PEAKS THAT ARE MARKED SPECIFICALLY BY SSDP1 104
FIGURE 4.1 – *LDB1* IS NOT EXPRESSED IN *LDB1* KNOCKOUT CELLS WITHOUT AN INDUCIBLE TRANSGENE AND *LDB2* IS NOT EXPRESSED IN ANY CONDITION 107
FIGURE 4.2 – CRITICAL MOTOR NEURON GENES ARE DEPENDENT ON LDB1 AND MNC INDUCTION 109
FIGURE 4.3 – LDB1 GENOTYPE DOES NOT AFFECT GENE EXPRESSION LEVELS IN EMBRYONIC STEM CELLS 110
FIGURE 4.4 – DOXYCYCLINE INDUCTION CHANGES A SMALL NUMBER OF GENE EXPRESSION LEVELS COMPARED TO GENERAL NEURONAL DIFFERENTIATION 111
FIGURE 4.5 – LDB1 KNOCKOUT AFFECTS THE EXPRESSION OF MANY GENES, BUT ONLY A SELECT NUMBER THAT ARE ALSO MOTOR NEURON INDUCED 112
FIGURE 5.1 – EXPERIMENTAL SETUP FOR IMAGING OF MOTOR NEURON COMPLEX COMPONENTS VIA 3D-SIM IMAGING 119
FIGURE 5.2 – 3D-SIM IMAGING OF MNC COMPLEX COMPONENTS 120
FIGURE 5.3 – NUCLEAR THRESHOLDING AND CENTROID CALLS FROM IMMUNOFLUORESCENCE FOCI 121
FIGURE 5.4 – DISTANCE DISTRIBUTIONS OF LIM-HD TF FOCI 122
FIGURE 5.5 – MEAN DISTANCE DISTRIBUTIONS OF LIM-HD TF FOCI 124
FIGURE 5.6 – ASSOCIATION OF MNX1 LOCUS AND ISL1-FLAG SIGNAL IN LDB1 +/+ CELLS VIA RASER-FISH 127
FIGURE 5.7 – LDB1 AND LHX3-HA FORM TIGHTLY ASSOCIATED CLUSTERS AS IMAGED WITH 3D-SIM 128
FIGURE 6.1 – ISL1-FLAG AND LHX3-HA INTERACT IN A LDB1-INDEPENDENT MANNER 132
FIGURE 6.2 – ISL1-FLAG AND SSDP1 CO-IMMUNOPRECIPITATE IN AN LDB1-DEPENDENT MANNER 134
FIGURE 6.3 – DUAL ISL1-FLAG AND LHX3-HA CO-IMMUNOPRECIPITATION DETECTS INTERACTIONS IN A LDB1- AND DOXYCYCLINE-DEPENDENT MANNER. 135
FIGURE 7.1 – DIAGRAMMATIC WORKFLOW OF HI-C AND HICHIP ASSAYS 138
FIGURE 7.2 – GENOMIC STRUCTURES MANIFESTED IN CONTACT MAPS 139
FIGURE 7.3 – 3D GENOME ORGANIZATION AROUND THE *MNX1* LOCUS 140
FIGURE 7.4 – 3D GENOME ORGANIZATION IN THE LARGER DOMAIN AROUND *MNX1* 142
FIGURE 7.5 - 3D GENOME ORGANIZATION AROUND THE *SLIT3* LOCUS 144
FIGURE 7.6 – GENOME ORGANIZATION AROUND THE *TMTC2* LOCUS 146
FIGURE 7.7 – GENOME-WIDE CONTACT ENRICHMENT AT LDB1 AND H3K27AC CHIP-SEQ PEAKS 148
FIGURE 7.8 –HI-CHIP OF LDB1- AND H3K27AC-MEDIATED CONTACTS AROUND THE *MNX1 LOCUS* 150
FIGURE 7.9 – GENOME-WIDE CONTACT ENRICHMENT AT LDB1 AND H3K27AC HICHIP PEAKS 152
FIGURE 7.10 – LDB1- AND H3K27AC-MEDIATED CONTACTS MAY BE REDUCED IN THE ABSENCE OF LDB1 154
FIGURE 7.11 – DESIGN OF CAPTURE-C STRATEGY FOR DETECTING LDB1-REGULATED AND CONTROL ELEMENTS 156
FIGURE 7.12 – GENE ONTOLOGY AND RELATIVE GENOMIC POSITIONS OF CAPTURE-C PROBES 157
FIGURE 7.13 – CAPTURE-C OF SELECTED REGIONS SHOWS SOME DOXYCYCLINE-INDUCED INCREASE IN CONTACT FREQUENCIES 158
FIGURE 8.1 – CONCLUSION: LDB1 REGULATES GENOMIC ARCHITECTURE AND ACTIVITY AT MOTOR NEURON GENES 162
FIGURE 8.2 – LDB1 IS A CENTRAL HUB FOR MULTIPLE MULTI-SCALE REGULATORY MECHANISMS GOVERNING CHROMATIN ARCHITECTURE AND DEVELOPMENT 172

FIGURE 8.3 – CHROMATIN INTEGRATOR PROTEINS CONDENSE EXTERNAL SIGNALS INTO GENE EXPRESSION RESPONSES 176

TABLE OF TABLES

TABLE 1 – LDB1 AND LDB2 FUNCTION IN DIVERSE TISSUES 39
TABLE 2 – LDB1 PARTNERS FUNCTION IN HUMAN DISEASES 48
TABLE 3 – LDB1-REGULATED AND DOX-INDUCIBLE GENES IN DAY 6 EBS 114

LIST OF APPENDICES

Appendix A: Equipment and Materials ……..182
Appendix B: Statistics and Controls ……..186
Appendix C: Code ……..186
Appendix D: Publications related to or including parts of this thesis ……..186

LIST OF ABBREVIATIONS

3D-SIM – Three-dimensional Structured Illumination Microscopy
ChIP – Chromatin Immunoprecipitation
DNA – deoxyribonucleic acid
DOX - Doxycycline
EB – Embryoid Body
ECR2 – Evolutionarily Conserved Region 2
EMSA – Electrophoretic Mobility Shift Assay
ESC – Embryonic Stem Cell
DE – Differentially Expressed
HA – Human influenza Agglutinin
HD - Homeodomain
IDR – Internally Disordered Region
Mb – Megabase
MN – Motor Neuron
MNC – Motor Neuron Complex
NIL – Ngn2/Isl1/Lhx3
PCR – Polymerase Chain Reaction
RNA – ribonucleic acid
SELEX – Systemic Evolution of Ligands by EXponential enrichment
TF – Transcription Factor
ZFN – Zinc Finger Nuclease

CHAPTER 1. INTRODUCTION

1.1 - MULTI-DIMENSIONAL GENOMIC ORGANIZATION

STRUCTURE-FUNCTION RELATIONSHIPS IN THE BIOLOGICAL INFORMATION SPACE

Biological structure and biological function are intrinsically inseparable. The eukaryotic genome is a prime example of this principle, as the information content of the genome is encoded in physical space, and can only be manifested in nature through a complex series of structural changes carried out by energy-dependent macromolecular complexes. Understanding the interplay of genomic structure and genomic function is one of the core aims in modern biology. The information contained in the "epigenetic" space, specifically beyond the linear genetic code and including the chemical, protein, and spatial contexts of genetic information, is both the link between the external environment and genome (Shapiro 1997), and the method of specifying cellular identity between genetically identical cells within multicellular organisms (Waddington 1957). The dynamic spatial organization of genetic information within the nucleus is critical to genome replication, repair, and division. It is also an essential aspect of eliciting specific responses in gene expression to developmental or functional cues from a vast array of relatively non-specific possible interactions between biologically relevant molecules. A detailed understanding of this process is essential for not only therapeutic interventions in defects of human development, adult-onset disease, and the delay of aging, but also for the fundamental understanding of biological information transfer, network complexity, and a basic comprehension of mechanisms that drive life.

The interphase eukaryotic genome is one of the most complex informational structures in nature. A substantial and unaccounted amount of that information is carried in the genome's spatial organization, and in its interactions with various proteins in the form of chromatin and the ribonucleo-protein complexes that act upon it. Genetic and molecular research in the 20th century culminated in a central dogma of molecular biology (Crick 1970) whereby information transfer from DNA, through RNA, to protein, was established. Later thinkers expanded this initial framework, recognizing reverse transcription, prions, interspecies horizontal transfer, and other non-linearities in the greater molecular biological informational ecosystem (Shapiro 2009). However, the underlying premise of many of these theories was that the information transferred was encoded in spatially contiguous DNA sequences and their resulting transcriptional products, including coding- and non-coding RNA species, and their subsequent protein products. This ignores the information contained in the literal physical organization of non-contiguous DNA sequences and their associated proteins, which is necessarily required for the storage, expression, and transmission of such genetic information.

The structural importance of biological information was recognised in early studies of cell division, showing that the nucleus was a distinct and essential subunit of the cell (Haeckle 1868); that a substance called 'nuclein', which we now understand to be chromatin, was a unique non-protein substance essential for heredity (Miescher 1884); and that the nucleus contained many structurally independent elements (chromosomes) that contained hereditary information (Flemming 1886). Further, even the structural organization of chromosomes

themselves was recognised in the 19th century, not only in their role as drivers of cell division (Boveri 1888), but also the spatial orientation of chromosomes within functioning cells was essential to their biological action (Rabl 1885). However, this inclusion of structural organization into our understanding of genetics, heredity, and cellular biology was neglected for many years until advances in microscopy made it apparent that the mammalian interphase nucleus was highly structured and much more complex than had been previously thought (Cremer et al. 1982; Strickfaden et al. 2010).

The spatial or 3D genome organization referred to throughout this thesis include specific enhancer-promoter interactions between regulatory sites, changes in chromatin architecture and remodeling that effect transcription, and large-scale domain organization, up to and including chromosomal and nuclear-level structures. These phenomena are inter-related, and the specific level of spatial genomic organization being referred to will depend on the broader biological context.

With the advances in microscopy, genetics, sequencing, and computation within the last decades now fully trained on understanding epigenetic and spatial genome organization, we are rediscovering and re-establishing what was initially hypothesised in the 19th century – that the structure of biological information is important to its function. New methods in imaging such as super-resolution and single-molecule imaging, and ligation- or proximity-based sequencing methods such as Hi-C and its many derivatives are now providing nearly molecular-resolution maps and mechanisms to explain the ever-expanding complexities of cellular fate choice and organismal development. These discoveries drive an

increasing appreciation of the spatial dimension in genome organization, nuclear architecture, and chromatin remodeling at every level of interaction and cell function – including mitosis (Naumova et al. 2013), transcription (Jackson et al. 1993) DNA replication (Cook and Lang 1984), DNA repair (Lemaître et al. 2014), developmental gene regulation (Kosak et al. 2002), and stress response (Brickner and Walter 2004). The importance of this underlying structural element – genomic and nuclear organization – to cell and molecular biology has been thoroughly reviewed in relation to both cell biology and human disease (Misteli 2007; 2010; Andrey and Mundlos 2017)

Certain changes in 3D genome organization are causative to several developmental malformations such as poly- and brachydactyly from genomic duplications or inversions resulting in erroneous enhancer targeting (Lupiáñez et al. 2015). The 3D genome disorganization in progeria is also causative to the in terms of disease etiology (Zheng et al. 2018). The spatial organization of non-coding regulatory DNA elements is critical to their function as regulators (enhancers or silencers) of non-contiguous genes in a wide variety of medically relevant biological models, but each element on its own, or its spatial regulation is not always causal to a given phenotype (Hnisz et al. 2015; Cremer and Cremer 2019). However, 3D genomic organization cannot be described as strictly "causative" to a given cellular process since it is a fundamental property of the cell. Further, the absolute spatial positions of genomic regions within a similar cell population is probabilistic (Bickmore 2013). Severe changes to chromosome organization can be tolerated (Ghavi-Helm et al. 2019), and the majority of individual loci experience a range of nuclear positions across a cell population

despite similar transcriptional states (Finn et al. 2019). Thus, much of the information accessible about the relationship between genome architecture and gene expression or cell behaviour is necessarily correlative. This information must be built on assaying changes in spatial genome organization, chromatin architecture, and gene expression simultaneously in response to perturbations of candidate factors, such as proteins or specific genetic elements that can be precisely manipulated, and correlating those together to determine which factors have roles in controlling both genome organization and cellular function.

The goal of this thesis is to identify, in a model of mouse MN differentiation and with a high-precision, multi-modal approach, the effects on gene expression, chromatin architecture, and spatial genome organization of a protein that plays key roles in development, transcription-factor binding, and 3D-genome interactions – LIM-Domain-Binding Protein 1 (LDB1).

Genomic Structure as a Driver of Cellular Identity in Development

The genetic identity of developing multicellular organisms is constant, yet each creates many distinct and finely differentiated tissues. Establishment of cell-fate identity in development and coordinated gene regulation is manifested through the 3D organization of the genome and its intrinsic chromatin signatures. The nuclear genome is structured hierarchically (Rocha et al. 2015; Gibcus and Dekker 2013). This hierarchy starts with the entire nucleus, a spherical object with a diameter on the order of 10 μm containing (in the diploid mouse cell) about 5400 Mb of DNA; this nucleus contains 20 pairs of chromosomes, with each chromosome ranging from ~50-250 Mb; chromosomes in interphase are compartmentalised on scales of tens of Mb that often correlate with genomic activity levels and hetero- or euchromatic states; within each compartment, the genome is physically organised into Mb-scale domains called Topologically Associated Domains (TADs), which are characterised by architectural proteins such as CTCF at their boundaries, and a significantly higher frequency of intra- than inter-TAD contacts; within TADs, there are substantial interactions between distal regulatory elements on the tens to hundreds of kb scale which regulates individual genes expressed through well-established transcriptional mechanisms on the level of kilobases to individual base pairs **(Figure 1.1)**.

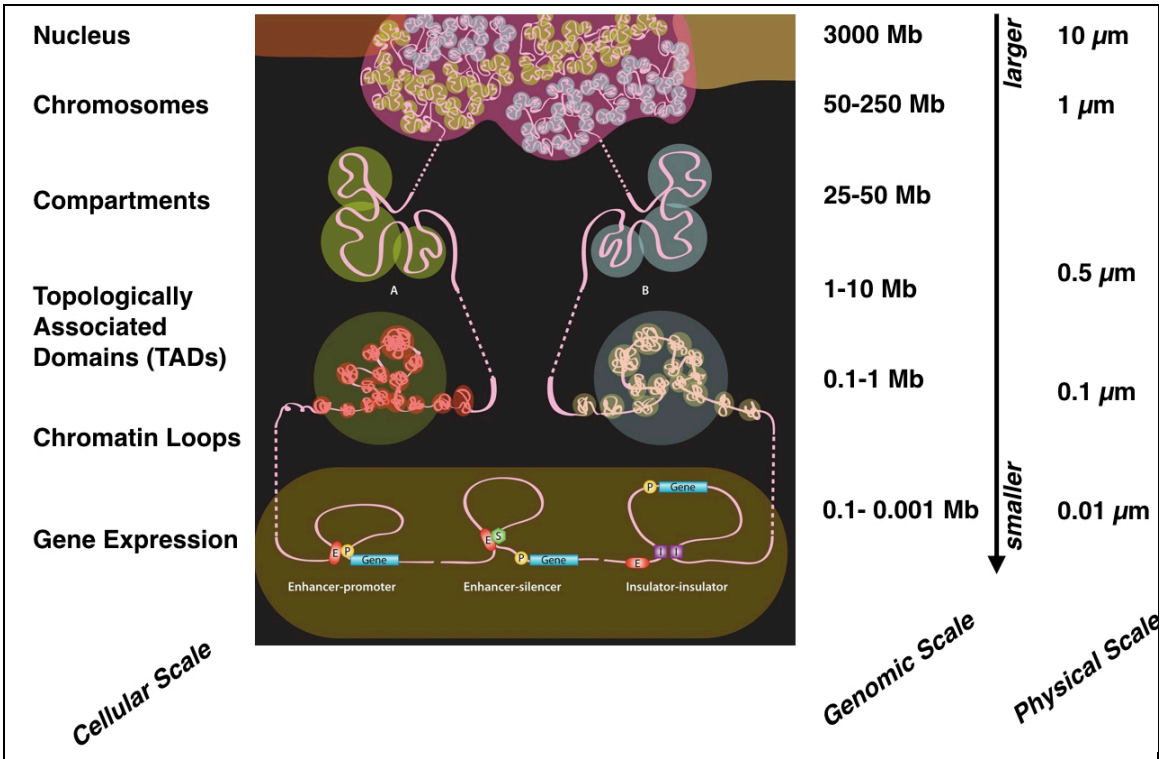


Figure 1.1 - The eukaryotic genome is intrinsically structured in three dimensions
A schematic diagram of the hierarchical spatial organization of genetic material in eukaryotes. Cellular scale is depicted on the left, while genomic (in millions of base pairs) and physical (in micrometers) scales are indicated. Various types of three-dimensional interactions between genes and regulatory elements are illustrated at the level of gene expression. The time axis is not indicated. Adapted from Fraser et al. 2015.

The structural, chemical, and molecular biology of transcription is well understood and too voluminous to be summarised in this thesis (Cramer 2019). However, the knowledge about genomic organization and processes at the higher-order, kb-Mb scales of chromatin remodeling and chromosome structure is much less developed than at other spatial and physical scales **(Figure 1.2 A)** (Marti-Renom and Mirny 2011). These nano- to meso-scale interactions of chromatin in space are currently understood as a complex and dynamic interplay of chromatin loops and active genomic regions organised around a macro-molecule-filled inter-chromosomal network, with transcriptionally inactive

chromatin condensing away from the more transcriptionally active milieu of the nucleus (**Figure 1.2 B)** (Cremer et al. 2000; 2015)**.**

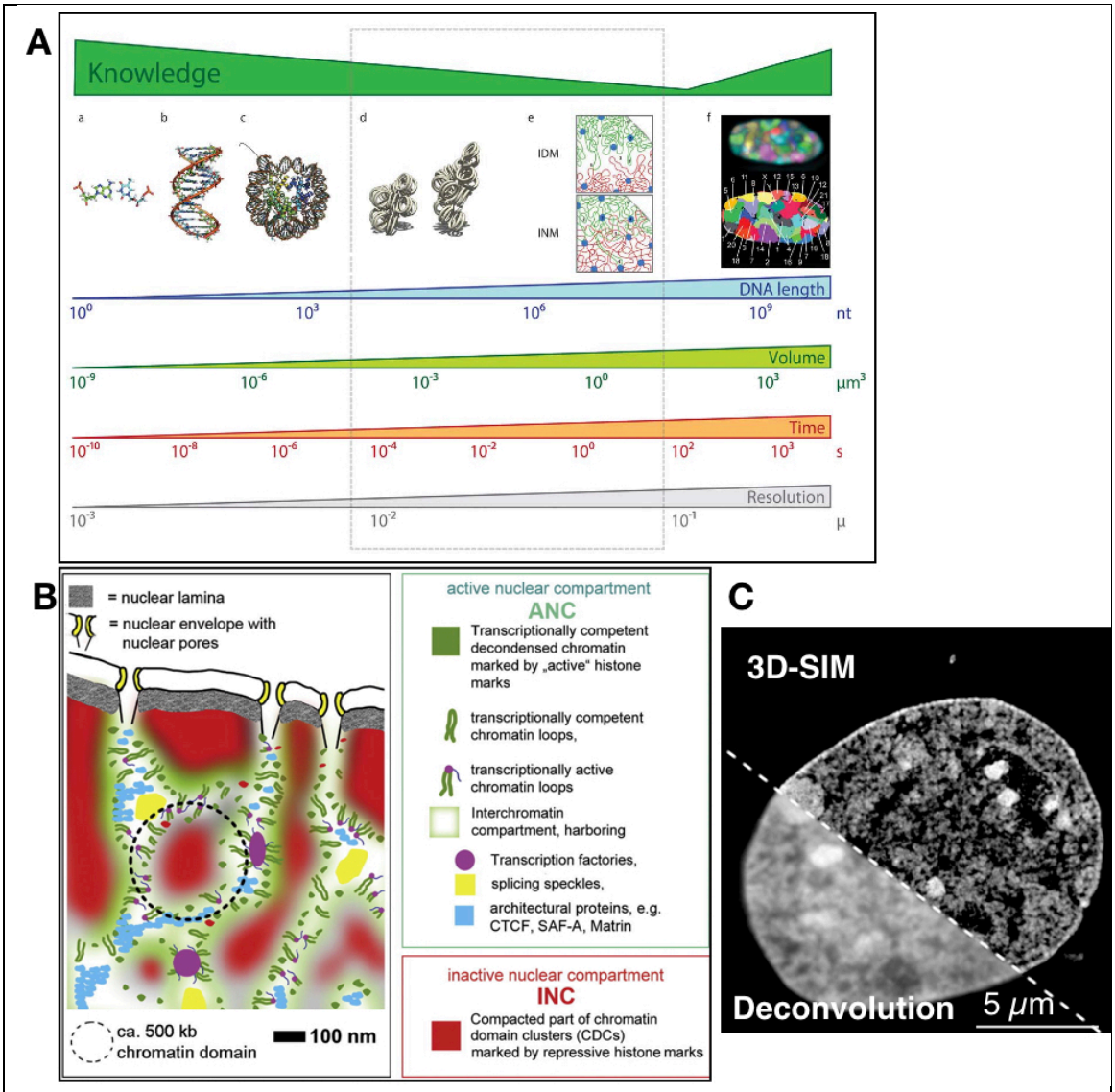


Figure 1.2 - Genome organization at the sub-nuclear scale is complex and highly structured
A) Schematic showing scales of nuclear organization in DNA length, volume, time, and resolution, overlaid with a theoretical amount of knowledge about each level of resolution. Adapted from Marti-Renom 2011. B) Diagrammatic view of the nuclear volume partitioned into compartments and interacting with various macromolecules in the nuclear space. Adapted from Cremer 2015 C) Example of structural complexity unveiled by advanced microscopy techniques, showing a C127 nucleus stained with DAPI either with widefield-deconvolution (left) or 3D Structured Illumination Microscopy (right). Adapted from Smeets 2014.

These interactions are of particular interest because they constitute the differential *cis*- and *trans*-interactions that drive cell-type specific gene expression. To understand and control the function of gene regulatory elements in both development and disease, it is essential to map and perturb the higher-order chromatin structures and the protein or nucleic acid factors that regulate these structures. While the importance of four-dimensional nuclear architecture and genome organization is well established (Bickmore and van Steensel 2013; Cavalli and Misteli 2013; Cremer et al. 2015; Norton and Phillips-Cremins 2017), the specific sets of dynamic 3D genome and epigenomic changes in each clinically relevant developmental and disease system is only now being determined systematically. This is made more challenging by the diverse array of protein, genetic, and biophysical factors influencing genomic organization, and the intrinsic degree of randomness that aspects of such a complex system will manifest.

There are several fundamental forces that shape genome organization (Rada Iglesias et al. 2018). These include transcription (Cook and Marenduzzo 2018), CTCF and cohesin function along with other architectural proteins (Ong and Corces 2014), phase separation (Erdel and Rippe 2018), and biophysical functional zonation (Miron et al. 2019). Topological control of gene expression is essential to multi-cellular life, but is also present in prokaryotes (Kobren and Singh 2019). The biophysical compartmentalization of the genome, and consequently its function, must also be considered in the context of soft-matter physics (Iborra 2007). Thus, a multi-disciplinary effort is needed to create therapeutically useful and theoretically meaningful understandings of genome

organization. There are now a number of excellent techniques for determining spatial genome architecture (Parnar, 2019). However, combining these diverse modalities into a research strategy that can assess and identify aspects of molecular structure and biological function at both single-cell and population resolutions, with sufficient spatial and temporal resolution, is a challenge even for the most well-studied genetic loci (Williamson et al. 2014). Thus, it is essential to expand theoretically the research space in which we explore genome organization, and to form both technical and theoretical connections between seemingly disparate areas **(Figure 1.3)**.

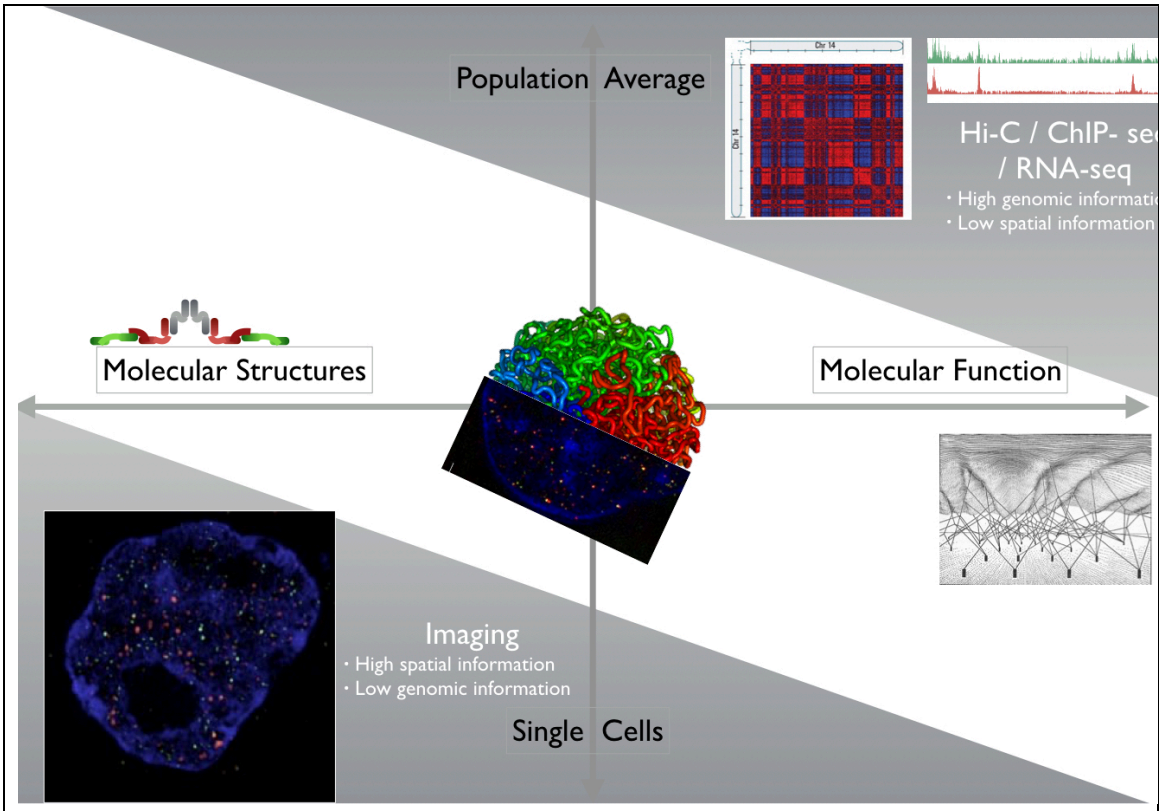


Figure 1.3 - The research space in which chromatin organization is understood

Theoretical schematic showing contrasts between single-cell techniques such as imaging and population-based techniques, with non-specific example data from both. The contrast is also drawn between delineating molecular structures, such as hypothetical transcription-factor complex on the left, and actual biological function of those molecules, as indicated by the Waddington fate diagram on the right. Adapted from Waddington 1957 and Aiden et al. 2009.

Fundamentally, we seek to know the X, Y, Z, and T positions of every macromolecule in the nucleus, and presume that if we know this we could infer the cell state, and direct manipulations of it. While this is still technically inconceivable, we can begin in small, well-defined model systems, testing theories of cooperative spatial regulation of the genome by candidate proteins that have a strong effect on cell fate choice. In this thesis, I perform such a study on the nuclear adaptor and chromatin integrator protein LDB1.

LDB1 is a critical regulator of developmental gene regulation in higher eukaryotes and a mediator of transcription-factor binding and function throughout the tree of life. Since its discovery 25 years ago, the role of LDB1 (and its human and mouse paralog LDB2) has been elucidated in a wide range of developmental contexts, organisms, and disease models. In addition to its initially appreciated roles in mediating TF interactions, recent work has shown LDB1 to be a critical player in long-range gene activation and 3D genome organization, a possible hub of phase-separated transcriptional regulation, and a target for next-generation methods of cellular reprogramming. Together, these functions can be assigned to a novel category of proteins that are capable of interacting with DNA-binding factors, transcriptional machinery, and architectural factors; we name them "chromatin integrator proteins", in the tradition of previous reviews (Matthews and Visvader 2003; Liu and Dean 2019) **(Figure 1.4**).

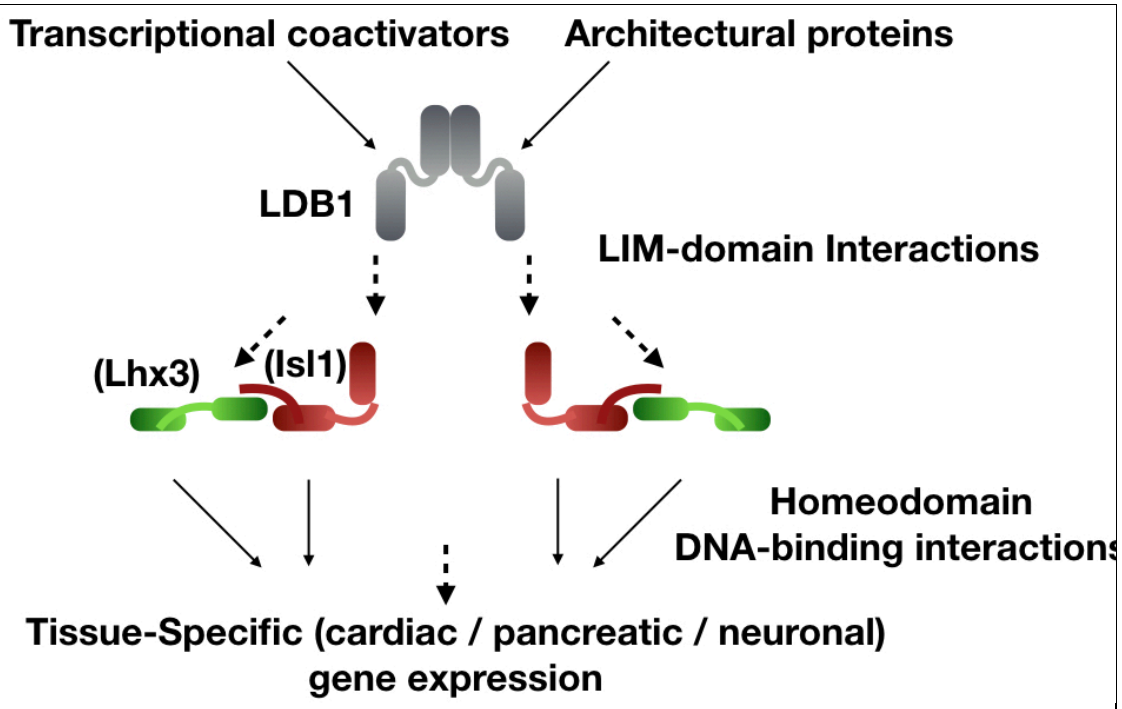


Figure 1.4 - LDB1 as a chromatin integrator protein
Schematic demonstration of LDB1 as a hub for interactions of varying specificity between biomolecules, leading to tissue-specific gene expression.

LDB1 occupies a critical point in gene activation networks because of its ability to interact with diverse protein partners, particularly linking sequence-specific transcription factors with broadly functioning architectural and transcriptional machinery, as well as forming self-association-based hubs. For these reasons, I propose to use a LDB1 knockout model of MN development to study the effects of a chromatin integrator protein on differentiation and spatial genome organization.

1.2 - LDB1 IS A CHROMATIN INTEGRATOR PROTEIN

The fundamental role of LDB1 in eukaryotic biology derives from its structure as a LIM-interacting protein. LIM domains are deeply conserved cysteine-rich zinc-finger protein-protein interaction domains, found in eukaryotes, bacteria, and plants (Dawid et al. 1998; Rétaux and Bachy 2002; Gill 1995). LDB1 is conserved from worms to humans, and the role of multi-functional LIM binding domains proteins is ancient (Matthews and Visvader 2003). Its domain structure includes: an N-terminal self-association or dimerization domain (DD) with internally disordered regions (IDRs); an Ldb1-Chip-conserved-domain (LCCD) that can bind SSDP1 among other proteins; an other-interacting domain (OID) that can bind CTCF among others; a nuclear localization signal (NLS); and a LIM-interacting domain (LID), combining several key interaction domains into a single multi-functional regulatory protein (Jurata and Gill 1997; Breen et al. 1998; Chen et al. 2002; Krivega et al. 2014; Lee et al. 2017) **(Figure 1.5**).

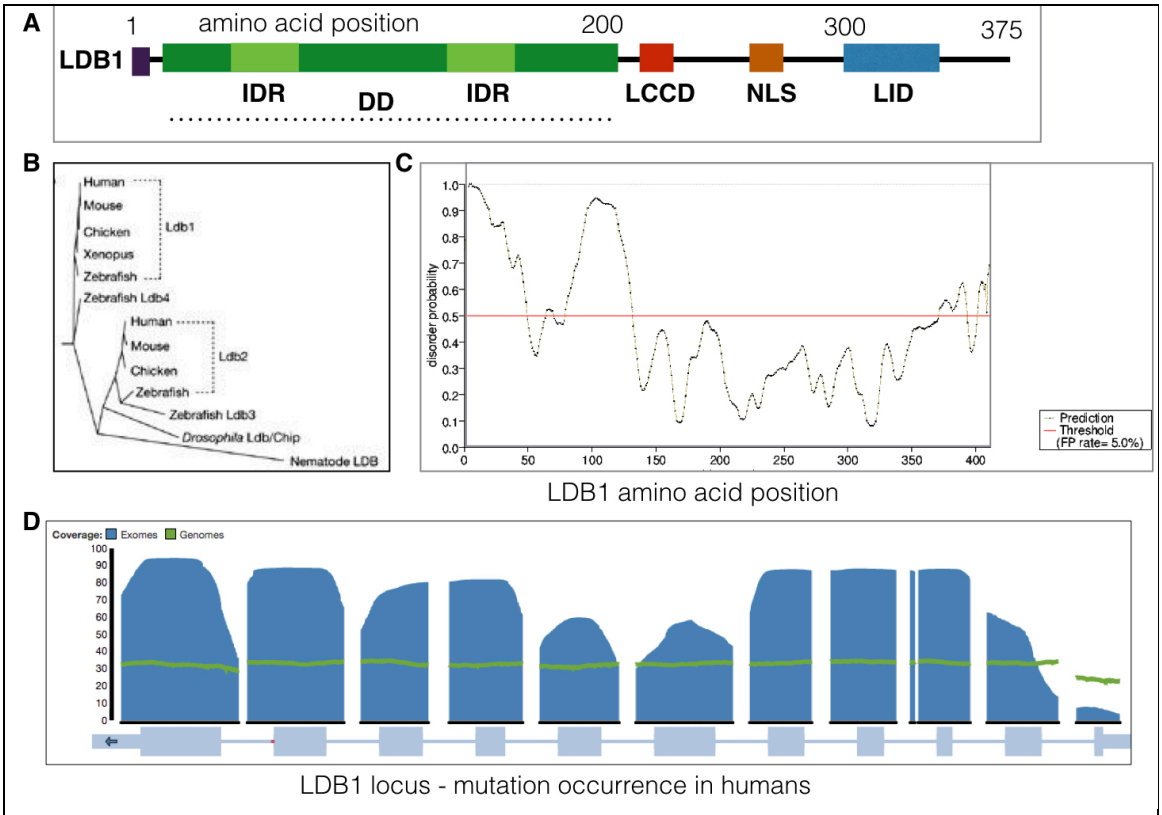


Figure 1.5 – LDB1 domain structure and evolutionary conservation
A) A non-exact diagrammatic schema of LDB1 protein structure, from N to C termini, with relevant domains indicated along with amino acid. **B)** Evolutionary diversion diagram of LDB1 and its paralogs in various species, taken from Matthews 2001. **C)** Graph of disorder probability on Y axis along the length of LDB1 from N to C termini, with high disorder probability aligning with the IDRs of the LDB1 dimerization domain. **D)** Screenshot of the gnomAD browser for the LDB1 gene showing non-synonymous mutations - a single splice acceptor mutation at the beginning of the first transcribed exon, which indicates that in the >100,000 individuals in the GnomAD database, all carry functional copies of LDB1.

LDB1, which was then known as NLI, and also as CLIM-2, was originally identified as an interactor of the LIM-domain containing *Rbtn2* (or LMO2) gene in mice (Jurata et al. 1996), and in screens for physical interactors of LIM-domain proteins in both yeast and mammalian cells (Bach et al. 1997; Agulnick et al. 1996). It was also identified as a binding partner of the haematopoietic regulator LMO2 (Visvader et al. 1997). The human orthologs were cloned, mapped, and originally identified as CLIM1 and CLIM2 (Semina et al. 1998; Ueki et al. 1999),

while the mouse gene was cloned contemporaneously (Yamashita et al. 1998). LDB1 is an essential gene for combinatorial transcription factor function, as it is ubiquitously expressed throughout development at low levels in all tissues listed in the GTex database, and mediates the interactions of many different types of LIM-domain proteins, as well as other transcription factors such as Otx and GATA, and architectural proteins such as CTCF and Mediator (Agulnick et al. 1996; Jurata et al. 1998; Breen et al. 1998; Bach et al. 1997).

The role of LDB1 in transcriptional regulation was suggested by its interactions with LIM-homeodomain transcription factors (LIM-HD TFs) Isl1, Lhx3, Lhx5, and others, and its correlation with Isl1 expression suggested a critical role in MN development, despite its near-universal expression and known roles in erythropoiesis (Jurata et al. 1996). This was confirmed through immunoprecipitation experiments showing LDB1 as a physical linker of Isl1 and Lhx3 molecules, critical drivers of MN specification (Jurata and Gill 1997; Breen et al. 1998; Jurata et al. 1998). The requirement of LDB1 for dimerization to interact with Lmx1 indicated that it also may be involved in forming higher-order protein complexes (Jurata and Gill 1997). LDB1 interacts with other LIM homeodomain TFs, having been identified in a tetramer of the Lhx1 LIM domain, and shown to bind LIM TFs such Lhx1, Lhx3, and LMO1, but not non-TF LIM proteins such as LMK, MLP, or Zyxin (Agulnick et al. 1996). LDB1 was also initially identified as a member of the erythroid-specifying GATA1/TAL1/E47/LMO2 complex (Wadman et al. 1997; Visvader et al. 1997). LDB1 and Isl1 function together in cardiomyocytes, likely mediating contacts between DNA sequences containing the Isl1 binding site and at least one other

trans-acting factor (Caputo et al. 2015), as well as in the pancreas, where Isl1 is an essential regulator of ß-cell formation. From its discovery, LDB1 was postulated to help coordinate the interaction of other LIM-homeodomain-containing proteins in a cell-type specific manner (Matthews and Visvader 2003).

The original mouse knockout of *Ldb1*, created with a targeting vector to replace exons 3-9 with a PGK-Neo cassette, did not result in any viable homozygotes at birth or at E16.5, and are likely dead by E9.5 (Mukhopadhyay et al. 2003). Embryos collected at days E7.5 or E8.5 showed massive pleiotropic abnormalities during gastrulation, including defects in heart formation, forebrain formation, haematopoietic development, extraembryonic tissue, anterior-posterior axis formation, and Wnt pathway expression (Mukhopadhyay et al. 2003) **(Figure 1.6)**.

The deletion mutants of LDB1 partner LMO4 are also embryonic lethal, exhibiting exencephaly and failure of neural tube closure (Lee et al. 2005a). Several *Ldb1* conditional knockouts are viable, and have been well characterised. Conditional deletion of *Ldb1* in Nkx2.1-expressing neurons of the telencephalon

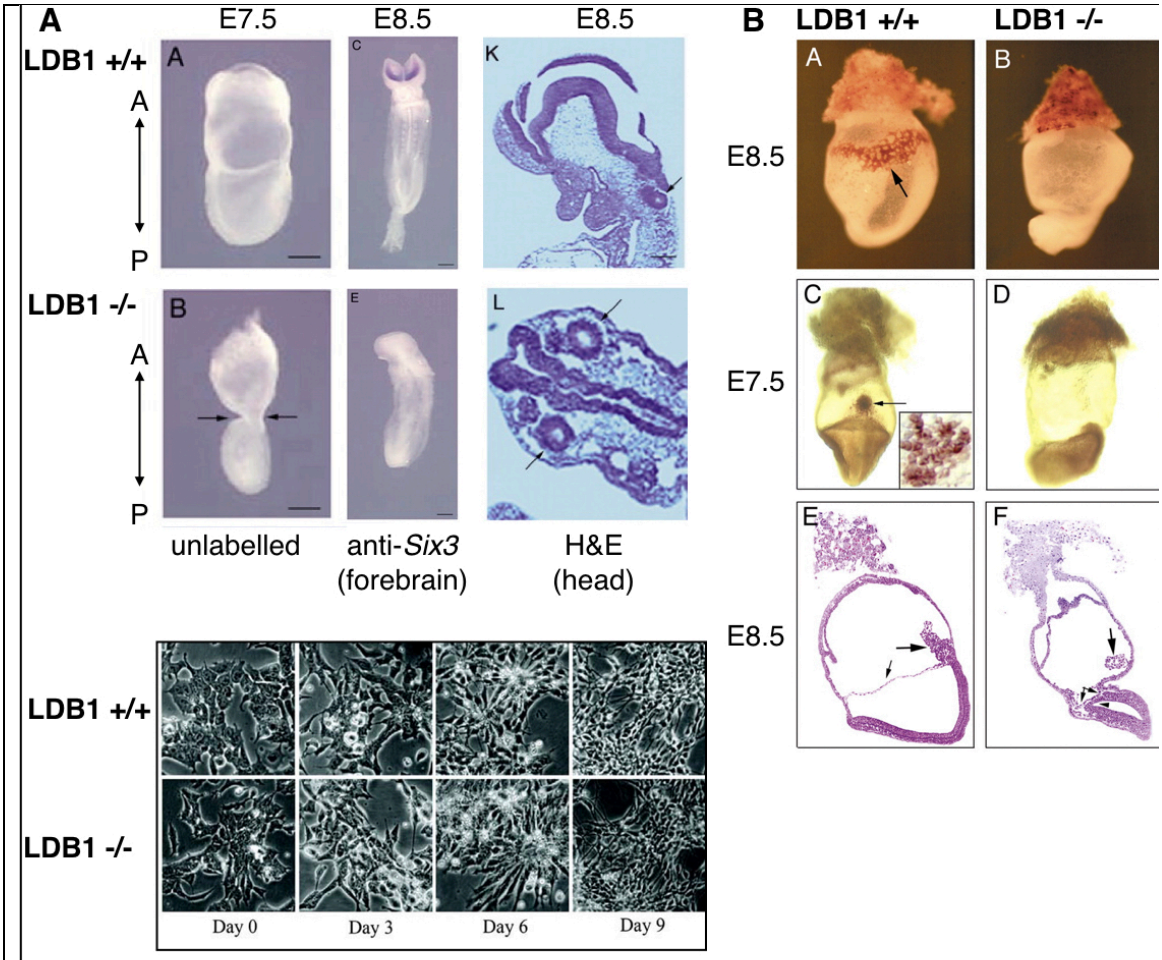


Figure 1.6 - Effects of LDB1 deletion on embryos and embryonic stem cells
A) Images of LDB1 +/+ and -/- embryos at the indicated days. **B)** Embryonic phenotypes showing defects in hemoglobin production (**A&B**), germ cells (**C&D**), and ectoderm and allantois (**E&F**). Both **A** and **B** from Mukhopadyay 2001. **C)** Brightfield images of embryonic stem cells with the indicated LDB1 genotype at various days of neural differentiation, showing minimal difference at this cell stage. Adapted from Hwang 2008.

caused a reduction of cholinergic neurons, cortical and hippocampal GABAergic interneurons, striatal parvalbumin-positive neurons, and expression changes in several marker genes characteristic of the medial ganglionic eminence, where

Nkx2.1 targets neuronal LIM-HD transcription factors LHX6 and LHX8 which require LDB1 to bind to the SBE3 enhancer of *Shh* (Zhao et al. 2014). The *engrailed-1* conditional knockout of LDB1 in the mouse showed major neurological defects and failed to generate several neuronal lineages (Kim et al. 2016b). In the developing retina, LDB1 interacts with Lhx2 to maintain the retinal progenitor cell population, and also shares some functional redundancy with LDB2 (Gueta et al. 2016). In the rat pituitary gland, Isl1 and Lhx3 work with SF-1 to target the gonadotropin-releasing hormone receptor (Granger et al. 2006). While LDB1 is dispensable for Neu-induced tumor progression and mammary development in mouse, it is required for the endothelial to mesenchymal transition and migration to secondary sites of tumor formation (Ahmed et al. 2018).

The *Drosophila* LDB1 ortholog *Chip* was discovered in a screen for mutations to reduce activity of the remote wing margin enhancer in the *cut* locus, indicating a future role in long-range chromatin organization (Morcillo et al. 1996). Shown to be essential and to interact with other LIM-domain proteins, LDB1 deletion caused a segmentation phenotype in the developing embryo (Morcillo et al. 1997), and was later shown to facilitate *apterous* (Ap) TF activity by forming a bridge between two Ap molecules (Milan and Cohen 1999; van Meyel et al. 1999), where it regulates axon guidance (van Meyel et al. 2000). LDB1 also has critical functions in worm development (Cassata et al. 2000), and has broad developmental expression in the zebrafish (Toyama et al. 1998). Further, the LDB1-mediated Isl1/Lhx3 complex was recently shown to exist in *C. elegans* neurons and function in a similar way to the mammalian complex (Bhati et al.

2017). Recently, the *Arabidopsis* homolog of LDB1, SEU, was shown to regulate both light and temperature-sensing pathways as a transcriptional mediator of H3K4me3 methylation (Huai et al. 2018). This reinforces the depth of conservation of LDB1's functional role in mediating protein interactions to drive transcriptional outcomes.

Interestingly, models of LDB1 -/- stem cells show that ESCs are normal, and that approximately normal neuronal differentiation is achievable with expression of MAP2K unaffected in an adherent cell culture model, despite neuronal expression differences using the five-step embryoid body hanging drop culture model (Hwang et al. 2008). This suggests that LDB1 is required not for self-renewal, but rather for consolidation of specific cell fates, perhaps through stabilizing transcriptional interactions to establish terminal cell fate gene expression programs. This role must be mediated through the three-dimensional space of the nucleus, the DNA/RNA/protein milieu, and with concomitant remodeling of chromatin. LDB1, with its ability to bind transcription factors, chromatin remodeling complex members, architectural proteins, and self-associate through IDR-mediated interactions (including phase separation mechanisms), is likely so conserved and responsible for a large variety of phenotypes because it is a fundamental regulator of developmental 3D genome and chromatin architecture remodeling. LDB1 has also long been analyzed as a critical long-range interacting factor in expression of the β-globin locus, and as a transcriptional "master regulator" of erythropoiesis in general, which has been reviewed comprehensively by Love and colleagues (Love et al. 2014b).

The functional domains of LDB1 also appear to have separable effects on long-range interactions between enhancers and promoters versus solely gene activation (Krivega et al. 2014), implicating LDB1 in more than simply bridging LIM-HD TF complexes, but also in establishing long-range contacts in the 3D genome and recruiting additional transactivators through its dimerization domain. This role may involve the regulation of LIM-domain-mediated TF binding through properties of the intrinsically disordered region (IDR) in LDB1's self-association domain (Robertson et al. 2018), which has similarities to domains shown to catalyze the formation of transcriptional "hubs" at enhancer elements (Boija et al. 2018; Chong et al. 2018; Shrinivas et al. 2019).

LDB1 is also an intriguing therapeutic target, often described as a "looping factor" to facilitate long-range regulation of specific enhancer-promoter pairs. This function has been most clearly demonstrated in erythropoiesis, where LDB1 forms complex with GATA, LMO2, TAL1, and others (Wadman et al. 1997; Visvader et al. 1997; Love et al. 2014b; Li et al. 2013; Meier et al. 2006). LDB1-fusion proteins can restore developmentally induced contacts between promoters and the globin locus control regions (LCRs, also called enhancers) and drive gene activation at both the β- and γ-globin loci, thus showing promise as therapeutic approaches to both haemoglobinopathies and sickle-cell anemia where production of haemoglobin is a point of therapeutic intervention (Deng et al. 2012; 2014; Breda et al. 2016). Since LDB1 is a pleiotropic factor that is utilised across tissue types, any therapeutic approach could not chemically target LDB1 throughout the body. Driving increased (or decreased) LDB1 function with its other TF partners to target a specific cell state could be achieved by

introducing fusion protein constructs to a critical population *ex-vivo* and then reintroducing the cells into the patient (Rivella et al. 2018), or by introducing a vector-deliver combination of transgenes to drive a particular cell identity. Further exploration of cell-based reprogramming strategies and gene therapy approaches is beyond the scope of this thesis, but this the most reasonable future application of LDB1 biology to human disease.

The relationship between the function of gene regulatory elements and the physical organization of the nucleus and chromatin is interdependent (Vermunt et al. 2019). Thus, to understand and manipulate the gene regulatory elements responsible for MN identity, the role of LDB1 as a mediator of nuclear organization is of prime interest. While it has been well established that LIM-HD:LIM-domain binding protein stoichiometries regulated transcriptional outputs and cell fate choice in the MN subtypes (Song et al. 2009), the actual mechanism of how the transcription factor complexes maintain and establish these outcomes is unknown. Further, homeodomain transcription factors are extremely promiscuous DNA binding proteins, and have degenerate motifs with up to millions of binding sites in the genome, requiring multi-protein cooperative complexes to impart specificity (Gehring et al. 1994; Berger et al. 2008a). Therefore, there is a missing link between the formalistic models of LIM-HD TF stoichiometry determining cell fate and the actual formation of stable gene expression networks at a small subset of MN genes. We hypothesize that LDB1 functions as an integrator of TF binding to promote gene expression at a small subset of potential TF sites, and is correlated with open chromatin and epigenetic marks such as H3K27ac as well as spatial association of regulatory elements.

While LDB1 has long been recognised as a critical multimerization factor that promotes association of LIM-HD TFs, and also facilitates specific enhancer-promoter contacts (Matthews and Visvader 2003; Liu and Dean 2019), this thesis places it in the context of larger-scale 3D genome organization and as a specific integrator of gene regulatory networks with the transcriptional machinery.

Recently, LDB1 was identified as a critical regulator of *trans* contacts in mouse olfactory sensory neurons (OSNs), in combination with the LIM-HD TF Lhx2, to specify the unique expression of one olfactory receptor in each OSN (Monahan et al. 2019). This is despite the general consensus that individual 3D genomic organization is highly variable between individual cells of the same identity (Stevens et al. 2017; Finn et al. 2019). While the olfactory receptor model is likely unique to that biological system, LDB1 may be also coordinating an ensemble of *cis* and/or *trans* interactions in MNs in cooperation with the LIM-HD TF's Isl1 and Lhx3, as well as other unknown factors.

The only current assessment of LDB1's role in genome-wide chromatin organization comes from Hi-C and ChIP-seq studies in a conditional knockout model in mouse olfactory sensory neurons (Monahan et al. 2019), where it shows a remarkable function in mediating long-range *trans* contacts between disparately located enhancers to specify singular expression of a given olfactory receptor gene. There is additional evidence of LDB1-associated changes in nuclear organization as the canonical LDB1 target human gene *Hlxb9* locus moves from the nuclear periphery to the nuclear interior upon *in vitro* differentiation (Leotta et al. 2014). This type of sub-nuclear repositioning of specific loci is a well-characterised type of developmentally induced genome reorganization (Kosak et

al. 2002; Demmerle et al. 2013). This suggests an important role for LDB1 in re-organizing nuclear architecture to facilitate *Mnx1* expression, and is similar to many developmental activation events involving 3D genome remodeling (Shachar and Misteli 2017). However, any effect of LDB1 on either genome-scale contact frequencies, or sub-nuclear repositioning of developmentally important loci is not known in a MN context.

There is a large body of work describing the organismal and deleterious effects of LDB1 (mis)function, but the actual mechanism of action, and the implications for chromatin remodeling in development, is not yet clear. These interactions in diverse systems can be summarised as collections of LIM-HD TFs and their associated co-factors but there is not a unified model of how LDB1 integrates genomic information, despite excellent reviews of its function in several developmental contexts (Love et al. 2014b; Matthews and Visvader 2003) (**Figure 1.7**).

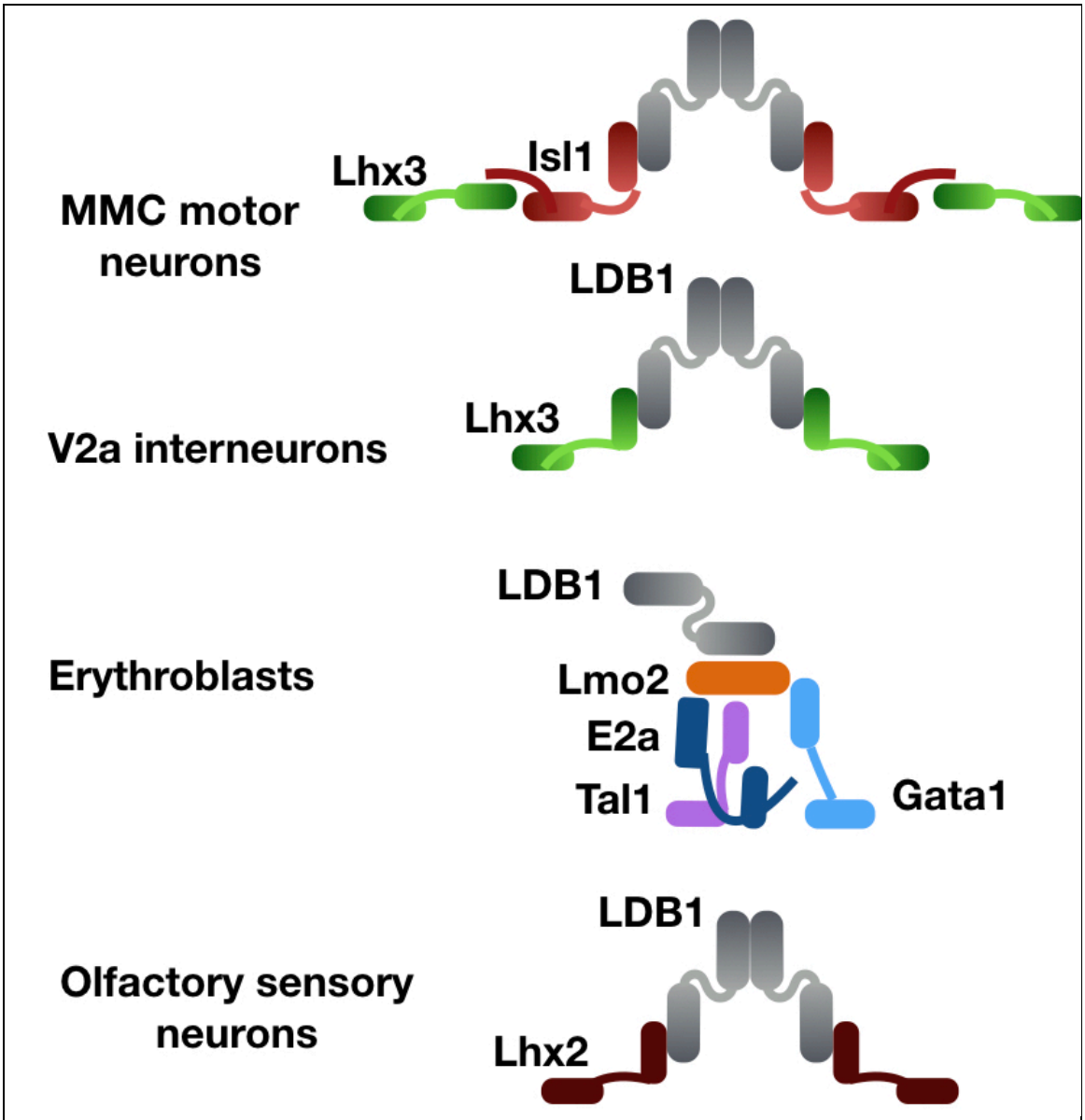


Figure 1.7 – Known roles of LDB1 in mediating transcription factor complexes
Schematic diagram of LDB1-involved complexes in a variety of transcriptional complexes from various tissue types, synthesised from various literature sources.

A summary of LDB1 and LDB2, with their partners in a range of tissues and organisms as determined from the literature may be found in **Table** 1.

Table 1 – LDB1 and LDB2 function in diverse tissues

Tissue	Partners	Organism	Reference

spinal motor neurons	*Isl1, Lhx3*	mouse, chicken, worm	(Thaler et al. 2002; Lee et al. 2008; Mazzoni et al. 2013; Rhee et al. 2016)
V2A interneurons	*Lhx3*	mouse, chicken	(Song et al. 2009; Thaler et al. 2002)
pancreas (β-cells)	*Lhx1*	mouse, rat, human	(Hunter et al. 2013; Ediger et al. 2016; Bethea et al. 2019)
pancreas (islet cells)	*Isl1*	mouse	(Hunter et al. 2013; Galloway et al. 2015)
CNS	*SSDP1*	zebrafish	(Toyama et al. 1998; Becker et al. 2002; Zhao et al. 2007; 2014)
pituitary glad	*SSDPs, Lhx2*	pig, mouse	(Susa et al. 2006; Cai et al. 2008; Zhang et al. 2015)
cornea	*Lhx2*	mouse	(Xu et al. 2007a)
fibroblasts	*SLK*	mouse	(Storbeck et al. 2009)
head development	*Otx1*	mouse	(Nishioka et al. 2005; Mukhopadhyay et al. 2003)
erythrocytes	*GATA1, GATA2, TAL1, KLF, LMO2*	mouse, human	(Meier et al. 2006; Song et al. 2012; Li et al. 2013; Soler et al. 2010; Lee et al. 2017; Song et al. 2007)
cortical motor neurons	*Ctip2*	mouse	(Leone et al. 2017)
cranial motor neurons	*Isl1, Lhx3, Lhx4*	rat	(Kim et al. 2016a)
glaucoma	*Lmxb1*	mouse	(Cross et al. 2014)
podocytes	*Pdx1*	mouse	(Suleiman et al. 2007)
cardiac	*Isl1*	mouse, fly	(Caputo et al. 2015; Werner et al. 2017)
retina	*Lhx2, Rnf12*	mouse	(de Melo et al. 2018; Gueta et al. 2016)
wing	*Ap*	fly	(Morcillo et al. 1997; van Meyel et al. 2003)

1.3 – LIM-HD TFs DRIVE MOTOR NEURON DIFFERENTIATION

The mammalian central nervous system (CNS) is a prime example of the importance of 3D genome reorganization in development (Bonev et al. 2017; Kishi and Gotoh 2018). In the developing spinal cord, diverse sets of motor neurons (MNs) innervate the body and provide electrical connectivity throughout the vertebrate body plan. The adult CNS is composed of many distinct subtypes of terminally differentiated neurons, each express different combinatorial codes of transcription factors (TFs) to drive stable cell identities (Francius and Clotman 2013). The diversity of MN subtypes in particular is coordinated by a "code" of LIM-domain (**L**in-11/**I**sl-1/**M**ec-3) (Gill 1995; 2003) containing transcription factors (Shirasaki and Pfaff 2002). This model of transcription-factor code-driven specification of the spinal cord is a well-established and elegant model of multi-level organization (Dawid et al. 1998; Thor et al. 1999; Pfaff and Macfarlan 2009). However, the mechanisms beyond transcription-factor binding, and the relationship of these TFs to the larger physical genome in development, remain largely unknown.

The diversity and timing of MN development is complex (Stifani 2014), but can be broken down to these fundamental processes. In developing MNs of the medial motor cortex, which innervates the torso and vital organs, retinoic acid (RA) and sonic hedgehog (Shh) signaling drive co-expression of the LIM homeodomain (LIM-HD) transcription factors Isl1 and Lhx3 (Jessell 2000; Lee and Pfaff 2001). Isl1 and Lhx3 bind each other through their LIM domains, and promote MN fate by binding to AT-rich sequences via their homeodomains, specifically activating expression of the "master regulator" MN transcription factor

Mnx1 (*Hlxb9* in human, also *Hb9* in mouse) (Thaler et al. 2002). The Mnx1/Hb9 protein itself functions as a transcriptional repressor of non-MN genes in postmitotic neurons (Arber et al. 1999). In contrast, expression of Lhx3 without Isl1 drives V2A interneuron identity instead of MN identity through activation of the V2A master regulator TF *Chx10* (also *Vsx2* in mouse) (Thaler et al. 2002; 2004). Later studies have shown the synergistic activity of Ngn2 (Lee and Pfaff 2003; Lee et al. 2005b), along with Isl1 and Lhx3, to specify MNs and induce MN identity through their expression and subsequent occupancy of key enhancer elements (Mazzoni et al. 2013). Expression of these TFs and others is sufficient to create induced MNs from fibroblasts, iPSCs, and ESCs (Hester et al. 2011; Son et al. 2011), and expression of an Isl1-Lhx3 fusion can induce MNs from ESCs (Lee et al. 2012). However these TF multimers are generally mediated through LIM-domain binding with adaptor proteins such as LDB1 (Thaler et al. 2002; Lee et al. 2008). While the enhancer-promoter interactions that occur throughout MN differentiation and direct transcriptional reprogramming of ESCs are highly dynamic (Rhee et al. 2016), there appear to be a small subset of essential MN genes driven by Isl1-Lhx3-LDB1 complexes (Lee et al. 2012; Mazzoni et al. 2013; Velasco et al. 2017). A subset of these genes are also bound by an Isl1-Onecut1 complex (Rhee et al. 2016), and the MN-specific Isl1-Lhx3- complex also targets its own expression (Erb et al. 2017) and may co-bind with transcription factors such as STAT3 (Lee et al. 2013) and adaptors such as SSDP proteins (Lee et al. 2016).

LDB1 is a critical component in the “LIM code” of transcription factor binding specification in the nervous system (Lumsden 1995; Gill 2003; Thor et al.

1999). The concept of combinatorial transcription factor regulation was refined through studying LIM-HD interactions in the developing spinal cord, an elegant model for cell fate choice driven by combinatorial expression of LIM-HD TFs (Tsuchida et al. 1994; Lee and Pfaff 2001; Shirasaki and Pfaff 2002). In MNs of the medial motor cortex, the LIM-HD TFs Isl1 and Lhx3 associated through LIM-LIM interactions and bind to AT-rich DNA binding sites through their HD, particularly activating the master regulator TF of terminal MN fate, *Mnx1* or *Hb9* in mouse, *Hlxb9* in humans, which functions as repressor to consolidate cell fate in post-mitotic neurons (Thaler et al. 2002; Arber et al. 1999). This LIM-HD TF cooperativity involves LDB1, the nature of the Isl1-Lhx3-LDB1 hexamer (**M**otor **N**euron **C**omplex, or **MNC**) being determined by chick electroporation assays showing that dominant-negative dimerization-domain mutants of LDB1 prevent formation of both interneurons, through tetramer formation with Lhx3, and MNs, through hexamer formation with Isl1 and Lhx3 (Thaler et al. 2002). Isl1 and Isl2 have also been shown to have pre- and post-mitotic roles in neuron specification, and Isl1 misexpression alone cannot determine a MN fate (Thaler et al. 2004).

The Isl1-Lhx3-LDB1 complex is essential for both *Mnx1* activation specifically, and for MN fate in general. Its components also interact with Ngn2 and NeuroM proteins for synergistic activation of MN specification (Lee and Pfaff 2003). In both MNs and interneurons, the requirement for LDB1 can be overridden by the artificial tethering of both Isl1 and Lhx3 homeodomain and LIM domains, suggesting that the primary role of LDB1 is to mediate protein-protein interactions between transcription factors (Thaler et al. 2002; Lee et al. 2012). The region significantly upstream of the *Mnx1* gene (originally called M250 and

derivatives, now referred to as ECR2) was identified as driving *Mnx1* expression, and taken to interact through a long-range mechanism with the proximal promoter around 10 kb in the 3' direction (Lee et al. 2004). Further, the relative stoichiometries of Isl1, Lhx3, LMO4, and LDB1 play an essential role in determining the MN vs. V2A interneuron fate choice (Song et al. 2009; Robertson et al. 2018)

In the established model of TF-induction-mediated reprogramming to neural cells, the Isl1 and Lhx3 proteins were mapped by ChIP-seq (Mazzoni et al. 2013), and shown to co-occupy tens of thousands of binding sites, including at the *Mnx1* ECR. While the genome-wide binding data confirmed MNC occupancy at critical MN genes, it did not resolve the nature of MNC binding, and whether the requisite Isl1 and Lhx3 motif sequences were distributed in any particular way, either at the critical regulatory regions or genome-wide. Subsequently, Isl1 and the associated neuronal TF Onecut were mapped (Rhee et al. 2016), confirming the transient nature of Isl1-binding sites throughout MN development, which complemented similar studies of transcriptional activation and chromatin remodeling throughout the MN developmental timecourse (Velasco et al. 2017). Many of these sites were also found in ChIP-seq data from an Isl1-Lhx3 fusion construct as well in ChIP-seq data from putative MNC complex member phosphatase (Lee et al. 2012; 2013),. While the ChIP-seq data is difficult to interpret, the Isl1-Lhx3 fusion protein was found to be sufficient to drive MN differentiation as measured by RNA-seq, Hb9 staining, luciferase reporter assays, and importantly *in-ovo* electroporation (Lee et al. 2012). Interestingly, the site of the fusion protein was determined by ChIP-seq and unsurprisingly found to

be more degenerate than an *in-vitro* target site (Lee et al. 2013). The Isl1-Lhx3 fusion ChIP data also claims to show that *Isl1, Lhx3*, and *Lmo4* are all targets of the MNC, implicating a self-regulating feedback loop that may contribute to establishing the MN-specific gene expression program (Erb et al. 2017).

The discrepancies between potential Isl1 and Lhx3 biding sites and the actual sites of MNC binding and function are not easily explained by existing data. The degeneracy of the short AT-rich HD motifs implies that specifying either pairs of canonical binding sites separated by a specific linker sequence, or stabilizing binding at low-affinity sites, is mediated by LDB1 and is essential for correct execution of the MN gene expression program. That HD TFs function through cooperative binding to specify target sites is long established (Wolberger 1996; Bobola and Merabet 2017), but the sequence-specificity implied by the existing Isl1/Lhx3 “hexamer” model for MN specification (Erb et al. 2017) is hard to reconcile with the absence of clear hexamer sites at the promoters of MN genes, or any preferential pairing between Isl1 and Lhx3 sites (see Chapter 3). The association kinetics of the entire Isl1/Lhx3/LDB1 hexamer was recently determined experimentally in an elegant study by Matthews and colleagues (Robertson et al. 2018). The DNA targets used to determine binding kinetics represent either the Isl1 binding site ATTA, the LHX3 site AATTAATTAA, or the longer “HxRE” motif with a combination of the two sites, were generated based on studies of the MNC complex at the canonical *Mnx1* proximal enhancer element ECR2. These targets were similar to SELEX results for either LIM-HD protein individually, or a non-LDB1-containing Isl1-Lhx3 fusion protein (Lee et al. 2008). These sequences were tested via EMSA and biophysical assays for

binding to their canonical target sequences (Robertson 2018), and shown to bind only in partnership.

Therefore, while LDB1 is well-established as a critical mediator of this transcriptional activating complex, the mechanism by which it functions and the discrepancy between functional binding and the wide number of possible binding sites is remains of interest. Arguably, the primary function of LDB1 as a transcriptional co-activator remains to be established. One model that could unify these disparate findings is a 3D-genome reorganization-based model whereby LDB1 stabilises LIM-HD TF binding to Isl1-Lhx3 sites that may or may not be linearly adjacent in the genome. Further, it may reinforce binding at a very restricted subset of MNC sites that have a sufficient richness of AT motifs. Independently of mediated LIM-HD TF binding, it may then recruit and self-associate through the IDR-containing N-terminal region to create a local concentration increase of chromatin remodelers, transcriptional machinery, and architectural ‘looping’ proteins to help establish higher-order organizational structures, potentially at transcriptional factories or nuclear speckles within the interchromatin space.

1.4 - Modes and motivations of understanding 3D genome organization

Many gene therapy approaches addressing haemoglobinopathies and sickle cell anemia have benefitted from research involving induced LDB1 tethering to target specific enhancer-promoter interactions (Breda et al. 2016; Rivella et al. 2018; Ghiaccio et al. 2019). At the same time, 3D-genome based methods are being pioneered as diagnostic tools (Hnisz et al. 2016; Blobel and Deng 2018; Hnisz et al. 2019). At their core, all transcription-factor-based cellular reprogramming strategies rely on “chromatin integrator” proteins such as LDB1. As many of these factors are seemingly ubiquitously expressed and required for survival, many TF-based and (and chemically induced) reprogramming strategies will succeed in a cellular (nuclear) milieu that is supportive of normal integrator protein function. These functions should include TF-binding dynamics such as dimer-of-dimer interactions in the Isl1/Lhx3/LDB1 axis in MN development, mediating higher-order transcriptional remodeling complexes, and contributing to the specific gene expression pattern of that cell type. Thus, each therapeutically desirable cellular state, or ‘reprogramming target’, is likely dependent on a specific and limited set of integrator proteins that cooperate with specific TFs to consolidate a particular cell fate. By mapping the effect of these proteins on the transcriptional, epigenetic, and 3D genomic outputs of differentiation, we may be able to map a localised, genetically defined, specific set of target-able elements, that when concomitantly activated or repressed through existing TF- or otherwise-mediated reprogramming and vector-delivery technologies, could dramatically improve both gene-editing and cellular reprogramming-based therapeutic outcomes. Further, new targets are being identified that drive neuron

regeneration and may be subject to regulation by LDB1 or 3D genome organization (Chen et al. 2018). The muscle LIM protein has also been shown to have function in the mouse axon, and drives regeneration after injury (Levin et al. 2019), indicating a potential future role in the therapeutic advantages of LIM-domain proteins. To illustrate the wide potential for targeting LDB1-related transcriptional complexes, the functional partners of LDB1 in a variety of human diseases, as determined from the literature, can be found in **Table 2**.

Table 2 – LDB1 partners in human diseases

Disease	**Organ**	**Partners**	**Reference**
nail-patella syndrome	podocytes	LMX1B	(Dreyer et al. 2000)
head and neck cancer	head and neck	SSDP1	(Simonik et al. 2016)
hemoglobinopathies	bone marrow	GATA / TAL / LMO4	(Song et al. 2007)
acute myeloid leukaemia	bone marrow	PU.1, SSBP2	(Schuetzmann et al. 2018) (Wang et al. 2010)
diabetes	β-cells	Isl1/2, Lhx1	(Hunter et al. 2013)
squamous cell carcinoma	oral cavity	LMO4	(Mizunuma et al. 2003)
diet-induced obesity	ß-cells	Isl1, Lhx1	(Loyd et al. 2017)
colorectal cancer	Colon	Wnt signaling	(García et al. 2016)

Further, the role of chromatin organization as a cause or consequence of gene transcriptional and developmental activation is still unclear. There is vigorous debate in the field about the fidelity of 3D interactions as determined by microscopy in contrast with population-averaged genome-wide data (Williamson et al. 2014), but while some underlying principles are clear, locus-specific

dynamics, causation, and the functional requirements for 3D genome organization remain to be determined. Since the underlying rules governing 3D genome control seem to be both locus and developmental context-specific, there is a need for further characterization of key 3D genome regulators like LDB1, and a need for cross-modal approaches to study genome organization, as has been called for (Dekker et al. 2013; Ay and Noble 2015; Marti-Renom et al. 2018).

1.5 - GAPS IN KNOWLEDGE OF LDB1 FUNCTION AND DEVELOPMENTAL CHROMATIN REMODELING

While homeodomain proteins bind target sites with high affinity and a degree of specificity, particularly across large families, these proteins have broadly similar binding sites centring on AT-rich sequences, which complicates the task of identifying their specific regulatory targets (Affolter et al. 2008). A comparison of most mouse homeodomains showed that while some homeodomain families have clear DNA-binding patterns, many others such as LIM-HD proteins in the Lhx family have degenerate and non-specific binding sites, and therefore rely on other factors than DNA-sequence specificity (Berger et al. 2008b). This level of non-specificity both expands the range of biological information accessible to each type of TF, and also provides flexibility to function across many developmental contexts (Chu et al. 2012). Thus, the ability of LIM-HD-interacting proteins like LDB1 to recruit a variety of different co-factors is essential to imparting developmental specificity on LIM-HD TFs. In the erythropoietic system, where LDB1 complexes with GATA1, LMO2, TAL1, and E2A (Ryan et al. 2007), GATA1 and LMO2 also compete for binding to each other, affecting the sequence specificity of GATA1 in complex with the LDB1-interactors (Wilkinson White et al. 2011).

In mouse MNs, direct immunoprecipitation experiments have also shown that the Isl1-Lhx3 complex, and by assumption LDB1, interact with p300 and RbBP5 (Seo et al. 2015). In zebrafish erythroleukemic cells, biotin-labeled LDB1 pulled down a complex including Eto-2, Cdk9, LMO4, and other known components of the mouse erythroid LDB1 complex when assayed by mass

spectroscopy and functional validation (Meier et al. 2006). LDB1 was shown early on to mediate homo- and hetero-dimerization of TFs (Jurata et al. 1998; van Meyel et al. 1999), which an essential property for many types of TF function and specificity (Milan and Cohen 1999) (Smith and Matthews 2016). LDB1 can form multimers while LDB2 can form octamers in *in vitro* condition (Cross et al. 2010), predicting a multimeric complex of TFs and associated proteins is assembled at sites of LIM-HD TF-mediated LDB1 occupancy, but none of these models has been demonstrated experimentally.

The protein structure interactions between LDB1 and its LIM-domain partners are challenging to study due to the intrinsically disordered nature of many of the domains. However, several crystallography studies have been performed on isolated fragments, such as between LIM-interacting domains (LIDs). The structural logic for association of the hexamer is that both Isl1 and LDB1 LIDs interact with Lhx3, thereby providing a surface of competition for either tetramer or hexamer formation (Bhati et al. 2008). Solution structures for some combinations of LIM- and LID- domains have been solved, such as between Lmo2-LIM and LDB1-LID (Dastmalchi et al. 2012), Lhx3-LIM and Isl1-LIM (Deane et al. 2003), Lmo4-LIM and LDB1-LID (Omari et al. 2011), and Lhx3-LIM and LDB1-LID (Bhati et al. 2008) among others. While diffraction data of an Isl1-Ldb1 fusion protein engineered to prevent insolubility of LIM-domain/LIM-only protein interactions was acquired, the complex structure could not be determined due to excessive radiation damage (Gadd et al. 2012), making the overall structure of the hexamer complex difficult to determine. Interestingly, the association kinetics of the LDB1 dimerization or self-association domain are

modest (10^{-5} M) (Cross et al. 2010), but the effects of binding kinetics on its partner LIM-HD TF's to DNA, but not to each other, are substantial and lead to a complex web of possible interactions leading to a DNA-bound Isl1-Lhx3-LDB1 complex (Robertson et al. 2018). Thus, identifying cofactors that may play a role in modulating the stability and formation of this complex is of interest.

An intriguing class of LDB1-interactors is the Single-Stranded Binding Protein (SSDP1 in humans or SSDP in mice) class, originally discovered as pyrimidine-tract-binding associated proteins at the promoter of the collagen gene in chick erythrocytes (Bayesian and Lukens 1996), This class of protein both interacts with and promotes the function of LDB1 and LDB2 in *Xenopus*, *Drosophila,* and mammals (Chen et al. 2002). SSDP1/SSBP3 also interact with LDB1/Isl1 complexes in the pancreas (Galloway et al. 2015). One of the critical functions of SSDP1/SSBP3 in the context of the erythroid enhancer complex is to protect LDB1 and its associated complex from ubiquitin-mediated degradation by inhibiting interactions with the ubiquitin ligase RLIM (Xu et al. 2007b), which is also a known LDB1-complex component and negative regulator of complex stability (Bach et al. 1999; Hiratani et al. 2003). Stabilization of LDB1 by SSDP1 from RLIM-mediated ubiquitination, even as LDB1 stabilised LIM-HD proteins, was clearly shown in a series of biochemical experiments in zebrafish (Güngör et al. 2007). In *Drosophila,* the LDB1-SSDP1 interaction was characterised in neural and wing tissues, and deletion of SSDP1 causes cell-lethal phenotypes (van Meyel et al. 2003). Additionally, the SSDP1 mutant mouse has major defects in head development mediated by the Lhx1/LDB1 complex (Nishioka et al. 2005). SSPB2 mutants in mice also cause an increase in carcinomas and B cell

lymphomas, and regulate LDB1 stability (Wang et al. 2010). In zebrafish, overexpression of SSDP1 increases *in* vivo levels of LDB1 while also hindering CNS development (Zhong et al. 2011). The haematopoietic defects were also noticed in terms of stem cell maintenance and stress response (Li et al. 2014). The ChIP(LDB1)/SSDP1 axis in flies is also conserved in several Wnt-dependent developmental pathways (Fiedler et al. 2015). In *Xenopus* LDB1 is actively degraded by Rnf12/RLIM to adjust LDB1:XLIM stoichiometries (Hiratani et al. 2003), so SSDP1-mediated protection of LDB1 from degradation may also play a critical role in adjusting LDB1::LIM stoichiometry in MNs.

This is in line with the conventional understanding of LDB1 mediating Isl1-Lhx3 dimers that bind to paired motifs, which thereby specify MN genes (Thaler et al. 2002; Lee et al. 2012; 2013; Mazzoni et al. 2013; Rhee et al. 2016). However, a bioinformatics analysis of the SELEX- and luciferase-determined hexamer response element (Lee et al. 2008), and a similarly derived tetramer site used in detailed kinetic studies of Isl1/Lhx3/LDB1 binding via EMSA (Robertson et al. 2018), shows only a marginal number of putative tetramer or hexamer sites occupied by an Isl1 or LDB1 ChIP signal (see Chapter 3). Since the HxRE upstream of *Mnx1* is reported in literature as the 'only known' motif with a symmetrical, directly adjacent set of two half sites (Lee et al. 2013), perhaps this is a motif arrangement unique to that locus, and other LDB1-bound regions may be regulated by sets of half-sites with variable spacing, or perhaps that are substantially separated from each other over linear genomic distances. Given the extremely common presence of canonical Isl1 and Lhx3 binding motifs, only some of which are paired or adjacent at LDB1 or Isl1-FLAG peaks, it is difficult to

claim that MN-specific regulatory regions could be predicted by specific orientation of Isl1 and Lhx3-binding sites. Fundamentally, the HxRE as determined by enhancer-bashing the *Mnx1* locus, SELEX with Isl1/Lhx3 LIM-HD TFs, or luciferase reporter assays, is sufficient to drive Isl1 and Lhx3 binding at its *Mnx1* location and to drive *Mnx1* transcription. However, HxRE placement is not informative for specifying the rest of the MN specific gene expression profile beyond *Mnx1*. Thus, there is a critical unmet need for new approaches in direct reprogramming, gene therapy, and understanding of the fundamental biology of MNs to progress, and the current models of transcriptional activation in this cell type are insufficient.

1.6 - AIMS OF THIS STUDY

To explore the function of LDB1 in motor neuron development, we have developed an embryonic stem cell (ESC)-based system to explore the function of spinal motor neuron specific LIM-HD complexes in the regulation of MN gene

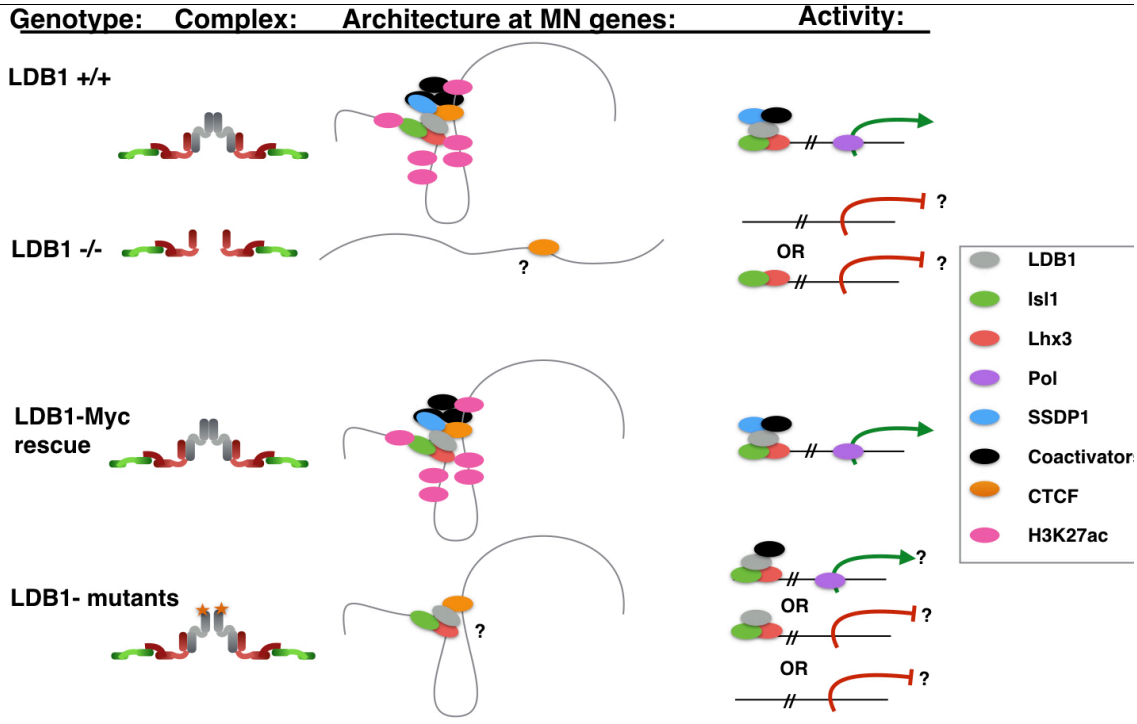


Figure 1.8 – Hypothesis: LDB1 regulates genomic architecture and activity at motor neuron genes
Schematic representation of hypothesis for LDB1's effects on MNC complex binding, genome organization, and gene expression at motor neuron genes.

expression and 3D chromatin architecture. We propose a relationship between LDB1 function, genome architecture, TF binding, and gene expression similar to **Figure 1.8**. We developed a transgenic ES cell line that upon doxycycline treatment expresses epitope-tagged equimolar transcripts of the pro-neural TF Ngn2 along with the MN-inducing LIM-HD transcription factors Isl-1 and Lhx3, (which form a hexameric complex with LDB1), in both *Ldb1* wildtype and *Ldb1* knockout backgrounds. We also create cell lines with doxycycline-inducible wildtype rescue and LCCD-domain-mutant forms of LDB1. These cell lines are

then differentiated into embryoid bodies, and assessed with a variety of epigenomic, transcriptomic, and functional assays. We aim to determine the effect of LDB1 on: Isl1 and Lhx3 genomic binding; H3K27ac occupancy and open chromatin; gene expression; association of Isl1 and Lhx3 with each other; and genomic organization both genome-wide and at H3K27ac and LDB1-bound regions. We seek to identify a minimal set of genes or regulatory regions that correlate with MN identity, as well as to establish the genomic and molecular associations of LDB1 and its partners within the context of MN development.

CHAPTER 2. METHODS

2.1 - BIOCHEMICAL, CELL, AND MOLECULAR BIOLOGY METHODS

CELL CULTURE

ESCs were derived from blastocysts isolated at day EA. from crosses between LDB1+/- (generated in the Macfarlan lab) and A2Lox.cre::rTTA mice (gift of the Kyba Lab).After colony expansion, feeder removal, and genotyping, *Ldb1 +/+* or *Ldb1 -/-* male cells were then exposed to 0.5 *μ*g / mL of doxycycline to induce Cre expression, and the ICE reaction was initiated by electroporation (Amaxa) of the p2Lox plasmid containing either the Ngn2/Isl1/Lhx3 (NIL) construct or the NIL construct with wildtype or SSDP1-mutant LDB1 as appropriate (see **Figure 3.2),** resulting in self-excision of Cre, integration of the transgene of interest, and activation of a *neo* resistance marker, all at the X-linked *HPRT* locus (Iacovino et al. 2011). Surviving cells were grown to form individual G418-resistant colonies, genotyped, and selected clones were expanded and passaged on TC-treated culture dishes coated with 0.1% gelatin in standard ESC media (DMEM with 15% FBS, 1% L-Glutamine, 1% HEPES, 1% NEAA, 1% Anti-anti, 0.1% BME, 0.01% mLIF). Cells were passaged with 1x TrypLE every other day, growing to a maximum confluency of 70%. Cell were treated with G418 (175 *μ*g/mL) as a selection agent to ensure the continued presence of the NIL cassette, and with Plasmocin (1:2000) as an anti-plasmodium agent 2-4 passages prior to differentiation, as necessary. Cells were grown at 37˚C in 5% CO_2.

Embryoid bodies (EBs) were grown by resuspending trypsinised ES cells in ADFNK media (DMEM-F12 and Neurobasal medium mixed equally with 10% KOSR, 1% L-Glutamine, 1% Anti-anti, 0.01%) BME at a density of 2 x 10^6 cells in

10 mL ADFNK in a 10 cm untreated Petri dish, or at 5 x 10^6 cells in 25 mL ADFNK in a 15 cm untreated Petri dish. Dishes were placed on a rotating platform inside a 37˚C and 5% CO_2 incubator, and rotated at approximately 50 RPM. After 48 hrs, EBs were collected by gravity, and the ADFNK media exchanged to contain 1.0 μM retinoic acid (RA) and 1.0 μM smoothened agonist (SAG). After 96 hrs, the media was exchanged again to contain 1 μM RA, 1.0 μM SAG, 1.0 μM doxycycline, and 2.0 μM of the Notch pathway inhibitor DAPT for another 48 hrs before collection at the 144 hr timepoint, referred to hereafter as Day 6.

Western Blots

EBs were collected, washed in PBS, and flash frozen on dry ice. Cells were lysed in 1% Igepal / 150mM NaCl / 10mM Tris pH 7.5 for 30 min on ice, then sonicated with a Bioruptor Plus (Diagenode) for seven cycles on high for 30 sec on / 30 sec off. Lysates were then spun at 10,000 x *g* for 15 min at 4˚C and the supernatant collected. Protein concentration was measured by bicinchoninic acid (BCA) assay (Pierce), and 10 μg of protein was diluted with LDS loading buffer, boiled for 10 min, and run on a 4-12% Bis-Tris gel (Invitrogen) for 35 min at 200 V in MES buffer (Invitrogen). Gels were then transferred to 0.45 μm nitrocellulose membrane (Invitrogen) and transferred in 1x Transfer buffer (Invitrogen) + 20% MeOH for 60 min at 30 V. Membranes were blocked in 5% milk/TBST for 1 hr, and incubated overnight at 4˚C with the appropriate antibody in 5% milk/TBST. Membranes were then washed 3 x 5 min in TBST, followed by incubation with the appropriate secondary antibody in 5% milk/TBST. HRP secondary was detected

with electrochemiluminescence (GE Healthcare), and autoradiography film (HyBlot) and a traditional developer, or an Azure c600 detection instrument was used for detection.

FACS

Samples for FACS were prepared from Day 6 embryoid bodies, dissociated with the Worthington Papain dissociation system as per manufacturer's instructions. At least 1 million cells were processed, and then counted for GFP+ fluorescence on a FACScalibur (BD Biosciences) with a single gate for live cells, and one channel (488nm) acquisition. Cell numbers were collected from doxycycline and non-doxycycline-treated EBs, and the absolute percentages of GFP+ cells were compared in FloJo 10.6.

IMMUNOFLUORESCENCE

EBs were collected at Day 6 in PBS and fixed in 4% EM-grade paraformaldehyde/PBS for 15 min, washed thrice in PBS-T (0.01% Tween-20 / 1x PBS) and collected by gravity at each step, permeabilised in 0.2% Triton-X-100 for 10 min, and blocked in 3% BSA/PBS for 1 hr, all at room temperature (RT) with nutation. Appropriate antibodies were added in 3% BSA/PBS, incubated for 1 hr at RT, and EBs were washed thrice with PBST before addition of the appropriate secondary antibody in 3% BSA/PBS and incubation for 1 hr at RT. EBs were then washed thrice in PBS-T, once in 0.2 μg/mL DAPI/PBS, and once in H_2O before being equilibrated in Vectashield and transferred onto a slide, excess media drained, and additional Vectashield added before sealing with a

coverslip and clear nail polish. Slides were imaged on a Leica CTR 600 at 20X magnification and a Hamamatsu ORCA-ER camera. Images were processed in Velocity 6.1.1.

2.2 - Sequencing Methods

ChIP-Seq

Chromatin immunoprecipitation and sequencing was performed as described with significant modifications (O'Geen et al. 2010). DAY 6 EBs were collected by gravity, washed once in 1X PBS, and fixed in 1% fresh methanol-free formaldehyde (Thermo) in 1x PBS for 10 min at 22˚C. Fixation was quenched with 2.0 M glycine to a final concentration of 0.4 M glycine for 10 min at 22˚C, and cells were spun at 435 x *g* for 5 min, before washing in ice-cold PBS once. The pellet was stored at -80˚C. Cells were thawed and incubated in cell lysis buffer (0.1% Igepal / 5mM PIPES / 85mM KCl / protease inhibitors) for 15 min nutating at 4˚C, and nuclei were isolated by douncing with a Type B pestle 20 times per sample, and spinning down at 435 x *g* for 5 min at 4˚C. Nuclei were resuspended in nuclear lysis buffer (1% SDS / 50mM Tris pH 8.0 / 10mM EDTA / protease inhibitors) and incubated for 30 min on ice. Sonication was performed on a Bioruptor Plus (Diagenode) on high with 30 sec on / 30 sec off for 30 cycles with a 5-minute break between each set of 15 cycles. The resulting chromatin was spun at 10,000 x *g* for 10 min and the supernatant was removed and frozen at -80˚C. Once thawed, chromatin was incubated with the appropriate antibody overnight at 4˚C. BSA-blocked Protein A or Protein G Dynabeads were added to the chromatin-antibody solution, and nutated at 4˚C for 2 hrs, before the beads

were isolated by magnet and washing initiated. Beads were washed twice with Wash Buffer 1 (1% Igepal, 0.25% Na-deoxycholate, 150mM NaCl, 50mM Tris pH 7.4, 10mM EDTA), thrice with Wash Buffer 2, (1% Igepal, 1% Na-deoxycholate, 500mM LiCl, 100mM Tris pH 9.0), and once with Wash Buffer 3 (1% Igepal, 1% Na-deoxycholate, 500mM LiCl, 150mM NaCl, 100mM Tris pH 9.0). Antibody-DNA complexes were removed from beads with ChIP Elution Buffer (1% SDS, 50mM $NaHCO_3$) and vigorous shaking for 30 min at 37˚C. Crosslinks were reversed by addition of NaCl to 0.54M final concentration, and incubation at 67˚C overnight, followed by RNAseA treatment at 37˚C for 20 min, and Proteinase K treatment at 55˚C for 1 hr. ChIP DNA was purified with either Qiagen MinElute or Zymo DNA Clean and Concentrator kits, and assayed via Qubit before qPCR or ChIP-seq library preparation. Library prep was performed with the Takara/Clontech DNA SMART kit, according to manufacturer specifications, and analyzed by Qubit and Bioanalyzer before being run at 50 bp single-end in on a HiSeq 2500 instrument (Illumina).

NUCLEAR RNA-SEQ AND RT-QPCR

EB and ESC samples were collected via gravity or trypsinization, respectively, washed in PBS, and flash-frozen. Total RNA was isolated with the RNeasy kit (Qiagen) according to manufacturer's instructions and using gDNA eliminator columns. RNA-seq libraries were constructed by the NICHD MGC core, with the TruSeq Stranded mRNA Library Prep Kit (Illumina) and poly-A enriched, before quantification with a Bioanalyzer. Samples were run as 50 bp paired-end on a HiSeq 2500 Instrument. cDNA for RT-qPCR was generated with the

SuperScriptIII kit (Invitrogen) using random hexamers. RT-qPCR was performed with EvaGreen mastermix and specified primers as per manufacturer's instructions for 40 cycles on a Biorad LightCycler 480 Instrument.

ATAC-SEQ

EBs were collected in PBS, lysed in HS Lysis buffer (0.05% Igepal / 50mM KCl / 10mM $MgSO_4$ / 5mM HEPES / 1x PI) for 10 min at 4˚C, and nuclei were isolated by douncing with a Type A pestle 10 times per sample, and spun at 1500 x *g* for 5 min at 4˚C. Nuclei were collected and washed twice in buffer RS and nuclei counted. For transposition, 5 x 10^5 nuclei were aliquoted into a reaction with 1 μL Illumina TN and incubated for 30 min at 37˚C with shaking at 300 rpm. Libraries were amplified with Illumina reagents and indexes per kit instructions for ten PCR cycles, cleaned, and quantified with Qubit and Bioanalyzer before being run as 50 bp single-end in on a HiSeq 2500 instrument (Illumina).

HI-C

In-situ HiC (Rao et al. 2014) was performed as described with some modifications. EBs were collected at Day 6, washed in PBS, and fixed in 1% EM-grade formaldehyde for 10 min at RT, quenched with 0.4 M glycine for 10 min, washed and flash-frozen. 10 x 10^6 cells were then subjected to the Arima Hi-C protocol (Arima Genomics), using the manufacturer's Mammalian Cells protocol with no modifications. Hi-C libraries were amplified with the KAPA Hyper Prep (KAPA/Roche) kit and IDT for Illumina UDI Indexes (Illumina) as per

manufacturer's instructions, quantified with KAPA Library Quantification Kit, dual size-selected with AmPure XP beads (Agencourt), and analyzed by Qubit and Bioanalyzer before being run as single-indexed, paired-end 75 bp on a HiSeq 2500.

HI-CHIP

HiChIP was performed for some samples as described (Mumbach et al. 2016) with some modifications. EBs were collected at Day 6, washed in PBS, and fixed in 1% EM-grade formaldehyde for 10 min at RT, quenched with 0.4 M glycine for 10 min, washed and flash-frozen. 1.5 x 10^7 cells were lysed in Hi-C Lysis Buffer (0.2% Igepal / 10mM Tris pH 8.0 / 10mM Naci / protease inhibitors) nutating for 30 min at 4˚C. Nuclei were isolated by douncing with a Type B pestle 20 times per sample, and spun at 2500 x *g* for 5 min at 4˚C, before lysis with 0.5% SDS for 10 min at 62˚C and quenching with 1% Triton-X-100 for 15 min at 37˚C. The *in-situ* Hi-C library was generated with 15 *μ*L of MboI digestion for 2 hrs. After MboI inactivation, restriction fragments were filled-in with biotin-dATP and unlabeled nucleotides with 50 U DNA PolI Large (Klenow) for 1 hr at 37˚C, and then proximity ligation was performed with 4000 U T4 DNA Ligase (NEB) for 4 hr at RT. Nuclei were then lysed in Nuclear Lysis Buffer (1% SDS / 50mM Tris pH 7.5 / 10mM EDTA / 1X PI), sheared with a Covaris ME220 (Duty Factor = 5, Avg Intensity = 3.75, CPB = 200, Time = 240 seconds), spun at 16000 x *g* for 15 min at 4˚C, and the supernatant removed. Chromatin was diluted with ChIP Dilution buffer (0.01% SDS / 1.1% Triton-X-100 / 167mM NaCl / 16.7mM Tris pH 7.5 / 1.2mM EDTA / 1X PI), pre-clearance of chromatin with Protein A or G beads for 1

hour at 4˚C, and incubation overnight at 4˚C with the appropriate antibody. The next day, Protein A or G beads were added, bound for 2 hrs at 4˚C, and washed three times each in Low Salt (01.% SDS / 1% Triton-X-100 / 150mM NaCl / 20mM Tris pH 7.5 / 2mM EDTA), High Salt (01.% SDS / 1% Triton-X-100 / 500mM NaCl / 20mM Tris pH 7.5 / 2mM EDTA), and LiCl (1% Igepal / 1% Na-deoxycholate / 250mM LiCl / 10mM Tris pH 7.5 / 1mM EDTA) buffers at RT. Hi-ChIP DNA was eluted with ChIP Elution Buffer (1% SDS, 50mM $NaHCO_3$) and vigorous shaking for 30 min at 37˚C, before Proteinase K treatment overnight at 55˚C, and reversal of crosslinks at 67˚C with shaking for at least 2 hrs. Hi-ChIP DNA was eluted with Zymo purification columns and quantified with Qubit. For transposase-based library construction, 50-150 ng of Hi-ChIP DNA was pulled down with Streptavidin C-1 beads, washed twice in Biotin Binding buffer (2M NaCl / 10mM Tris pH 7.5 / 1mM EDTA), twice in Tween Wash Buffer (0.05% Tween-20 / 1M NaCl / 5mM Tris pH 7.5 / 0.5mM EDTA), and once in TD Buffer (20% DMF / 20mM Tris pH 7.5 / 10mM $MgCl_2$) before on-bead transposition with 1–4μL of Illumina Nextera Tn5, depending on input mass, and Illumina transposase buffer, with transposition for 10 min at 55˚C and interval shaking. Samples were then inactivated with 50mM EDTA for three washes, first at 50˚C for 30 min, 3 min twice, and two washes with Tween Buffer, and once with 10mM Tris pH 7.5. Transposed on-bead libraries were then amplified by PCR, using Nextera amplification reagents and indexes, for 7-10 cycles, depending on DNA input amount. Amplified libraries were dual-size selected with SPRI beads, and eluted in water before quantification with Qubit and Bioanalyzer. Libraries were sequenced as dual-indexed 100 bp paired-end reads on a HiSeq 2500

instrument. Alternatively, Arima Hi-C libraries were generated (see above) with the Mammalian Cell Lines protocol as per manufacturer's instructions, but prior to cross-link reversal, samples were integrated into the Hi-ChIP protocol above, with identical library construction procedures.

CAPTURE-C

Capture-C was performed as described (Davies et al. 2016) with some modifications. Briefly, 2 x 10^7 EBs were collected at Day 6, washed in PBS, and fixed in 2% EM-grade formaldehyde for 10 min at RT, quenched with 0.4 M glycine for 10 min, lysed, washed and flash-frozen. Cells were thawed, lysed again, and dounced with a Type B pestle 60 times per sample before nuclei were collected by spinning 140000 rpm for 5 min at RT, permeabilised in 1% SDS, and stopped with 1.67% Triton-X-100. 3C libraries were constructed by digestion with a total of 1500 U of DpnII (NEB), added stepwise, shaking at 1400 rpm overnight at 37˚C, with appropriate digestion/ligation controls. Enzyme was then inactivated at 65˚C for 20 min, and proximity ligation was performed with 240 U T4 ligase (Thermo) and shaking at 1400 rpm overnight at 16˚C. Crosslinks were reversed at 65˚C overnight, with simultaneous Proteinase K treatment. Capture-C and control DNA was phenol/chloroform purified, and run on an agarose gel assess digestion and ligation efficiencies. For samples that were digested and ligated appropriately, 5-6 μg of 3C library was sonicated on a Covaris ME220 (Duty Factor = 10, Avg Intensity = 17.5, CPB = 200, Time = 360 seconds), and SPRI cleaned. Library construction was performed with NEBNext DNA Library Prep Kit according to manufacturer's instructions with SPRI cleanup after each step, and

Herculase (Nimblegen) was used for six cycles of PCR addition of indices with NEBNext Multiplex Oligos. Sequence-specific capture of indexed 3C libraries was performed with Nimblegen SeqCap EZ Hybridization/Wash and Accessory kits, with SeqCap EZ HE-oligo kits for blocking and indexing oligos. Capture-C oligos were generated as 5'-biotinylated Xgen Lockdown Probes (IDT), and were resuspended so that each of the 220 individual oligos was at 2.9 μM. The blocking, indexing, and capture oligos were added to the 3C libraries with hybridization components as per manufacturer's instructions, and incubated at 47˚C for 72 hrs. Biotinylated oligos and their hybridization partners from the 3C library were then pulled down with M270 Streptavidin beads and washed as per manufacturer's instructions. Capture-C libraries were then subjected to 14 cycles of PCR using KAPA HiFi ReadyMix and SeqCap indexing oligos, before SPRI cleaning, and 2 μg of amplified captured material was hybridised for a second round with the same combination of blocking, indexing, and capture oligos, this time for 24 hr at 47˚C. The entire biotin pull-down, washing, and PCR amplification step was performed again, before SPRI cleanup, and quantification with Qubit, KAPA Library Quantification Kit, and Bioanalyzer. Pooled samples were then sequenced as a dual-index, paired-end 75 bp on a HiSeq 2500 instrument.

2.3 - Super-Resolution Microscopy Methods

Immunofluorescence for Structured Illumination Microscopy

Immunofluorescence for Structured Illumination Microscopy was performed as described (Kraus et al. 2017), with minor modifications. Doxycycline induced or uninduced ESCs were grown on high-precision coverslips, before washing with

PBS and fixation with 4% EM-grade formaldehyde/PBS for 15 min. Coverslips were washed thrice with PBS-T, permeabilised with 0.2% Triton-X-100/PBS, and blocked with 1x MaxBlock for 30 min at RT before incubation with appropriate primary antibody in MaxBlock for 1 hr at RT, washing, and secondary antibody incubation in MaxBlock for 1 hr at RT. Coverslips were then washed thrice in PBS-T, incubated in 0.2 μg/mL DAPI, and pre-equilibrated in Vectashield before careful transfer to slides and mounting in Vectashield and sealing with nail polish. Coverslips were thoroughly cleaned with 70% EtOH and chloroform prior to imaging.

Structured Illumination Microscopy

Super-resolution 3D-SIM images were acquired with the OMX V3 Blaze system (Applied Precision, GE Healthcare) using an UPLSAPO 60X/1.42NA oil immersion objective lens (Olympus) and pco.edge 5.5 sCMOS cameras (PCO). Acquisitions were reconstructed with SoftWoRx 6.5.2 (Applied Precision / GE Healthcare). Images were collected according to best practices with appropriate calibration controls (Demmerle et al. 2017). SIMcheck was used for quality control (Ball et al. 2015).

Structured Illumination Microscopy Analysis

3D-SIM super-resolution images were reconstructed from the 16-bit raw data with SoftWoRx 6.5.2 using a Weiner Filter setting of 0.002 and apodization filter of 1.0 with channel-specific OTF sets as described in (Demmerle et al. 2017). Channel

registration was performed with OMXeditor software (gift from John Sedat lab) on multi-colour Tetraspeck beads. 16-bit unthresholded images were then analyzed with the “nearest neighbor analysis” or “NNA” script written in Octave (Eaton et al. 2014) by David Pinto at Micron Oxford Advanced Bioimaging Unit. Some additional analysis was performed in R. Image stacks were created using SIMcheck and FIJI (Schindelin et al. 2012).

2.4 – Quantification and statistical analysis methods

ChIP-seq analysis.

After quality control with FastQC, data was processed from 50 bp single-end reads using in-house scripts on the NIH Biowulf cluster (Martin 2011). The first 3 3’ bases were removed as per library kit instructions. Reads were then aligned to the mm9 genome using Bowtie*2* (Langmead and Salzberg 2012) using default parameters. After alignment and sorting, duplicates were removed with samtools (Li et al. 2009), and the SMARTcleaner algorithm for removal of Clontech SMART library polyA/T duplicates was run with minimal modifications (Zhao and Zheng 2018). Coverage of reads was calculated using deeptools BamCoverage function using default parameters (Ramirez et al. 2016). Peaks were called using MACS2 (Zhang et al. 2008), after which peak centres were expanded by 200 bp in each direction and converted to FASTA with the getfasta Bedtools utility (Quinlan and Hall 2010), and motifs were called using MEME-ChIP (Ma et al. 2014). Biological replicates were merged at the .bam stage and converted into coverage tracks for each sample.

RNA-SEQ ANALYSIS

Untrimmed data was aligned with STAR (Dobin et al. 2013) to the mm9 genome using default parameters, and converted to count tables with the featureCounts (Liao et al. 2013a) module of Subread (Liao et al. 2013b) using default parameters, or indexed and converted to bigwig format with bedtools (Quinlan and Hall 2010). Differential expression was determined with DEseq2 (Love et al. 2014a). Biological replicates were merged at the .bam stage and converted into coverage tracks for each sample.

ATAC-SEQ ANALYSIS

After quality control with FastQC, Illumina adaptors were trimmed using CutAdapt (Martin 2011) and TrimGalore, and single-end 50 bp reads were aligned with Bowtie2 using default parameters (Langmead and Salzberg 2012). Coverage of reads was calculated using the deeptools BamCoverage function using default parameters (Ramirez et al. 2016). Biological replicates were merged at the .bam stage and converted into coverage tracks for each sample.

HI-C ANALYSIS

Hi-C raw data was processed with Hi-C Pro, version 2.10 (Servant et al. 2015) using the Arima Hi-C restriction enzyme file, and aligned with standard bowtie settings to the mm9 genome. Data was converted to .hic format with the hicpro2juicebox.sh function in HiC-Pro, and visualised in the Juicebox application (Durand et al. 2016a). Loop calls were made with the HICCUPS module of Juicer

Tools (Durand et al. 2016b) with parameters –m 1024 –r 5000 –k KR –f .1 –p 4 –l 7 –d 20000, and the “Ignore_sparsity” flag.

Replicates were generated by separate HiC-Pro runs, and then merged at the .allValidPairs stage before conversion to .hic format with the hicpro2juicebox.sh utility, and then conversion into cooler (Abdennur and Mirny 2019) format with the hic2cool script provided by the 4DN consortium. All contact-based maps were visualised in HiGlass (Kerpedjiev et al. 2018). Pileup analysis was made with the APA function of Juicebox using default parameters, or the coolpup.py and plotpup.py tools (https://github.com/Phlya/coolpuppy).

HI-CHIP ANALYSIS

Paired-end reads were aligned to mm9 with Hi-C Pro (Servant et al. 2015) using an MboI restriction enzyme digest file, and standard Bowtie settings. Data was converted to .hic format with the hicpro2juciebox.sh function in HiC-Pro, and visualised in HiGlass (Kerpedjiev et al. 2018).. Loop calls were made with the HICCUPS module of Juicer Tools (Durand et al. 2016b) with parameters –m 1024 –r 5000 –k KR –f .1 –p 4 –l 7 –d 20000, and the “Ignore_sparsity” flag.

CHAPTER 3. LDB1 REGULATES LIM-HD TF OCCUPANCY AND H3K27AC DEPOSITION IN COOPERATION WITH SSDP1

3.1 – LDB1 AND INDUCTION OF ISL1-FLAG AND LHX3-HA REGULATE MNX1 EXPRESSION IN AN IN-VITRO MODEL OF MOTOR NEURON DIFFERENTIATION

The study of motor neurons (MNs) is inhibited by the intrinsic difficulty in obtaining tissue samples and isolating enriched population of relevant neuronal subtypes (Davis-Dusenbery et al. 2014). The physical structure of the spinal cord presents an inability to biopsy live patients or animals, and while GFP-labeled MNs can be isolated from mice (Erb et al. 2017), many epigenetic assays require much larger numbers of cells than can be feasibly isolated from reasonable number of animals.

Therefore, we developed a drug-inducible epitope-tagged expression system to drive MN identity, with strong similarity to MMC-type neurons (Wichterle et al. 2002). We crossed A2lox ICE-containing mice, who can accept a designed transgene cassette (Iacovino et al. 2011), with LDB1 heterozygotes created by a traditional targeting vector eliminating Exons 2-11 of *Ldb1* **(Fig 3.1)** and isolated blastocysts at Day 3.5.

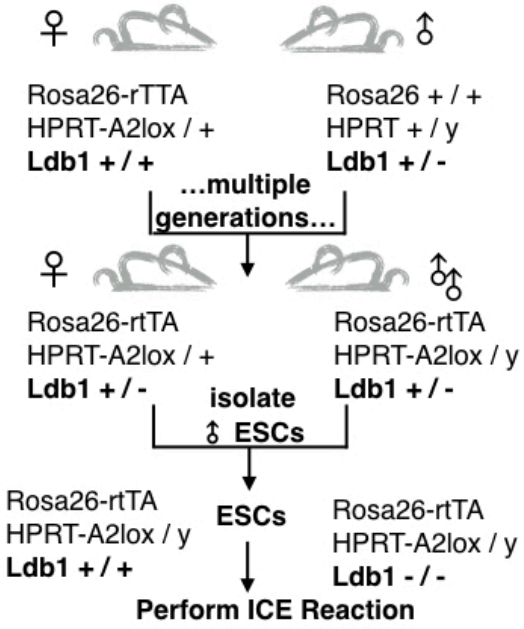


Figure 3.1 – Mouse crossing scheme for LDB1 -/- inducible ESCs
Schematic of mouse crossing and genotypes, resulting in LDB1 +/+ and -/- ESCs with the Inducible Cassette Exchange (ICE) system to accommodate a transgene of choice, as well as the Rosa26-rtTA construct.

After culturing ESCs, doxycycline was added to activate the tetracycline responsive element (TRE) driving Cre expression, and the p2Lox plasmid containing either the Ngn2-Isl1-Lhx3 transgene or the Ngn2-Isl1-Lhx3-LDB1 (mutant) transgene **(Fig. 3.2 A)** was transfected after 24 hrs of doxycycline exposure. These cells were cultured under G418 selection and individual clones were isolated and gelatin-adapted. Cell lines were then genotyped and, subsequently, litter-matched male cells were established to ensure single-copy presence of the introduced transgene at the X-linked *HPRT* locus. Proliferating ESCs were then subjected to a differentiation protocol with two days of embryoid body generation in –LIF, low-serum media, followed by two days of RA and SAG stimulation, followed by two days of induction with doxycycline as well as RA, SAG, and DAPT, to form embryoid bodies (EBs) **(Fig. 3.2 B)**.

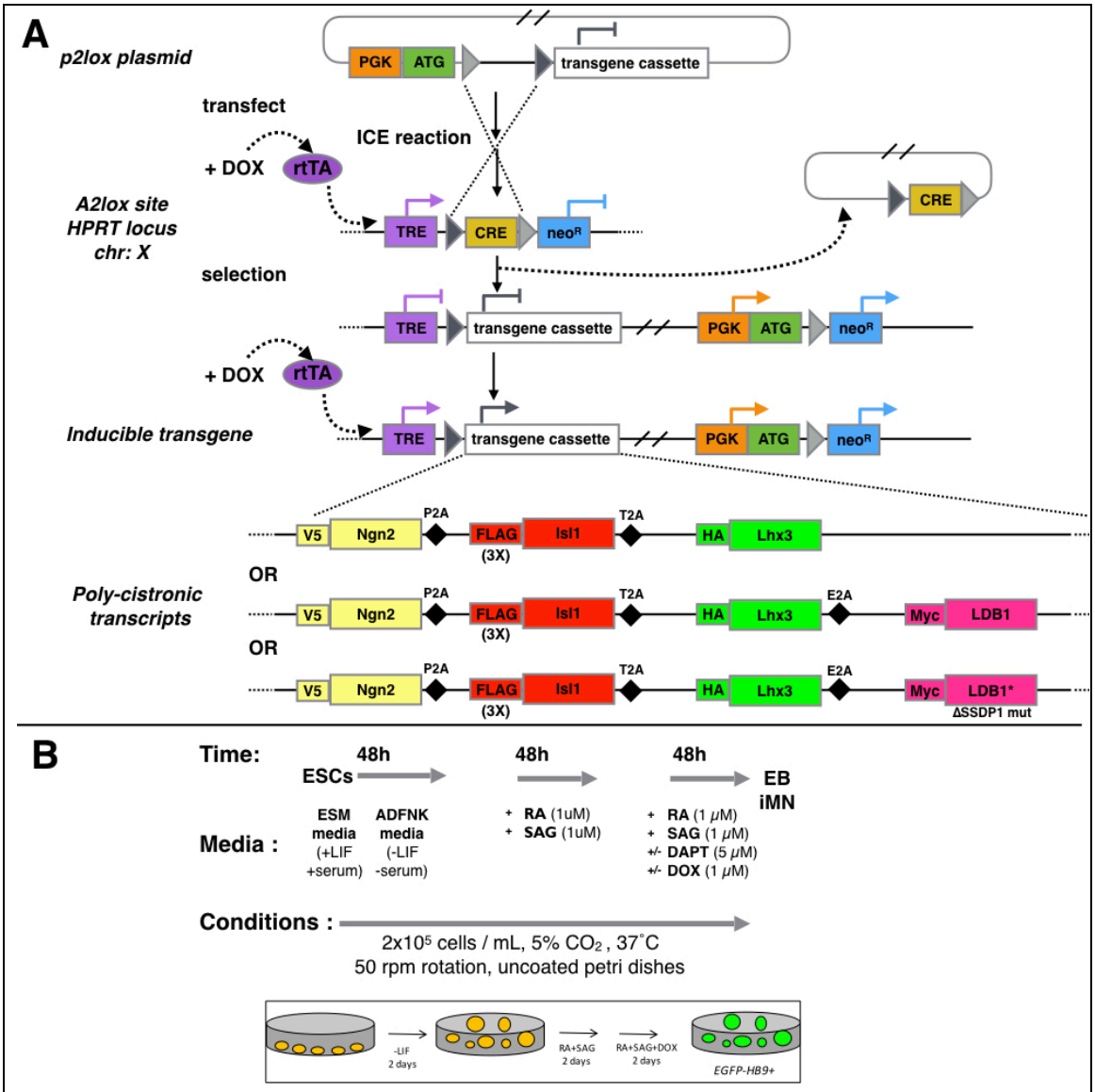


Figure 3.2 – Inducible MN system design and differentiation protocol
Diagram of all ICE system components, starting with the p2lox plasmid, allowing for doxycycline-induced Cre-mediated recombination at the *Hprt* locus. Upon successful recombination and cassette insertion, the Cre is self-excised and neomycin resistance is expressed. Once clones are selected, the second addition of doxycycline will drive expression of a single mRNA from the transgene. The transgene contains either N-terminally epitope-tagged Ngn2, Isl1, and Lhx3, with T2A/P2A/E2A translational termination signals, or those three factors plus a wild-type or SSDP1-binding mutant form of LDB1 with an N-terminal Myc tag. **B)** Schematic of differentiation assay, showing time after removal from ESC conditions, the media composition, and growth conditions. ESM media contains serum while ADFNK contains Knock-Out Serum Replacement (KOSR). Retinoic acid (RA), Sonic hedgehog agonist (SAG), doxycycline (DOX), and the γ-secretase inhibitor DAPT are labeled with their final concentration. Bottom panel shows induction scheme in *EGFP-Mnx1* EBs resulting in green EBs at Day 6. Unless otherwise noted, all cells in this study are Day 6 EBs. These lines were generated in the Macfarlan lab by CJ Miller.

The MN induction transgene (NIL) was then expressed by doxycycline treatment in either a *Ldb1 +/+* or *Ldb1 -/-* background, and in some cases a normal or mutant Myc-tagged version of the *Ldb1* gene was expressed on the transgene in addition. SSDP proteins can regulate the abundance of LIM-binding proteins (Xu et al. 2007b) through RLIM-mediated ubiquitylation or recruitment of HDACs to LIM-mediated genes (Bach et al. 1999), and are required for LIM-mediated aspects of chromatin mark deposition in mouse and chicken (Lee et al. 2016). We created a transgene including a mutation in the SSDP-binding LCCD domain of LDB1 (**Figure 1.4**) to convert five amino acid residues to alanine, which ablates SSDP1-LDB1 interaction (Chen et al. 2002).. We then expressed the NIL construct with either a WT or SSDP1-binding mutant (ΔS) of LDB1 in *Ldb1 -/-* Day 6 EBs, or the accompanying *LDB1* +/+ EBs. We then performed a variety of assays to determine the epigenetic profile, protein association, RNA expression, and 3D genome organization of these cells. **(Figure 3.3)**

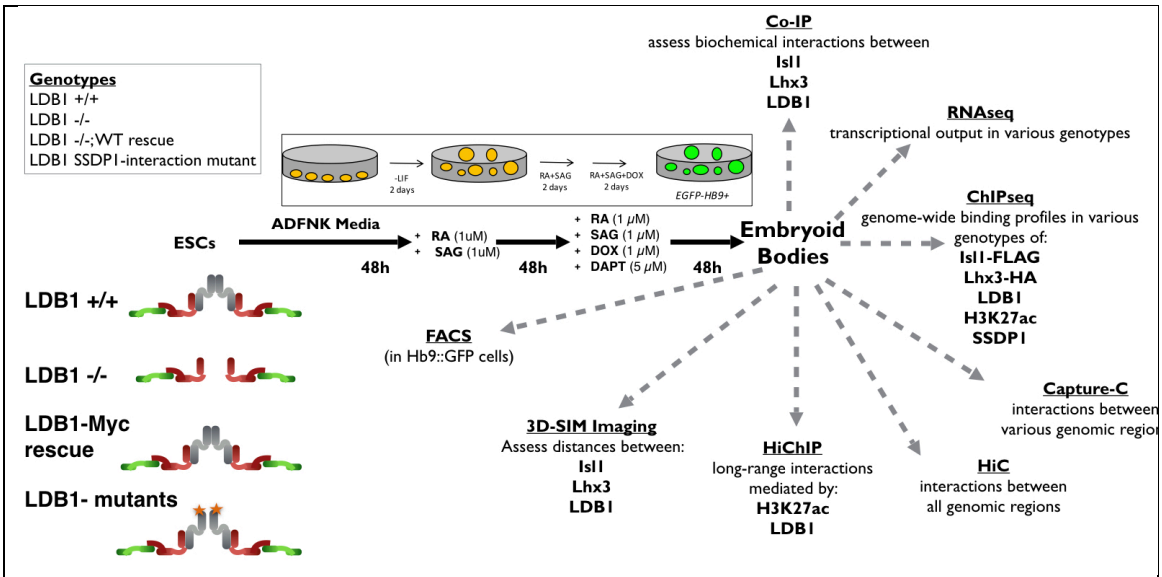


Figure 3.3 – Experimental setup to determine effects of LDB1 knockout
Diagrammatic overview of cell types, differentiation strategy, and technical approaches to understand LDB1 function.

Upon doxycycline induction of Isl1-FLAG, Lhx3-HA, and Ngn2-V5, and in some cases LDB1-Myc, we observe, at the Day 6 EB stage, significant upregulation of the transgene-encoded proteins FLAG-Isl1 and HA-Lhx3, but not in the absence of doxycycline, as assayed by Western blot **(Figure 3.4 A),** and by immunofluorescence of whole EBs **(Figure 3.4 B)**. Importantly, we observe the Mnx1 protein in the protein lysate only in the presence of both doxycycline and LDB1 **(Figure 3.4 A, B).**

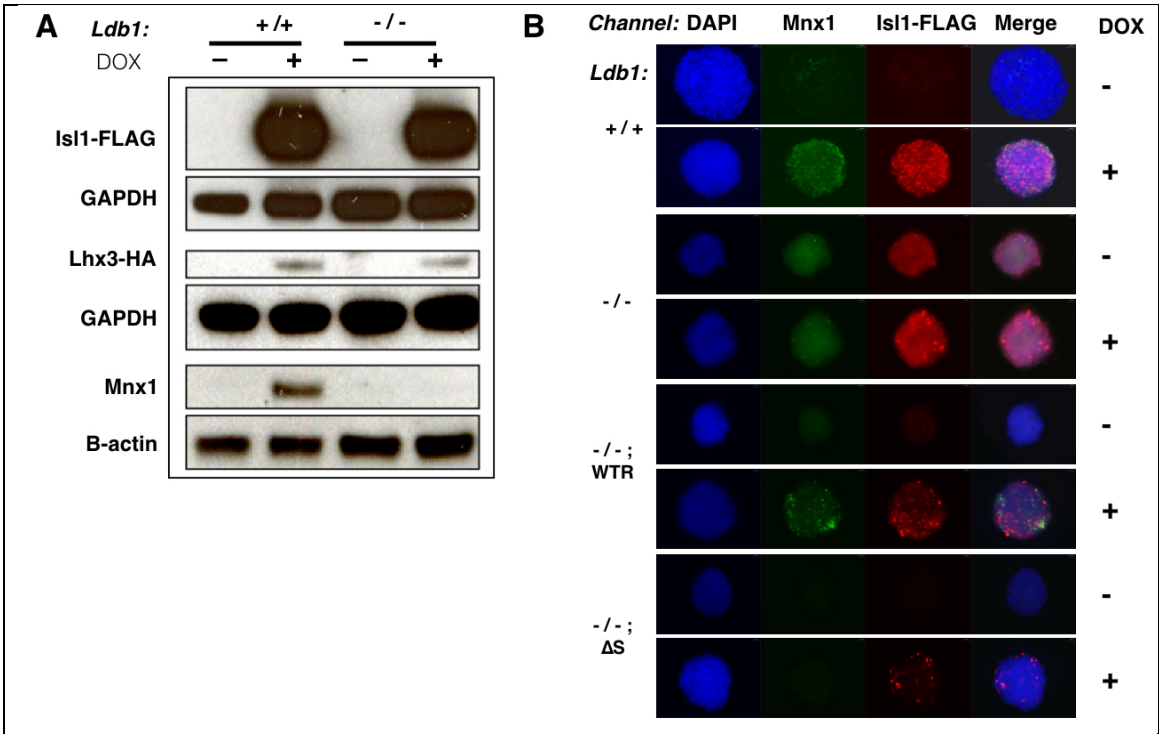


Figure 3.4 – Mnx1 is induced at the protein level by Isl1-FLAG and Lhx3-HA in a LDB1-dependent manner
A) Western blots for the antibody indicated on the left, in either LDB1 +/+ or -/- genotypes, with or without doxycycline. GAPDH or B-actin included as loading controls. Image courtesy of CJ Miller. **B)** Immunofluorescence on Day 6 EBs with or without doxycycline, as indicated on the right, stained for the antibody or DNA stain as indicated in each channel, and in LDB1 wildtype, knockout, wildtype rescue, or SSDP1-interacting mutant genotypes as indicated on the left.

In addition to the LDB1 and doxycycline-dependence of Mnx1 expression observed by immunofluorescence, we also see that Mnx1 expression can be rescued by induction of the wildtype LDB1-Myc transgene, but not with the accompanying ΔS mutant which cannot bind SSDP1 **(Figure 3.5**).

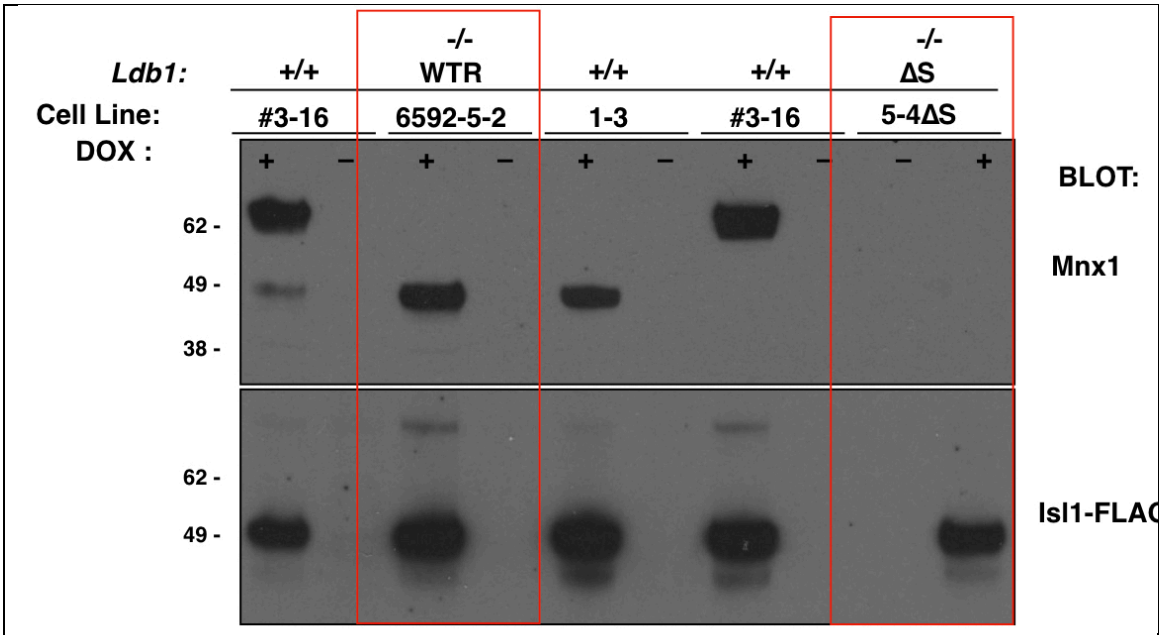


Figure 3.5 - Mnx1 expression depends on the LDB1-SSDP1 interaction
Western blot for Mnx1 or Isl1-FLAG expression, as indicated on the right, with lysates from various cell lines with or without doxycycline, indicated on the top. Cell line #3-16 has a GFP-tagged endogenous Mnx1, resulting in the larger band size. The red boxes indicate comparison between a DOX-induced wildtype isoform of LDB1 (WTR), or the SSDP1-interacting mutant LDB1, showing that Mnx1 expression is ablated in ΔS mutant but not in the wildtype rescue.

This general expression profile, specific to embryoid bodies and in doxycycline-induced, LDB1-, and LDB1-SSDP1-dependent manners, is observed in RNA-seq data, thus verifying expression of the induced transcripts and their main target (**Figure 3.6**).

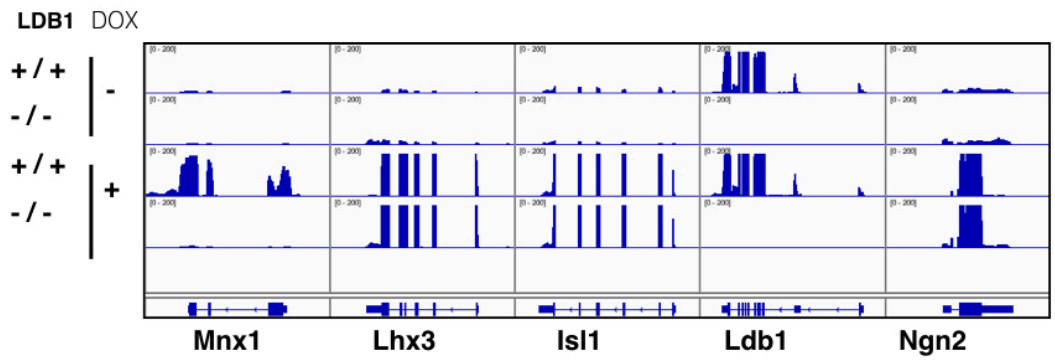


Figure 3.6 – LDB1- and doxycycline- dependent induction of Mnx1 and target LIM-HD TFs.
Browser screenshot of RNAseq data from either LDB1 wild type or knockout genotypes, as indicated, with or without doxycycline, as indicated. Scales are identical throughout.

Data regarding further exploration of gene expression changes upon induction MN differentiation and LDB1 deletion can be found in Chapter 4. As embryoid bodies are a relatively artificial system, the lab also generated ESCs in an *Ldb1* +/+; A2Lox.cre background that had a GFP reporter knocked into the endogenous Mnx1 locus. These EBs were then analyzed for GFP signal by FACS at Day 6 of differentiation. They show a reproducible induction of Mnx1::GFP expression in 50-70% of cells (data not shown). While we did not construct a Mnx1::GFP reporter line in *Ldb1* -/- cells, we can establish that doxycycline induction provides a reasonable conversion to a MN-like cell identity. Taken together, this evidence suggests LDB1 is a required positive regulator of *Mnx1* expression in a TF-induced model of mouse MN development.

3.2 – LDB1 DELETION OR MUTATION OF ITS SSDP1-INTERACTING DOMAIN REDUCES LIM-HD TF OCCUPANCY AT KEY MOTOR NEURON GENES AND GENOME-WIDE

While LDB1 is known to form hexameric complexes with Isl1 and Lhx3, specifically at the ECR regions ~10 kb upstream of the *Mnx1* locus, its genome-wide positioning, and the extent of its co-occupancy of Isl1 and Lhx3 binding sites in MNs is unknown. Studies of LDB1 binding in other developmental systems indicate global binding of LDB1 to enhancer elements along with a cohort of cell-type specific TFs and co-activators (Soler et al. 2010; Li et al. 2013), but LDB1 binding in a MN differentiation is unknown. The LIM-HD TFs Isl1 and Lhx3 have been previously mapped in a different model of MN differentiation (Mazzoni et al. 2013; Rhee et al. 2016), but neither their relationship to LDB1 binding, nor the effect of LDB1 knockout on Isl1 and Lhx3 binding are known. Hypothesizing that ChIP-seq of LDB1 would reveal the cell-type specific TF binding sites of Isl1-FLAG and Lhx3-HA, we performed ChIP-seq for LDB1 in *Ldb1* +/+ Day 6 EBs. We also hypothesised that Isl1-FLAG and Lhx3-HA binding would be altered by LDB1 deletion, so we performed ChIP-seq for Isl1-FLAG and Lhx3-HA in both *Ldb1* +/+ and *Ldb1 -/-* Day 6 EBs, as well as Day 6 EBs containing the wildtype of SSDP1-mutant forms of LDB1 on a *Ldb1 -/-* background. We also performed ChIP-seq against H3K27ac and SSDP1 in each of the respective *Ldb1* genotypes, as well as ATAC-seq, which maps accessible chromatin (Buenrostro et al. 2013). ChIP-seq of Day 6 EBs confirms previous studies showing Isl1/Lhx3 co-occupancy at the nearby distal evolutionarily conserved regions (ECRs) of the Mnx1 gene, in agreement with previously published description of Isl1 and Lhx3 binding in similar yet distinct models of MNs (Mazzoni et al. 2013; Erb et al. 2017;

Lee et al. 2012; Rhee et al. 2016) The LDB1 ChIP confirms a high level of co-occupancy at Isl1-FLAG and Lhx3-HA peaks, notably in exactly alignment with FLAG-Isl1 and HA-Lhx3 peaks at the *Mnx1* ECRs, These are also marked by SSDP1 peaks, high ATACseq signal, and high H3K27ac signal. Signals for FLAG-Isl1, HA-Lhx3, H3K27ac, and SSDP1 binding at the *Mnx1* ECRs were substantially dependent on LDB1 and on its interaction with SSDP1, being reduced at those cites in either *Ldb1 -/-* or SSDP1-interactor mutant cells. ATAC-seq signal indicating open chromatin was reduced, but not eliminated, in the absence of *Ldb1* and its SSDP1 interaction **(Figure 3.7**).

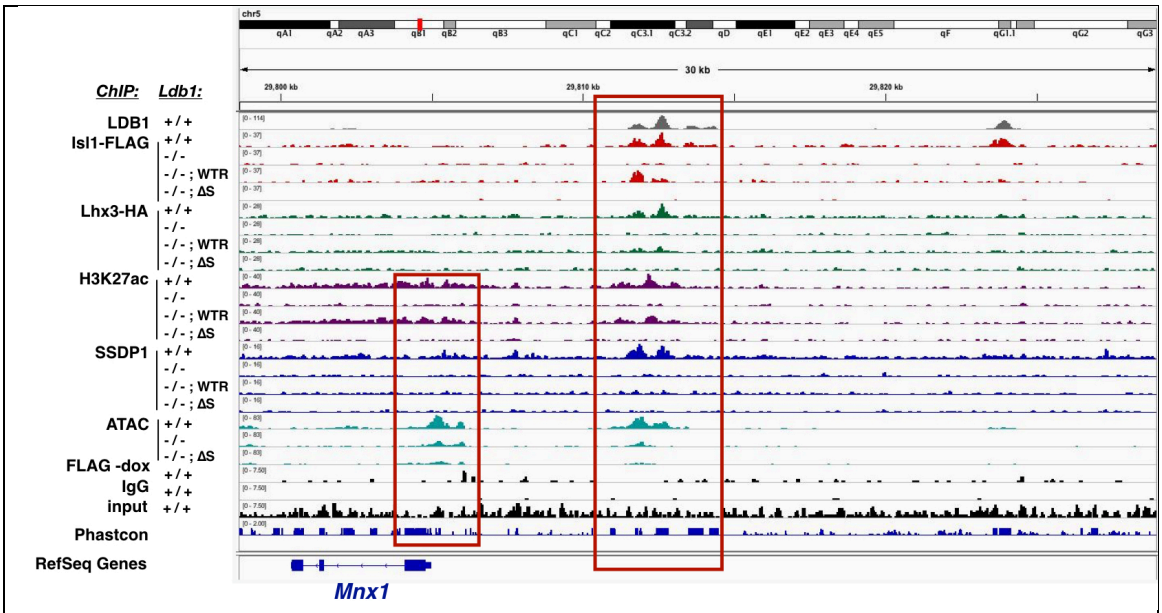


Figure 3.7 – LDB1 regulates occupancy of LIM-HD TFs, SSDP1, H3K27ac, and open chromatin at the *Mnx1* locus
ChIP-seq and ATAC-seq tracks of a 30-kb region surrounding the *Mnx1* locus and its upstream enhancer elements on chr:5, in *Ldb1* +/+ , *Ldb1* -/-, or rescue and mutant constructs in Day 6 EBs. LDB1 signal is dark grey, Isl1-FLAG ChIP signal is red, Lhx3-HA signal is green, H3K27ac is purple, SSDP1 is blue, ATAC-seq is teal, and input, IgG, and FLAG control (anti-FLAG ChIP from cells with no doxycycline induction) signals are in black, with Phastcons 30-way mammalian conservation and Refseq genes in blue at the bottom. Scales are noted and consistent within each IP. Areas of LDB1-mediated binding at the upstream element (ECR2) and the *Mnx1* promoter are highlighted in red boxes.

We observe similar characteristic patterns surrounding LDB1 peaks at several other critical MN loci. *Nrp1*, a neuronal guidance peptide-encoding gene, exhibits similar patterns at its promoter and at what may be a distal regulatory element near the end of the gene **(Figure 3.8)**.

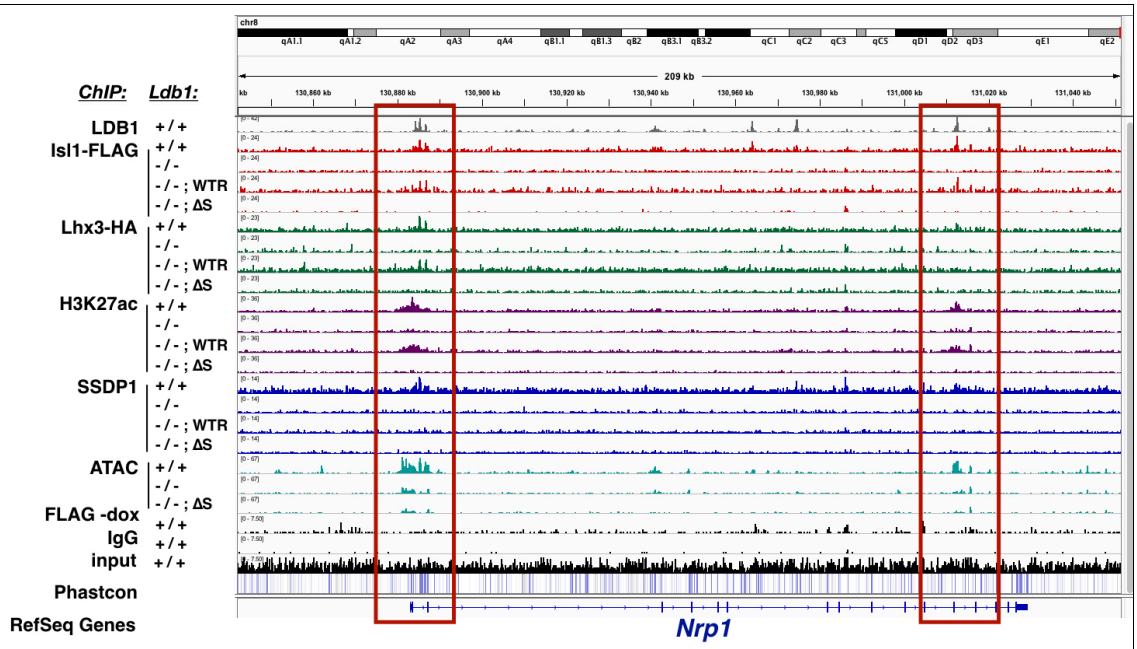


Figure 3.8 – LDB1 regulates occupancy of LIM-HD TFs, SSDP1, H3K27ac, and open chromatin at the *Nrp1* locus
ChIP-seq tracks of a 200-kb region surrounding the *Nrp1* locus on chr:8, showing Isl1-FLAG, Lhx3-HA, and LDB1 ChIP signal in *Ldb1* +/+ or *Ldb1* -/- Day 6 EBs. LDB1 signal is dark grey, Isl1-FLAG ChIP signal is red, Lhx3-HA signal is green, H3K27ac is purple, SSDP1 is blue, ATAC-seq is teal, and input, IgG, and FLAG control (anti-FLAG ChIP from cells with no doxycycline induction) signal is in black, with Phastcons 30-way mammalian conservation in blue at the bottom. Scales are noted and consistent within each IP. Areas of LDB1-mediated binding at the promoter and another site in a distal exon are highlighted in red.

This is contrasted at the gene *Slit3*, which encodes neural guidance molecules. There is a single intronic LDB1- and SSDP1-dependent peak (**Figure 3.9)**.

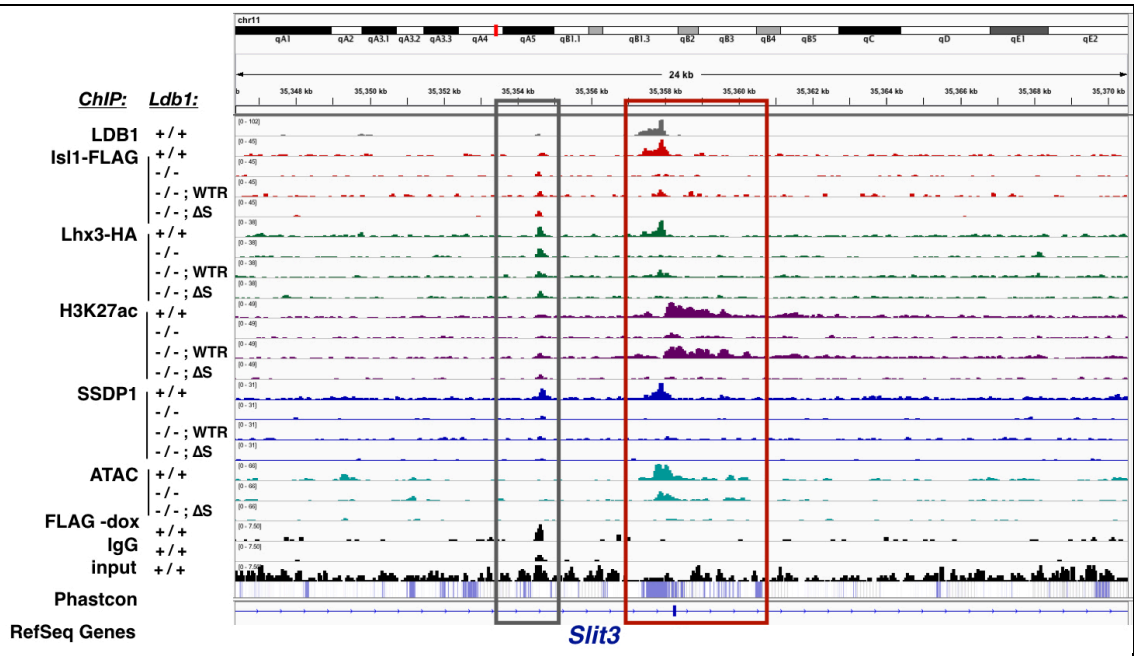


Figure 3.9 – LDB1 regulates occupancy of LIM-HD TFs, SSDP1, H3K27ac, and open chromatin at the *Slit3* locus
ChIP-seq tracks of a 24-kb region within the *Slit3* locus on chr:11, showing Isl1-FLAG, Lhx3-HA, and LDB1 ChIP signal in *Ldb1* +/+ or *Ldb1* -/- Day 6 EBs. LDB1 signal is dark grey, Isl1-FLAG ChIP signal is red, Lhx3-HA signal is green, H3K27ac is purple, SSDP1 is blue, ATAC-seq is teal, and input, IgG, and FLAG control (anti-FLAG ChIP from cells with no doxycycline induction) signal is in black, with Phastcons 30-way mammalian conservation in blue at the bottom. Scales are noted and consistent within each IP. Areas of LDB1-mediated binding at an intergenic region are highlighted in red, while non-specific background signal is highlighted in grey.

A similar pattern to that observed at *Mnx1* and *Nrp1* is observed at the *Arhgap36* locus promoter. It is also faintly observed at what may be its nearby regulatory elements **(Fig. 3.10**). These peaks are distinct from replicable background peaks, such as the intergenic region downstream of *Arhgap36*, which encodes a MN-induced GTPase(Nam et al. 2019).

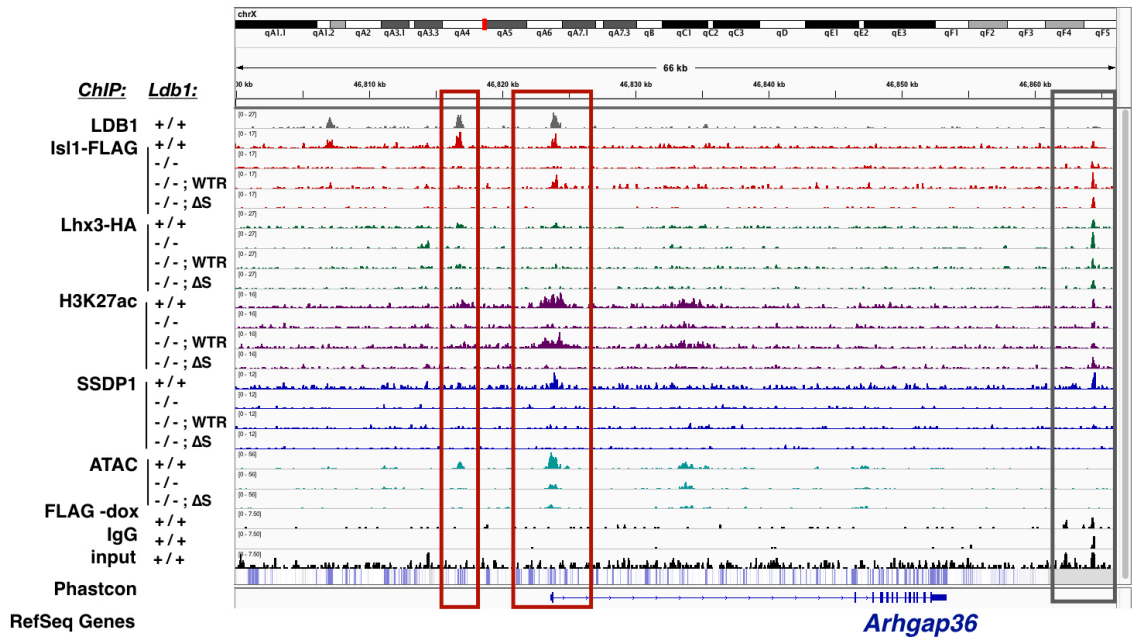


Figure 3.10 – LDB1 regulates occupancy of LIM-HD TFs, SSDP1, H3K27ac, and open chromatin at the Arhgap36 locus
ChIP-seq tracks of a 66-kb region surrounding the *Arhgap36* locus on chr:X showing Isl1-FLAG, Lhx3-HA, and LDB1 ChIP signal in *Ldb1* +/+ or *Ldb1* -/- Day 6 EBs. LDB1 signal is dark grey, Isl1-FLAG ChIP signal is red, Lhx3-HA signal is green, H3K27ac is purple, SSDP1 is blue, ATAC-seq is teal, and input, IgG, and FLAG control (anti-FLAG ChIP from cells with no doxycycline induction) signal is in black, with Phastcons 30-way mammalian conservation in blue at the bottom. Scales are noted and consistent within each IP. Areas of LDB1-mediated binding at the promoter and an upstream region are highlighted in red, while non-specific background signal is highlighted in grey.

Thus, at this subset of specific MN related genes, we can observe a clear pattern of LDB1, Isl1-FLAG, Lhx3-HA, and SSDP1 binding that is associated with open chromatin and H3K27ac deposition. We also observe that binding at those regions is dependent on both LDB1 and its interaction with SSDP1.

When studied genome-wide, nearly all top-ranking LDB1, Isl1-FLAG, and Lhx3-HA peaks show strong binding with the other two factors within a window. Open chromatin, as profiled by ATAC-seq, and SSDP1 ChIP-seq signal also selectively mark the top LDB1 peaks **(Figure 3.11)**. Notably, when peaks are sorted based on LDB1 P-value **(Figure 3.11,** top panel), Isl1-FLAG, Lhx3-HA,

ATAC, and SSDP1 signals are clearly enriched at the top LDB1 peaks, but enrichment of H3K27ac is not as evident. However, when sorted by signal strength, while the previously mentioned LIM-HD (Isl1-LAG and Lhx3-HA), accessibility (ATAC-seq), and cofactor (SSDP1) marks are still clustered with the top LDB1 peaks, it becomes clear that there is a subset of H3K27ac peaks, distributed around the centre of a subset of perhaps 200 LDB1 peaks, that are offset from the peak centre as marked by MNC complex components but are markedly enriched upstream and downstream of the centres **(Figure 3.11**, bottom panel).

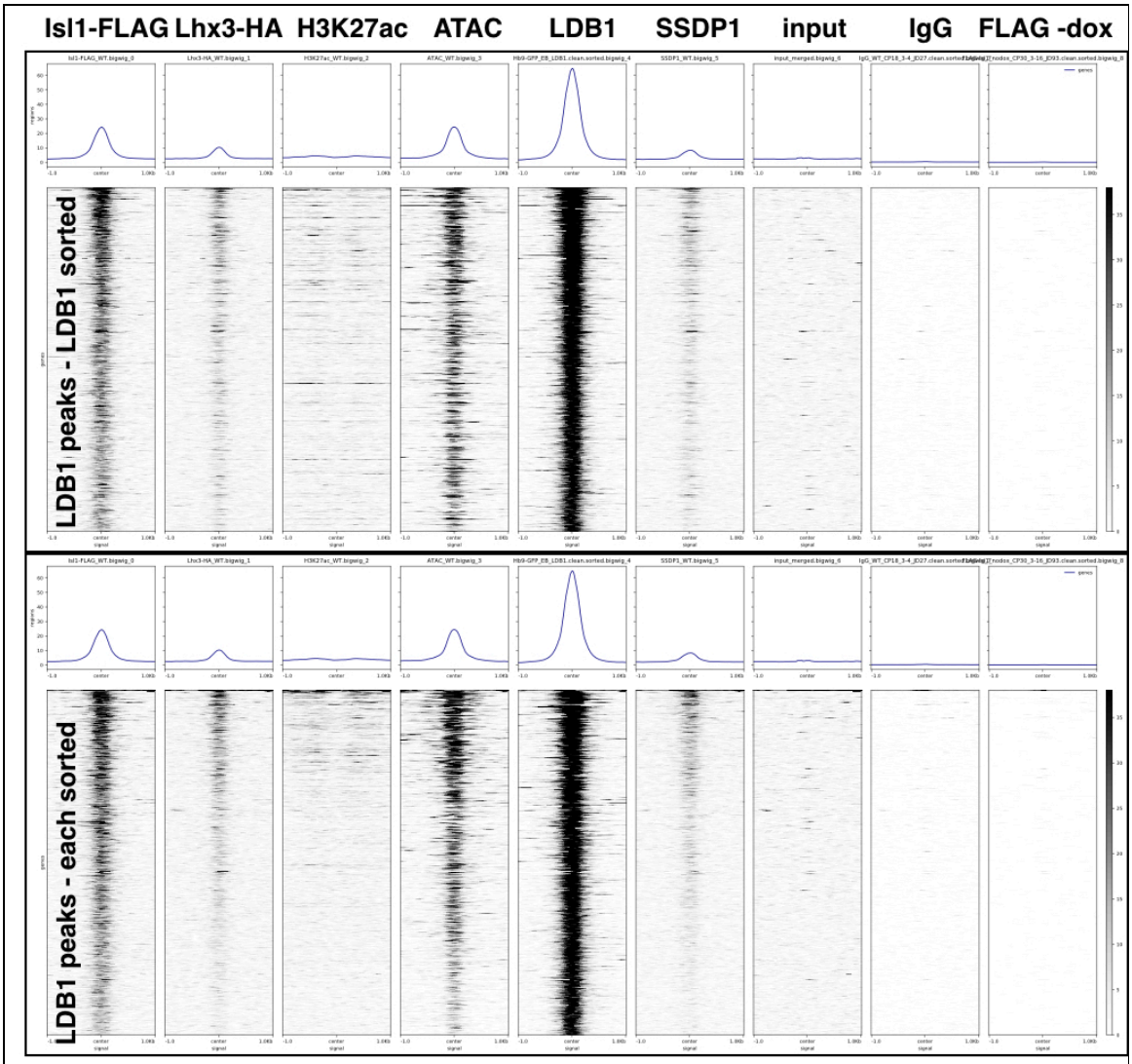


Figure 3.11 – LDB1 associates with LIM-HD TFs, open chromatin, H3K27ac, and SSDP1

Heatmaps of normalised ChIP-seq coverage of various factors from Day 6 doxycycline-induced EBs in *Ldb1 +/+* cells. Each of 1000 rows is a LDB1 peak sorted by P value (top panel), or sorted by signal strength of the respective assay plotted (bottom panel). Note the close correlation between sorting methods. The antibody or method is indicated above each column, and represents normalised coverage from two biological replicates from ChIP for Isl1-FLAG, Lhx-HA, H3K27ac, SSDP1, LDB1, IgG, FLAG with no induction (-dox), or ATAC-seq. Signal intensity is plotted across a 2 kb region centreed at the middle of each peak region.

When these marks are assayed in the *Ldb1 -/-* Day 6 EBs at peaks marked by wildtype LDB1, the signals of LIM-HD TFs are substantially reduced, the cofactor SSDP1 is entirely absent, and accessibility is decreased but still

substantial. H3K27ac binding is still visible at some LDB1 peaks but has lost the characteristic phased pattern in relation to the peak centre **(Figure 3.12)**.

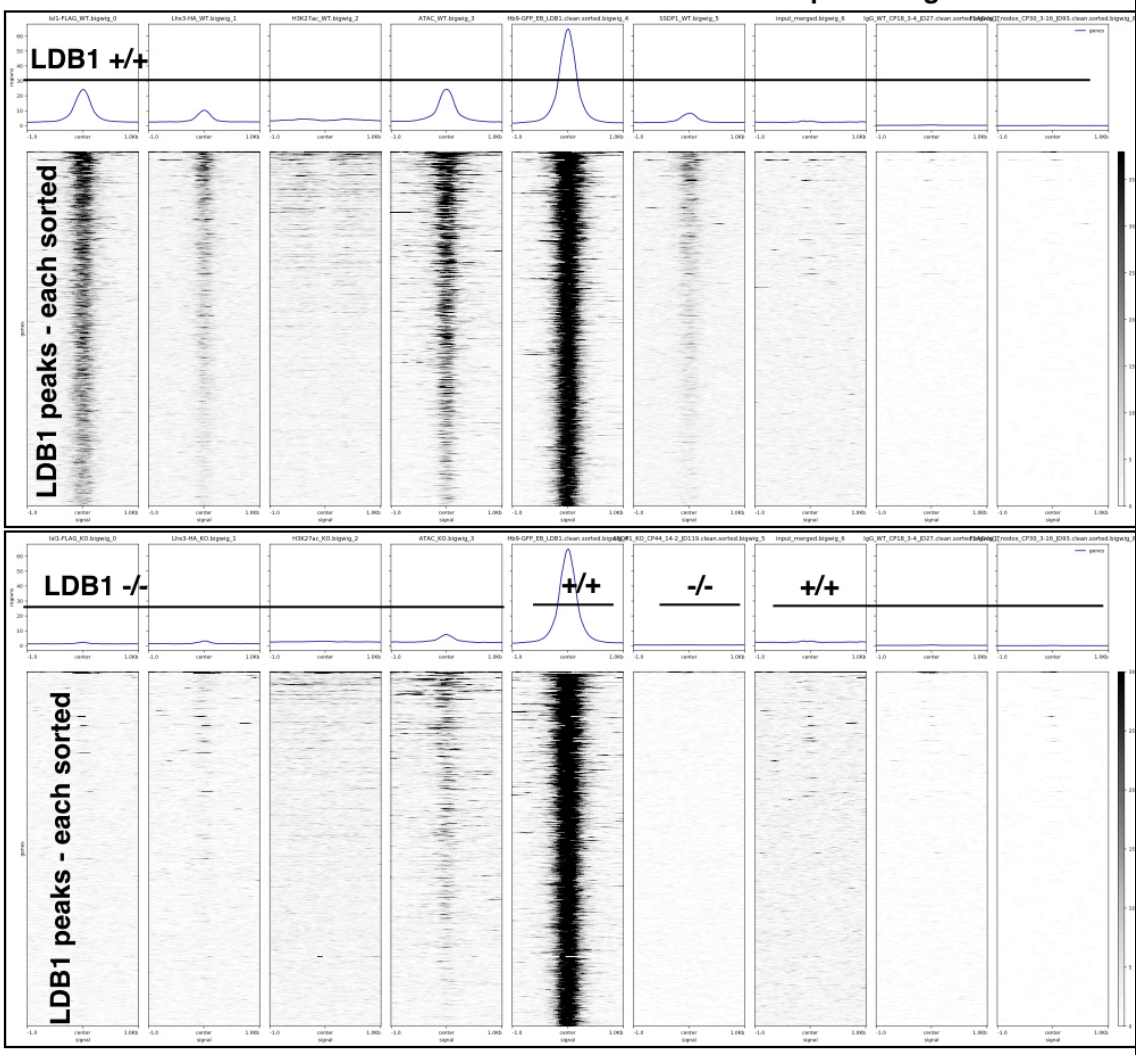


Figure 3.12 – MNC components, H3K27ac, and open chromatin are reduced at LDB1 peaks in the absence of LDB1
Heatmaps of normalised ChIP-seq coverage of various factors from Day 6 doxycycline-induced EBs in either *Ldb1 +/+ (*top row) or *Ldb1 -/- (*bottom row) cells. Each of 1000 rows is a *Ldb1 +/+* sorted by P value. The antibody or method is indicated above each column, and represents normalised coverage from two biological replicates from ChIP for Isl1-FLAG, Lhx-HA, H3K27ac, SSDP1, LDB1, IgG, FLAG with no induction (-dox), or ATAC-seq. Signal intensity is plotted across a 2 kb region centred at the middle of each peak region, and then sorted by signal intensity. In the bottom column the LDB1 ChIP-seq, input, IgG and FLAG-dox signals are from *Ldb1 +/+* cells for comparison, while the other ChIP and ATAC signals are from *Ldb1 -/- cells.*

Importantly, the LDB1 motif generated from this ChIP-seq data is similar to two spaced LIM-HD motifs, as determined through many studies of LIM-HD binding and by the motif results of both Isl1-FLAG and Lhx3-HA. These motifs are subsequently eliminated from LIM-HD TF ChIP-seq signal in *Ldb1 -/-* Day 6 EBs, leaving only poly-A motifs **(Figure 3.13)**. The poly-A nature of the knockout motif calls is most likely not specific re-targeting of LIM-HD TFs to specific polynucleotide repeats in the genome, but is rather an artifact of ChIP-seq library construction with such low-complexity libraries (Zhao and Zheng 2018). These poly-A motifs may also possibly reflect a general tendency of LIM-HD TFs to occupy AT-rich genomic regions, likely due to their promiscuous low-affinity binding and very high number of potential target sites when not being selectively targeted in multi-protein complexes.

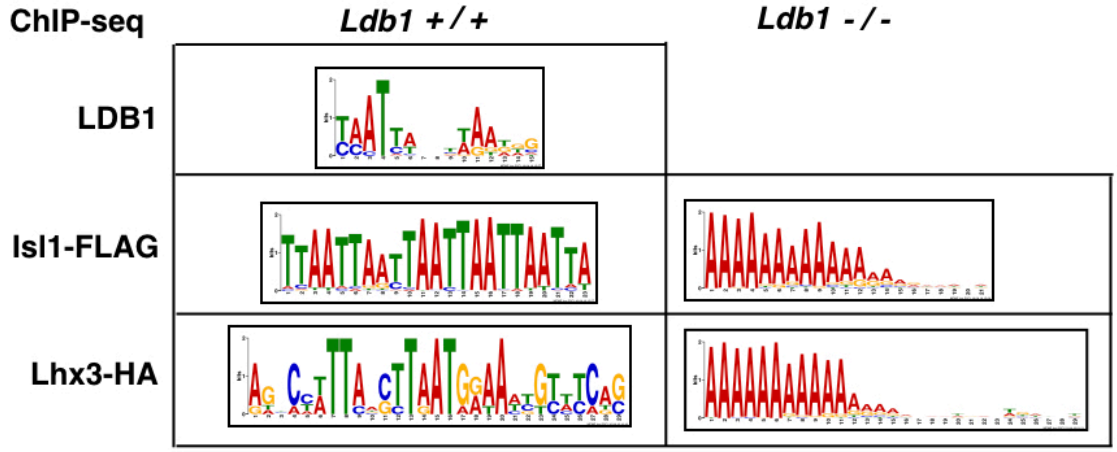


Figure 3.13 – LDB1 binding motifs are similar to pairs of LIM-HD motifs, which are ablated in the absence of LDB1
MEME results showing top motif occurrences in the regions surrounding 200 bp of the centre of the top 1000 ChIP-seq peaks from the antibody listed on the left. The LDB1 genotype is indicated at the top. Poly-A repeat motifs may be artifactual but underscore the lack of a strong consensus of LIM-HD binding in *Ldb1* knockout cells.

The *Ldb1 -/-* Day 6 EBs also produced consistently low ChIP yields of the LIM-HD TFs. This indicates that either cross-linking efficiency or overall DNA-binding of these TFs is reduced in the absence of LDB1. That is consistent with EMSA results showing affinity of LIM-HD TFs for more complex binding sites is only increased when multiple binding domains are present, and that single LIM-HD domains display weak binding, even to their consensus target sites (Robertson et al. 2018). ChIP-seq for Ngn5 or Ngn2-V5 yielded poor results, likely due to the more transient nature of Ngn2 binding at “early motor neuron” promoter-proximal E-box elements (Velasco et al. 2017).

One theory of LDB1 function asserts that it directs Isl1 and Lhx3 to paired binding sites called hexamer response elements (HxRE), and that this orientation is key for complex function at MN genes (Lee et al. 2012). However, we observe no genome-wide pairing of “HxRE” motifs, as might be expected if this exact motif arrangement was critical to specify MNC binding, and that the canonical Isl1-Lhx3 sequence (DNAt) (Robertson et al. 2018) is only found at 70 sites genome-wide, of which only a few are LDB1-bound when compared to LDB1 ChIP-seq data (**Figure 3.14**).

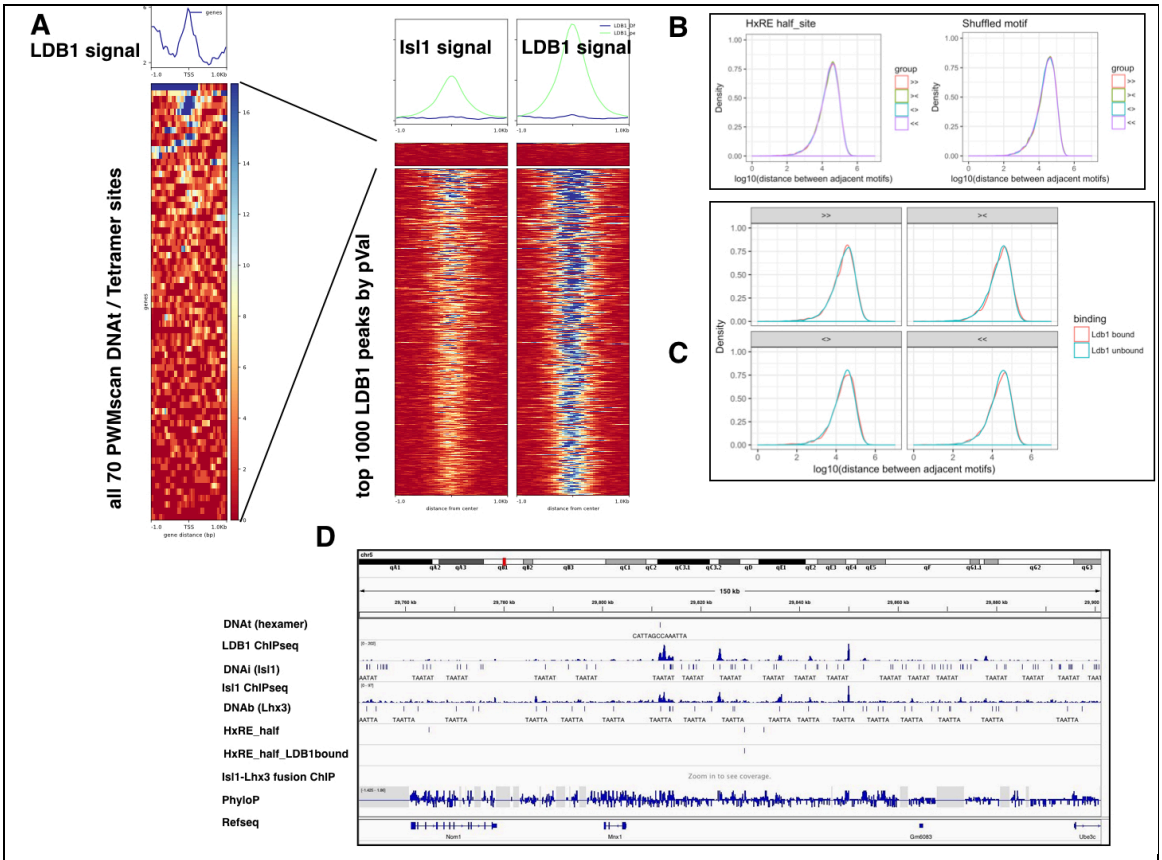


Figure 3.14 - Proposed MNC target sites are unpaired in the genome, and not indicative of LDB1 binding

A) All FIMO-determined hexameter sites from Robertson 2018 plotted against our LDB1 ChIP-seq from Day 6 EBs, with Isl1-FLAG plotting as comparison. **B)** HxRE half sites (Isl1+Lhx3) determined from Lee 2013, compared to a shuffled motif and plotted with respect to their orientation (>> = in same orientation, >< = facing, etc.) across the genome. **C)** LDB1-bound or unbound HxRE sites plotted by orientation. Analysis by M. Sun. For **B)** and **C),** plots are of density (out of 1.00 arbitrary units) in the Y axis against the log10 distance between adjacent motifs (in base pairs) on the X axis. While all four orientations have a separate colour, they are overlaid closely **D)** Browser shot at the *Mnx1* upstream regulatory region of this study's LDB1 ChIP-seq data, Isl1 ChIP-seq (Mazzoni 2013), and Isl1-Lhx3 fusion ChIP-seq (Lee 2013), compared to experimentally determined binding sites (Lee 2012, Robertson 2018), showing different variations of LIM-HD motifs and their relation to their respective binding sites. Philpot conservation and Refseq genes are shown below

Further, the extent of Isl1 and Lhx3 potential binding sites make any prediction as to their relevant targets near impossible because of their deeply non-specific preference for AT-rich binding motifs.

Taken together, these results show that LDB1 regulates Isl1-FLAG and Lhx3-HA binding at a genome-wide level, that all three factors co-occupy a large number of peaks in induced MNs, and that in the absence of *Ldb1* LIM-HD TFs are not able to bind to their tissue-specific regulatory targets. Further, SSDP1 marks the strongest LDB1 peaks and is the most affected by the *Ldb1* deletion. Chromatin accessibility is reduced overall in the absence of LDB1 but remains present at low levels. H3K27ac marks are weakly present at LDB1 peaks regardless of *Ldb1* knockout, but cluster around a subset of LDB1 peaks only in the presence of LDB1. We hypothesize that LDB1 is directing, with the LIM-HD TFs and critically SSDP1 as part of the MNC, not only binding to these sites but also H3K27ac-mediated remodeling of a critical subset of the genome. This subset of regions likely comprises, or are regulatory elements for, the critical developmentally regulated subset of MN-specific genes.

3.3 – LIM-HD TF AND SSDP1 OCCUPANCY, AND CHROMATIN ACCESSIBILITY AT LDB1 PEAKS DEPENDS ON LDB1 AND ITS INTERACTION WITH SSDP1

Cofactors such as SSDP1 are important in stabilizing binding partners such as LDB1 from degradation, and LDB1 is critical to stabilizing LIM-HD TF complex formation and targeting. Therefore, we compared ChIP-seq data from Isl1-FLAG, Lhx3-HA, H3K27ac, and SSDP1 at LDB1 wildtype peaks in Day 6 EBs that were either *Ldb1* +/+, *Ldb1* -/-, or that were *Ldb1* -/- but contained either wildtype (WTR) or SSDP1-interacting mutant forms of LDB1 expressed as part of the same transgene cassette from which the LIM-HD TFs are induced in response to doxycycline. ChIP signals from Isl1-FLAG, Lhx3-HA, SSDP1, and H3K27ac (at a select subset of LDB1 peaks) are all reduced in both a LDB1- and LDB1-SSDP1 interaction-dependent manner **(Figure 3.15).**

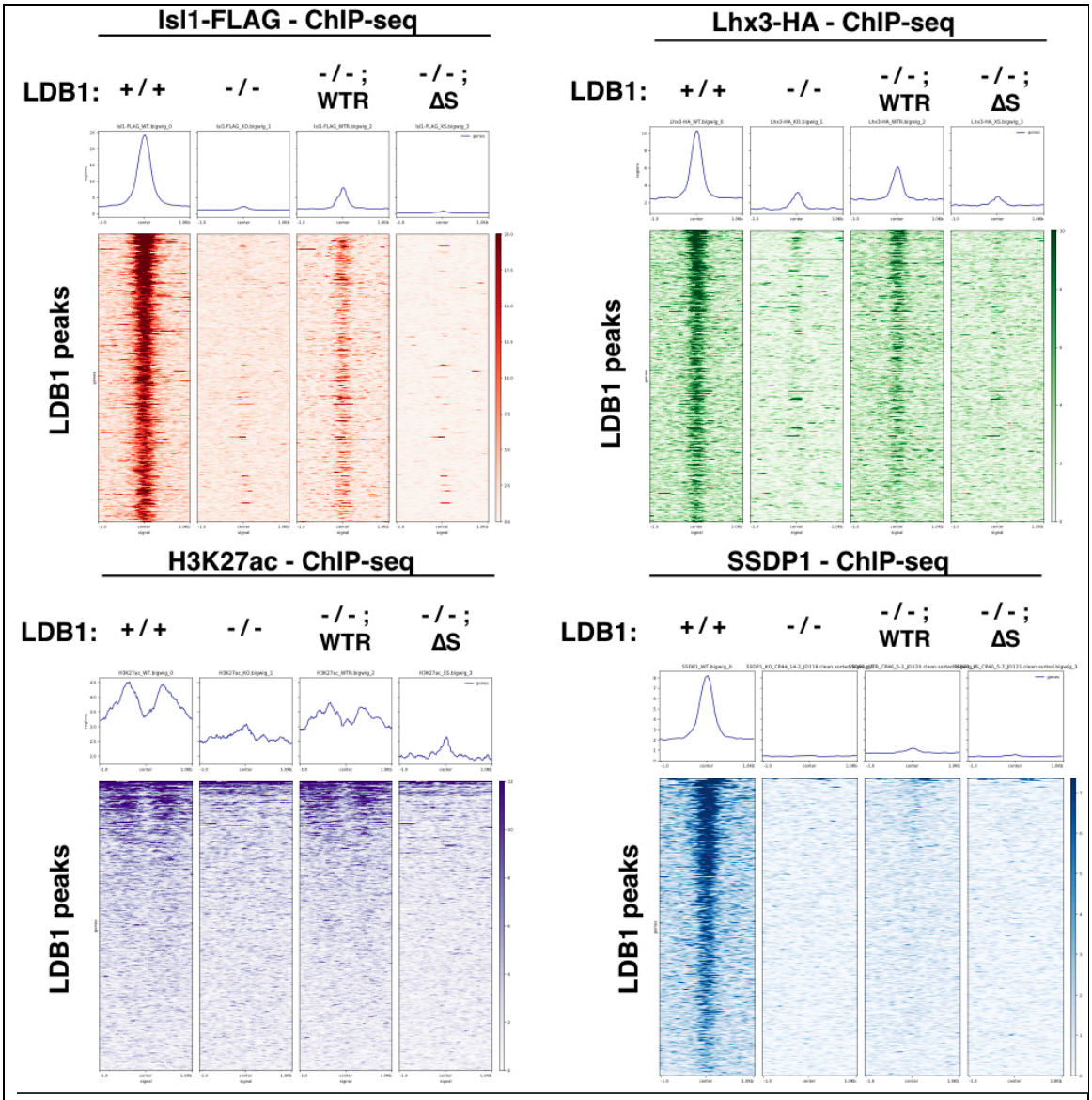


Figure 3.15 – LDB1 and its SSDP1 interaction regulates occupancy of Isl1-FLAG, Lhx3-HA, H3K27ac, and SSDP1
Heatmaps of normalised ChIP-seq coverage of Isl1-FLAG, Lhx3-HA, H3K27ac, and SSDP1 from DAY 6 doxycycline-induced EBs in either *Ldb1 +/+ , Ldb1 -/-* or respective rescue and mutant cell lines in each column. Each of 1000 rows is a *Ldb1 +/+* peak sorted by P value. The is indicated above each set of columns, and represents normalised coverage from two biological replicates from ChIP for Isl1-FLAG, Lhx3-HA, H3K27ac, or SSDP1. Signal intensity is plotted across a 2 kb region centred at the middle of each peak region and sorted by signal intensity.

Given the substantial overlap between LDB1 and Isl1-FLAG peaks, we also asked if the same pattern of LDB1 and SSDP1-dependency was observed at

sites of Isl1-FLAG binding in LDB1 wildtype cells. We observe a similar pattern with Isl1-FLAG ChIP-seq signal, although the differential nature of the H3K27ac peak association is reduced **(Figure 3.16)**.

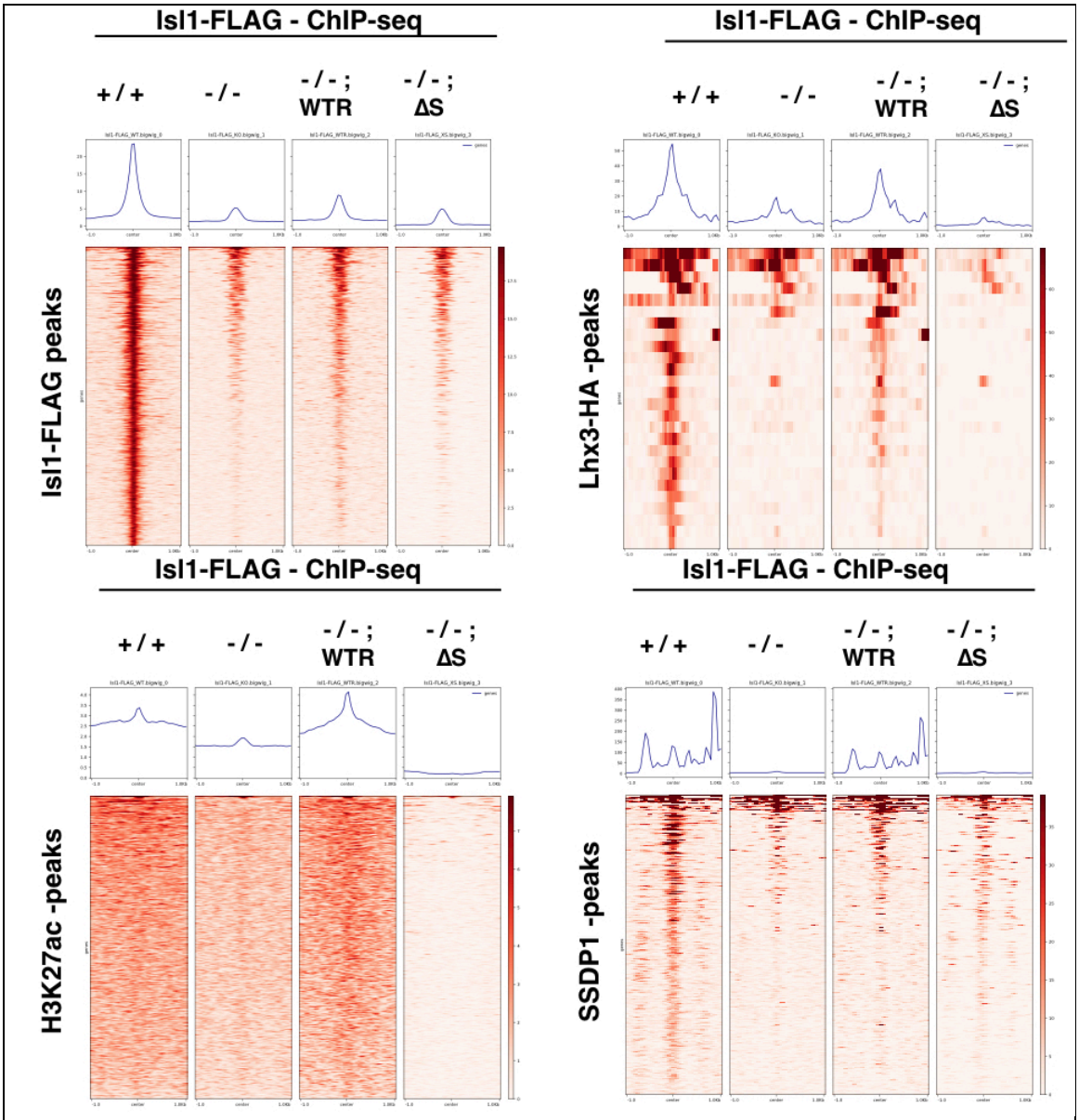


Figure 3.16 – Isl1-FLAG co-occupancy with Lhx3-HA, H3K27ac, and SSDP1 is dependent on LDB1 and its SSDP1-interacting domain
Heatmaps of normalised Isl1-FLAG ChIP-seq coverage from Day 6 EBs in different *Ldb1* genotypes, indicated above each panel, plotted against peaks called from Isl1-FLAG itself (top left), Lhx-HA (top right), H3K27ac (bottom left), or SSDP1 (top right). Peak regions have been sorted by P value and top 1000 for each protein plotted, with the exception of Lhx3-HA which had <1000 peaks. Signal intensity is plotted across a 5 kb region centred at the middle of each peak region, and then sorted for signal intensity. Note the decrease of Isl1-FLAG signal across all other peak regions in the absence of LDB1.

It is also clear that a substantial fraction of the strongest Isl1-FLAG peaks are LDB1-independent, and associate to some degree with other Lhx3-HA-bound

regions and possibly a few SSDP1-bound regions. But, there is also a strong likelihood that these signals are from background regions amplified during Isl1-FLAG ChIP-seq library construction, or from genomic AT-rich repeat signals that are not MN specific.

3.4 - H3K27AC OCCUPANCY IS REDUCED AT KEY MN GENES IN THE ABSENCE OF LDB1

There are an abundance of H3K27ac peaks throughout the genome, and the histone mark is often used as a canonical marker of open chromatin and enhancer/promoter/active gene identity (Ji et al. 2016). Given LDB1's role as a transcriptional co-activator, it seemed likely that H3K27ac marks at enhancers and promoters of MN-involved genes may be affected in the absence of LDB1. We performed cross-linking ChIP-seq against H3K27ac in all *Ldb1* genotypes as described. When plotted again LDB1 peaks, there is a clear preference for areas ~1 kb to either side of LDB1 peaks, but this pattern is only revealed when sorting peak regions by H3K27ac signal intensity, whereas when sorted by LDB1 peak significance the pattern is evident at the top LDB1 peaks but much more distributed throughout the top 1000 LDB1-bound sites. **(Figure 3.17)**. This binding pattern could be largely rescued by re-introducing wildtype LDB1 through doxycycline induction, but was absent in the SSDP1-binding LDB1 mutant construct with ChIP-seq signal levels comparable to LDB1 knockouts.

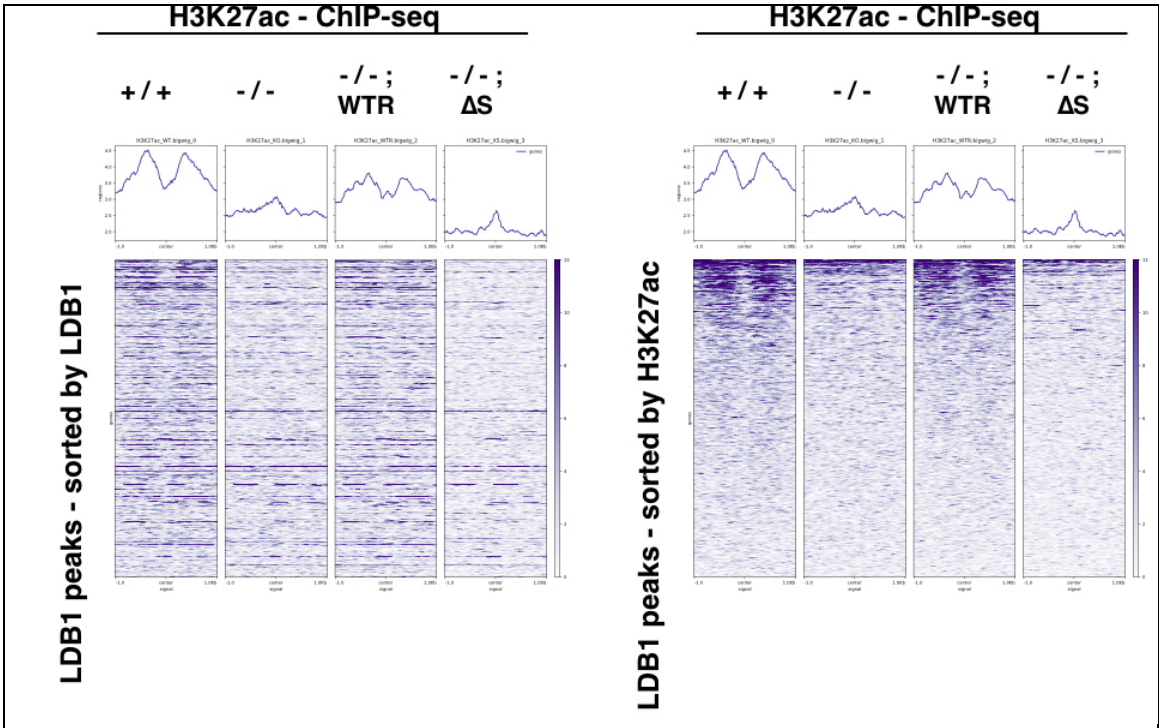


Figure 3.17 – H3K27ac is differentially regulated on a subset of LDB1 peaks
Heatmaps of normalised H3K27ac ChIP-seq coverage from Day 6 doxycycline-induced EBs in either *Ldb1 +/+ , Ldb1 -/-* or respective rescue and mutant cell lines in each column. The left set is sorted by LDB1 peak P value, while the right set is sorted by H3K27ac ChIP signal intensity.

The specificity in H3K27ac occupancy appears specific to LDB1 rather than LIM-HD TFs, as the association between Isl1-FLAG peaks and H3K27ac-occupied regions is not LDB1-dependent. When Isl1-FLAG ChIP-seq signal in any LDB1 genotype is plotted against H3K27ac peaks, there is a strong enrichment in every condition, likely the result of non-specific homeodomain binding at any AT-rich region of open chromatin. **(Figure 3.18).**

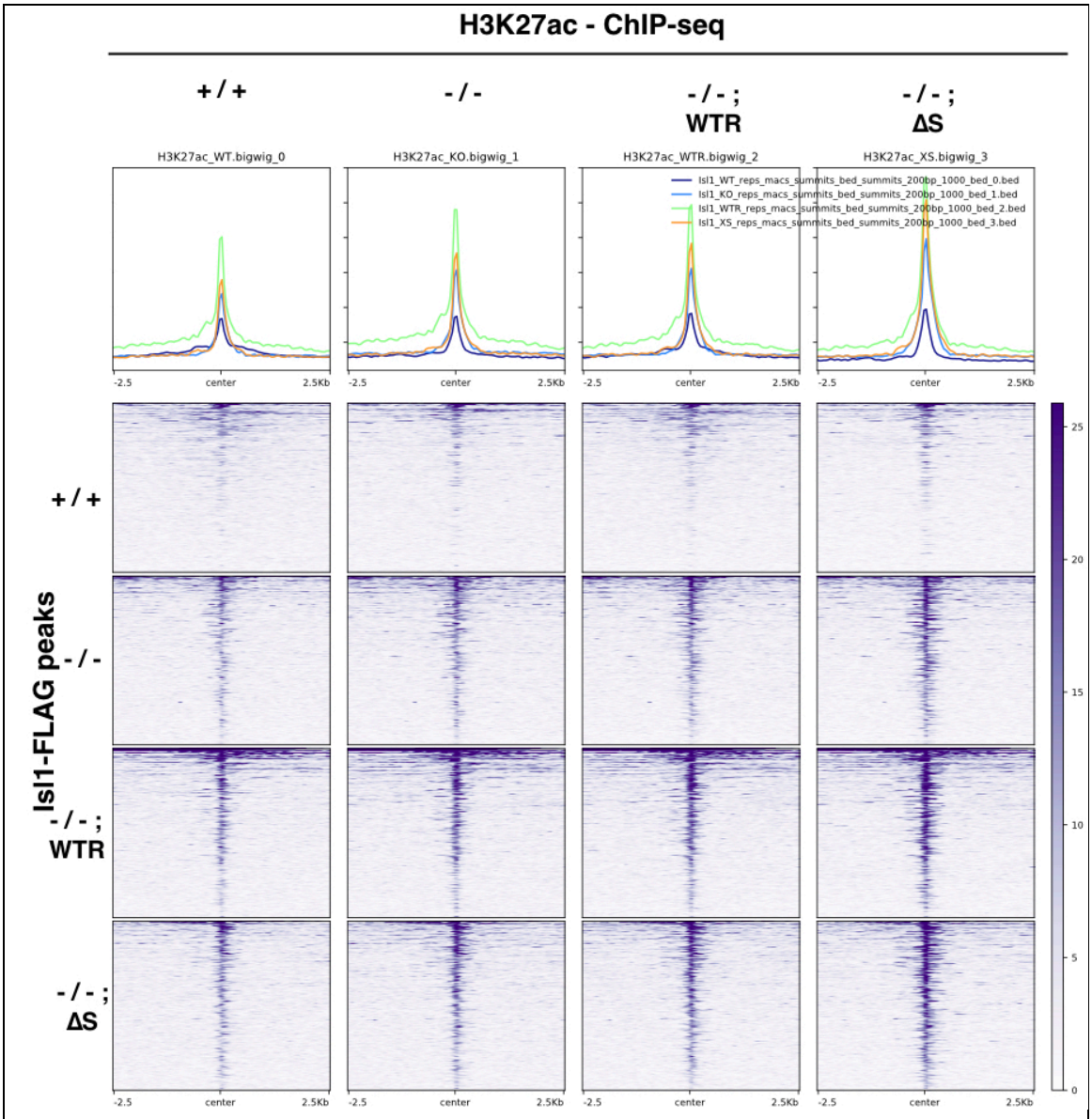


Figure 3.18 – H3K27ac retains occupancy at Isl1-FLAG sites independently of LDB1 function
Heatmaps of normalised H3K27ac ChIP-seq coverage from Day 6 doxycycline-induced EBs in either *Ldb1 +/+ , Ldb1 -/-* or respective rescue and mutant cell lines in each column. Each set of regions is the top 1000 Isl1-FLAG peaks in each LDB1 genotype, respectively.

We hypothesize that in the absence of LDB1, the promiscuous homeodomain of Isl1 will likely recognise and bind with low affinity to any accessible or open chromatin in the accessible nuclear space, which is generally marked by H3K27ac. All attempts to call motifs from any subset of LDB1-independent Isl1-

FLAG peaks resulted in poly-A/T regions such as in the *Ldb1 -/-* Isl1-FLAG data as a whole **(Figure 3.13)**, so while we cannot exclude that it is functionally binding to another set of regions, we were not able to determine those regions as separate from any library-construction-induced noise.

3.5 – LDB1 MARKS SITES OF OPEN CHROMATIN VIA ATAC-SEQ

LDB1 is thought to drive transcriptional activation, which includes histone remodeling and generation of more accessible nucleosome-free regions. Therefore, it was logical that would we seek to assay chromatin accessibility in our mutant embryoid bodies. Using the ATAC-seq assay (Buenrostro et al. 2015), we sequenced open chromatin in LDB1 wildtype, LDB1 knockout, and ΔS LDB1-mutant cell lines, and see a markedly LDB1- and SSDP1-interaction-dependent reduction of chromatin accessibility at LDB1 peaks **(Figure 3.19**).

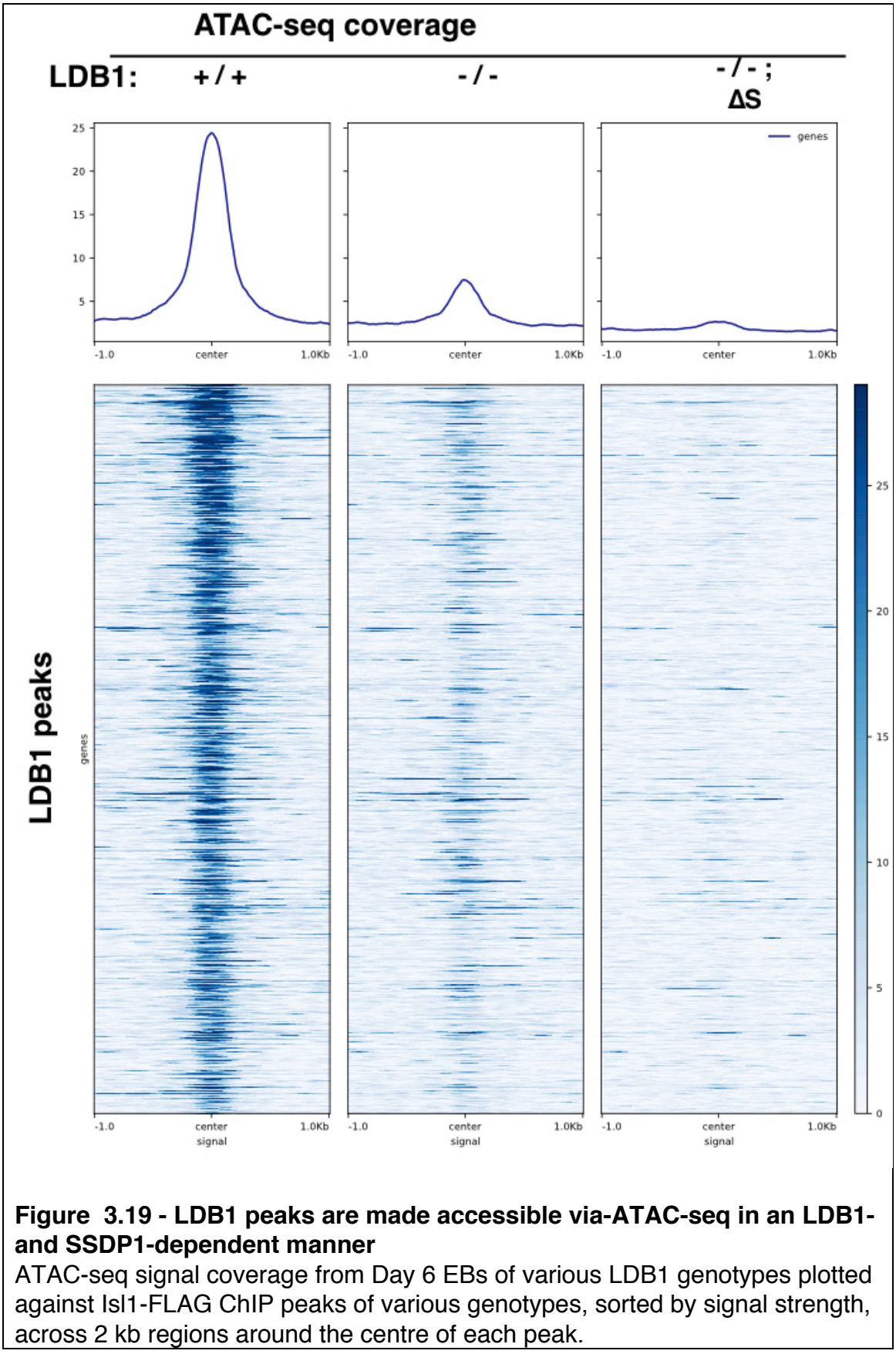


Figure 3.19 - LDB1 peaks are made accessible via-ATAC-seq in an LDB1- and SSDP1-dependent manner
ATAC-seq signal coverage from Day 6 EBs of various LDB1 genotypes plotted against Isl1-FLAG ChIP peaks of various genotypes, sorted by signal strength, across 2 kb regions around the centre of each peak.

This is consistent with our previous results indicating LDB1 is stabilized through its interaction with SSDP1 and directs chromatin remodeling to the sites of LIM-

HD TFs bound to multiple genomic locations. This pattern is evident from mapping ATAC-seq reads to Isl1-FLAG peaks in multiple LDB1 genotypes. There is differential accessibility of Isl1-FLAG peaks that are LDB1-dependent, while the majority of Isl1-FLAG peaks in contexts without LDB1 are not particularly accessible **(Figure 3.20)**.

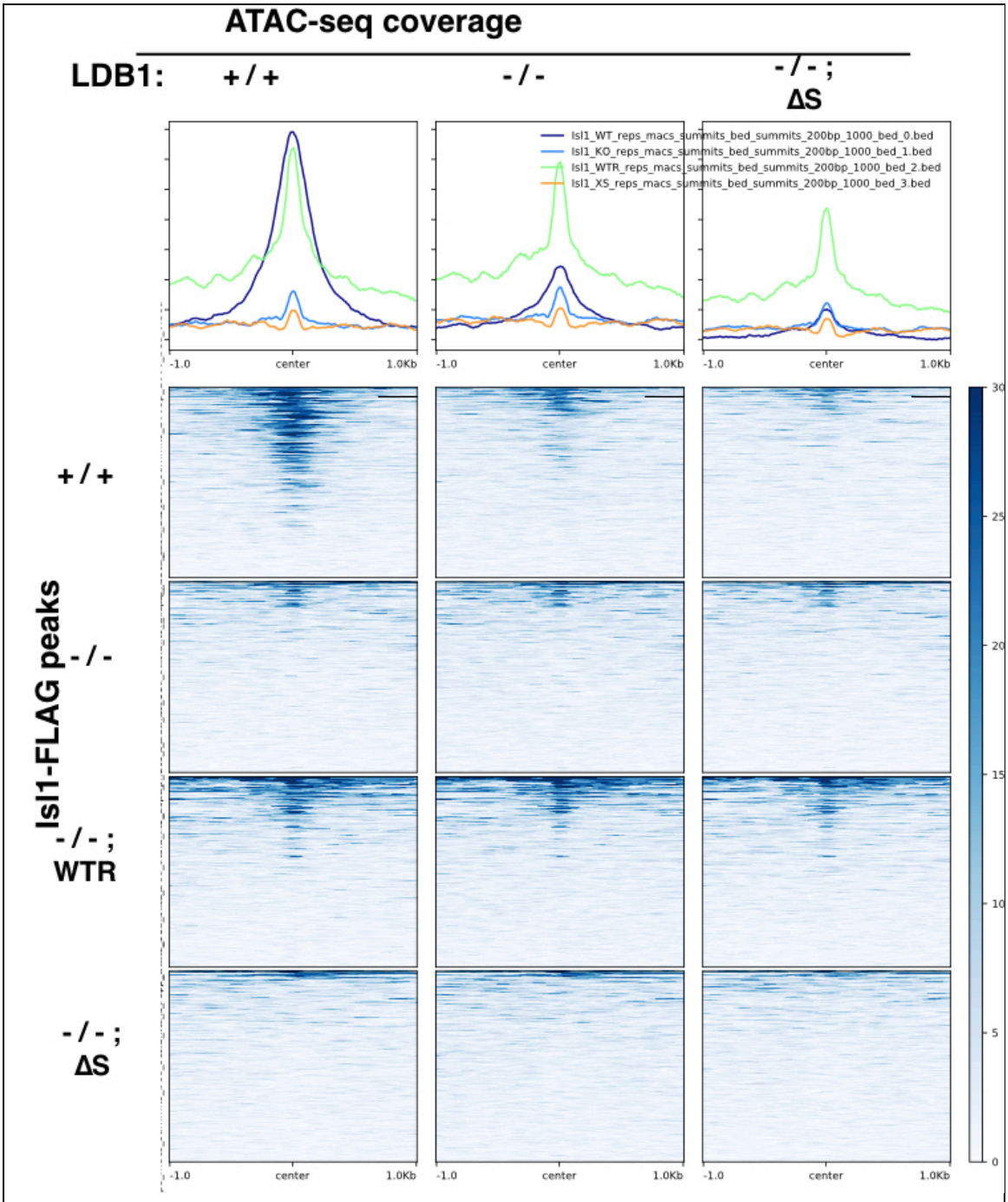


Figure 3.20 – Isl1-FLAG peaks are made accessible via-ATAC-seq in an LDB1- and SSDP1-dependent manner
ATAC-seq signal coverage from Day 6 EBs of various LDB1 genotypes plotted against Isl1-FLAG ChIP peaks of various genotypes, sorted by signal strength, across a 2 kb region around the centre of each peak.

By combining ATAC-seq with the matched ChIP-seq datasets, it is evident that chromatin accessibility is more generalized than the H3K27ac phased occupancy

observed at LDB1 sites, but that they are closely related, and indicate locations where LDB1 is directing occupancy of LIM-HD TFs in an SSDP1-dependent manner.

3.6 – SSDP1 MARKS A SUBSET OF LDB1 AND LIM-HD TF BOUND SITES THAT ARE SPECIFICALLY ACCESSIBLE AND H3K27AC-OCCUPIED

When we observe LDB1, Isl1-FLAG, Lhx3-HA, H3K27ac, and SSDP1 ChIP-seq and ATAC-seq peaks from LDB1 wildtype Day 6 EBs in aggregate, and plot their reciprocal signal intensities, several notable patterns emerge **(Figure 3.21)**.

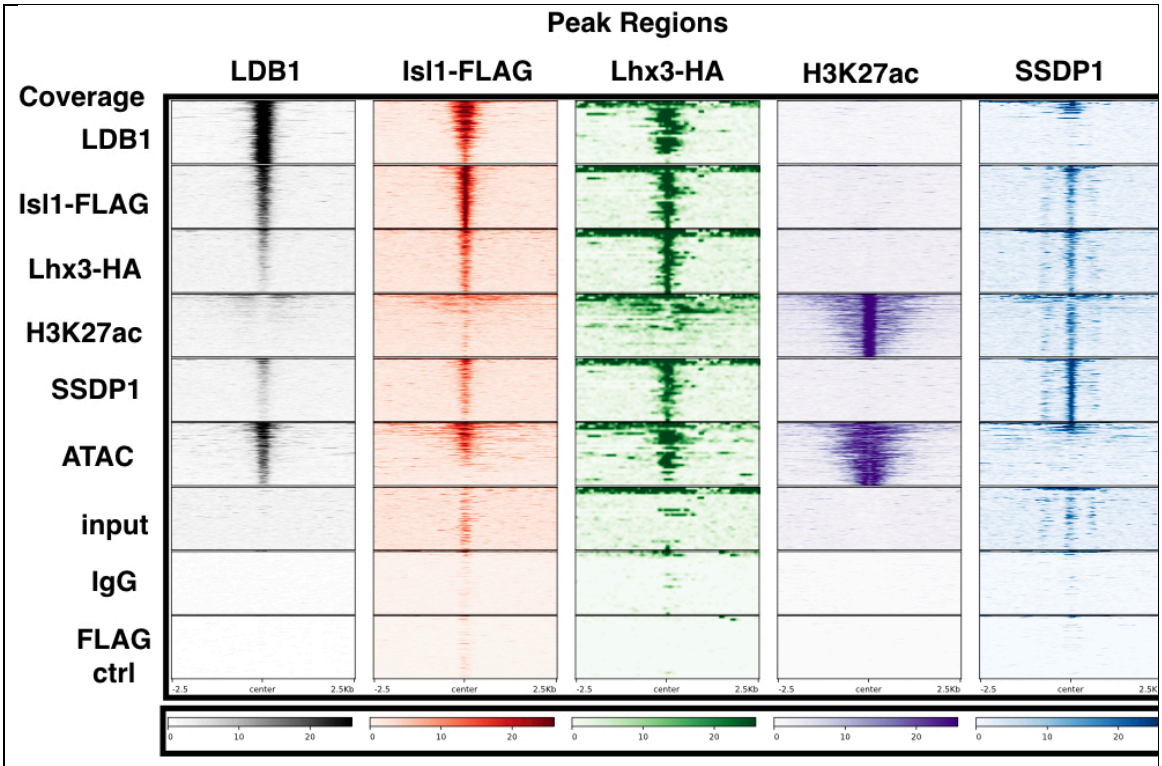


Figure 3.21 - LDB1 associates with a select group of peaks that are marked specifically by SSDP1
ChIP-seq or ATAC-seq coverage from LDB1 wildtype Day 6 EBs. Signal coverage from respective ChIP-seq experiments is displayed in each colour, with each column being the same ChIP signal. Peak regions are the top 1000 regions for each indicated ChIP or ATAC experiment and cover a 5 kb region around the centre of each peak.

In this plot, peak regions from the same target are colour-coded and sorted by column, while coverage from each respective assay is plotted in each row. As expected we observe a strong signal in the centre of each plot along the left-to-right diagonal, indicating that the signal from each assay is indeed localizing to

the peak regions called from that signal. Second, a genome-wide basis the most accessible and H3K27ac-occupied regions are not the same as the LDB1- and LIM-HD TF-marked regions of interest. Third, SSDP1 peaks, which associate to some degree with all the factors assayed, include a small subset of peaks that are specifically LDB1-bound and have accessible chromatin. This subset of peaks is of particular interest, as the large-scale overlap of LIM-HD TFs with LDB1 and other markers of accessible chromatin is likely due to a high amount of non-specific or promiscuous low-affinity binding of LIM-HD TFs to non-relevant target sites.

We interpret these results to mean that without SSDP1, LDB1-mediated Isl1-Lhx3 complexes are not binding to MN specific sites, possibly because of ubiquitin-mediated degradation of LDB1 in the absence of SSDP1 binding. In aggregate, this data indicates a direct relationship between the Isl1-Lhx3-LDB1 complex, SSDP1 and H3K27ac deposition at a subset of LDB1-specific regions.

CHAPTER 4. LDB1 REGULATES LIM-HD TF-INDUCED GENE EXPRESSION PROGRAMS

4.1 – LDB1 IS RELIABLY INDUCED AND IS NOT COMPENSATED FOR BY LDB2 IN THE MOTOR NEURON INDUCTION SYSTEM

Fundamentally, any changes in chromatin architecture, genome occupancy by transcription factors, cofactors, or histone modifications are only biologically relevant if they have some effect on cell identity, ultimately through transcriptional output of RNA. Because there were substantial changes in TF binding in the absence of LDB1, and because LDB1 mutations are lethal in the embryo, we hypothesized that a large range of genes would be mis-expressed in Day 6 EBs. To test this, we performed nuclear mRNA-seq on EBs in multiple LDB1 genotypes with and without doxycycline induction and then assessed the level of *Ldb1* and *Ldb2* transcripts in all samples (**Figure. 4.1**.)

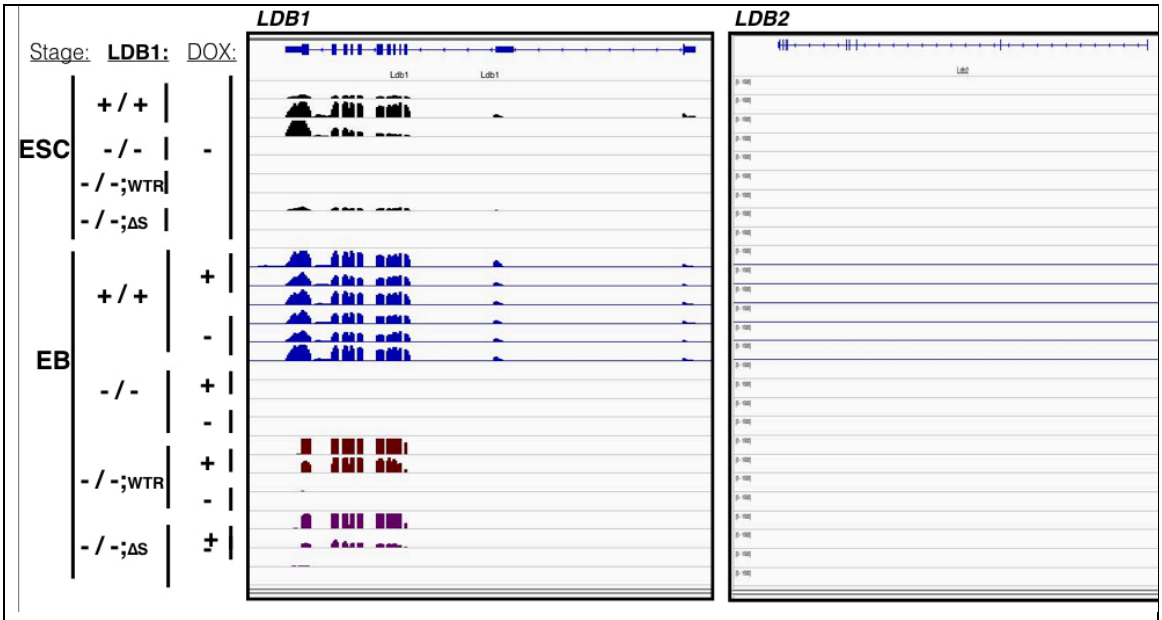


Figure 4.1 – *Ldb1* is not expressed in *Ldb1* knockout cells without an inducible transgene and *Ldb2* is not expressed in any condition
RNA-seq snapshots of the *Ldb1* and *Ldb2* gene loci, showing 2 or 3 biological replicates of both ESCs and Day 6 EBs with *Ldb1* genotype and doxycycline induction status indicated on the left. ESC transcripts are black, while *Ldb1 +/+*. EBs are blue, *Ldb1 -/-*. EBs are green, *Ldb1 -/-* with a wildtype rescue (WTR) transgene are in burgundy, and *Ldb1 -/-* with a SSDP1-binding domain mutation transgene (ΔS) are in magenta. For the *Ldb1* locus, scale for ESCs and *Ldb1 +/+* EBs is 1500, while for *Ldb1 -/-* and *Ldb1 -/-* transgene EBs it is 500. For the *Ldb2* locus, scale for all samples is 1500.

Ldb1 is expressed at moderate but variable levels in *Ldb1* +/+ ESCs. It is more strongly and consistently expressed among replicates upon neuronal induction both with and without doxycycline. Importantly, no *Ldb1* transcripts were detected at appreciable levels in the *Ldb1* -/- cells either at the ESC, uninduced EB or doxycycline-induced EB stage, and the *Ldb1* transgene expression was detected only in the presence of doxycycline. Additionally, although *Ldb2* can be compensatory and have overlapping function with *Ldb1* in the retina (Gueta et al. 2016), there is no *Ldb2* expression at appreciable levels in any of the cell types studied (**Figure 4.1**).

4.2 – LDB1 REGULATES GENE EXPRESSION OF SELECT MOTOR NEURON GENES

We first checked the LDB1- and doxycycline-dependent RNA expression levels of essential regulators of MN identity. There is a strong up-regulation of Mnx1 transcription only in doxycycline-treated (induced) Day 6 EBs that have homozygous wildtype *Ldb1* alleles, or contain the wildtype *Ldb1* transgene, while negligent expression of *Mnx1* mRNA is seen in both *Ldb1 -/-* and SSDP1-domain-binding mutant transgene expressing EBs **(Figure 4.2)**. At other essential MN genes, including *Slit2*, which is known to be regulated by Isl1-containing complexes (Kim et al. 2016a), we also observe a doxycycline- and LDB1-dependent expression profile **(Figure 4.2)**.

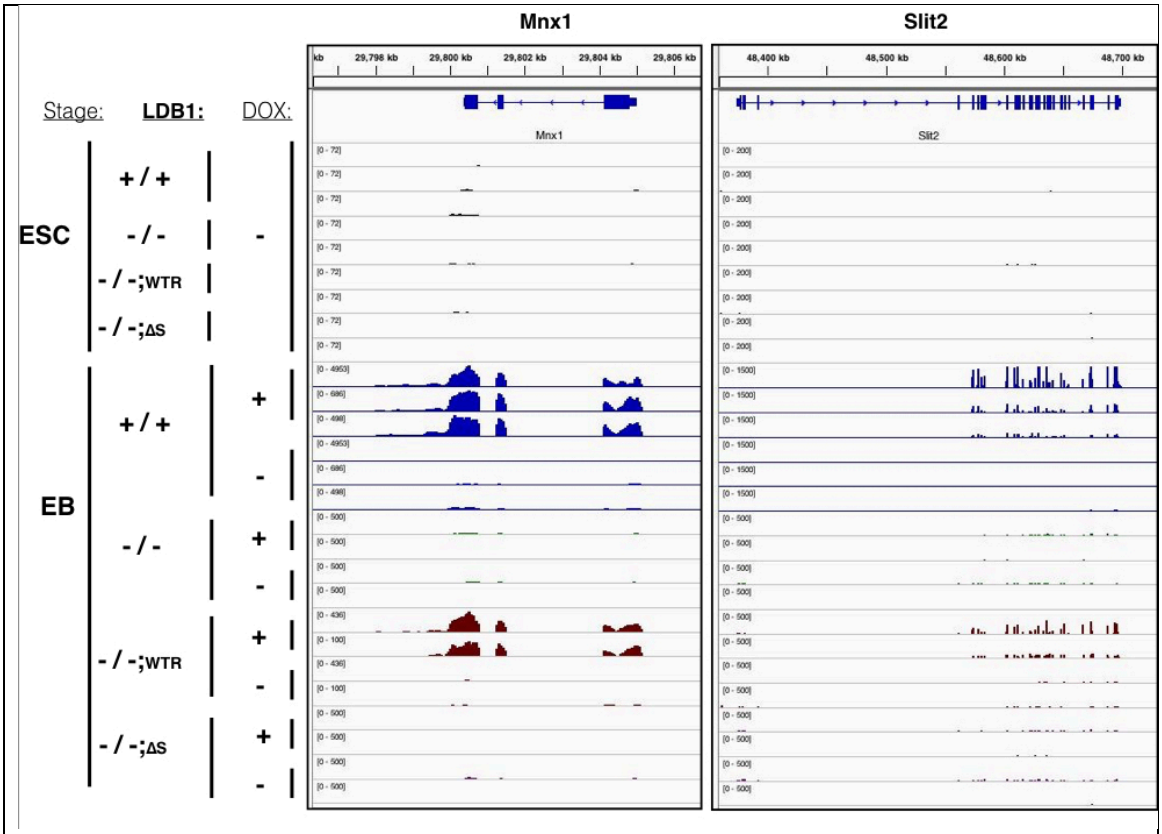


Figure 4.2 – Critical motor neuron genes are dependent on LDB1 and MNC induction
RNA-seq snapshots of the Mnx1 and Slit2 gene loci, showing 2 or 3 biological replicates of both ESCs and Day 6 EBs with both Ldb1 genotype and doxycycline induction status indicated on the left. ESC transcripts are black, while Ldb1 +/+. EBs are blue, Ldb1 -/-. EBs are green, Ldb1 -/- with a wildtype rescue (WTR) transgene are in burgundy, and Ldb1 -/- with a SSDP1-binding domain mutation transgene (ΔS) are in magenta. Scales are indicated for each sample.

4.3 – LDB1 DELETION AFFECTS GENERAL DIFFERENTIATION AND LIM-HD TF-INDUCED MOTOR NEURON PROGRAMS

Embryonic stem cells have intrinsic variability in gene expression levels due to genetic background, culture conditions, and age (Guo et al. 2016). To establish that the ground-state embryonic stem cells from which our EB cultures were derived were not substantially divergent based on *Ldb1* genotype, we performed RNA-seq using two biological replicates for each of the four LDB1 genotypes analyzed. There were no significant changes in expression between these ESC lines as analyzed by DEseq2 (**Figure 4.3)**.

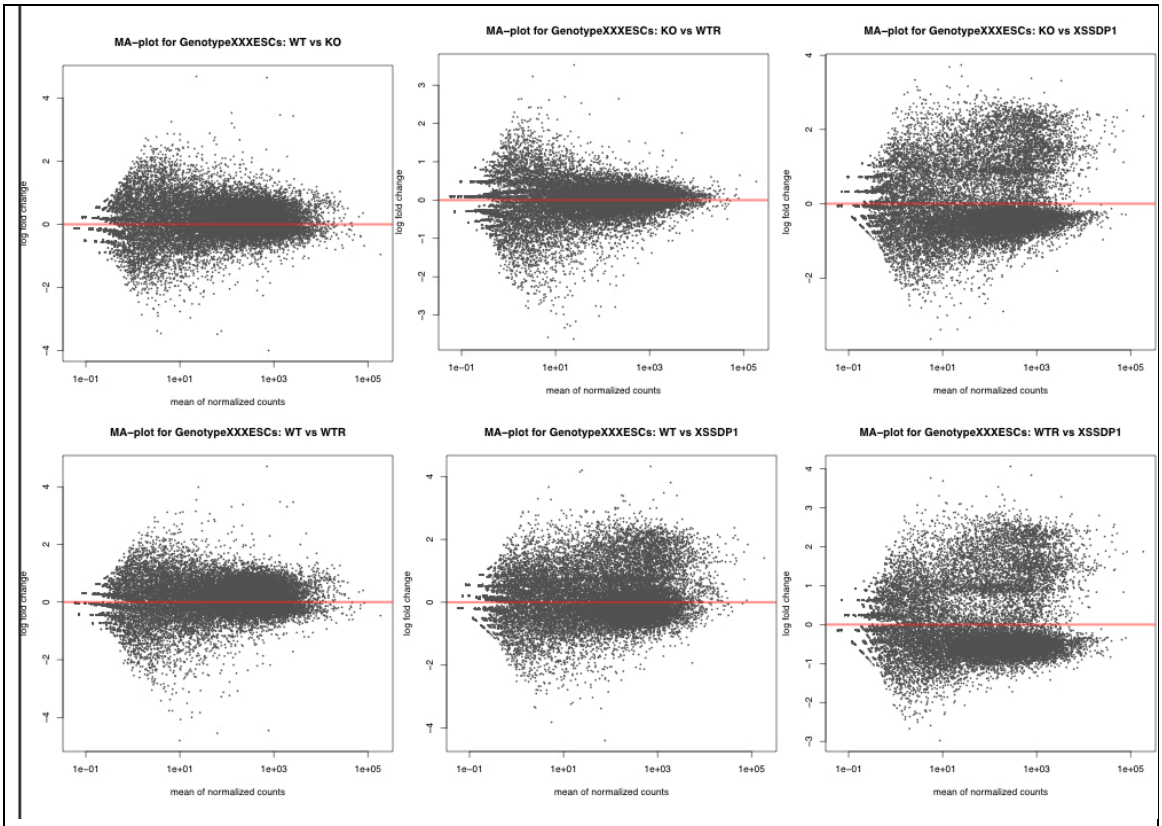


Figure 4.3 – Ldb1 genotype does not affect gene expression levels in embryonic stem cells
MA plots of comparisons between *Ldb1* genotypes in ESCs. Titles are formatted as "control vs. test". The Y axis indicates log2 fold change for each gene, while the X axis indicates mean average expression level for each gene. Genes are coloured red if they are significantly up- or down-regulated. Note that there are no significantly up- or down-regulated genes.

We next compared overall gene expression levels between doxycycline-induced (+DOX) versus solely retinoic acid- and smoothened agonist-patterned EBs (-DOX). We compared these both in relation to ESCs and to each other. Notably, the effect of generalized neuronal differentiation, with 4,898 or 7,038 differentially expressed genes (corrected P value < 0.05) between undifferentiated ESCs and Day 6 EBs without or with doxycycline. In contrast, doxycycline induction alone had a much more specific effect, with 87 differentially regulated genes **(Figure 4.4).**

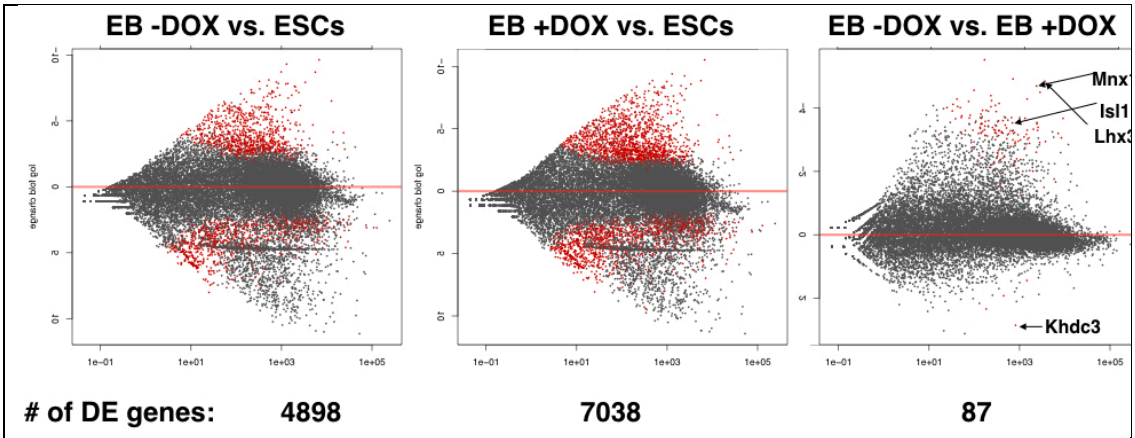


Figure 4.4 – Doxycycline induction changes a small number of gene expression levels compared to general neuronal differentiation
MA plots of comparisons with combined replicates from *Ldb1+/+* cell lines from each developmental stage: ESC, EBs with no doxycycline induction, or EBs with doxycycline induction. Here doxycycline induces the V5-Ngn2/FLAG-Isl1/HA-Lhx3 cassette, and does not include any wildtype or mutant forms of LDB1. Titles are formatted as "control vs. test". The Y axis indicates log2 fold change for each gene, while the X axis indicates mean average expression level for each gene. Genes are coloured red if they are significantly up- or down-regulated. The number of differentially expressed (DE) genes with an adjusted P value of 0.05 or lower are indicated below each graph. Genes of interested are indicated with arrows.

This indicates that changes in gene expression between ESCs and neuronally-induced cells, whether they have a MN identity or not, involves nearly two orders of magnitude more genes than the difference between induced MNs and non-

specifically differentiated neural cells. Importantly, this analysis identifies a minimal subset of <100 genes responsive to the Isl1-FLAG and Lhx3-HA LIM-HD TFs. Among these DOX-induced genes, in *Ldb1 +/+* Day 6 EBs, we observe upregulation of several canonical markers of MN identity, including: *Isl1, Lhx3, Mnx1, Slit2, Fgf9, Ngfr, Nrxn1, Megf11, Pcdh9, Onecut3, Isl2, Irx1,* and many *Slc* genes, while there is downregulation of only 3 markers, *Gbx2, Hlx, and Khdc3.*

We next compared doxycycline-induced Day 6 EBs with and without LDB1. As expected with a fundamental regulator as LDB1, there were 1,363 differentially expressed genes **(Figure 4.5 A)**.

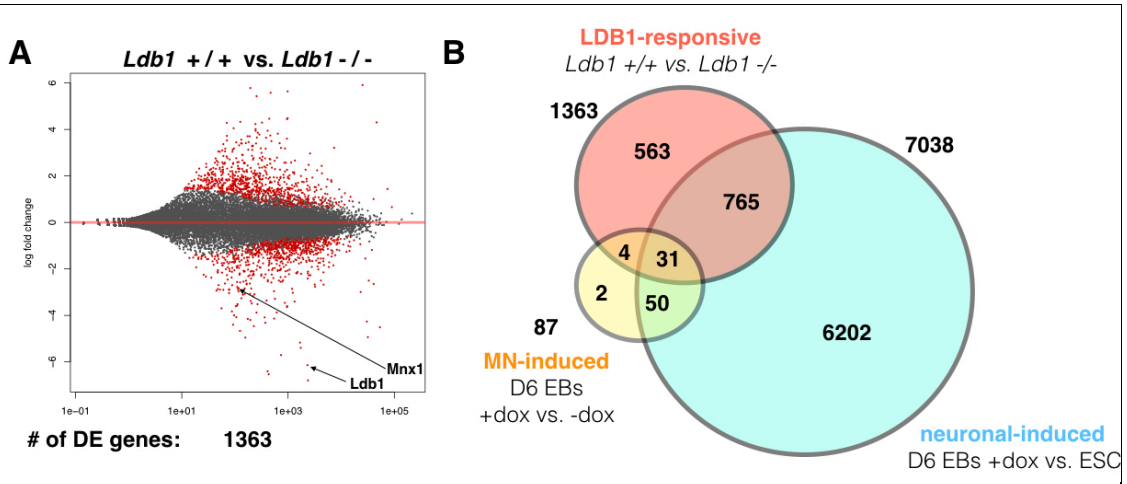


Figure 4.5 – LDB1 knockout affects the expression of many genes, but only a select number that are also motor neuron-induced
A) MA plot of LDB1 knockout compared to wildtype gene expression levels. The Y axis indicates log2 fold change for each gene, while the X axis indicates mean average expression level for each gene. Genes are coloured red if they are significantly up- or down-regulated. The number of differentially expressed (DE) genes with an adjusted P value of 0.05 or lower is indicated below each graph. Genes of interested are indicated with arrows. **B)** Venn diagram representation of the extent of overlap of differentially regulated genes between combined replicates from different conditions. Proportions are not to scale, and absolute numbers of overlapping DE genes are printed inside each class, while total numbers of DE genes from each condition are outside each class.

In the *Ldb1* knockout compared to wildtype in Day 6 EBs, there is significant downregulation of the expected *Ldb1 and Mnx1,* along with a wide range of other factors including *Srgn, GATA4,* many laminin and collagen-related genes as well as neuronal markers such as SLC-family genes. By comparison there were also many genes up-regulated, including *Lhx5*, imprinted genes *H19 and Peg10,* and a wide range of transcription-related genes. Gene ontology analysis returned many different classifications for these DE genes, suggesting an endodermal fate but without a clear alternative fate path. Rather than diverting to another specific neural lineage such as interneurons, which also require LDB1 function, these cells appear to not have a clearly identifiable fate identity as determined by gene ontology.

4.4 – IDENTIFICATION OF A MINIMAL MOTOR NEURON GENE EXPRESSION PROGRAM

An overarching goal of studying LDB1 in MN differentiation is to identify a subset of genes necessary for MN identity. By comparing the list of NIL-induced genes with the list of LDB1-induced genes **(Figure 4.5 B)**, we define a list of 35 genes which we report here in **Table 3**, all of which are upregulated with the exception of *Khdc3* and *Gbx2*. The upregulated genes include several members of the solute carrier family, aquaporins, as well as the expected *Mnx1, Slit2, and Isl1 and 2*, and several other neuronal genes such as *Megf11, Fgf9,* and *Rxrg*.

Table 3 – LDB1-regulated and DOX-inducible genes in Day 6 EBs

ENTREZ GENE ID	Name
18022	nuclear factor, erythroid derived 2 – (Nfe2)
11829	aquaporin 4 – (Aqp4)
18018	nuclear factor of activated T cells, cytoplasmic, calcineurin dependent 1 – (Nfatc1)
243616	solute carrier family 6 (neurotransmitter transporter, GABA), member 11 – (Slc6a11)
319764	RIKEN cDNA A730046J19 gene – (A730046J19Rik)
75552	progestin and adipoQ receptor family member IX – (Paqr9)
214058	multiple EGF-like-domains 11 – (Megf11)
109594	LIM domain only 1 – (Lmo1)
63993	solute carrier family 5 (choline transporter), member 7 – (Slc5a7)
16773	laminin, alpha 2 – (Lama2)
54153	RAS p21 protein activator 4 – (Rasa4)
71306	microfibrillar-associated protein 3-like – (Mfap3l)
12291	calcium channel, voltage-dependent, T type, alpha 1G subunit – (Cacna1g)
15452	hypoxanthine guanine phosphoribosyl transferase – (Hprt)
74563	RasGEF domain family, member 1C – (Rasgef1c)
56175	beta-site APP-cleaving enzyme 2 – (Bace2)
20183	retinoid X receptor gamma – (Rxrg)

84004	melanoma cell adhesion molecule – (Mcam)
109254	androgen dependent TFPI regulating protein – (Adtrp)
319807	NACHT and WD repeat domain containing 2 – (Nwd2)
104360	insulin related protein 2 (islet 2) – (Isl2)
13599	endothelin converting enzyme-like 1 – (Ecel1)
240843	bone morphogenic protein/retinoic acid inducible neural-specific 2(– Brinp2)
76157	solute carrier family 35, member D3 – (Slc35d3)
24012	regulator of G protein signaling 7 – (Rgs7)
66991	KH domain containing 3, subcortical maternal complex member – (Khdc3)
14472	gastrulation brain homeobox 2 – (Gbx2)
20563	slit homolog 2 (Drosophila) – (Slit2)
72685	DnaJ heat shock protein family (Hsp40) member C6 – (Dnajc6)
231290	solute carrier family 10 (sodium/bile acid cotransporter family), member 4 – (Slc10a4)
20508	solute carrier family 18 (vesicular monoamine), member 3 – (Slc18a3)
15285	motor neuron and pancreas homeobox 1 – (Mnx1)
75404	Rho GTPase activating protein 36 – (Arhgap36)
14180	fibroblast growth factor 9 – (Fgf9)
244425	RIKEN cDNA A730069N07

4.5 - Discussion

Taken together, this collection of RNA-seq data shows a specific subset of neuronally induced genes that are: 1) responsive to patterning factors RA, SAG, and DAPT; 2) responsive to doxycycline-induced TF reprogramming with LIM-HD TFs FLAG -Isl1, HA-Lhx3, and V5-Ngn2; 3) are responsive to *Ldb1* deletion. This constitutes an overlapping set of minimal genes that are differentially expressed specifically upon both MN induction and LDB1 deletion. These genes include transcription factors, growth factors, signaling and membrane transporter proteins, and other regulators of cell physiology that are downstream targets of the MN-induction program and LDB1 function. There may also be many other relevant genes that fall slightly outside these particular significance and fold-change criteria. Responsiveness to both MN-induction with inducible TFs and deletion of *Ldb1* cannot differentiate between which upstream genes regulate the progression of the MN-induction program, and which downstream genes are mis-expressed. However, the identification of canonical MN factors, along with a number of downstream cell physiology, transporter, and signaling genes, further defines the transcriptome of induced MNs and identifies potential markers of MN function for future investigation.

Additionally, by identifying the LDB1-responsive genes that are MN-induction specific, we have a narrowed list of candidate regions at which to look for LDB1-related changes in 3D genome organization in future experiments. These regions are some unknown number of steps downstream of the experimental manipulation, but with the knowledge that LDB1 is essential for cooperative binding of LIM-HD TFs at several key MN-specific genes such as

Mnx1, Nrp1, Slit2, and others, we can use these genes as starting points when considering genome-wide interaction data.

There are also several additional analyses possible with these RNA-seq datasets, such as clustering studies to identify the neuronal subtype *Ldb1 -/-* cells may assume, and further classification of gene programs responsive to *Ldb1* deletion but not MN induction. This would further define the range of LDB1 downstream targets in relation to the critical components of the MN differentiation program.

CHAPTER 5. 3D-SIM IMAGING OF LDB1 AND LIM-HD TFS

5.1 – 3D STRUCTURED ILLUMINATION MICROSCOPY OF MNC COMPONENTS

We hypothesized that if LDB1, Isl1, and Lhx3 were forming discrete transcriptional complexes at a subset of genomic sites it should be possible to visualize these clusters, and that if LDB1's main function was to bridge Isl1-Lhx3 dimers, we could observe changes in the spatial distributions of Isl1 and Lhx3 signals in the absence of LDB1. Due to the scale of these interactions, in the tens to hundreds of nm, and the need to image in multiple wavelengths simultaneously in intact nuclei, we elected to use 3D-Structured Illumination Microscopy (3D-SIM) (Schermelleh et al. 2008). This technique has, in theory, an optimal combination of multi-channel acquisition, spatial resolution in all three dimensions, and acquisition- and dye-specific characteristics that make it well-suited for studying nuclear structures (Smeets et al. 2014; Schermelleh et al. 2010; Demmerle et al. 2015; Schermelleh et al. 2019).

We adapted a modified version of our MN induction protocol, wherein the ESCs were not differentiated into EBs, but induced with the same chemical mixture, with or without doxycycline for the same 48 hrs as in previous experiments **(Figure 5.1)**.

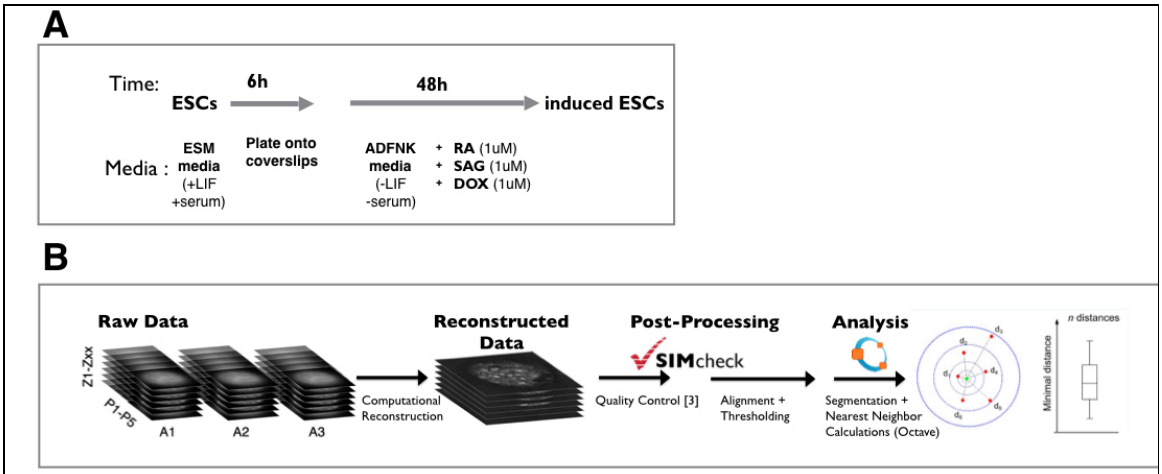


Figure 5.1 – Experimental setup for imaging of MNC components via 3D-SIM imaging.
A) Schematic of differentiation assay, showing time after removal from ESC conditions, the media composition, and growth conditions. ESM media contains serum while ADFNK contains Knock-Out Serum Replacement (KOSR). Retinoic acid (RA), Sonic hedgehog agonist (SAG), doxycycline (DOX) are labeled with their final concentration **B)** Schematic of data processing pipeline for determination of nearest-neighbor distances from 3D-SIM data, and the various quality control and processing steps from the raw data to nearest-neighbor distances.

This allowed improved imaging of individual nuclei, as dissociated EBs created samples with excessive background for effective super-resolution imaging. We then acquired calibrated and QC-optimized (Demmerle et al. 2017) 3D-SIM data for both Isl1-FLAG and Lhx3-HA with and without LDB1, using either antibodies against the epitope tags, or the native proteins. We reconstructed and post-processed the data according to standard practices (Ball et al. 2015; Kraus et al. 2017). In the reconstructed images there are clearly distinct foci of LIM-HD TFs throughout the nucleus, generally in the inter- or peri-chromatin space, as would

be expected for activating marks (Miron et al. 2019), that do not visibly change their distributions in the absence of LDB1 **(Figure 5.2)**.

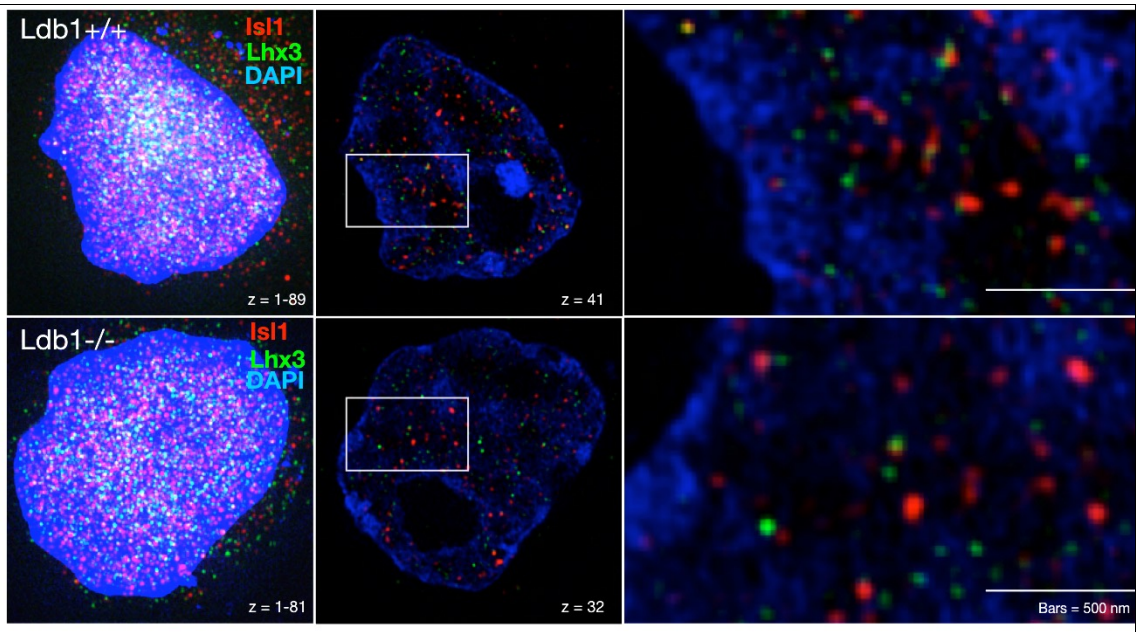


Figure 5.2 – 3D-SIM Imaging of MNC complex components
Reconstructed, channel-registered 3D-SIM image showing a maximum projection of the entire Z-stack (left panel), a single optical section (middle panel) or an inset of the middle panel (right panel) of a single nucleus of an induced ESC, showing anti-FLAG-Isl1 staining in red, anti-HA-Lhx3 staining in green, and DAPI-stained DNA in blue. LDB1 genotype is indicated for each row. Z-planes are indicated in bottom right of the first two panels. Scale bar indicates 500 nm.

5.2 – Analysis of LIM-HD TF Foci positions

While super-resolution imaging is informative, robust quantification of the newly discovered spatial relationships is essential and remains a technical challenge (Durisic et al. 2014). After image collection, we performed a watershed segmentation and nearest-neighbor calculation on each foci of Isl1, Lhx3, or LDB1 signal to determine inter-foci distances with the "NNA" script, which generates nuclear-masked centroid calls of multiple channels, and then calculates the nearest centroid distances of the opposite channel for each of the two channels, as well as summary statistics (**Figure 5.3**).

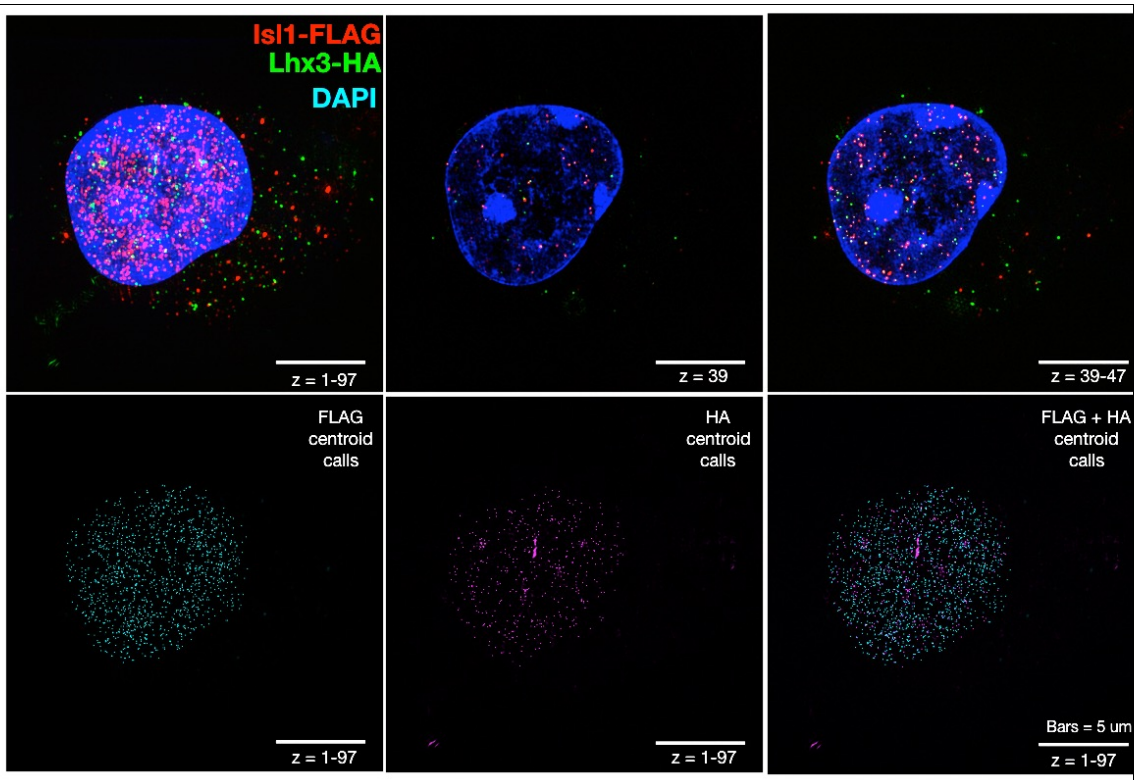


Figure 5.3 – Nuclear thresholding and centroid calls from immunofluorescence foci
Reconstructed, channel-registered 3D-SIM images showing a maximum projection of the entire Z-stack (top left panel), a single optical section (top middle panel) or middle stack of sections (top right panel) of a single nucleus of an induced ESC, showing anti-FLAG-Isl1 staining in red, anti-HA-Lhx3 staining in green, and DAPI-stained DNA in blue, in LDB1 +/+ cells. "nna" script centroid calls for each channel are displayed separately or together in the bottom row,

showing the maximum projection of the entire Z-stack. Scale bar indicates 5000 nm.

The utility in the approach can be seen in the substantial differences in distance distributions between induced and un-induced cells, where antibody staining against LIM-HD TFs in un-induced cells provide a baseline distribution of background signal (which is in our hands antibody-specific). Upon induction, the distance distributions condensed substantially **(Figure 5.4)**.

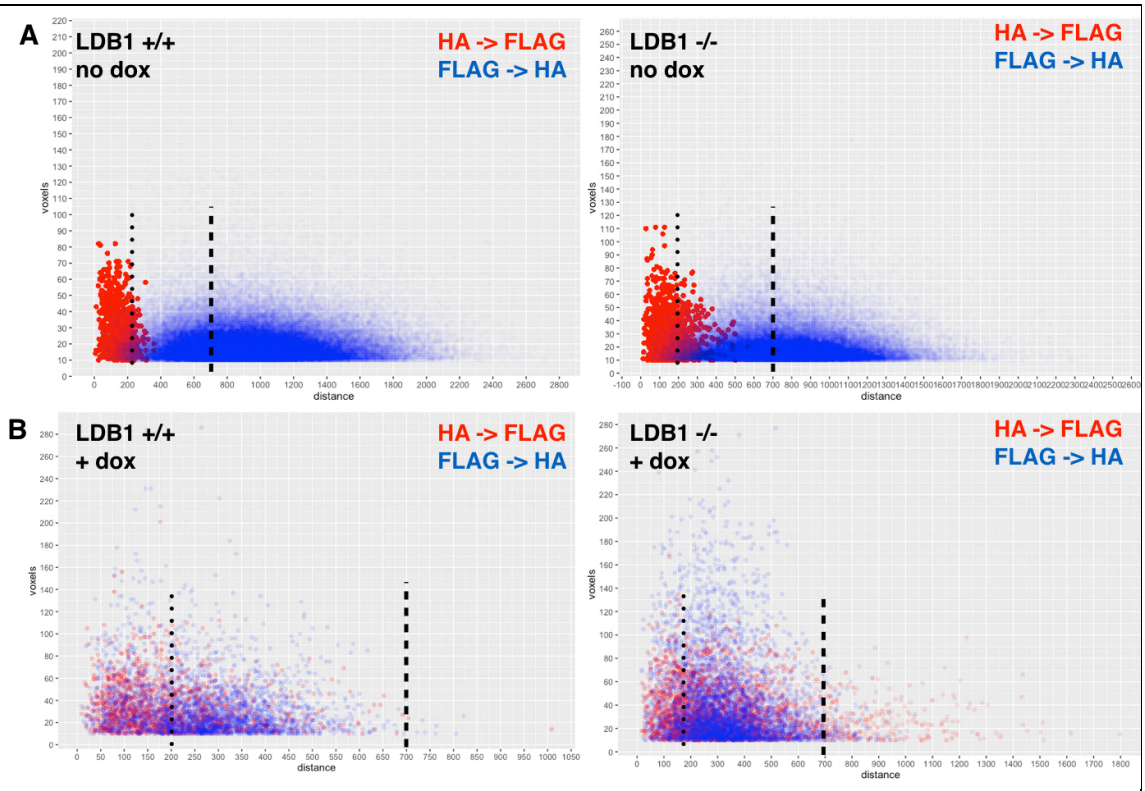


Figure 5.4 – Distance distributions of LIM-HD TF foci
Results of nearest-neighbor analysis from single representative cells plotted with voxel size of the foci on the Y axis, and distance to the nearest neighbor of the opposite channel, in nanometers, on the X axis. Top row shows distances from un-induced cells, and represent background. Each column shows the respective *Ldb1* genotype. Vertical dotted line indicates 200 nm, while the dashed line indicates 700 nm. Other dashed lines indicate the scale of distance ranges plotted. Distances from Isl1-FLAG to Lhx3-HA foci are in blue, while distances from Lhx3-HA foci to Isl1-FLAG foci are in red.

The decrease in distance distributions upon doxycycline induction could be an effect of either the drug itself making the antibody signal more closely associated, or causing a decrease in nuclear volume. Since we did not perform controls with doxycycline induction of cells lacking the transgene, nor measure nuclear volume as a consequence of drug treatment, we cannot rule out these effects. Further, the effect of formaldehyde fixation on foci distribution was not analyzed, but is a common confounding factor in microscopy studies and could cause a MN-induction-dependent effect on foci association in combination with the relative density of the nucleus. However, this control would have required generation of fluorescently-tagged cell lines for each protein of interest and was beyond our resources.

When stained for either FLAG-Isl1 and HA-Lhx3 with antibodies against the epitope tags, or for Isl1/2 and Lhx3 with antibodies against the proteins themselves, no difference in staining pattern could be detected between *Ldb1 +/+* and *Ldb1 -/-* cells.Quantification of these inter-foci distances did not show any statistically significant differences between *Ldb1 +/+* and *Ldb1 -/-* cells for either epitope-tag or native-protein antibody stainings **(Figure 5.5)**. However, co-staining of Isl1(-FLAG) and Lhx3-(HA) in multiple conditions yielded distance distributions (mean of mode inter-foci distances) between 175 nm and 225 nm, below the diffraction limit of conventional microscopy, and greatly less than distances calculated from foci of background staining in each channel, validating our use of 3D-SIM methodology and demonstrating for the first time individual protein-complex association measurements of these LIM-HD TF complexes in individual nuclei **(Figure 5.5)**.

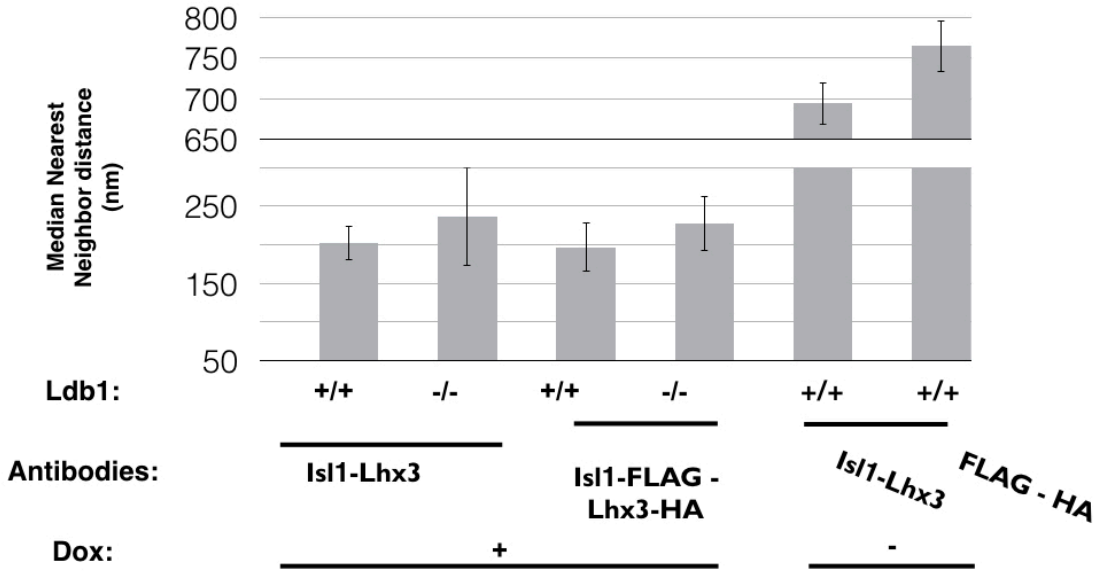


Figure 5.5 – Mean distance distributions of LIM-HD TF foci
Quantification of nearest-neighbor distances. Each bar represents the mean of mode distances from nearest-neighbor analysis from 5 cells for each condition. Distances are expressed in nm on the Y axis, which is discontinuous to include the non-doxycycline-induced samples for background distance controls. LDB1 genotype is indicated for each condition. Antibody pairs are indicated, with Isl1-Lhx3 representing antibodies against Isl1/2 or Lhx3 protein itself, while Isl1-FLAG – Lhx3-HA represent antibodies against FLAG or HA epitopes. Nearest neighbor distances from the same antibody stainings in uninduced ESCs is also plotted. No comparisons between genotypes were significant by an unpaired t-test. Bars represent S.E.M.

There are several aspects of these experiments, including use of primary-secondary antibody complexes, limitation in the number of 3D-SIM quality channels, and assumptions for thresholding, that make an absolute determination of differences in spatial separation of LDB1-Isl1-Lhx3 complex components difficult. Further, the nature of the MNC complex as a trimer of dimers, or a dimer of trimers, confounded the analysis because we did not have a system that could differentiate between inter- and intra-complex fluorophores. Thus, any overall change in distance between non-LDB1-coordinated Isl1-Lhx3 dimers may

obscure the contribution in the signal from the absence of functional MNC complexes. However, these data illustrate a distinct spatial pattern of complex formation and co-occurrence in three-dimensionally intact nuclei, and strengthen the case for distinct, spatially regulated foci of genome organization and chromatin remodeling mediated by this complex.

5.3 – Association of Isl1-FLAG and Mnx1 loci

A truly functional measure of whether the LIM-HD TF foci we observe are functionally activating MN specific genes would be to test for spatial co-occupancy of both the MNC complex components and the loci of interest. We tested this approach on a small scale by performing a non-denaturing fluorescence *in-situ* hybridization assay, RASER-FISH, which utilizes UV-induced DNA strand breakage and exonuclease activity to expose genomic DNA to hybridization probes, thus reducing the ultra-structural damage caused by traditional heat denaturation methods (Brown et al. 2018). In LDB1 wildtype cells, after restricting the image to the planes in which the *Mnx1* BAC probe was visible, there was a clear spatial proximity of Isl1-FLAG signal to the probe, which seemed to be clearly above background anti-FLAG staining levels **(Figure 5.6)**.

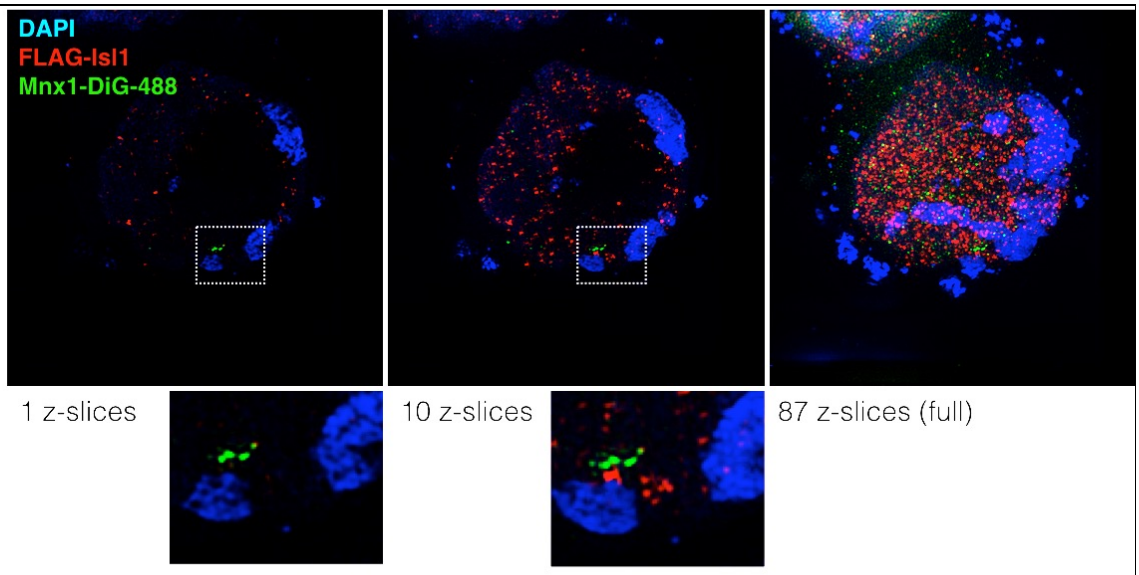


Figure 5.6 – Association of Mnx1 locus and Isl1-FLAG signal in LDB1 +/+ cells via RASER-FISH
Reconstructed, channel-registered 3D-SIM image of a single nucleus in an induced ESC, treated with RASER-FISH for a 200-kb BAC targeting the *Mnx1* locus, in greed, and Isl1-FLAG staining red, and DAPI-stained DNA in blue. Each panel shows, from left to right, a single slice, a maximum intensity projection of 10 z-slices, or the entire image projected. Insets of similar intensity projections are shown below at the site of the *Mnx1* BAC signal. Image has been manually scaled for contrast.

While we did not establish the percentage of cells with *Mnx1* signal or a statistical measure of association with Isl1-FLAG, and we did not extend this analysis to LDB1 knockout cells or other candidate MNC complex components, this assay provides additional evidence of spatially organized MNC-target locus conformations in individual nuclei, which has not been shown previously.

5.4– LDB1 ASSOCIATES WITH LHX3-HA IN CLEAR, POSSIBLY MULTIMERIC FOCI

We also wondered whether LDB1 was nucleating LIM-HD TFs within the nuclear space. While this was not studied extensively, and there are many alternative and advantageous strategies for observing TF clustering, particularly in live cells (Karbassi et al. 2019), we see a clear pattern of LDB1 foci associating with bright, possibly multimeric foci of Lhx3-HA **(Figure 5.7)**.

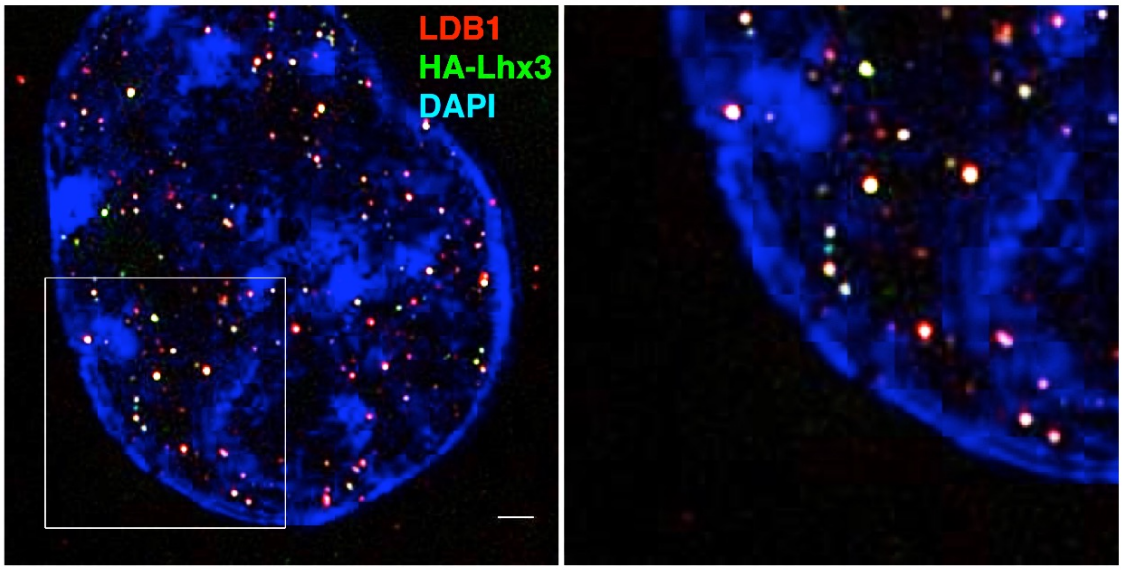


Figure 5.7 – LDB1 and Lhx3-HA form tightly associated clusters as imaged with 3D-SIM
Reconstructed, channel-registered 3D-SIM image of a single optical section of a single nucleus in an induced ESC, showing anti-LDB1 staining in red, anti-HA-Lhx3 staining in green, and DAPI-stained DNA in blue. The right panel is the inset indicated by white box in the first panel. Image has been manually scaled for contrast. Scale bar indicates 500 nm.

While the number of Lhx3-HA foci per cell is relatively high, we observe a relatively limited number of LDB1 foci in individual nuclei, and the majority of LDB1 signal is very tightly associated with Lhx3-HA signal in the *Ldb1 +/+* cells. This level of association was less than, but similar to, a co-occupancy / antibody-labeling control, where the same secondary antibody against the HA tag was

labeled in both green and red channels (data not shown). Further, LDB1 and HA-Lhx3 co-staining showed a recurrent pattern of nearby co-occurring clusters (data not shown), which might indicate a recurrent spatial pattern of LDB1-mediated LIM-HD TF clusters.

5.5 – Discussion

Fundamentally, the effort to measure intra-complex distances with nanometer precision and test these distances as a result of LDB1 function did not succeed. Recently, new methods and algorithms for detecting fluorophore distances at the single-molecule level, often with non-super-resolution microscopy techniques, have been developed (Niekamp et al. 2019), and these may be better suited for such an approach. Further, live-cell imaging of LDB1, its cofactors, and other markers may well be utilized to define clusters of transcriptionally active sites, as has been tried with other transcriptional factors (Cho et al. 2018).

However, the existing data, with additional thresholding or statistical approaches, may reveal a significant difference in spatial distance distributions as a result of LDB1. Further, this approach, when refined, could be used to look at clustering and distributions of a wide range of transcription factors and associated proteins, such as chromatin integrator proteins like LDB1. If these proteins are indeed functionally condensing dispersed TFs and cofactors into hubs or at specific loci, then an imaging-based approach measuring complex component distances would provide powerful support of these proposed functions.

CHAPTER 6. INTERACTIONS BETWEEN MNC COMPLEX COMPONENTS

6.1 – LIM-HD TFS DO NOT REQUIRE LDB1 TO INTERACT

Given the role of LDB1 in mediating formation of dimerized TF complexes through its LDB1-dimerization function, we asked whether the LIM-mediated interactions of the TFs LDB1 coordinates would be changed in the absence of LDB1. We observe via co-immunoprecipitation-Western blot of Day 6 EBs induced with doxycycline, in either *Ldb1 +/+* or -/- backgrounds, that association of Isl1-FLAG and Lhx3-HA is independent of LDB1 (**Figure 6.1**).

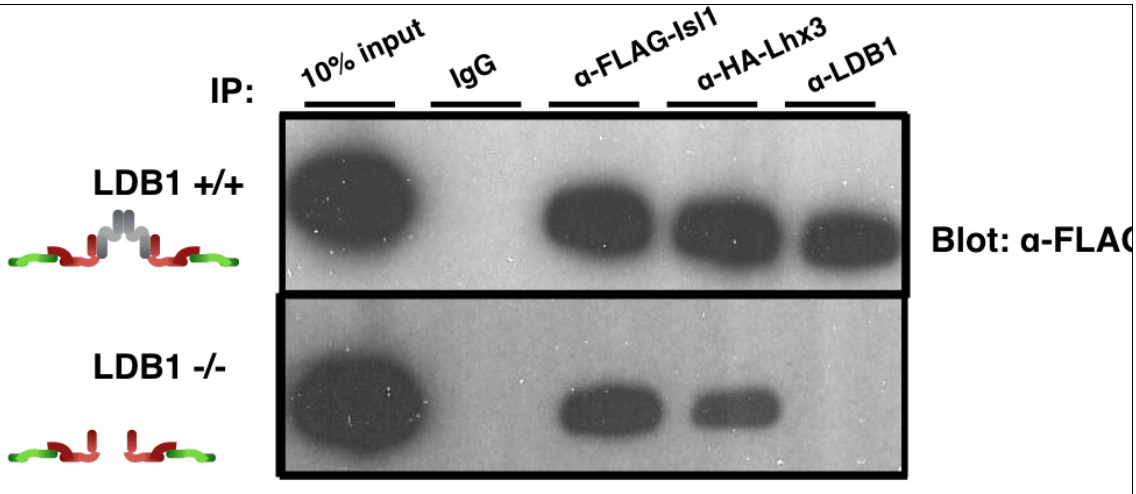


Figure 6.1 – Isl1-FLAG and Lhx3-HA interact in a LDB1-independent manner
Co-immunoprecipitation Western Blot of FLAG-Isl1 interacting with FLAG-Isl1, HA-Lhx3, or LDB1 in *Ldb1* +/+ or *Ldb1* -/- Day 6 EBs. The immunoprecipitating antibody is noted for each lane of the gel, which shows the signal for anti-FLAG-Isl1 for each immunoprecipitation.

This indicates that the interaction of LDB1 with Isl1 via its LIMULI interaction does not regulate the LIM-based interactions that cause Isl1-Lhx3 dimer formation, which have a relatively high protein-binding affinity for each other but a lower affinity for DNA and are transcriptionally inactive (Robertson et al. 2018). The fact that LIM-HD can heterodimerize without LDB1 is already established (Jurata et

al. 1998; Matthews et al. 2008), but we confirm that in the MN-induction system, removal of LDB1 is not affecting the ability of Isl1-FLAG and Lhx3-HA to biochemically interact.

6.2 – LDB1-SSDP1 INTERACTION IS NECESSARY FOR ISL1-FLAG RECRUITMENT TO THE COMPLEX

We also hypothesized that without LDB1 or SSDP1 function, formation of the MNC would be disrupted. We observe that Isl1-FLAG can co-immunoprecipitate with LDB1 and SSDP1 in *Ldb1 +/+*, or LDB1 rescue cells, but not in *Ldb1 -/-* cells, or in LDB1-SSDP1 interaction mutant cells where there is some detection of induced SSDP1-interaction mutant LDB1 with Isl1-FLAG, but SSDP1 is clearly not recruited to the complex **(Fig 6.2)**.

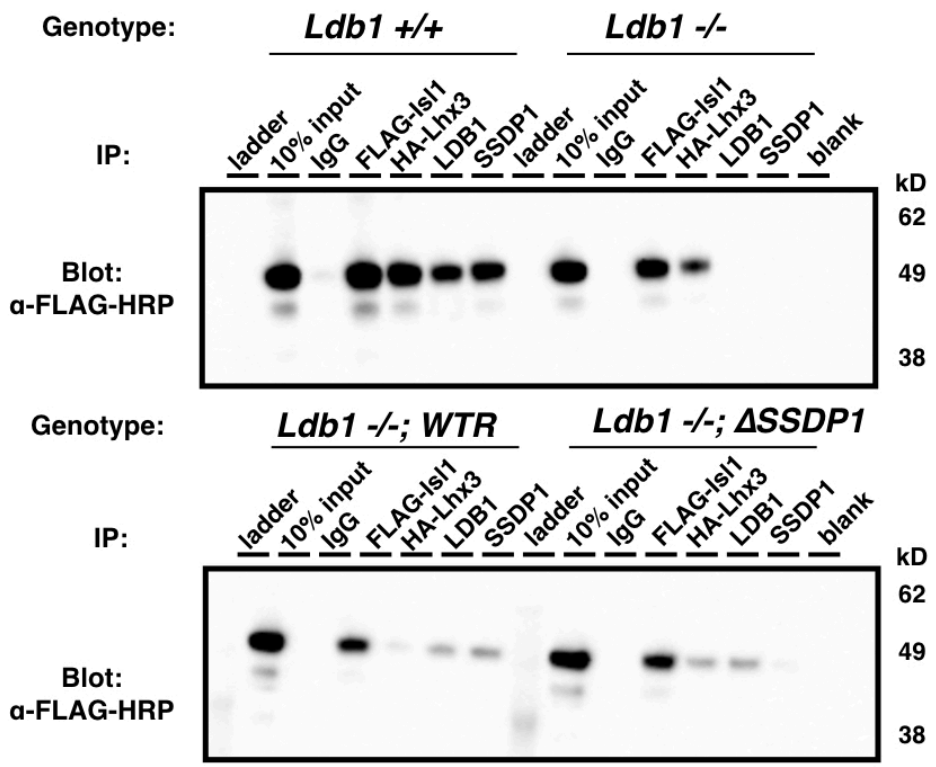


Figure 6.2 – Isl1-FLAG and SSDP1 co-immunoprecipitate in an LDB1-dependent manner
Co-immunoprecipitation of FLAG-Isl1 with various components of the motor neuron complex. Each lane is either 10% of the immunoprecipitated protein or was incubated with the indicated antibody. Samples were then run on a Western blot and probed with an anti-FLAG antibody. Size in kid is indicated on the right.

We conclude that LDB1 is necessary for SSDP1 association with Isl1-FLAG. Additionally, we performed a dual-immunoprecipitation approach using anti-FLAG and anti-HA antibodies, eluted by their respective peptides, and then silver-stained, in induced and un-induced Day 6 EBs both with and without LDB1. In this experiment there are LDB1- and induction-specific bands at a variety of molecular weights, indicating the biochemically verifiable presence of an LDB1-mediated complex containing both Isl1-FLAG and Lhx3-HA **(Figure 6.3**).

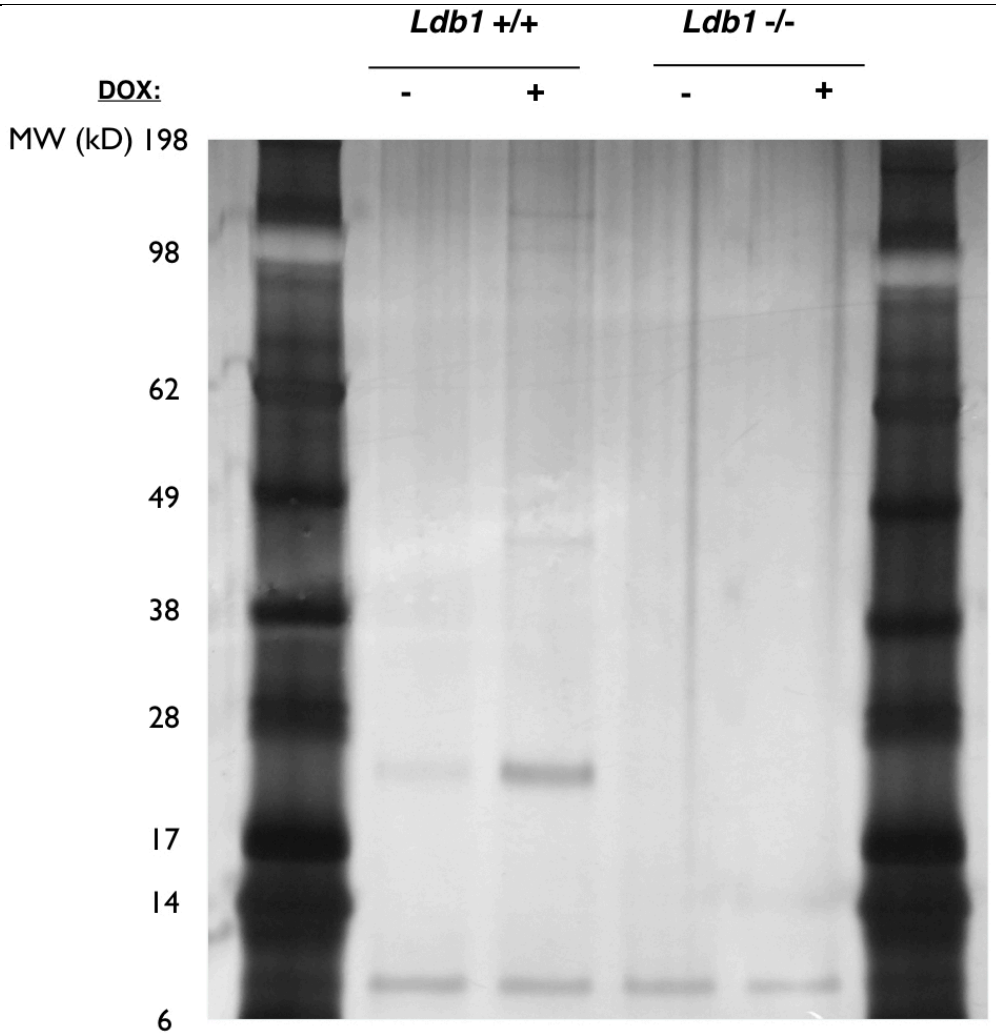


Figure 6.3 – Dual Isl1-FLAG and Lhx3-HA co-immunoprecipitation detects interactions in a LDB1- and doxycycline-dependent manner.
Sequential Isl1-FLAG and Lhx3-HA co-immunoprecipitation from Day 6 EBs with or without doxycycline in *Ldb1 +/+* or -/- cells, as indicated above. The leftmost and rightmost lanes are ladders. Eluate was run on a 4-12% Bis-Tris gel, and then subjected to a silver stain. Note the presence of bands at multiple molecular weights only in the presence of both LDB1 and doxycycline-induced LIM-HD TFs.

Future experiments to determine the components of the MNC, particularly by IP-mass spectroscopy, would be promising, although complicated by the

likelihood that in the absence of LDB1 no complex beyond LIM-mediated Isl1-Lhx3 interactions would be purified.

CHAPTER 7. LDB1 REGULATES GENOMIC ORGANIZATION AT A SELECT NUMBER OF REGIONS

7.1 – LDB1 DELETION REGULATES 3D GENOMIC ORGANIZATION WITH LIM-HD TF AND H3K27AC OCCUPANCY AND GENE EXPRESSION AT SOME MN-INDUCED GENES

We hypothesize that LDB1 mediates long-range interactions to establish MN specific gene regulation at a subset of genomic regions. To test this hypothesis, we isolated genome-wide contacts in both *Ldb1* +/+ and *Ldb1* -/- doxycycline-induced Day 6 EBs with an *in situ* Hi-C protocol (Rao et al. 2014). Briefly, in Hi-C the genome is digested by restriction enzymes in intact nuclei, after which any cut ends are biotinylated ligated together, creating a library of biotin-labeled chimeric fragments from spatially associated DNA sequences. When paired-end sequenced at a high depth, the pair frequencies generate a map of contact probability in a genome-wide, all-vs-all manner **(Figure 7.1A)**. We also detected LDB1- or H3K27ac-associated genomic interactions with the Hi-ChIP assay (Mumbach et al. 2016) in *Ldb1* +/+ Day 6 EBs. HiChIP, by contrast, utilizes a Hi-C library that is then immunoprecipitated with an antibody against a protein of interest. The protein-associated chimeric fragments are then amplified off of beads with tagmentation and processed similarly to Hi-C **(Figure 7.1 B)**. When sequenced, the contact probability allows a genome-wide, all-vs-all approach to 3D genome contacts, but targeted to regions that also interact with the protein of interest.

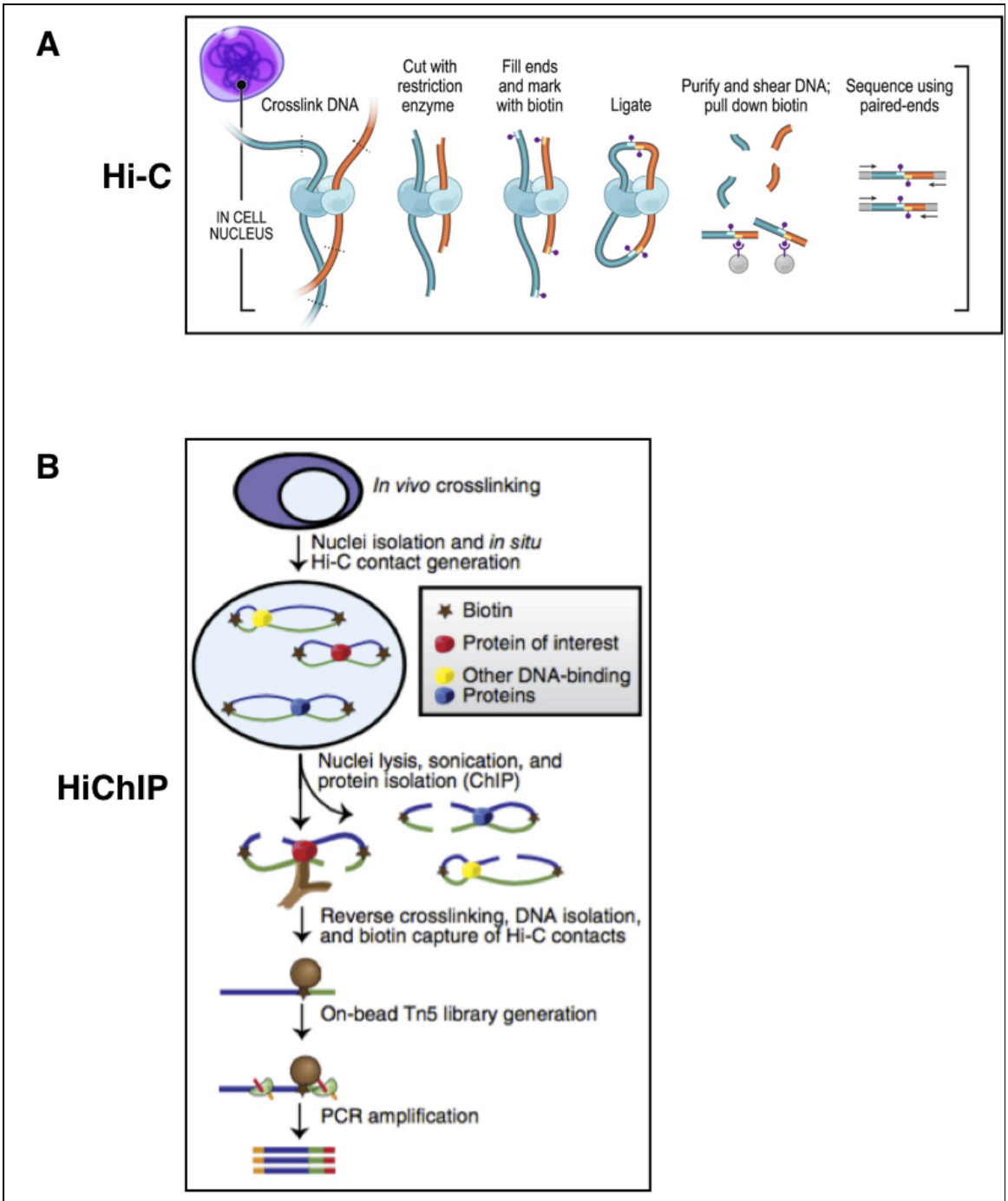


Figure 7.1 – Diagrammatic workflow of Hi-C and HiChIP assays
A) Scheme for generation of Hi-C contacts genome-wide. **B)** Scheme for generation of HiChIP contacts at sites with a protein of interest. Adapted from (Rao et al. 2014; Mumbach et al. 2016).

The resulting graphs should be read as an increase in intensity corresponding to an increase in contact frequency between two points separated along a diagonal (all graphs are mirror images along the diagonal). The strongest contact probability signal will always be along the diagonal, as nearby sequences interact more frequently than distant ones. These assays can detect a variety of genomic structures through contact mapping, such as domains, loops, and stripes at sites of spatial genome organization **(Figure 7.2**).

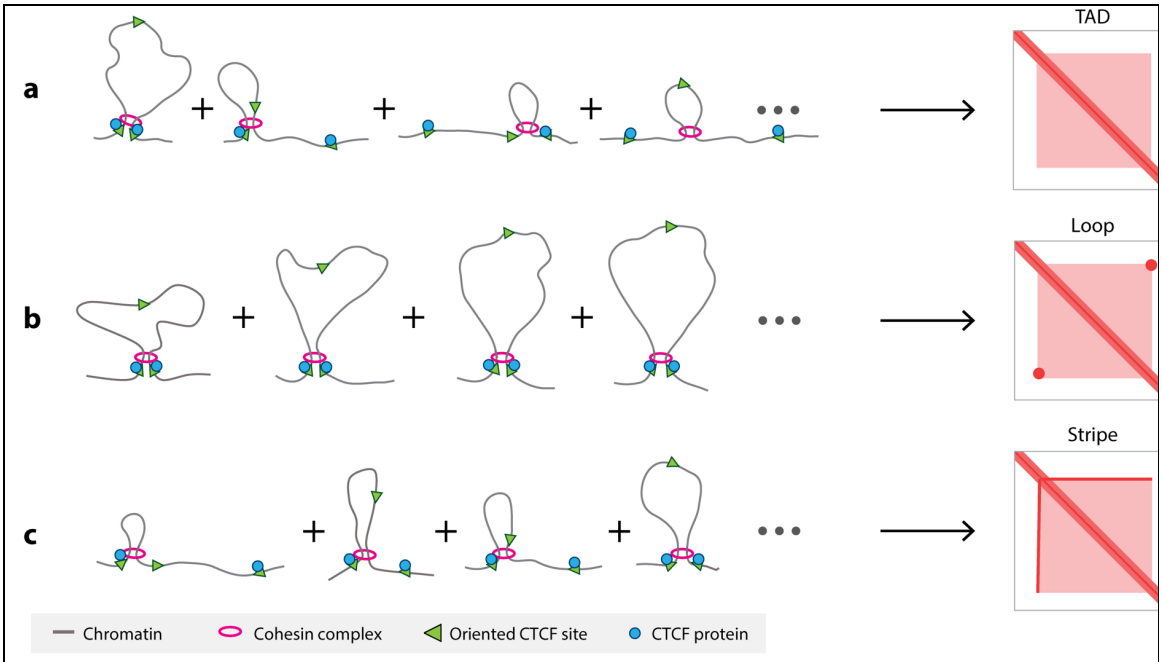


Figure 7.2 – Genomic structures manifested in contact maps
Diagrammatic explanation of how a population of chromatin structures will emerge in a contact probability map. These structures include TADs **(A)**, loops **(B)**, and stripes **(C)**. From (Parmar et al. 2019).

By plotting these datasets together, along with datasets for LDB1, Isl1-FLAG, SSDP1, and H3K27ac ChIP-seq, as well as RNA-seq from the coordinated datasets, we observe a pattern of LDB1- and SSDP1-mediated 3D

genome organization affecting transcription and H3K27ac patterns as well as gene expression at certain distal element or genes.

The first example is the well-studied *Mnx1* locus **(Figure 7.3)**.

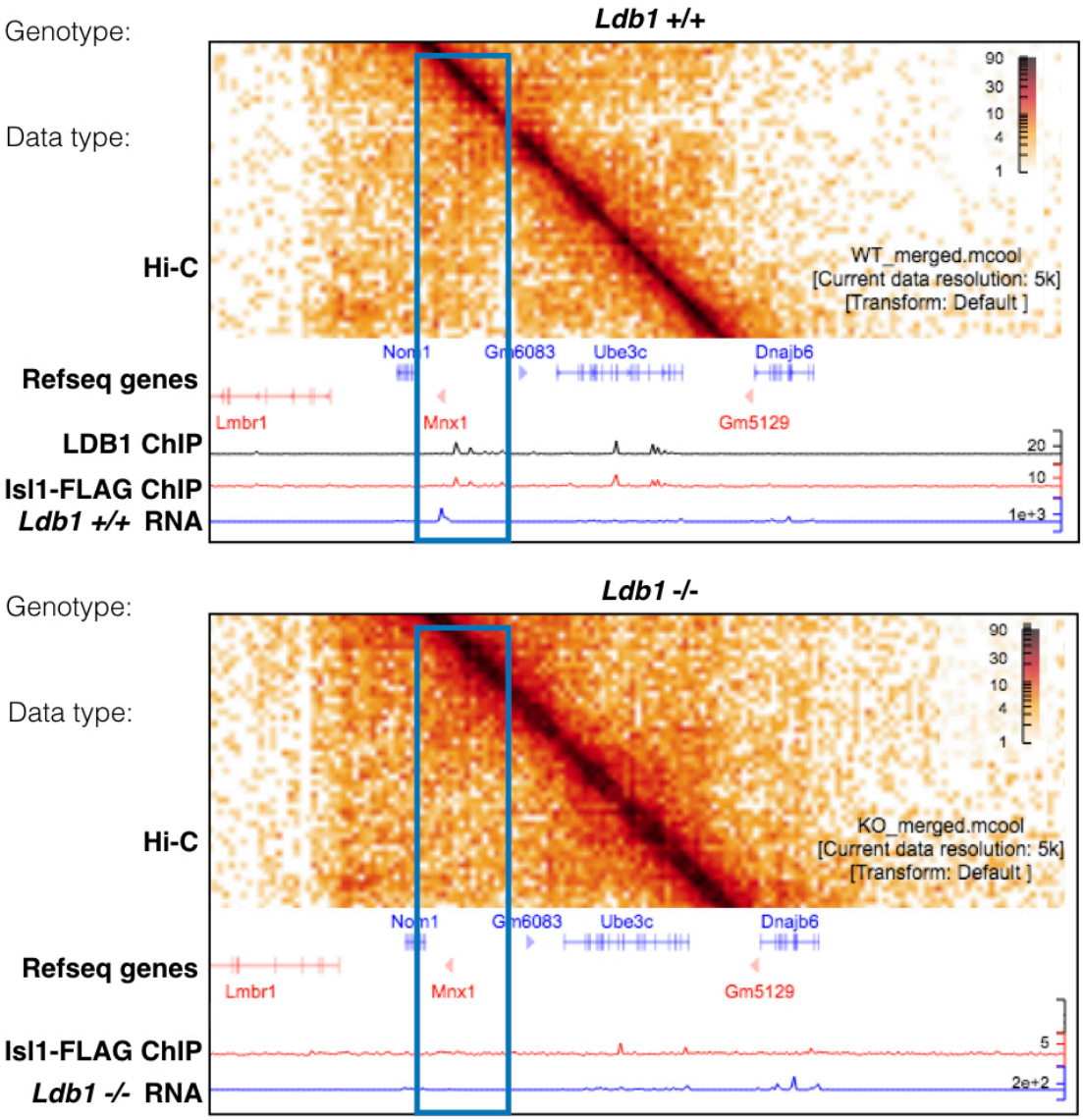


Figure 7.3 – 3D genome organization around the *Mnx1* locus
Hi-C signal in *Ldb1 +/+* **(top panel)** and *Ldb1 -/-* **(bottom panel)** Day 6 doxycycline-induced EBs, with ChIP-seq tracks of LDB1 binding in black, Isl1-FLAG binding in red, and RNA expression in blue for the respective genotype. Refseq gene annotation tracks are below the Hi-C signal, and intensity scales are indicated on the right. Equivalent regions of interest are indicated in blue boxes.

At this resolution we observe no clearly visible difference in large scale genome organization with Hi-C between *Ldb1* genotypes, but do note the existence of an enriched structure at the *Mnx1* locus in particular, coincident with LDB1 and Isl1-FLAG ChIP-seq peaks. By taking a larger-scale view of the *Mnx1* locus **(Figure 7.4)**, we observe the TAD containing *Mnx1*, with clearly defined structures that seem reduced in *Ldb1 -/-* data at the same intensity.

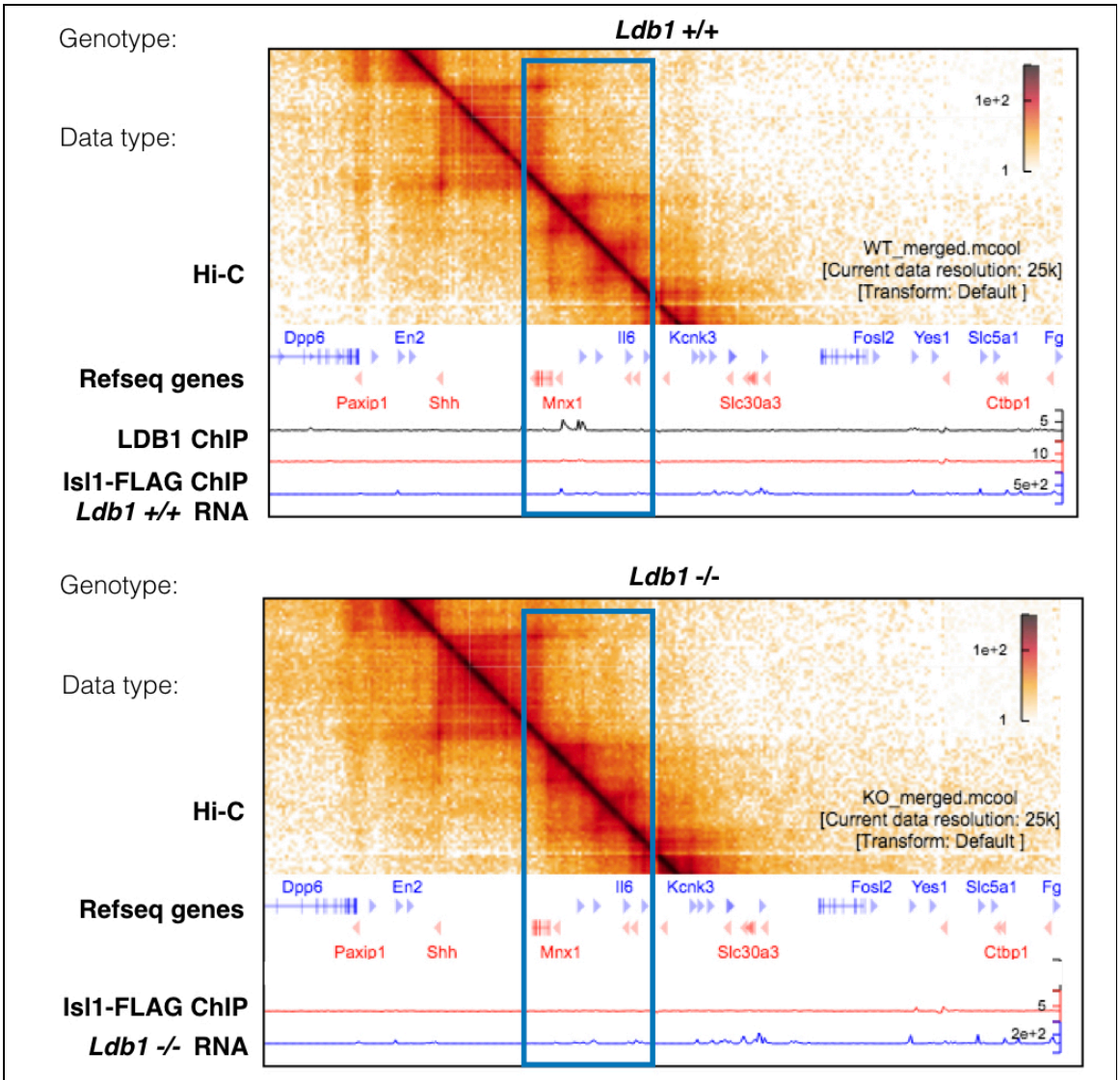


Figure 7.4 – 3D genome organization in the larger domain around *Mnx1*
Hi-C signal in *Ldb1 +/+* **(top panel)** and *Ldb1 -/-* **(bottom panel)** Day 6 doxycycline-induced EBs, with ChIP-seq tracks of LDB1 binding in black, Isl1-FLAG binding in red, and RNA expression in blue for the respective genotype. Refseq gene annotation tracks are below the Hi-C signal, and intensity scales are indicated on the right. Equivalent regions of interest are indicated in blue boxes.

We also observe the well-reported interaction between the *Shh* enhancer and its regulatory ZRS element, which is in an intron of the *Lmbr* gene adjacent to *Mnx1*. Further, the *Shh* locus, which is just 5 Mb distal from *Mnx1*, shows a distinct chromatin topology, characterised by long-range interactions with the

ZRS enhancer (Williamson et al. 2019). This is likely pre-formed early in neural or ectodermal lineages (Brown et al. 2018; Paliou et al. 2019) and is clearly LDB1-independent, as expected.

At another classical MN gene, *Slit3*, in the absence of LDB1 we observe changes in the larger domain containing the large *Slit3* locus, and the loss of an extension stripe emanating from the LDB1 peak at its putative intronic enhancer element **(Figure 7.5)**.

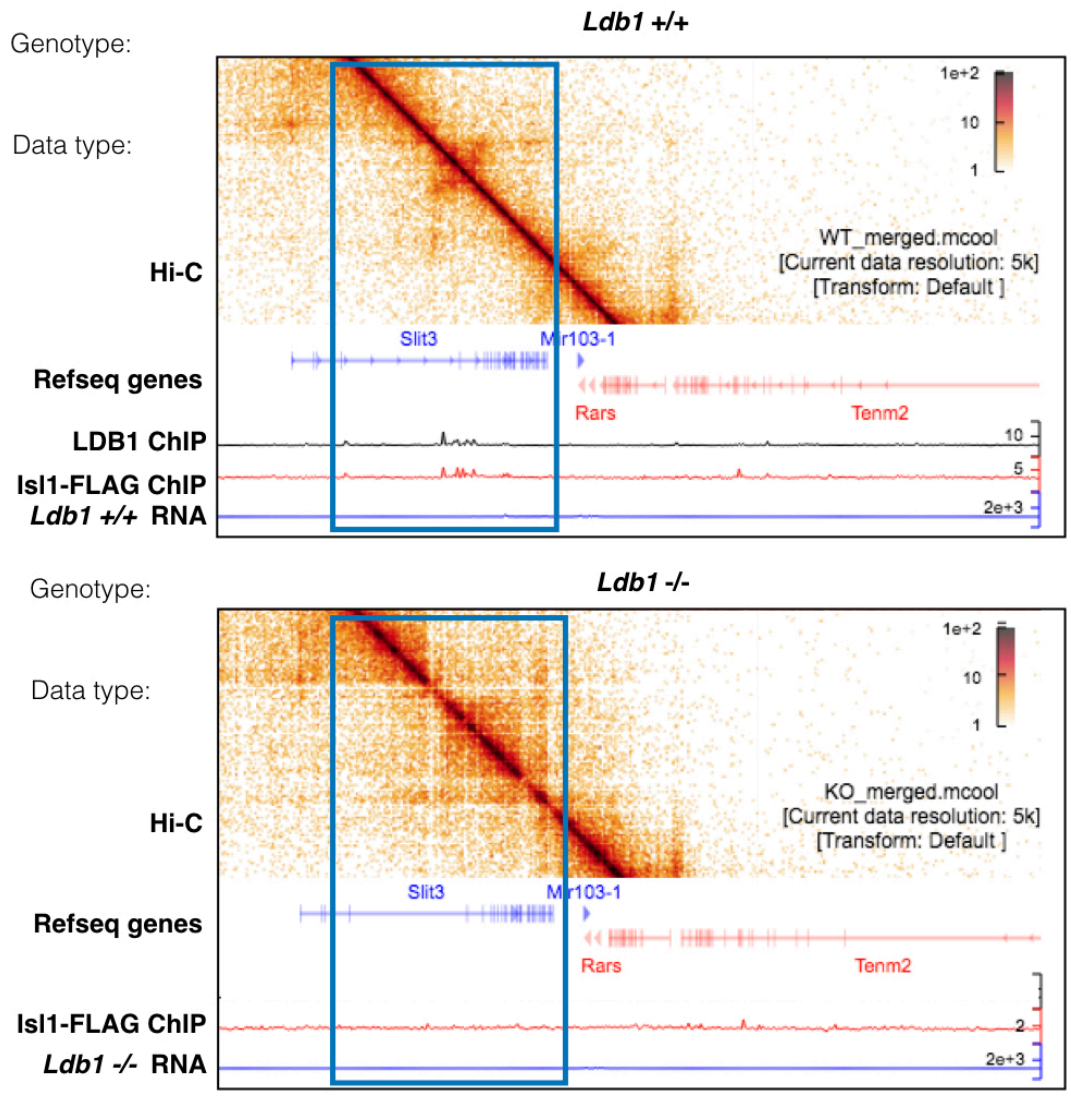


Figure 7.5 - 3D genome organization around the *Slit3* locus
Hi-C signal in *Ldb1 +/+* **(top panel)** and *Ldb1 -/-* **(bottom panel)** Day 6 doxycycline-induced EBs, with ChIP-seq tracks of LDB1 binding in black, Isl1-FLAG binding in red, and RNA expression in blue for the respective genotype. Refseq gene annotation tracks are below the Hi-C signal, and intensity scales are indicated on the right. Equivalent regions of interest are indicated in blue boxes.

These are potentially the stripes characteristic of loop extrusion models in data from other systems (Kragesteen et al. 2018; Kraft et al. 2019), which also appear in the LDB1 and H3K27ac HiChIP data at the *Mnx1* locus (see Chapter 7.2).

At the calcium-related ER transmembrane protein *Tmtc2*, there is a clear loss in the *Ldb1* knockout HiC of a long-range contact present in both LDB1 and

H3K27ac HiChIP, between two boundary elements associated with the TSS or proximal promoter, where there is a distinct LDB1 peak, and an intragenic region that clearly results in a loss of expression of some isoform of the gene in the LDB1 knockout **(Figure 7.6)**.

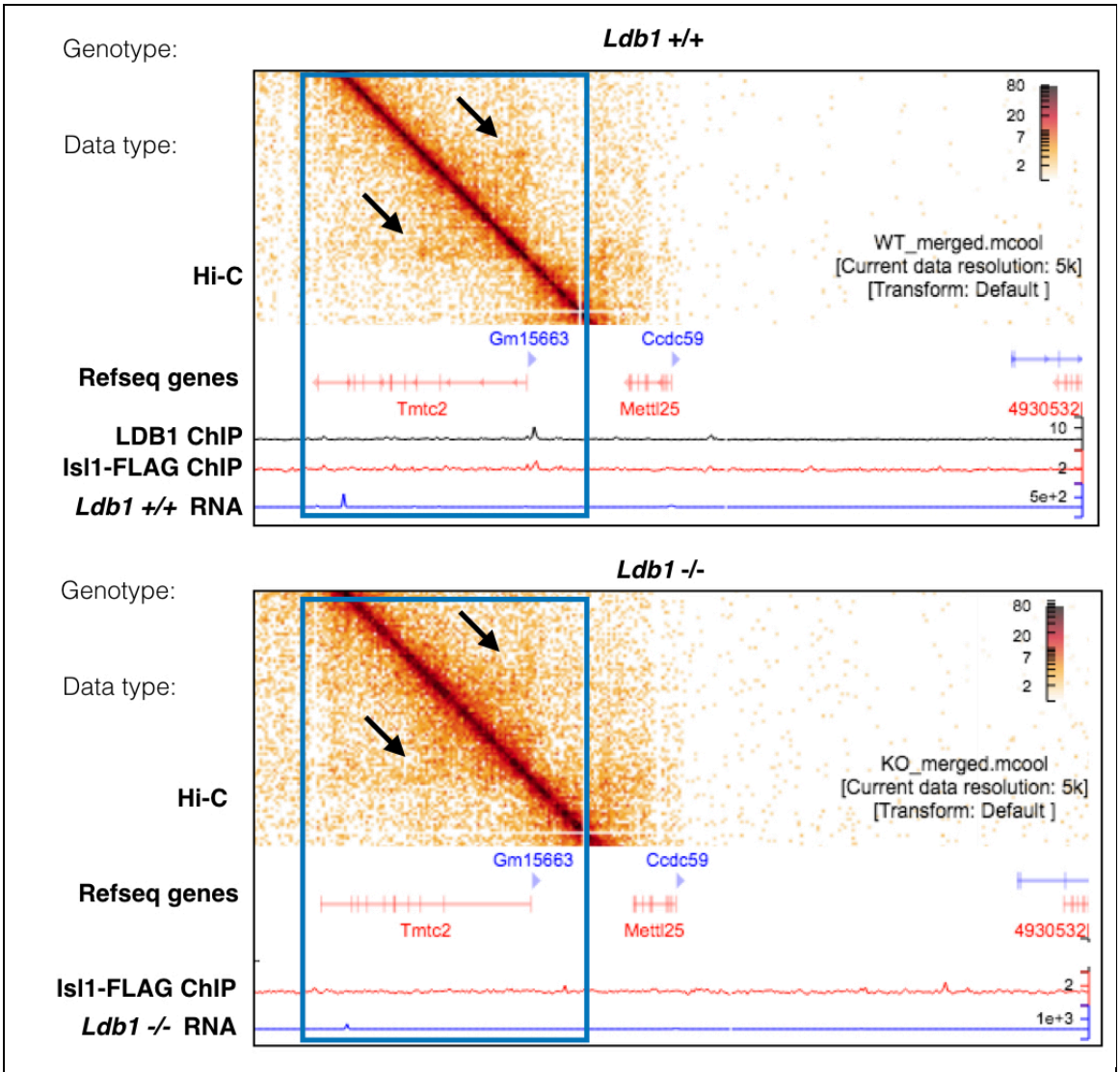


Figure 7.6 – Genome organization around the *Tmtc2* locus
Hi-C signal in *Ldb1 +/+* **(top panel)** and *Ldb1 -/-* **(bottom panel)** Day 6 doxycycline-induced EBs, with ChIP-seq tracks of LDB1 binding in black, Isl1-FLAG binding in red, and RNA expression in blue for the respective genotype. Refseq gene annotation tracks are below the Hi-C signal, and intensity scales are indicated on the right. Equivalent regions of interest are indicated in blue boxes and with arrows.

Both *Slit3* and *Tmtc2* are examples of genes that are coordinately regulated by LDB1, LIM-HD TF induction, and experience changes in H3K27ac occupancy and 3D genome architecture in their immediate genomic vicinity in the absence of LDB1. Some genes, presumably downstream of LDB1/Isl1/Lhx3 activation, do not have observable changes in 3D genome organization in the

absence of LDB1, and further analysis should further delineate these genes depending on their respective combinations of LDB1, LIM-HD TF, and SSDP1 occupancy, differential H3K27ac deposition, and differential gene expression.

Genome-wide, Hi-C signal is reduced surrounding LDB1 ChIP-seq peaks in the absence of LDB1, while the same Hi-C signal is high at sites of H3K27ac ChIP-seq peaks in both *Ldb1 +/+* and *Ldb1 -/-* doxycycline-induced Day 6 EBs **(Figure 7.7).**

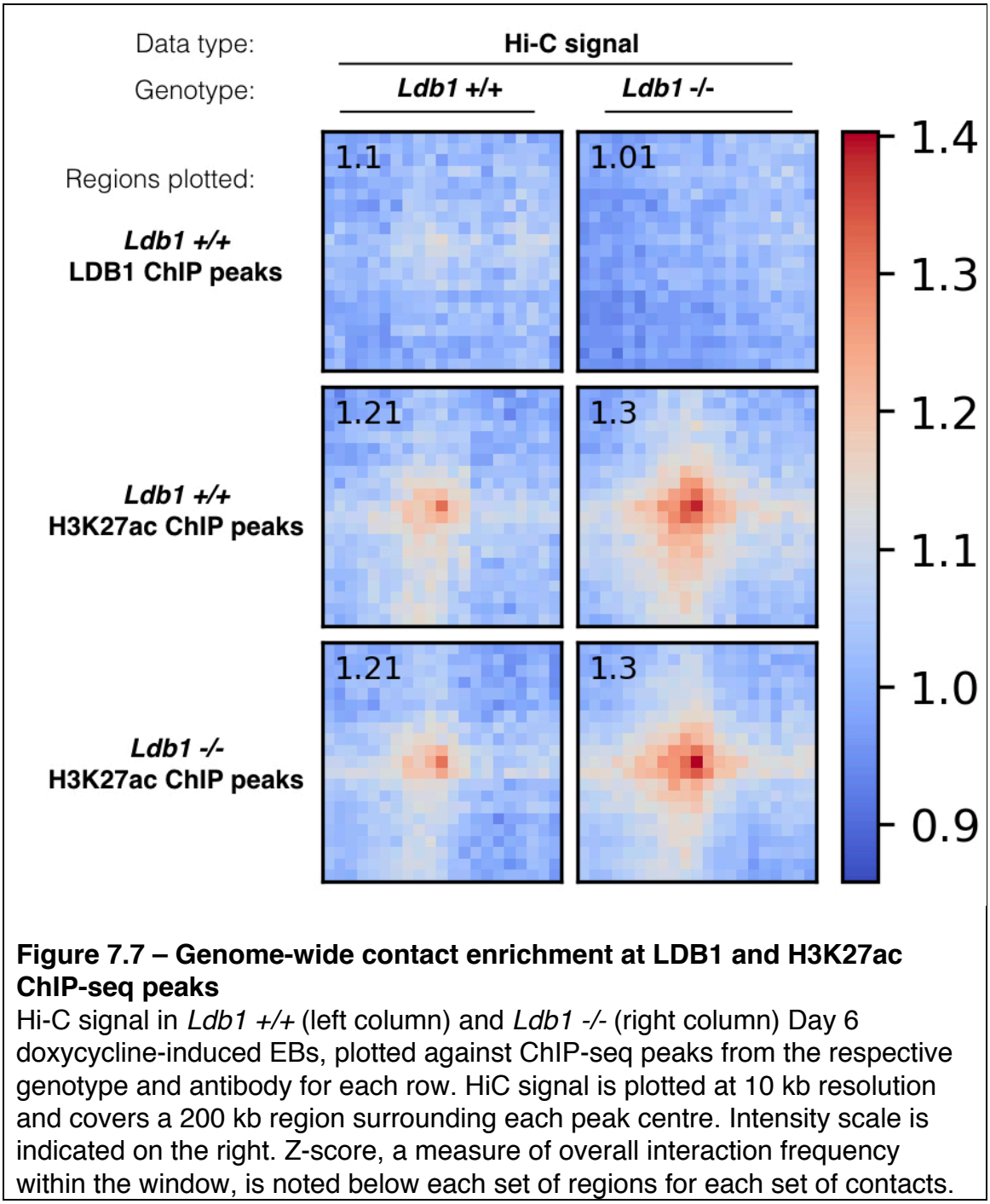


Figure 7.7 – Genome-wide contact enrichment at LDB1 and H3K27ac ChIP-seq peaks
Hi-C signal in *Ldb1 +/+* (left column) and *Ldb1 -/-* (right column) Day 6 doxycycline-induced EBs, plotted against ChIP-seq peaks from the respective genotype and antibody for each row. HiC signal is plotted at 10 kb resolution and covers a 200 kb region surrounding each peak centre. Intensity scale is indicated on the right. Z-score, a measure of overall interaction frequency within the window, is noted below each set of regions for each set of contacts.

While further analysis of these datasets should include motif analysis of differential loops, and further comparison of Hi-C data at subsets of LDB1-bound regions as identified by ChiP-seq, we can determine a subset of LDB1-mediated changes in 3D genome organization at both canonical and novel MN-related genes.

7.2 – LDB1 DELETION REDUCES LOOP FORMATION AT A SUBSET OF LDB1- AND H3K27AC MEDIATED GENOMIC INTERACTION SITES.

To detect a more relevant subset of genome-wide contacts, we performed HiChIP for both LDB1 and H3K27ac, and note both the distinct alignment of HiChIP signal to the respective ChIP-seq tracks, their demarcation of domains and structures also visible in Hi-C libraries, and the similarity of the overall patterns between LDB1 and H3K27ac **(Figure 7.8)**.

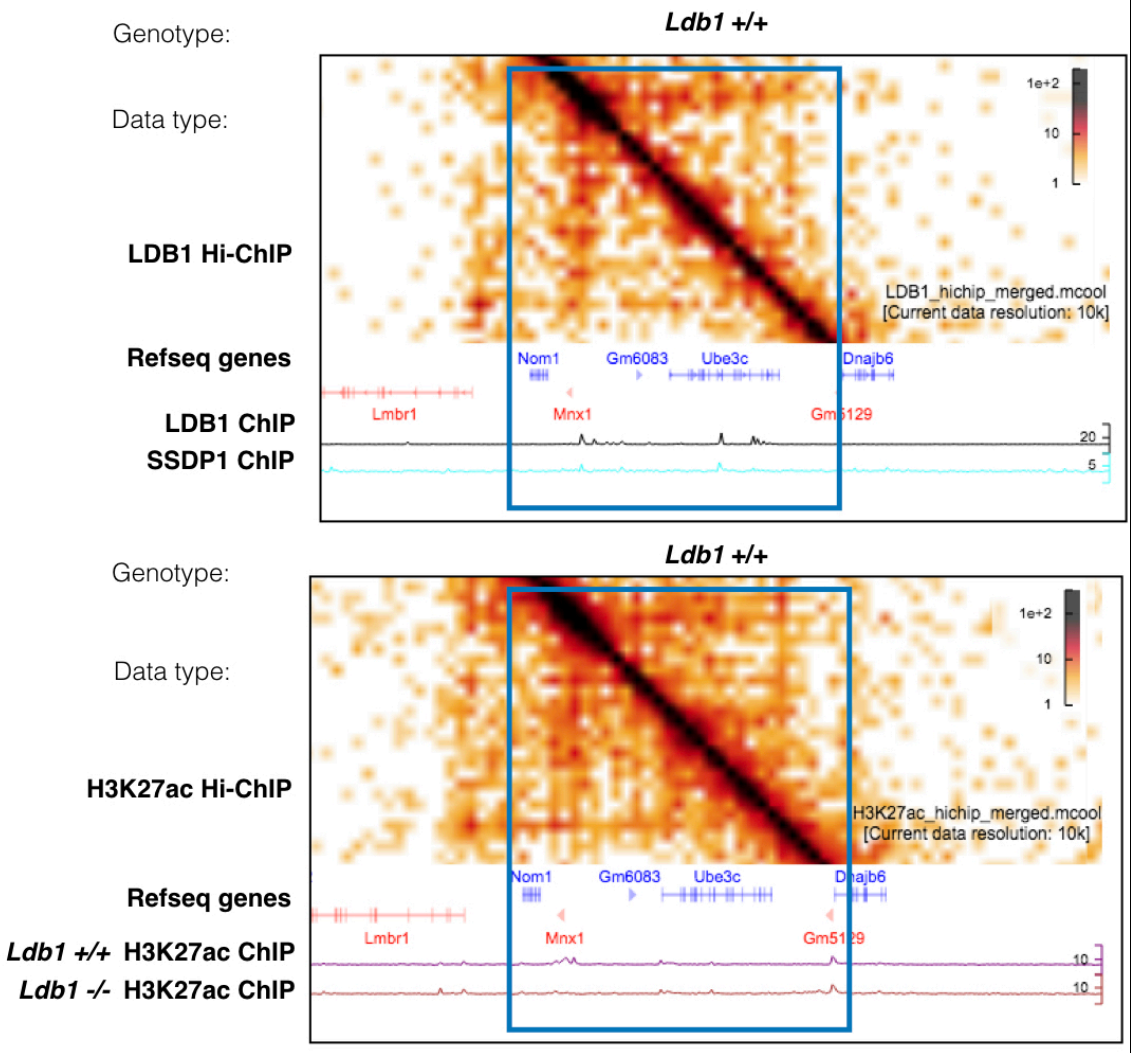


Figure 7.8 –Hi-ChIP of LDB1- and H3K27ac-mediated contacts around the *Mnx1 locus*
Top Panel: HiChIP against LDB1 in *Ldb1 +/+* Day 6 EBs, with ChIP-seq tracks of *Ldb1 +/+* LDB1 binding in black, and *Ldb1 +/+* SSDP1 binding in teal.
Bottom Panel: HiChIP against H3K27ac in *Ldb1 +/+* Day 6 EBs, with ChIP-seq tracks of *Ldb1 +/+* H3K27ac binding in purple, and *Ldb1 -/-* H3K27ac binding in brown. Refseq gene annotation tracks are below the Hi-ChIP signal, and intensity scales are indicated on the right. Equivalent regions of interest are indicated in blue boxes.

We also observe an interesting peak of LDB1, SSDP1, and H3K27ac occupancy that is LDB1-dependent in the *Ube3c* locus, which warrants further investigation.

At other loci, such as at the MN-regulated, LDB1-bound GTPase *Arhgap36* locus, the change in Hi-C determined domain structure is not notably different in the absence of LDB1, but the LDB1 and H3K27ac HiChIP signals show structural arrangements that may indicate the existence of previously unknown *cis*-regulatory elements of these genes (data not shown).

The similarity of LDB1 and H3K27ac HiChIP signals may be due to a genuine overlap in signals, a high degree of similar background, or evidence that LDB1 pulls down H3K27ac-marked regions in the HiChIP assay. But when plotted against genome-wide contact data from Hi-C and HiChIP assays, loops determined by HiChIP are enriched in both Hi-C and HiChIP signal, with a reduction of LDB1-mediated loops in the *Ldb1 -/-* Hi-C signal **(Figure 7.9)**.

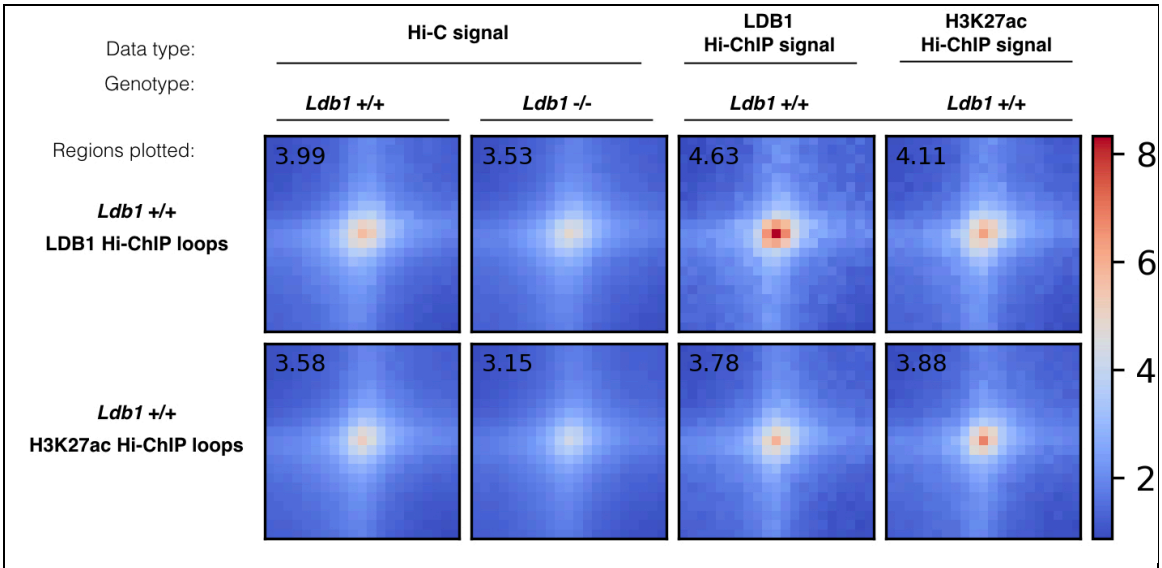


Figure 7.9 – Genome-wide contact enrichment at LDB1 and H3K27ac HiChIP peaks
Hi-C or HiChIP signal from either *Ldb1 +/+* or *Ldb1 -/-* Day 6 EBs (as indicated by genotype) in each column, plotted against regions generated by HICCUPS loop calls on either LDB1 or H3K27ac HiChIP data in *Ldb1 +/+* Day 6 EBs in each row. Hi-C or Hi-ChIP signal is plotted at 10 kb resolution and covers a 200 kb region surrounding the centre of each loop, which have been normalised for distance. Intensity scale is indicated on the right. Z-score, a measure of overall interaction frequency within the window, is noted below each set of regions for each set of contacts.

However, when combined with the orthogonal datasets, together they still help to define the location and nature of LDB1-mediated 3D genome organization. A next clear step would be to compare H3K27ac HiChIP from *Ldb1* knockout cells and compare any changes in H3K27ac HiChIP signal with changes seen at the HiC, ChIP, and gene expression levels. SSDP1 HiChIP would also be highly informative, especially as a control in *Ldb1* knockout cells would be possible (unlike with LDB1 HiChIP itself), but the assay is highly antibody-dependent, particularly with low-peak-number or highly promiscuous TFs and cofactors like SSDP1 and Isl1-FLAG, respectively. Further, we did not specifically look at inter-chromosomal contacts, which in some systems are

importantly regulated by LDB1. We expect to investigate the possibilities of these changes in future analyses.

We also performed an initial analysis of overall HiC contact frequencies across either HiC-derived or LDB1 or H3K27ac HiChIP-derived loop calls on a subset of our data, and see a selective decrease of contact frequency at the HiChIP-derived loops in the LDB1 knockout HiChIP data, whereas HiC-derived LDB1 wildtype loops were not as strongly affected by LDB1 knockout, supporting our hypothesis that LDB1 specifies a particular subset of misregulated contacts but that genome wide much of the 3D architecture is unaffected **(Figure 7.10)**.

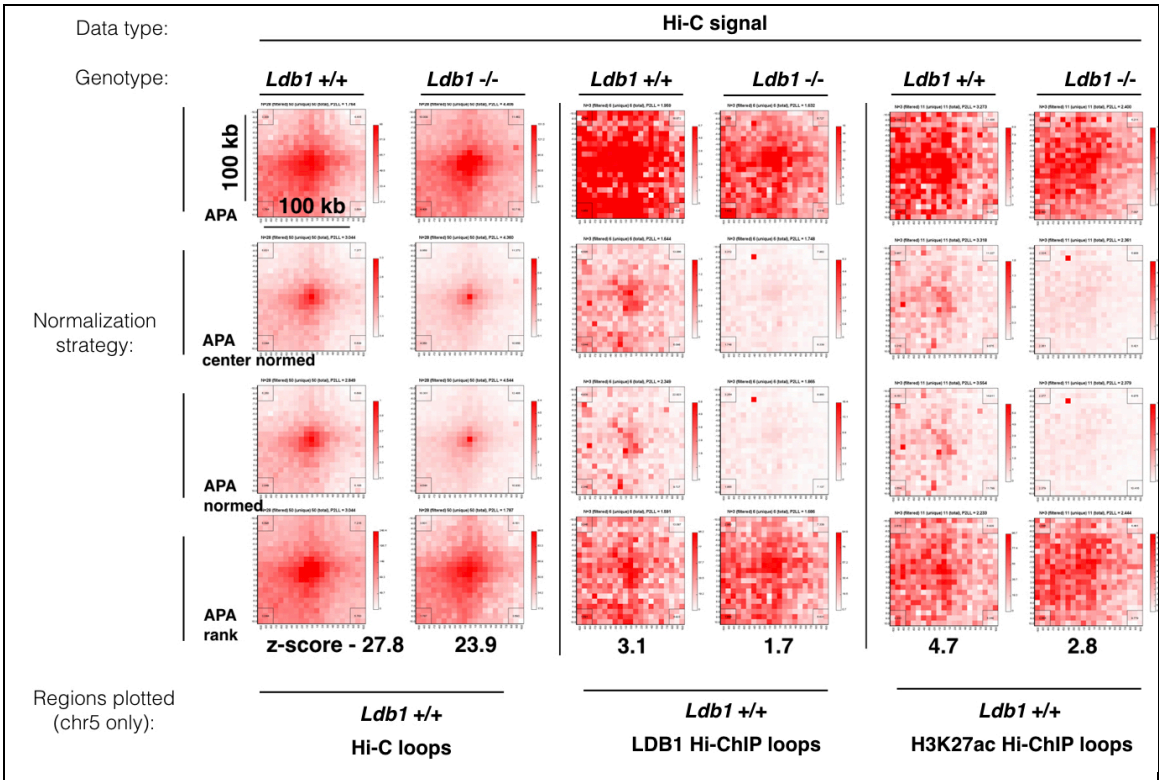


Figure 7.10 – LDB1- and H3K27ac-mediated contacts may be reduced in the absence of LDB1
HiC data from either *Ldb1 +/+* or *Ldb1 -/-* Day 6 EBs, plotted over a sub-selection of regions generated by HICCUPS loop calls on either WT HiC data (left two columns), LDB1 HiChIP loops (centre two columns), or H3K27ac HiChIP loops (right two columns) on chromosome 5, where *Mnx1* is located. Each pixel/bin is a 10-kb region, and window size is 10 bins from the loop in each direction. Each column is a different normalization strategy used by the APA plotting function in the Juicer software. Z-score, a measure of overall interaction frequency within the window, is noted below each set of regions for each set of contacts. Intensity scales are matched within regions.

This analysis only observes loops on chromosome 5, but we are now extending our analysis to loop calls across the entire genome, and comparing differentially called loops between LDB1 genotypes in the HiC experiments to selectively enriched loops in the HiChIP experiments to further delineate the LDB1-responsive elements of 3D genome organization.

7.3 – Design and initial results of Capture-C targeting motor neuron and LDB1-related elements

While genome-level (Hi-C), and protein-specified (Hi-ChIP) contact information is essential to determine the role of LDB1 in spatial genome organization, given the material and resources generally available they cannot give the <5 kb genomic resolution necessary to identify specific contacts, and they cannot have the comparable depth at a particular genomically specified region of interest as a locus-driven approach (Simonis et al. 2007; Davies et al. 2017). As the minimal subset of LDB1- and NIL-regulated genes described above may account for (or be the targets of) the most important regulatory elements in an Isl1-Lhx3-driven MN, LDB1 would thus be the prime target for regulating their systematic association with any distal genomic elements they may regulate or be regulated by. We thus designed Capture-C (Hughes et al. 2013; Oudelaar et al. 2017) probes based on manual curation of LDB1, Isl1, and H3K27ac ChIP-seq, pre-existing knowledge of MN genes, and a variety of control regions. We included regions of known LDB1 binding such as the *Mnx1* ECR regions that are identified as critical MN targets in our data and others. We also included a variety of peaks that are bound most strongly by Isl1, most strongly by LDB1, or cooperatively with Isl1, LDB1, and H3K27ac. We also included a subset of regions to account for AT-rich segments of the genome, and certain open-chromatin regions of the *Myc* gene as a technical control for contact generation. We see that across our regions of interest there is decreased H3K27ac signal in *Ldb1 -/-* Day 6 EBs **(Figure 7.13).**

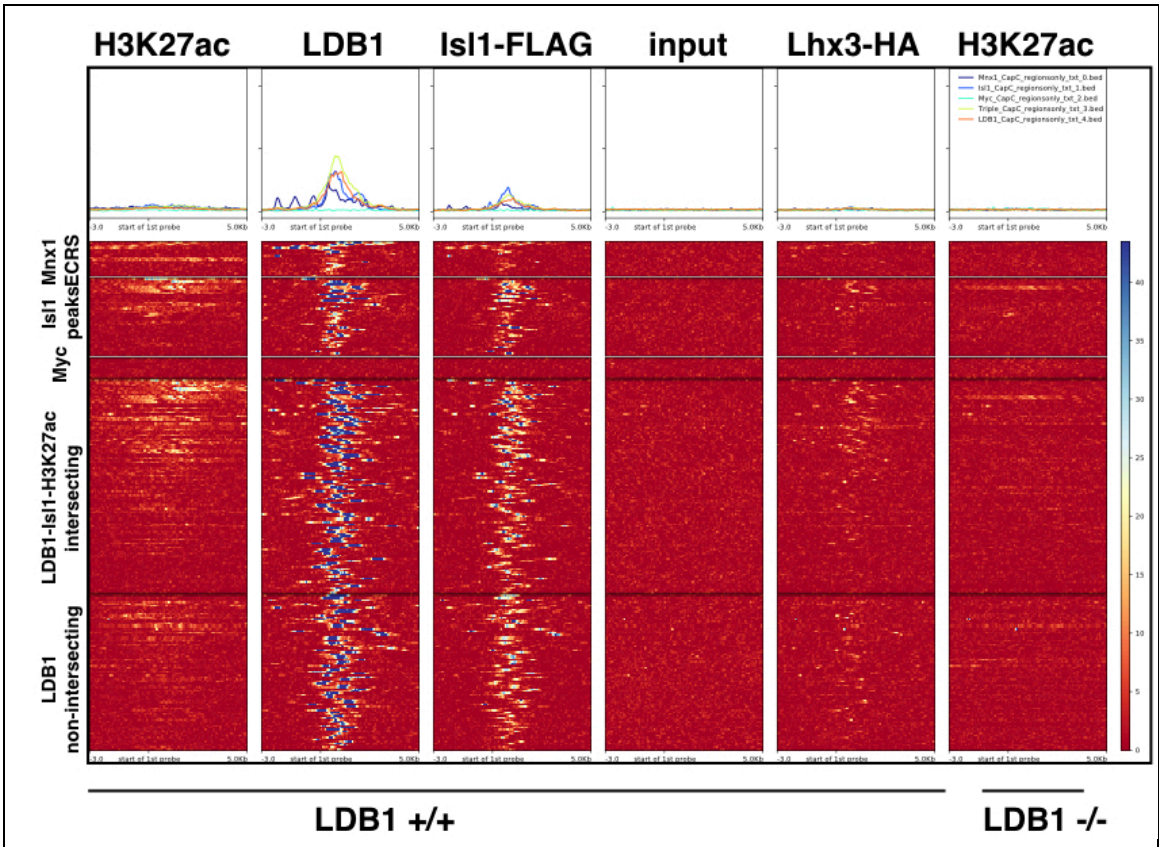


Figure 7.11 – Design of Capture-C strategy for detecting LDB1-regulated and control elements
Heatmaps showing occupancy of H3K27ac, LDB1, FLAG-Isl1, or HA-Lhx3 ChIP-seq signal at regions where Capture-C probes were designed, in either *Ldb1* +/+ or *Ldb1* -/- Day 6 EBs**.** Each set of regions represents up- or down-stream from the centre of a oligo region containing Dpn1 sites, that was selected based on varying criteria. From top down regions are selected as: belonging to the *Mnx1* ECR regions (Mnx1 ECRS); in the top 200 (by corrected P value) Isl1-FLAG peaks that were not LDB1 or H3K27ac peaks (Isl1 peaks); covering open chromatin regions (determined by ATAC-seq) at the Myc locus as a positive control for long-range interactions (Myc); coming from the overlap of LDB1, Isl1, and H3K27ac ChIP-seq datasets (LDB1-Isl1-H3K27ac intersecting); in the top 200 (by corrected P value) LDB1 peaks, but not Isl1 or H3K27ac. ChIP signal for the indicated protein and *Ldb1* genotype is plotted based on average signal strength. Scales for heatmap values are on the right**.**

Importantly, these regions are also enriched for GO terms related to MN development **(Figure 7.14 A),** are enriched for being 50-500 kb away from the nearest gene TSS **(Figure 7.14 B)**, and are also distributed nearly equally away from the nearest genes in either direction **(Figure 7.14 C)**.

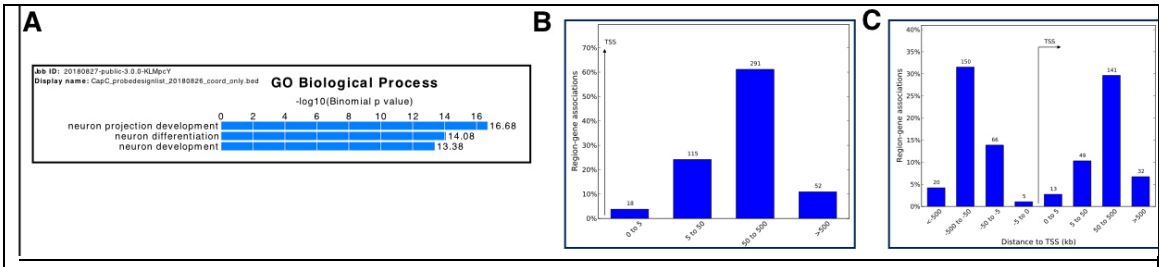


Figure 7.12 – Gene ontology and relative genomic positions of Capture-C probes
Output from GREAT software showing: A) gene ontology; B) distance from region to nearest gene in bins by distance; and C) distance and orientation from regions to the nearest TSS in bins by distance; for the selected Capture-C probe set.

There is substantial evidence that LDB1 is involved in mediating higher-order chromatin architecture and genome organization. We hypothesized that LDB1 would regulate long-range (from the kb- to Mb- range) interactions between promoters, enhancers or regulatory regions of a subset of genes important for MN specification, in concert with Isl1/Lhx3 binding and H3K27ac deposition.

Our initial results show that In *Ldb1 +/+* cells, the *Mnx1* ECRs interact slightly more with distal regions, relative to non-doxycycline induced EBs of the same genotype. (**Figure 7.15**). Among several capture regions near the *Mnx1* locus, at the canonical *Slit2* locus, and at candidate MN regulators *Zfp68* and *Pcdh9,* there is a notable increase in signal across a 0.2-1.0 Mb window size. In comparison, the *Myc* locus, included as a control for capture efficiency, was not substantially altered.

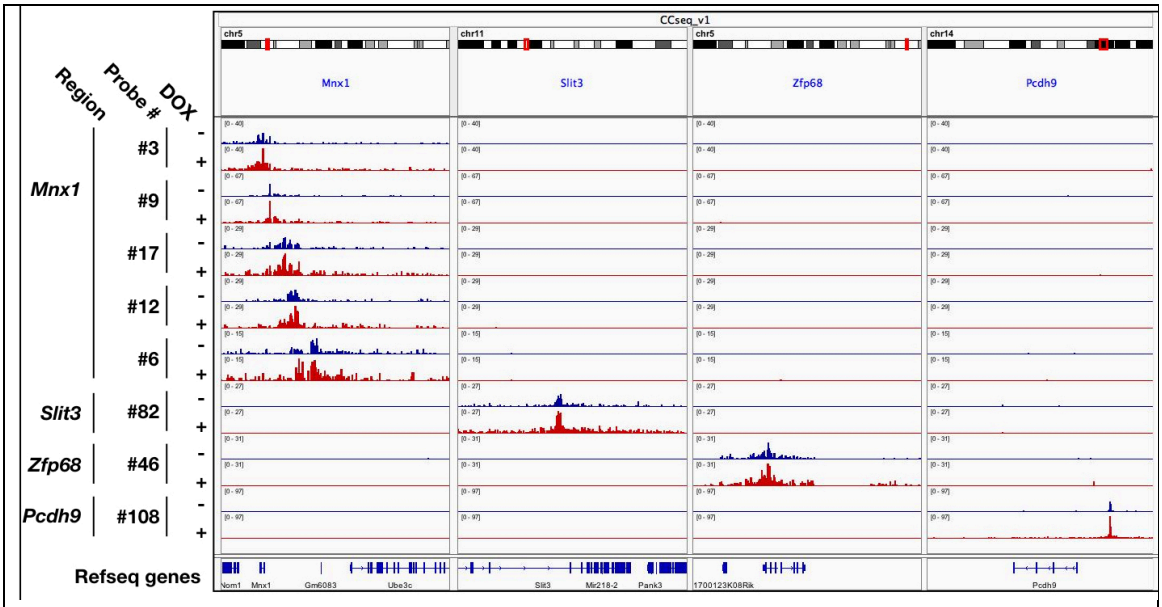


Figure 7.13 – Capture-C of selected regions shows some doxycycline-induced increase in contact frequencies
Raw Capture-C read profiles at *Mnx1, Slit3, Zfp68,* and *Pcdh9* loci in *LDB1 +/+,* doxycycline-induced and doxycycline-uninduced Day 6 EBs. Doxycycline-induced (DOX +) reads are shown in red, unindexed (DOX -) reads are show in blue. Each respective gene and probe ID are indicated on the left. Chromosomal position is indicated on the top. Refseq genes in the area of interest are indicated on bottom. Genomic scales are all in the 250 kb - 1 Mb range. Intensity scales are matched between probe set, and indicated for each track.

While this analysis must be expanded substantially in terms of genotype, to include *Ldb1 -/-* cells, and replicates, such that statistically significant peak calling is a possibility (Oudelaar et al. 2018), we believe this assay will identify a substantial proportion the NIL-induced and LDB1-dependent distal regulatory regions that specify MN identity.

7.4 - Discussion

Taken together, we find support for our hypothesis that LDB1 regulates 3D genomic architecture, with associated changes in chromatin architecture and gene expression, at a subset of specifically MN-induced genes. We find that on the scale of TADs, large scale domains, and overall genomic architecture, LDB1 does not substantially affect genome organization, in line with previous observations (Monahan et al. 2019). However, our first reported LDB1 HiChIP illustrated the detailed domain organization mediated by this chromatin integrator, and its association with similarly structured H3K27ac patterns, that are in some cases altered at the HiC level by LDB1 deletion.

While the direction of CTCF binding does synchronise with larger TAD structures in the HiChIP and HiC data, the differentially regulated H3K27ac / LDB1/Isl1 peaks do not appear to be predominantly coincident with CTCF sites, at least in datasets from other cell types (data not shown). This reinforces the model that while LDB1 does interact with CTCF in some situations (Lee et al. 2017), other LDB1 functions, particularly long-range interactions and transcriptional co-factor recruitment, are CTCF independent (Krivega and Dean 2017). Other architectural proteins, such as SMC1A HiChIP, show hyper-connected cliques of 3D interactions (Petrovic et al. 2019), or with CTCF, illustrate domain-scale loop architectures (Rowley et al. 2017). However, our LDB1 and H3K27ac HiChIP implied a more individualized function for these players, where sub-TADs or specific interactions seemed to be localized at relatively small subset of genetic elements. Therefore, at a subset of LDB1-regulated and DOX-induced genes, as well as at specific LDB1 and SSDP1-

marked putative enhancer elements, there can be subtle but distinctive variation in genomic architecture, including TAD fusion, preferential anchor choice, and loss of looping structures. When the Hi-C signal on regions centred on a subset of loop calls were aggregated there was a loss of signal representing genomic interactions at these regions in the *Ldb1 -/-* EBs relative to wildtype EBs, although overall loops were relatively unaffected, showing how HiChIP data can narrow down the potentially large list of relevant interactions for further study. Additionally, we implement Capture-C in our system and show a subjective increase in contact with distal elements at some MN-related loci upon NIL induction in LDB1 wildtype cells.

Taken together, this provides evidence for a dependence on LDB1 for long-range interactions at the promoter and enhancer elements of a select number of MN-indicative genes. To determine the functional consequence of these elements, they would need to be removed via genome editing, and the same 3D genome organization, ChIP-seq, and RNA-seq assays performed in the new cell line. This would be important future work once a more refined list of these potential elements is generated. However, in coordination with the previous experiments, this defines the effect of LDB1 on three-dimensional genome organization during MN development, and establishes a minimal set of genomic elements that respond to LDB1's transcriptional and chromatin-remodeling activities, in concert with the LIM-HD TFs Isl1 and Lhx3, and the LDB1-regulator SSDP1.

CHAPTER 8. DISCUSSION

8.1 – FUNCTION OF LDB1

In this study, we show that LDB1, with its partner SSDP1, is required for Isl1/Lhx3 complex binding to a set of MN genes and subsequent transcriptional activation, including the master MN regulator Mnx1. We also show that LDB1 mediates H3K27ac-demarcated long-range chromatin contacts at a subset of LDB1-regulated MN genes. We demonstrate that LDB1 is a critical factor in establishing a MN-specific chromatin state and 3D genome organization, and map the MN 3D genome in mice for the first time.

The results show that LIM-HD TF-driven MN induction results in a rapid up-regulation of *Mnx1* in an LDB1-dependent manner. We show that Isl1, Lhx3, and LDB1 co-occupy a large number of regions via ChIP-seq, and that Isl1, Lhx3, and H3K27ac occupancy of those regions is LDB1-dependent. By comparing induced, un-induced, and LDB1 knockout embryoid bodies via RNA-seq we determine that MN induction alters a narrow complement of LDB1-dependent genes, while LDB1 knockout itself changes expression of a wide range of genes. We further analyze the distribution and association of Isl1, Lhx3, and LDB1 with 3D structured Illumination microscopy to determine the spatial differences in complex formation, which do not change markedly with LDB1 deletion, nor does Isl1-Lhx3 association via co-immunoprecipitation. Additionally, we perform Hi-C, Hi-ChIP, and Capture-C experiments to determine the extent to which MN-specific 3D genome organization is controlled by the LIM-HD TFs, and how LDB1 knockout affects these changes. Thus, our original model of LDB1 function

(Figure 1.9) can now be adjusted to highlight which outcomes we see reflected in our combined datasets **(Figure 8.1**).

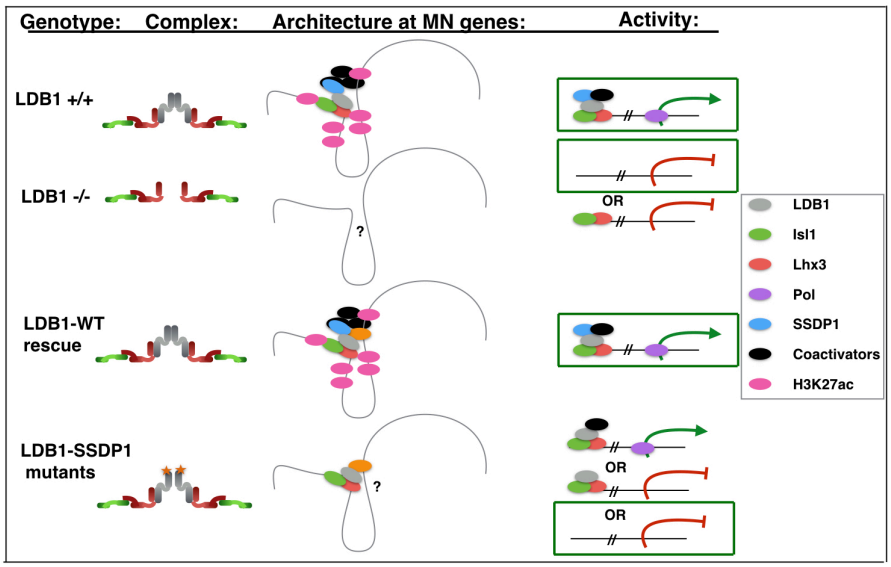


Figure 8.1 – Conclusion: LDB1 regulates genomic architecture and activity at motor neuron genes
Schematic representation of our conclusions for the effects of LDB1 on MNC complex binding, genome organization, and gene expression at motor neuron genes. Architecture at some motor neuron genes is disrupted in LDB1-dependent manners, but global domain organization is not substantially affected and our assays are not necessarily at the resolution to provide detailed looping topologies, so architecture in *Ldb1 -/- and* SSDP1-interaction mutants remains somewhat unanswered. Some possibilities raised in Figure 1.9 concerning LIM-HD TF binding and gene activity have been highlighted as confirmed by our data, in green boxes.

Binding of Isl1-FLAG, Lhx3-HA, SSDP1, and open chromatin at LDB1 sites is ablated in the absence of LDB1. We show also that LDB1, and importantly, its binding partner SSDP1 delineate sites of H3K27ac deposition, open chromatin, and in many cases 3D genome reorganization. However, a large fraction of Isl1-FLAG sites occupy sites of H3K27ac regardless of LDB1 genotype, just as the majority of H3K27ac peaks are unaffected by LDB1 deletion.

The *Ldb1 -/-* Day 6 EBs also yielded consistently low ChIP yields of the LIM-HD TFs, indicating that cross-linking efficiency or overall DNA-binding of these TFs is reduced in the absence of LDB1. ChIP-seq for Ngan or Ngn2-V5 yielded poor ChIP results, likely due to the more transient nature of Ngan binding at “early motor neuron” promoter-proximal E-box elements (Velasco et al. 2017). Taken together, these results show that LDB1 regulates Isl1-FLAG and Lhx3-HA binding genome-wide, that all three factors co-occupy a large number of peaks in induced MN’s, and that in the absence of LDB1, LIM-HD TFs are not able to bind to their tissue-specific regulatory targets.

Further, these ESCs were grown in serum-containing media, and the difference between 2i and serum-containing media has been shown to have serious implications for our understanding of “pluripotency” (Schlesinger and Meshorer 2019), which is critical to any downstream differentiation assay. We did not rigorously test many of these assays in EBs from 2i-grown ESCs, but in assays of Mnx1::GFP reporter activation there was no significant difference (data not shown). Thus, the exact type of pluripotency is perhaps less important than the specific subset of terminal differentiation genes being prompted.

In kind, while thousands of genes are regulated by general neuronal differentiation, transcription factor induction, or LDB1 deletion, only a subset of dozens are both Isl1-Lhx3-Ngn2 responsive and LDB1 responsive. These constitute several known key regulators of MN fate, as well as several other new candidates.

While some of these genes are clearly regulated by Isl1/ LDB1 binding nearby, there is another relatively small subset of regions, likely the LDB1/H3K27ac

dependent peaks described above, that display differential 3D genome organization in an LDB1-dependent manner. Additionally, this is the first demonstrated ChIP-seq dataset for the SSDP1 protein, and it appears to be an even more specific identifier of these genes among the large number of possible homeodomain or dual-homeodomain sites.

This evidence is in line with elegant biophysical studies showing that decreased off-rate kinetics of the Isl1/Lhx3/LDB1 complex is essential for binding to the traditional tetrameric response element. If SSDP1 plays a critical role in protein stability, it is perhaps increasing the potential dwell time and energetic stability of a large pool of Isl1-Lhx3-LDB1 complexes, assembled from LDB1-independent LIM-domain binding between Isl1 and Lhx3, into larger multimeric hubs that catalyze large scale chromatin remodeling and genome architecture changes.

While the original discovery of LDB1 implicated its role in long-range enhancer interactions in the fly, it took two decades for the arrival of chromosome-conformation-capture technology to reveal the specific instances of LDB1-driven chromatin organization. LDB1 has also been used as a “forced looping” protein when fused to a sequence-specific ZFN (Deng et al. 2014). But, as the functions of LDB1 in erythropoietic systems were dissected, some domains are dispensable for dimerization, but essential for target gene activation at the ß-globin locus (Krivega et al. 2014), implying that “bridging” and “looping” functions are separable, as may be expected if stabilization of LIM-HD-TF multimers is performed through LID-mediated binding, while possible

multimerization and/or phase-separation activities are mediated through the IDRs of the LDB1 dimerization domain.

In this study, we characterise the genomic distribution and interactions of transcription factors, the mRNA transcriptome, and the 3D genome organization of induced mouse MNs in the presence and absence of the nuclear adaptor protein LDB1. The collected evidence shows that LDB1 is essential for the formation of a stable TF and cofactor complex that drives H3K27ac deposition and subsequent transcriptional output at a subset of essential MN-related genes, particularly the master regulator of neuron identity, *Mnx1*. LDB1 co-occupies sites of Isl1/Lhx3 binding *in vivo*, and directly affects H3K27ac deposition at MN genes. Mutation of the SSDP1-binding domain of LDB1 results in a genomic phenotype indistinguishable from the straight LDB1 knockout. This establishes a model whereby LDB1 directs formation of a transcriptional activating complex, independently of Isl1-Lhx3 LIM-mediated dimer formation, at the site of Isl1-Lhx3 dimer binding sites, which directs H3K27ac deposition and subsequent transcriptional activation in an indirect SSDP1-dependent manner. Absence of LDB1 in either the knockout, or through likely RLIM-mediated ubiquitin-directed degradation of LDB1 when the SSDP1-binding site is mutated, results in Isl1-Lhx3 dimers that can no longer establish transcriptional activation regardless of whether they bind their correct target sequences.

Further studies should investigate the molecular mechanisms of LDB1 turnover and SSDP1-mediated protection in this system, along with similar functional dissection of the other domains of LDB1. In addition, several other complex components may be present at LDB1-mediated transcriptional hubs at

this set of MN-specific genes, including STAT3 (Lee et al. 2013). The role of associated STAT3- or phosphoSTAT3-bound sites in establishing motor-neuron identity may provide another TF-based specification mechanism for MN identity, and proteomic analysis of the LDB1-containing complex as has been already performed in erythroid development (Meier et al. 2006; Soler et al. 2010) would help to elucidate all the factors in this presumably large multi-protein complex that establishes MN identity. However, the combined data here suggest a critical role for LDB1 in nucleating these sites of active transcription, highlight the importance of chromatin regulatory proteins to conferring specificity on LIM-HD TFs, and suggest a role of LDB1 in mediating these events through the newly discovered mechanisms of phase-transitioned enhancer condensates that mediate transcriptional activity at discrete subnuclear hubs.

8.2 - MECHANISMS OF LDB1 FUNCTION IN MOTOR NEURON SPECIFICATION

The structural basis of the LIM-domain interactions has been well characterised for Lhx3-LDB1 (Bhati et al. 2012), Isl1-LDB1 (Gadd et al. 2012) and Isl1-Lhx3 (Bhati et al. 2008) and a number of other complexes, particularly those containing LMO2. However, LDB1 as a full-length protein will not crystallize under most tested conditions (Gadd et al. 2012; 2013) likely due to its large intrinsically disordered region, and so it has been difficult to assess the role of LDB1 in mediating Isl1-Lhx3 interactions and deterring the structural requirements for dimerization. EMSA assays (Robertson et al. 2018) show that Isl1, Lhx3, and the Isl1-Lhx3 complex can bind DNA in a sequence-specific manner without cofactors, or without a LIM domain at all (Bridwell et al. 2001), and our *in-vivo* data shows that while TF-dimer formation is certainly possible, it is likely that this dimer does not associate with chromatin in a way that can be detected by ChIP, and is transcriptionally inert without its LDB1-mediated interaction with another Isl1-Lhx3 dimer.

Importantly, our results show that the traditional view of Isl1-Lhx3-Ldb1-SSDP1 complexes finding classical hexameric Isl1-Lhx3 DNA motif to target specific MN genes does not explain the distribution of sites at non-classical motifs, neither in terms of motif spacing nor in terms of ChIP-signal occupancy, or 3D-interaction frequency. We propose a new model whereby LDB1 stabilizes formation of Isl1-Lhx3 binding to sites containing AT-rich potential LIM-HD binding sites, but not limited to classical hexamers in the traditional spacing. There is no reason why, for example, two hexamer sites bound by respective Isl1-Lhx3 dimers at two distant regions could not be brought together at one of

these LDB1-mediated clusters. However, evidence that looping between hexamer ‘half-sites’ is possible has not yet been presented.

These complexes are stabilized by SSDP1, because when LDB1-SSDP1 binding is eliminated through mutation of LDB1, deposition of both MN TFs Isl1 and Lhx3, and the activating mark H3K27ac is substantially reduced at critical MN genes. Admittedly, any loops or TADs mediated by LDB1 in terminally differentiated MNs approximated by our Day 6 EB model would be dynamic structures (Hansen et al. 2018), and therefore we may be missing transient, highly dynamic, or low-affinity interactions, perhaps mediated by the “early” or “mid” enhancer elements delineated in alternative models of MN development (Velasco et al. 2017). However, the multiple orthologous techniques used in this study establish the correlative importance and identity of these critical elements.

Assuming that we have identified the LDB1-bound, 3D-interacting set of elements that likely regulate the core ensemble of motor-neuron specific genes, this presents an interesting therapeutic target. As an alternative approach to either induced-TF reprogramming or direct transplantation approaches, perhaps a refined set of elements and their subsequently regulated genes could be activated (or repressed) via a CRISPR/dCas9 activation/inactivation strategy to induce formation of new MMC-like MNs, either as laboratory models of MN-related disease or as a direct intervention.

Throughout these experiments, the effect of the wildtype LDB1-Myc rescue is notably less than the actual wildtype, whether this was in gene expression, LIM-TF binding, or active chromatin marks. This may be due to the inducible cassette system, stochasticity of expression within the population of

cells, or perhaps due to the Myc epitope tag, which is on the N-terminus of LDB1, the side of the protein with the internally disordered regions and dimerization domain. If self-association and potential transcriptional hub formation is a critical feature of LDB1 function, then perhaps the presence of the Myc epitope at that position reduces its ability to form hubs. This could be tested with a C-terminal Myc fusion, which may be useful given the variability of commercially available antibodies for LDB1.

8.3 - Chromatin Integrator Proteins as critical hubs for development and disease intervention

The clustering of heterochromatin, CG islands, telomeres, centromeres, TF hubs, enhancer clusters, etc., has been recently proposed to yield phase-separated domains in the nuclear space (Liu et al. 2018). There is direct evidence that low-complexity-domains (from the FUSS domain but with similarities to the LDB1 IDR), form local “interaction” hubs in high concentrations that regulate transcription, stabilize DNA-binding of TFs in a dynamic binding-partner-specific manner (Chong et al. 2018). Further, the activation domains of canonical TFs such as OCT. can also form phase-separated clusters with Mediator, or “transcriptional condensates” (Boija et al. 2018), and this can also happen at enhancer elements (Shrinivas et al. 2019). These interactions are postulated to occur via “fuzzy binding” of hydrophobic clouds of disordered protein moieties rather than with sequence-specific binding logic (Hahn 2018). This contrasts with LIM domain-mediated binding, but may be applicable to IDR-mediated interactions in the dimerization domains of LDB1. Further, dynamic multifactor hubs are pioneered by DNA binding TFs, such as Zelda in flies (Mir et al. 2018). Live-cell imaging of enhancer clusters (Liu et al. 2014) show that local concentrations of TFs and co-associated proteins, similar to clusters observed in fixed cells with LDB1, can provide an interphase tuning mechanism for gene activation (Mir et al. 2018; Dufourt et al. 2018). Thus, manipulation of these mechanisms, centred on chromatin integrator proteins, may provide a more specific intervention than small-molecule inhibitors of epigenetic enzymes, while also being more flexible than direct gene therapy approaches. Many gene

therapy approaches addressing haemoglobinopathies and sickle cell anemia have benefitted at the research level from induced LDB1 tethering to target specific enhancer-promoter interactions (Breda et al. 2016; Blobel and Deng 2018). At the same time, 3D-genome based diagnostic tools are being patented as diagnostic tools (Blobel and Deng 2018; Hnisz et al. 2016; 2019). A continued focus on chromatin integrator proteins as hubs to both identify and control minimal subsets of therapeutically or developmentally relevant genes should be a priority throughout the biomedical sciences. LDB1 is a prime example of these targets importance because it sits at the intersection of multiple critical informational axes **(Figure 8.2)**.

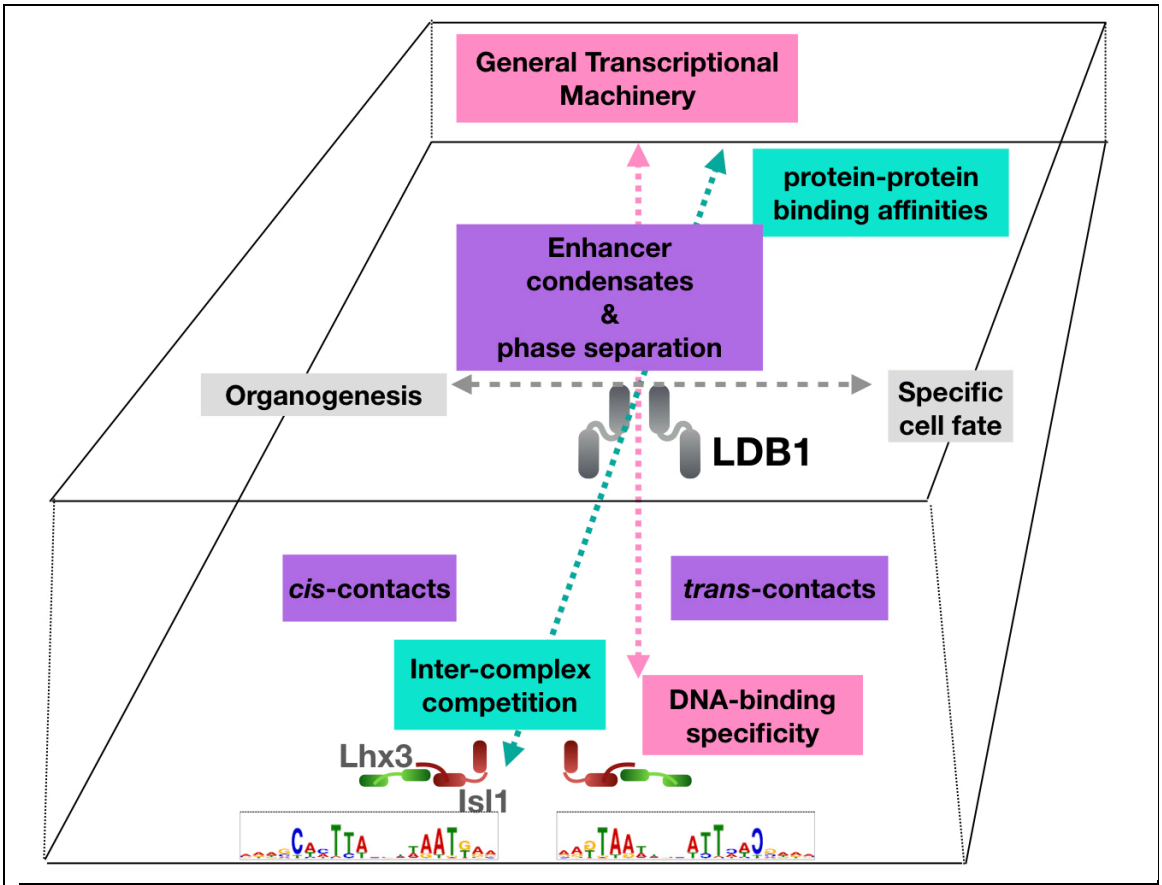


Figure 8.2 – LDB1 is a central hub for multiple multi-scale regulatory mechanisms governing chromatin architecture and development
Theoretical schematic positioning LDB1 at the centre of multiple axes of information. Processes linked through LDB1-mediated interactions are highlighted in similar colours, while general molecular mechanisms LDB1 is either demonstrated or hypothesized to centrally regulate are highlighted in purple. Protein schematic and DNA sequence motif images are for illustrative purposes only.

Fundamentally, all transcription-factor-based cellular reprogramming strategies rely on “chromatin integrator” proteins such as LDB1. Further, many human diseases can be linked to fundamental co-factors of LDB1 in a variety of organisms. As many of these factors are seemingly ubiquitously expressed and required for survival, many TF-based and chemically induced reprogramming strategies will succeed in a cellular (nuclear) milieu that is supportive of normal integrator protein function. Thus, each therapeutically desirable cellular state, or

‘reprogramming target’, is likely dependent on a specific and limited set of integrator proteins. By mapping the effect of these proteins on the transcriptional, epigenetic, and 3D genomic outputs of differentiation, we may be able to map a localized, genetically defined, specific set of target-able elements, that when concomitantly activated or repressed through TF- or otherwise-mediated existing reprogramming/vector-delivery technologies, could dramatically improve both gene editing and cellular reprogramming-based therapeutic outcomes.

8.4 - INTEGRATING GENOMIC AND SPATIAL MAPS

The original goal of this project was to create cross-correlated maps of spatial and genomic positions of all the players in LDB1-driven MN differentiation. However, the challenges of imaging with our antibody-based approach, especially on a multimeric construct, prevented us from constructing a rigorously defined spatial relationship between multiple partners, or between individual partners and genomic loci. Highly sophisticated computational methods for analyzing 3D-SIM data has revealed fundamental aspects of chromatin organization and nuclear structure (Miron et al. 2019), but combining spatially-resolved single-cell data about most macromolecules in the nucleus with the necessary genomic coverage is not yet technically feasible. Cross-modal techniques that can capture the same molecular interactions in the same cells, such as ATAC-see (Chen et al. 2016) could point to a way forward. As the flexibility and resolution of imaging-based approaches to chromatin conformation continue to improve (Bintu et al. 2018; Mateo et al. 2019; Cardozo Gizzi et al. 2019), there may be a way to simultaneously combine and probe information from both spatial and genomic perspectives. Orthogonal techniques for probing nuclear organization (Beagrie et al. 2017; Quinodoz et al. 2018) may also be integrated with imaging-based methods in the future to provide a full spatial, biochemical, and genomic readout of the 3D epigenome. However, this is a fundamental challenge in informational theory, with substantial contributions from soft-matter physics, molecular modeling, and matrix theory needing to be combined, and is thus well outside the scope of the thesis. We have attempted to create a series of datasets where these approaches might be fruitful, but to be

successful more care must be taken in designing uniquely identifiable markers for different subpopulations of molecules, in additional to designating their absolute genomic and spatial positions.

8.5 - Conclusions

The most recent revolution in the understanding of cell biology is perhaps the role of soft-matter phase separation, membranelles organelles, and dynamic local concentrations of macromolecules (Boeynaems et al. 2018). The rapidly expanding field of phase-separation biology and biomolecular condensates is still developing, and has many specific considerations outside the scope of this thesis (Alberti et al. 2019). The phase transition mechanism is approximated as a first order transition driven by increasing concentrations (Narayanan et al. 2019), and is fundamental to determining interphase chromatin organization (Strickfaden et al. 2019). LDB1 certainly increases concentrations of its binding partners at specific genomic loci, although not at a level that would be currently measurable, or perhaps detectable in an induction-based system such as ours. Further, loose networks of chromatin have been shown to congregate around hubs of active RNAPII (Nagashima et al. 2019), in confirmation with long-established theories of transcription factories (Papantonis and Cook 2013). We did not investigate whether these condensations occur with or are dependent on LDB1, but a connection between them is promising. The theme of these observations is that the diverse activities need to be brought together in space, somehow. This provides a framework for studying specific phenomena, and also ties back to the earliest theories of epigenetics.

The metaphor of cellular states as “sinks” or “nodes” in an epigenetic landscape as per Waddington’s original diagrams is still apt. However, the traditional image of an epigenetic landscape leading to an array of defined cell state nodes is only half of the original concept, which connected that landscape to the upper nodes of environmental inputs **(Figure 8.3**).

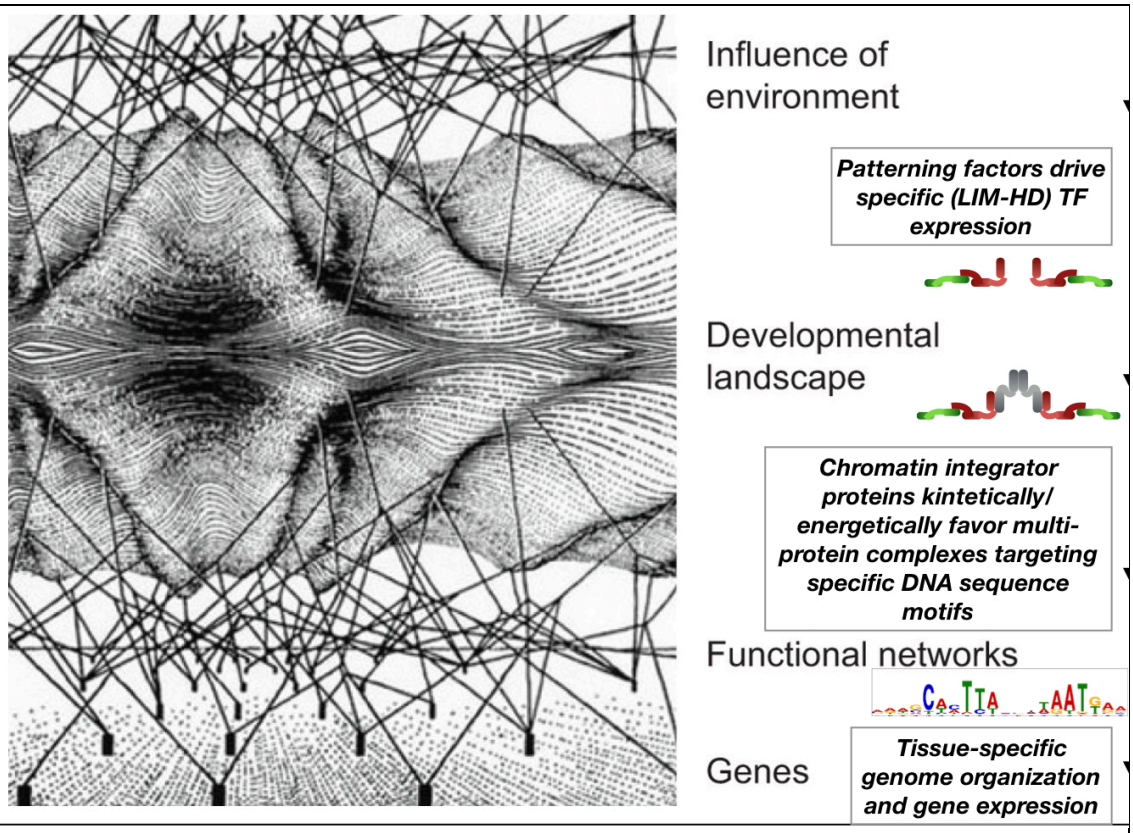


Figure 8.3 – Chromatin integrator proteins condense external signals into gene expression responses
A view of how chromatin integrator proteins shape the epigenetic, chromatin, and developmental landscapes. Figure originally from Waddington (1957).

If the entire space of the nucleus, with all its regulatory elements, is the epigenetic landscape, then perhaps transcription factors (which require a fine balance of specificity and redundancy for both molecular and evolutionary reasons) are the mediators between the environmental and/or cellular nodes and

the landscape, and chromatin integrator proteins such as LDB1 provide the gravity, or the glue, for multiple mediators to work together towards a common cellular fate. This mechanism is entirely independent of the actual regulatory sequences, or the cell state, (with some exceptions for particularly unique biologies requiring specialized mechanisms, like OSNs) as seen by the common re-usage of TFs and chromatin integrators like LDB1, across tissue types within an organism. We should perhaps consider LDB1, with its molecular characteristics, as a core regulator of network modularity, which is a useful concept to approach not only structure-function relationships in the brain, where it was originally developed (Sizemore et al. 2018), but also how the many defined fates of a cell can be derived from the same physical genomic object.

We posit that LDB1-mediated transcriptional hubs may be focal points for tissue-specific 3D genome organization and chromatin remodeling, and that these hubs may constitute the now widely postulated ‘enhancer condensates’ and provide a mechanism linking LIM-HD TF function, enhancer-promoter interactions, and liquid-liquid phase separation in the interchromatin space as critical drivers of development in mammals. We propose that LDB1, and other proteins like it containing a combination of TF-interacting, self-association, and IDR domains, regulate the transposition of genomic information into the folding of the physical genome. By coordinating promiscuous transcription factors at a minimal number of genomic elements and creating higher-order linkages with transcriptional machinery, as well as likely physically positioning these regions nearer the interchromatin space, with its higher concentration of RNA and protein factors. While LDB1 is not known to be a RNA-binding protein, it could well have

interactions with lncRNAs, which are also critical spatial regulators of the nuclear environment (Engreitz et al. 2016). This ‘hub-formation’ is well described in several specific instances, but it is likely a fundamental property through which the physical and genomic information sets are linked, loosely, like a multiple set of gears.

Thus, LDB1 and proteins like it may constitute a class of multi-scale regulators, bridging TF-DNA interactions with co-factor recruitment, enrichment of histone transferases / chromatin remodeling complexes, long- and medium-range *cis*- and *trans*- interactions of 3D genomic organization, and dynamic nuclear structure through formation of phase-separated microdomains. These ‘integrators of genomic architecture’, exemplified by LDB1, should be considered a prime area of theoretical and translational exploration as new techniques in epigenomic mapping, structural biology, nuclear phase separation biology, genome editing, and super-resolution imaging reach maturity.

APPENDIX A: EQUIPMENT AND MATERIALS

Below is a table of reagents used in this study, separated by type, with accompanying manufacturer and catalog information.

Name	Supplier	Cat No:
REAGENTS		
Dynabeads Protein A beads	Thermo	10002D
Dynabeads Protein G beads	Thermo	10004D
Dynabeads MyOne Streptavidin C1 beads	Thermo	65001
dNTP set	NEB	N0446S
M270 Streptavidin Dynabeads	Thermo	65305
Mouse Cot-1 DNA	Invitrogen	18440016
AmPure XP Beads	Agencourt	A63881
Protease Inhibitor Tablets (Complete, EDTA-free)	Roche	11873580001
High Prep PCR Beads	MagBio	AC-60005
KAPA Pure Beads	KAPA / Roche	KK8000 / 07983271001
T1 MyOne Streptavidin Dynabeads	Thermo	65897
CHEMICALS		
N, N-Dimethylformamide	Sigma	D4551-250ML
MES Running Buffer 20X	Thermo	NP0002
Transfer Buffer 50X	Thermo	NP0006
Methanol	JT Baker	JT5370
Ethanol	Fisher	BP2818100
Retinoic Acid	Sigma	R2625
DAPT	Selleckchem	S2215
Doxycycline	Sigma	D9891
SAG	Millipore / Sigma	566660
PD0325901	Selleckchem	S1036
CHIR	Selleckchem	S2924
UltraPure Water	KD Medical	RGF-3410
PCR Grade Water	Thermo	AM9932
DMSO	Sigma	D2650
PIPES	Sigma	P6757
5M NaCL	KD Medical	RGF-3270
2.5M LiCl	Sigma	L9650
1M NaHCO3	Sigma	S5761
3M NaOAc	Corning	46-033-4
0.5M EDTA	Amresco	15575-038
1M Tris-HCl pH 7.4	Amresco	119-069-101
1M Tris-HCl pH 8.0	KD Medical	RGF-3360
1M Tris-HCl pH 9.0	Amresco	E694
1M Sodium Bicarbonate	Sigma	S5761
HEPES	Thermo	15630-080
Glutamax	Thermo	35050-061

B-ME	Thermo	21985-023
MEM NEAA	Thermo	11140-050
mLIF	Millipore	ESG1107
DMEM	Thermo	11995-065
Neurobasal	Thermo	21103-049
Advanced DMEM / F12	Thermo	12634-101
ES FCS	Gemini	100-510
KOSR	Thermo	10828028
EmbryoMax 0.1% Gelatin / UltraPure Water	Millipore	ES-006-B
DPBS (1X)	Thermo	14190-144
TrypLE	Thermo	A12177
Antibiotic-Antimycotic	Thermo	15240
Trypan Blue	BioWhitaker	1794-26
Igepal (CA-630)	Sigma	13021
Triton-X-100	Roche / Sigma	10789704001
SDS	Thermo	28312
Glycine	JT Baker	4059-02
Sodium deoxycholate	Sigma	D6750
BSA solution	Ambion	AM2616
chloroform	MACRON	4440-04
PCI 24:25:1	Roche / Sigma	3117987001
Glycogen	Ambion	9510
LE Agarose	BioLink	11705
ENZYMES		
DpnII	NEB	R0543M
T4 Ligase, High Concentration	Thermo	EL0013
Herculase II Fusion Polymerase	Agilent	600677
RNAse A	Thermo	EN0531
Proteinase K	Roche	3115836001
MboI	NEB	R0147M
Klenow fragment (Large)	NEB	M0210
ANTIBODIES		
Goat polyclonal anti-LDB1	Santa Cruz	sc-11198
Mouse monoclonal anti-LDB1	Abcam	sc-98977
Mouse monoclonal anti-FLAG	Sigma	F1804
Rabbit polyclonal anti-GAPDH	Novus	NB200-322
Rabbit polyclonal anti-HA	Abcam	ab9110
Mouse monoclonal anti-H3K27ac	Active Motif	39685
c-Myc antibody 9E10 X	Santa Cruz	sc-40 X
Rabbit polyclonal SSDP1/SSBP3	Abcam	ab83815
Mouse IgG	Invitrogen	100005292
Rabbit IgG	Invitrogen	100005291
HA peptide	Sigma	32149
FLAG peptide	Sigma	32540

COMMERCIAL ASSAYS		
Clontech SMRT Library Construction Kit	Takara/Clontech	634865
Qubit HS DNA Quantification	Thermo	Q32854
Qubit BR DNA Quantification	Thermo	Q32850
Bioanalyzer DNA 1000 Assay	Agilent	5067-1505
Bioanalyzer High Sensitivity DNA Assay	Agilent	5067-4626
TruSeq RNA Library Construction Kit	Illumina	RS-122-2001
PROTSIL2 Silver Stain Kit	Sigma	PROTSIL2-1KT
RNEasy RNA Prep Kit	Qiagen	74106
Amersham ECL Detection	GE Healthcare	RPN2232
MinElute PCR Purification Kit	Qiagen	28006
Zymo Genomic DNA Clean & Concentrator Kit	Zymo	D4011
MinElute DNA Kit	Qiagen	28006
High Sensitivity DNA Kit	Agilent	5067-4626
Pierce BCA Assay	Thermo	23225
Nextera Library Prep Kit	Illumina	FC-121-1030
NuPage 4-12% Bis-Tris gel	Thermo	NP0321
Nimblegen SeqCap EZ Hybridisation and wash Kit	Roche	5634261001
Nimblegen SeqCap EZ Accessory Kit v2	Roche	06953212001
Nimblegen SeqCap EZ HE-oligo Kit A	Roche	6777287001
Nimblegen SeqCap EZ HE-oligo Kit B	Roche	06777317001
Library Concentration Kit (Universal)	Roche	KK4824 / 07960140001
Kapa Hyperprep Library Construction Kit	Roche	07962312001
NEBNext DNA Library Prep Reagent Set	NEB	E6040S
NEBNext Multiplex Oligos for Illumina Primer set 2	NEB	E7500S
Papain Dissociation System	Worthington	LK003150
PCR Purification Kit	Qiagen	28106
Arima Genomics Hi-C Kit	Arima	N/A
IDT Xgen Lockdown probes Mini pool	IDT	N/A
IDT for Illumina Universal Adaptors	IDT	20020590
EQUIPMENT		
XCell II Blot Module	Invitrogen	090707-132
XCell II Transfer Module	Invitrogen	E10951

microTUBE AFA Fiber Pre-Slit Snap-Cap 6x16mm	Covaris	520045
Nitrocelluse Membranes (0.45 μm)	Whatman	LC2006
Corning 500 mL Erlenmeyer flask - flat / vent cap	Capitol Scientific	431145
Corning 250 mL Erlenmeyer flask - flat / vent cap	Capitol Scientific	431144
Corning 125 mL Erlenmeyer flask - flat / vent cap	Capitol Scientific	431143
Nexcelom Cell Counter	Nexcelom	Cellometer Vision
Microfuge	Tomy	HF-120
Vortex	Daigger	2222OA
Benchtop centrifuge (RT)	Eppendorf	5434
Benchtop Centrifuge (4 deg)	Eppendorf	5417R
Bucket Centrifuge (4 deg)	Eppendorf	5702
10 cm TC dishes	Falcon	08-772E
15 cm TC dishes	Falcon	353025
10 cm Petri dishes	Falcon	351029
15 cm Petri dishes	Falcon	50-114-9058
5 mL Pipettes	Alkali Scientific	SP205
10 mL Pipettes	Alkali Scientific	SP210
25 mL Pipettes	Alkali Scientific	SP225
96-well qPCR plates	Bio-Rad	HSP9601
96-well flat bottom dishes	Falcon	08-772-2C
2.0 mL Protein LoBind Eppendorf tubes	Eppendorf	022431102
1.5 mL DNA LoBind Eppendorf tubes	Eppendorf	022431102
1.5 mL Eppendorf tubes	Axygen	MCT-150-C
0.5 mL PCR strip tubes	Genemate	490004
15 mL TPX sonication tubes	Diagenode	C30010009
Thermomixer	Eppendorf	Model R
Rotary Shaker (TC)	New Brunswick	Gyrotory G2
Nutator	Becton Dickenson	421105
Thermocycler	Bio-Rad	C1000
Biological Safety Cabinet	Thermo	Herasafe KS NSF A2
Cell Culture Incubator	Thermo	3310 Steri-Cult

Appendix B: Statistics and Controls

All sequencing experiments were replicated at least twice. For the specific number of replicates see either the figure legends or specific methods section. No aspect of this study was performed blinded. Sample size was not predetermined, and no outliers were excluded.

Appendix C: Code

Many of the analysis were enabled by the scripts *.msh* and *.sb,* created by Ming-an Sun for batch submission, and available upon request

The distance analysis script ".nna" was developed by David Pinto at Micron Oxford and is not publicly available, but the mathematical details are described in further detail in an unpublished manuscript available upon request. Distance datasets were also plotted with a custom R script using simple ggplot2 parameters, but no additional analysis was performed.

While no original code was developed as a part of this project, several workflows were established for analysis pipelines, including the use of the SMARTcleaner algorithm (Zhao and Zheng 2018), which had to be lightly modified to adapt to other parts of the pipeline.

All analysis scripts and codes are available upon request, and will be made publicly available at github.com/demmerle

Appendix D: Publications related to or including parts of this thesis

- Chapter 1 contains material including in a review manuscript on LDB1 function, currently in preparation for submission.
- Chapter 5 contains material previously published during my work in Oxford on technical aspects of structured illumination microscopy which were not included in this DPhil thesis. Images have been marked as "adapted from" the appropriate reference, and no text has been directly reproduced.
- Chapters 1-8 contain material that will be published as a research paper, currently in preparation for submission.

The following publications were produced during this DPhil.

Demmerle, J., Wegel, E., Schermelleh, L. and Dobbie, I. M. **(2015)**. Assessing resolution in super-resolution imaging. *Methods* 88, 3–10.

Ball, G., Demmerle, J., Kaufmann, R., Davis, I., Dobbie, I. M. and Schermelleh, L. **(2015)**. SIMcheck: a Toolbox for Successful Super-resolution Structured Illumination Microscopy. *Sci. Rep.* 5, 15915.

Demmerle J, Schermelleh L. **(2016)**. Xist at the edge: silencing at the nuclear periphery extends to X inactivation. *Translational Cancer Research* 5(6):S1189-S1191.

Demmerle, J., Innocent, C., North, A. J., Ball, G., Müller, M., Miron, E., Matsuda, A., Dobbie, I. M., Markaki, Y. and Schermelleh, L. **(2017)**. Strategic and practical guidelines for successful structured illumination microscopy. *Nat Protoc* 12, 988–1010.

Kraus, F., Miron, E., Demmerle, J., Chitiashvili, T., Budco, A., Alle, Q., Matsuda, A., Leonhardt, H., Schermelleh, L. and Markaki, Y. **(2017)**. Quantitative 3D structured illumination microscopy of nuclear structures. *Nat Protoc* 12, 1011–1028.

REFERENCES

Abdennur N, Mirny L. 2019. Cooler: scalable storage for Hi-C data and other genomically-labeled arrays. *bioRxiv* 557660.

Affolter M, Slattery M, Mann RS. 2008. A lexicon for homeodomain-DNA recognition. *Cell* **133**: 1133–1135.

Agulnick AD, Taira M, Breen JJ, Tanaka T, Dawid IB, Westphal H. 1996. Interactions of the LIM-domain-binding factor Ldb1 with LIM homeodomain proteins. *Nature* **384**: 270–272.

Ahmed S, Pryce BR, Al-Zahrani KN, Sabourin LA. 2018. The LIM domain binding protein 1, Ldb1, has distinct roles in Neu-induced mammary tumorigenesis. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*. **1865**: 1590-1597.

Alberti S, Gladfelter A, Mittag T. 2019. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**: 419–434.

Andrey G, Mundlos S. 2017. The three-dimensional genome: regulating gene expression during pluripotency and development. *Development* **144**: 3646–3658.

Arber S, Han B, Mendelsohn M, Smith M, Jessell TM, Sockanathan S. 1999. Requirement for the homeobox gene Hb9 in the consolidation of motor neuron identity. *Neuron* **23**: 659–674.

Ay F, Noble WS. 2015. Analysis methods for studying the 3D architecture of the genome. *Genome Biol* **16**: 1306.

Bach I, Carrière C, Ostendorff HP, Andersen B, Rosenfeld MG. 1997. A family of LIM domain-associated cofactors confer transcriptional synergism between LIM and Otx homeodomain proteins. *Genes & Development* **11**: 1370–1380.

Bach I, Rodriguez-Esteban C, Carrière C, Bhushan A, Krones A, Rose DW, Glass CK, Andersen B, Izpisúa Belmonte JC, Rosenfeld MG. 1999. RLIM inhibits functional activity of LIM homeodomain transcription factors via recruitment of the histone deacetylase complex. *Nat Genet* **22**: 394–399.

Ball G, Demmerle J, Kaufmann R, Davis I, Dobbie IM, Schermelleh L. 2015. SIMcheck: a Toolbox for Successful Super-resolution Structured Illumination Microscopy. *Sci Rep* **5**: 15915.

Bayarsaihan D, Lukens LN. 1996. Single-strand-DNA-binding factors specifically recognise the pyrimidine element in the chick α2(I) collagen gene promoter. *Biochem J* **314**: 293–296.

Beagrie RA, Scialdone A, Schueler M, Kraemer DCA, Chotalia M, Xie SQ, Barbieri M, de Santiago I, Lavitas L-M, Branco MR, et al. 2017. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**: 519–524.

Becker T, Ostendorff HP, Bossenz M, Schlüter A, Becker CG, Peirano RI, Bach I. 2002. Multiple functions of LIM domain-binding CLIM/NLI/Ldb cofactors during zebrafish development. *Mechanisms of Development* **117**: 75–85.

Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Peña-Castillo L, Alleyne TM, Mnaimneh S, Botvinnik OB, Chan ET, et al. 2008a. Variation in Homeodomain DNA Binding Revealed by High-Resolution Analysis of Sequence Preferences. *Cell* **133**: 1266–1276.

Bethea M, Liu Y, Wade AK, Mullen R, Gupta R, Gelfanov V, DiMarchi R, Bhatnagar S, Behringer R, HABEGGER KM, et al. 2019. The islet-expressed Lhx1 transcription factor interacts with islet-1 and contributes to glucose homeostasis. *Am J Physiol Endocrinol Metab* **316**: E397–E409.

Bhati M, Lee C, Gadd MS, Jeffries CM, Kwan A, Whitten AE, Trewhella J, Mackay JP, Matthews JM. 2012. Solution Structure of the LIM-Homeodomain Transcription Factor Complex Lhx3/Ldb1 and the Effects of a Pituitary Mutation on Key Lhx3 Interactions ed. R. Dobson. *PLoS ONE* **7**: e40719.

Bhati M, Lee C, Nancarrow AL, Lee M, Craig VJ, Bach I, Guss JM, Mackay JP, Matthews JM. 2008. Implementing the LIM code: the structural basis for cell type-specific assembly of LIM-homeodomain complexes. **27**: 2018–2029.

Bhati M, Llamosas E, Jacques DA, Jeffries CM, Dastmalchi S, Ripin N, Nicholas HR, Matthews JM. 2017. Interactions between LHX3- and ISL1-family LIM-homeodomain transcription factors are conserved in Caenorhabditis elegans. *Sci Rep* **7**: 4579.

Bickmore WA. 2013. The spatial organization of the human genome. *Annu Rev Genomics Hum Genet* **14**: 67–84.

Bickmore WA, van Steensel B. 2013. Genome Architecture: Domain Organization of Interphase Chromosomes. *Cell* **152**: 1270–1284.

Bintu B, Mateo LJ, Su J-H, Sinnott-Armstrong NA, Parker M, Kinrot S, Yamaya K, Boettiger AN, Zhuang X. 2018. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**.

Blobel G, Deng W. 2018. Controlled Gene Expression Methods. *US Patent Office*.

Bobola N, Merabet S. 2017. Homeodomain proteins in action: similar DNA binding preferences, highly variable connectivity. *Curr Opin Genet Dev* **43**: 1–8.

Boeynaems S, Alberti S, Fawzi NL, Mittag T, Polymenidou M, Rousseau F, Schymkowitz J, Shorter J, Wolozin B, Van Den Bosch L, et al. 2018. Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol* **28**: 420–435.

Boija A, Klein IA, Sabari BR, Dall'Agnese A, Coffey EL, Zamudio AV, Li CH, Shrinivas K, Manteiga JC, Hannett NM, et al. 2018. Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175**: 1842–1855.e16.

Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, Xu X, Lv X, Hugnot J-P, Tanay A, et al. 2017. Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**: 557–572.e24.

Boveri T. 1888. *Die Befruchtung und Teilung des Eies von Ascaris megalocephala*.

Breda L, Motta I, Lourenco S, Gemmo C, Deng W, Rupon JW, Abdulmalik OY, Manwani D, Blobel GA, Rivella S. 2016. Forced chromatin looping raises fetal hemoglobin in adult sickle cells to higher levels than pharmacologic inducers. *Blood* 2016–01–691089.

Breen JJ, Agulnick AD, Westphal H, Dawid IB. 1998. Interactions between LIM domains and the LIM domain-binding protein Ldb1. *J Biol Chem* **273**: 4712–4717.

Brickner JH, Walter P. 2004. Gene recruitment of the activated INO1 locus to the nuclear membrane. *PLoS Biol* **2**: e342.

Bridwell JL, Price JR, Parker GE, Schiller AM, Sloop KW, Rhodes SJ. 2001. Role of the LIM domains in DNA recognition by the Lhx3 neuroendocrine transcription factor. *Gene* **277**: 239–250.

Brown JM, Roberts NA, Graham B, Waithe D, Lagerholm C, Telenius JM, De Ornellas S, Oudelaar AM, Scott C, Szczerbal I, et al. 2018. A tissue-specific self-interacting chromatin domain forms independently of enhancer-promoter interactions. *Nat Commun* **9**: 3849.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**: 1213–1218.

Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. 2015. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**: 21.29.1–9.

Cai Y, Xu Z, Nagarajan L, Brandt SJ. 2008. Single-stranded DNA-binding proteins regulate the abundance and function of the LIM-homeodomain transcription factor LHX2 in pituitary cells. *Biochem Biophys Res Commun*

373: 303–308.

Caputo L, Witzel HR, Kolovos P, Cheedipudi S, Looso M, Mylona A, van IJcken WFJ, Laugwitz KL, Evans SM, Braun T, et al. 2015. The Isl1/Ldb1 Complex Orchestrates Genome-wide Chromatin Organization to Instruct Differentiation of Multipotent Cardiac Progenitors. *Cell Stem Cell* **17**: 287–299.

Cardozo Gizzi AM, Cattoni DI, Fiche J-B, Espinola SM, Gurgo J, Messina O, Houbron C, Ogiyama Y, Papadopoulos GL, Cavalli G, et al. 2019. Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *Mol Cell* **74**: 212–222.e5.

Cassata G, Röhrig S, Kuhn F, Hauri HP, Baumeister R, Bürglin TR. 2000. The Caenorhabditis elegans Ldb/NLI/Clim orthologue ldb-1 is required for neuronal function. *Dev Biol* **226**: 45–56.

Cavalli G, Misteli T. 2013. Functional implications of genome topology. *Nature Structural & Molecular Biology* **20**: 290–299.

Chen B, Li Y, Yu B, Zhang Z, Brommer B, Williams PR, Liu Y, Hegarty SV, Zhou S, Zhu J, et al. 2018. Reactivation of Dormant Relay Pathways in Injured Spinal Cord by KCC2 Manipulations. *Cell* **174**: 521–535.e13.

Chen L, Segal D, Hukriede NA, Podtelejnikov AV, Bayarsaihan D, Kennison JA, Ogryzko VV, Dawid IB, Westphal H. 2002. Ssdp proteins interact with the LIM-domain-binding protein Ldb1 to regulate development. *Proc Natl Acad Sci USA* **99**: 14320–14325.

Chen X, Shen Y, Draper W, Buenrostro JD, Litzenburger U, Cho SW, Satpathy AT, Carter AC, Ghosh RP, East-Seletsky A, et al. 2016. ATAC-see reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat Methods*.

Cho W-K, Spille J-H, Hecht M, Lee C, Li C, Grube V, Cissé II. 2018. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**: 412–415.

Chong S, Dugast-Darzacq C, Liu Z, Dong P, Dailey GM, Cattoglio C, Heckert A, Banala S, Lavis L, Darzacq X, et al. 2018. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* **361**.

Chu SW, Noyes MB, Christensen RG, Pierce BG, Zhu LJ, Weng Z, Stormo GD, Wolfe SA. 2012. Exploring the DNA-recognition potential of homeodomains. *Genome Res* **22**: 1889–1898.

Cook PR, Lang J. 1984. The spatial organization of sequences involved in initiation and termination of eukaryotic DNA replication. *Nucleic Acids Res* **12**: 1069–1075.

Cook PR, Marenduzzo D. 2018. Transcription-driven genome organization: a model for chromosome structure and the regulation of gene expression tested through simulations. *Nucleic Acids Res* **46**: 9895–9906.

Cramer P. 2019. Organization and regulation of gene transcription. *Nature* **224**: 1–10.

Cremer M, Cremer T. 2019. Nuclear compartmentalization, dynamics, and function of regulatory DNA sequences. *Genes Chromosom Cancer* **58**: 427–436.

Cremer T, Cremer C, Baumann H, Luedtke EK, Sperling K, Teuber V, Zorn C. 1982. Rabl's model of the interphase chromosome arrangement tested in Chinese hamster cells by premature chromosome condensation and laser-UV-microbeam experiments. *Hum Genet* **60**: 46–56.

Cremer T, Cremer M, Hübner B, Strickfaden H, Smeets D, Popken J, Sterr M, Markaki Y, Rippe K, Cremer C. 2015. The 4D nucleome: Evidence for a dynamic nuclear landscape based on co-aligned active and inactive nuclear compartments. *FEBS Lett* **589**: 2931–2943.

Cremer T, Kreth G, Koester H, Fink RH, Heintzmann R, Cremer M, Solovei I, Zink D, Cremer C. 2000. Chromosome territories, interchromatin domain compartment, and nuclear matrix: an integrated view of the functional nuclear architecture. *Crit Rev Eukaryot Gene Expr* **10**: 179–212.

Crick F. 1970. Central dogma of molecular biology. *Nature* **227**: 561–563.

Cross AJ, Jeffries CM, Trewhella J, Matthews JM. 2010. LIM domain binding proteins 1 and 2 have different oligomeric states. *Journal of Molecular Biology* **399**: 133–144.

Cross SH, Macalinao DG, McKie L, Rose L, Kearney AL, Rainger J, Thaung C, Keighren M, Jadeja S, West K, et al. 2014. A dominant-negative mutation of mouse Lmx1b causes glaucoma and is semi-lethal via LDB1-mediated dimerization [corrected]. ed. G.S. Barsh. *PLoS Genet* **10**: e1004359.

Dastmalchi S, Wilkinson White L, Kwan AH, Gamsjaeger R, Mackay JP, Matthews JM. 2012. Solution structure of a tethered Lmo2LIM2/Ldb1LID complex. *Protein Science* **21**: 1768–1774.

Davies JOJ, Oudelaar AM, Higgs DR, Hughes JR. 2017. How best to identify chromosomal interactions: a comparison of approaches. *Nat Methods* **14**: 125–134.

Davies JOJ, Telenius JM, McGowan SJ, Roberts NA, Taylor S, Higgs DR, Hughes JR. 2016. Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat Methods* **13**: 74–80.

Davis-Dusenbery BN, Williams LA, Klim JR, Eggan K. 2014. How to make spinal

motor neurons. *Development* **141**: 491–501.

Dawid IB, Breen JJ, Toyama R. 1998. LIM domains: multiple roles as adapters and functional modifiers in protein interactions. *Trends Genet* **14**: 156–162.

de Melo J, Clark BS, Venkataraman A, Shiau F, Zibetti C, Blackshaw S. 2018. Ldb1 and Rnf12-dependent regulation of Lhx2 controls the relative balance between neurogenesis and gliogenesis in retina. *Development* **145**: dev.159970.

Deane JE, Mackay JP, Kwan AHY, Sum EYM, Visvader JE, Matthews JM. 2003. Structural basis for the recognition of ldb1 by the N-terminal LIM domains of LMO2 and LMO4. *The EMBO Journal* **22**: 2224–2233.

Dekker J, Marti-Renom MA, Mirny LA, mirny. 2013. Exploring the three-dimensional organization of genomes: Interpreting chromatin interaction data. *Nature Reviews Genetics* **14**: 390–403.

Demmerle J, Innocent C, North AJ, Ball G, Müller M, Miron E, Matsuda A, Dobbie IM, Markaki Y, Schermelleh L. 2017. Strategic and practical guidelines for successful structured illumination microscopy. *Nat Protoc* **12**: 988–1010.

Demmerle J, Koch AJ, Holaska JM. 2013. Emerin and histone deacetylase 3 (HDAC3) cooperatively regulate expression and nuclear positions of MyoD, Myf5, and Pax7 genes during myogenesis. *Chromosome Res* **21**: 765–779.

Demmerle J, Wegel E, Schermelleh L, Dobbie IM. 2015. Assessing resolution in super-resolution imaging. *Methods* **88**: 3–10.

Deng W, Lee J, Wang H, Miller J, Reik A, Gregory PD, Dean A, Blobel GA. 2012. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* **149**: 1233–1244.

Deng W, Rupon JW, Krivega I, Breda L, Motta I, Jahn KS, Reik A, Gregory PD, Rivella S, Dean A, et al. 2014. Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell* **158**: 849–860.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.

Dreyer SD, Morello R, German MS, Zabel B, Winterpacht A, Lunstrum GP, Horton WA, Oberg KC, Lee B. 2000. LMX1B transactivation and expression in nail–patella syndrome. *Hum Mol Genet* **9**: 1067–1074.

Dufourt J, Trullo A, Hunter J, Fernandez C, Lazaro J, Dejean M, Morales L, Nait-Amer S, Schulz KN, Harrison MM, et al. 2018. Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs. *Nat Commun* **9**: 5194.

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. 2016a. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Systems* **3**: 99–101.

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016b. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Systems* **3**: 95–98.

Durisic N, Cuervo LL, Lakadamyali M. 2014. Quantitative super-resolution microscopy: pitfalls and strategies for image analysis. *Current Opinion in Chemical Biology* **20**:22-28

Eaton JW, Bateman D, Hauberg S, Wehbring R. 2014. *GNU Octave Version 3.8.1 manual: a high-level interactive language for numerical computation.* CreateSpace Independent Publishing Platform.

Ediger BN, Lim H-W, Juliana C, Groff DN, Williams LT, Dominguez G, Liu J-H, Taylor BL, Walp ER, Kameswaran V, et al. 2016. LIM domain–binding 1 maintains the terminally differentiated state of pancreatic β cells. *J Clin Invest* **127**.

Engreitz JM, Ollikainen N, Guttman M. 2016. Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nat Rev Mol Cell Biol* **17**: 756–770.

Erb M, Lee B, Yeon Seo S, Lee JW, Lee S, Lee S-K. 2017. The Isl1-Lhx3 Complex Promotes Motor Neuron Specification by Activating Transcriptional Pathways that Enhance Its Own Expression and Formation. *eNeuro* **4**: ENEURO.0349–16.2017.

Erdel F, Rippe K. 2018. Formation of Chromatin Subcompartments by Phase Separation. *Biophys J* **114**: 2262–2270.

Fiedler M, Graeb M, Mieszczanek J, Rutherford TJ, Johnson CM, Bienz M. 2015. An ancient Pygo-dependent Wnt enhanceosome integrated by Chip/LDB-SSDP. *Elife* **4**: :e09073

Finn EH, Pegoraro G, Brandão HB, Valton A-L, Oomen ME, Dekker J, Mirny L, Misteli T. 2019. Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell* **176**: 1502–1515.e10.

Francius C, Clotman F. 2013. Generating spinal motor neuron diversity: a long quest for neuronal identity. *Cell Mol Life Sci* **71**: 813–829.

Gadd MS, Jacques DA, Guss JM, Matthews JM. 2012. Crystallization and diffraction of an Isl1-Ldb1 complex. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **68**: 1398–1401.

Gadd MS, Jacques DA, Nisevic I, Craig VJ, Kwan AH, Guss JM, Matthews JM. 2013. A structural basis for the regulation of the LIM-homeodomain protein

islet 1 (Isl1) by intra- and intermolecular interactions. *J Biol Chem* **288**: 21924–21935.

Galloway JR, Bethea M, Liu Y, Underwood R, Mobley JA, Hunter CS. 2015. SSBP3 Interacts With Islet-1 and Ldb1 to Impact Pancreatic β-Cell Target Genes. *Mol Endocrinol* **29**: 1774–1786.

García SA, Swiersy A, Radhakrishnan P, Branchi V, Kanth Nanduri L, Győrffy B, Betzler AM, Bork U, Kahlert C, Reißfelder C, et al. 2016. LDB1 overexpression is a negative prognostic factor in colorectal cancer. *Oncotarget* **7**: 84258–84270.

Gehring WJ, Affolter M, Bürglin T. 1994. *Homeodomain proteins*.

Ghavi-Helm Y, Jankowski A, Meiers S, Viales RR, Korbel JO, Furlong EEM. 2019. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat Genet* **51**: 1272–1282.

Ghiaccio V, Chappell M, Rivella S, Breda L. 2019. Gene Therapy for Beta-Hemoglobinopathies: Milestones, New Therapies and Challenges. *Mol Diagn Ther* **381**: 142.

Gibcus JH, Dekker J. 2013. The Hierarchy of the 3D Genome. *Mol Cell* **49**: 773–782.

Gill GN. 2003. Decoding the LIM development code. *Trans Am Clin Climatol Assoc* **114**: 179–189.

Gill GN. 1995. The enigma of LIM domains. *Structure* **3**: 1285–1289.

Granger A, Bleux C, Kottler M-L, Rhodes SJ, Counis R, Laverrière J-N. 2006. The LIM-homeodomain proteins Isl-1 and Lhx3 act with steroidogenic factor 1 to enhance gonadotrope-specific activity of the gonadotropin-releasing hormone receptor gene promoter. *Molecular Endocrinology* **20**: 2093–2108.

Gueta K, David A, Cohen T, Menuchin-Lasowski Y, Nobel H, Narkis G, Li L, Love P, de Melo J, Blackshaw S, et al. 2016. The stage-dependent roles of Ldb1 and functional redundancy with Ldb2 in mammalian retinogenesis. *Development* **143**: 4182–4192.

Guo G, Pinello L, Han X, Lai S, Shen L, Lin T-W, Zou K, Yuan G-C, Orkin SH. 2016. Serum-Based Culture Conditions Provoke Gene Expression Variability in Mouse Embryonic Stem Cells as Revealed by Single-Cell Analysis. *Cell Reports* **14**: 956–965.

Güngör C, Taniguchi-Ishigaki N, Ma H, Drung A, Tursun B, Ostendorff HP, Bossenz M, Becker CG, Becker T, Bach I. 2007. Proteasomal selection of multiprotein complexes recruited by LIM homeodomain transcription factors. *Proc Natl Acad Sci USA* **104**: 15000–15005.

Hahn S. 2018. Phase Separation, Protein Disorder, and Enhancer Function. *Cell* **175**: 1723–1725.

Hansen AS, Cattoglio C, Darzacq X, Tjian R. 2018. Recent evidence that TADs and chromatin loops are dynamic structures. *nucleus* **9**: 20–32.

Hester ME, Murtha MJ, Song S, Rao M, Miranda CJ, Meyer K, Tian J, Boulting G, Schaffer DV, Zhu MX, et al. 2011. Rapid and Efficient Generation of Functional Motor Neurons From Human Pluripotent Stem Cells Using Gene Delivered Transcription Factor Codes. *Mol Ther* **19**: 1905–1912.

Hiratani I, Yamamoto N, Mochizuki T, Ohmori S-Y, Taira M. 2003. Selective degradation of excess Ldb1 by Rnf12/RLIM confers proper Ldb1 expression levels and Xlim-1/Ldb1 stoichiometry in Xenopus organizer functions. *Development* **130**: 4161–4175.

Hnisz D, Abraham B, Lee TI, Young RA. 2016. Super-enhancers and methods of use thereof. *US Patent Office*.

Hnisz D, Schuijers J, Lin CY, Weintraub AS, Abraham BJ, Lee TI, Bradner JE, Young RA. 2015. Convergence of Developmental and Oncogenic Signaling Pathways at Transcriptional Super-Enhancers. *Mol Cell* **58**: 362–370.

Hnisz D, Young RA, Borges-Rivera D, Weintraub AS, Ji X, Dadon DB, Fan ZP, Lee TI, Ihn T. 2019. Chromosome neighborhood structures and methods relating thereto. *US Patent Office*.

Huai J, Zhang X, Li J, Ma T, Zha P, Jing Y, Lin R. 2018. SEUSS and PIF4 Coordinately Regulate Light and Temperature Signaling Pathways to Control Plant Growth. *Molecular Plant* **11**: 928-942.

Hughes JR, Lower KM, Dunham I, Taylor S, De Gobbi M, Sloane-Stanley JA, McGowan S, Ragoussis J, Vernimmen D, Gibbons RJ, et al. 2013. High-resolution analysis of cis-acting regulatory networks at the α-globin locus. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**: 20120361–20120361.

Hunter CS, Dixit S, Cohen T, Ediger B, Wilcox C, Ferreira M, Westphal H, Stein R, May CL. 2013. Islet α-, β-, and δ-cell development is controlled by the Ldb1 coregulator, acting primarily with the islet-1 transcription factor. *Diabetes* **62**: 875–886.

Hwang M, Gorivodsky M, Kim M, Westphal H, Geum D. 2008. The neuronal differentiation potential of Ldb1-null mutant embryonic stem cells is dependent on extrinsic influences. *Stem Cells* **26**: 1490–1495.

Iacovino M, Bosnakovski D, Fey H, Rux D, Bajwa G, Mahen E, Mitanoska A, Xu Z, Kyba M. 2011. Inducible cassette exchange: a rapid and efficient system enabling conditional gene expression in embryonic stem and primary cells. *Stem Cells* **29**: 1580–1588.

Iborra FJ. 2007. Can visco-elastic phase separation, macromolecular crowding and colloidal physics explain nuclear organisation? *Theor Biol Med Model* **4**: 843.

Jackson DA, Hassan, AB, Errington RJ, Cook PR. 1993. Visualization of Focal Sites of Transcription Within Human Nuclei. *The EMBO Journal* **12**: 1059–1065.

Jessell TM. 2000. Neuronal specification in the spinal cord: inductive signals and transcriptional codes. **1**: 20–29.

Ji X, Dadon DB, Powell BE, Fan ZP, Borges-Rivera D, Shachar S, Weintraub AS, Hnisz D, Pegoraro G, Lee TI, et al. 2016. 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. *Cell Stem Cell* **18**: 262–275.

Jurata LW, Gill GN. 1997. Functional analysis of the nuclear LIM domain interactor NLI. **17**: 5688–5698.

Jurata LW, Kenny DA, Gill GN. 1996. Nuclear LIM interactor, a rhombotin and LIM homeodomain interacting protein, is expressed early in neuronal development. *Proc Natl Acad Sci USA* **93**: 11693–11698.

Jurata LW, Pfaff SL, Gill GN. 1998. The nuclear LIM domain interactor NLI mediates homo- and heterodimerization of LIM domain transcription factors. *J Biol Chem* **273**: 3152–3157.

Karbassi E, Rosa-Garrido M, Chapski DJ, Wu Y, Ren S, Wang Y, Stefani E, Vondriska TM. 2019. Direct visualization of cardiac transcription factories reveals regulatory principles of nuclear architecture during pathological remodeling. *J Mol Cell Cardiol* **128**: 198–211.

Kerpedjiev P, Abdennur N, Lekschas F, McCallum C, Dinkla K, Strobelt H, Luber JM, Ouellette SB, Azhir A, Kumar N, et al. 2018. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol* **19**: 125.

Kim K-T, Kim N, Kim H-K, Lee H, Gruner HN, Gergics P, Park C, Mastick GS, Park H-C, Song M-R. 2016a. ISL1-based LIM complexes control Slit2 transcription in developing cranial motor neurons. *Sci Rep* **6**: 36491.

Kim S, Zhao Y, Lee J-M, Kim WR, Gorivodsky M, Westphal H, Geum D. 2016b. Ldb1 Is Essential for the Development of Isthmic Organizer and Midbrain Dopaminergic Neurons. *Stem Cells and Development* **25**: 986–994.

Kishi Y, Gotoh Y. 2018. Regulation of Chromatin Structure During Neural Development. *Front Neurosci* **12**: 39.

Kobren SN, Singh M. 2019. Systematic domain-based aggregation of protein structures highlights DNA-, RNA- and other ligand-binding positions. *Nucleic Acids Res* **47**: 582–593.

Kosak ST, Skok JA, Medina KL, Riblet R, Le Beau MM, Fisher AG, Singh H. 2002. Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* **296**: 158–162.

Kraft K, Magg A, Heinrich V, Riemenschneider C, Schöpflin R, Markowski J, Ibrahim DM, Acuna-Hidalgo R, Despang A, Andrey G, et al. 2019. Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and congenital malformations. *Nat Cell Biol* **21**: 305-310.

Kragesteen BK, Spielmann M, Paliou C, Heinrich V, Schöpflin R, Esposito A, Annunziatella C, Bianco S, Chiariello AM, Jerković I, et al. 2018. Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. *Nat Genet* **50**: 1463–1473.

Kraus F, Miron E, Demmerle J, Chitiashvili T, Budco A, Alle Q, Matsuda A, Leonhardt H, Schermelleh L, Markaki Y. 2017. Quantitative 3D structured illumination microscopy of nuclear structures. *Nat Protoc* **12**: 1011–1028.

Krivega I, Dale RK, Dean A. 2014. Role of LDB1 in the transition from chromatin looping to transcription activation. *Genes & Development* **28**: 1278–1290.

Krivega I, Dean A. 2017. LDB1-mediated enhancer looping can be established independent of mediator and cohesin. *Nucleic Acids Res* **45**: 8255–8268.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359.

Lee B, Lee S, Agulnick AD, Lee JW, Lee S-K. 2016. Single-stranded DNA binding proteins are required for LIM complexes to induce transcriptionally active chromatin and specify spinal neuronal identities. *Development* **143**: 1721–1731.

Lee J, Krivega I, Dale RK, Dean A. 2017. The LDB1 Complex Co-opts CTCF for Erythroid Lineage-Specific Long-Range Enhancer Interactions. *Cell Reports* **19**: 2490–2502.

Lee S, Cuvillier JM, Lee B, Shen R, Lee JW, Lee S-K. 2012. Fusion protein Isl1-Lhx3 specifies motor neuron fate by inducing motor neuron genes and concomitantly suppressing the interneuron programs. *Proceedings of the National Academy of Sciences* **109**: 3383–3388.

Lee S, Lee B, Joshi K, Pfaff SL, Lee JW, Lee S-K. 2008. A Regulatory Network to Segregate the Identity of Neuronal Subtypes. *Dev Cell* **14**: 877–889.

Lee S, Shen R, Cho H-H, Kwon R-J, Seo SY, Lee JW, Lee S-K. 2013. STAT3 promotes motor neuron differentiation by collaborating with motor neuron-specific LIM complex. *Proceedings of the National Academy of Sciences* **110**: 11445–11450.

Lee S-K, Jurata LW, Funahashi J, Ruiz EC, Pfaff SL. 2004. Analysis of

embryonic motoneuron gene regulation: derepression of general activators function in concert with enhancer factors. *Development* **131**: 3295–3306.

Lee S-K, Jurata LW, Nowak R, Lettieri K, Kenny DA, Pfaff SL, Gill GN. 2005a. The LIM domain-only protein LMO4 is required for neural tube closure. *Molecular and Cellular Neuroscience* **28**: 205–214.

Lee S-K, Lee B, Ruiz EC, Pfaff SL. 2005b. Olig2 and Ngn2 function in opposition to modulate gene expression in motor neuron progenitor cells. *Genes & Development* **19**: 282–294.

Lee S-K, Pfaff SL. 2003. Synchronization of neurogenesis and motor neuron specification by direct coupling of bHLH and homeodomain transcription factors. *Neuron* **38**: 731–745.

Lee SK, Pfaff SL. 2001. Transcriptional networks regulating neuronal identity in the developing spinal cord. *Nat Neuro* **4**: 1183–1191.

Lemaître C, Grabarz A, Tsouroula K, Andronov L, Furst A, Pankotai T, Heyer V, Rogier M, Attwood KM, Kessler P, et al. 2014. Nuclear position dictates DNA repair pathway choice. *Genes & Development* **28**: 2450–2463.

Leone DP, Panagiotakos G, Heavner WE, Joshi P, Zhao Y, Westphal H, McConnell SK. 2017. Compensatory Actions of Ldb Adaptor Proteins During Corticospinal Motor Neuron Differentiation. *Cereb Cortex* **27**: 1686–1699.

Leotta CG, Federico C, Brundo MV, Tosi S, Saccone S. 2014. HLXB9 Gene Expression, and Nuclear Location during In Vitro Neuronal Differentiation in the SK-N-BE Neuroblastoma Cell Line *PLoS ONE* **9**: e105481.

Levin E, Leibinger M, Gobrecht P, Hilla A, Andreadaki A, Fischer D. 2019. Muscle LIM Protein Is Expressed in the Injured Adult CNS and Promotes Axon Regeneration. *Cell Reports* **26**: 1021–1032.e6.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Li J, Kurasawa Y, Wang Y, Clise-Dwyer K, Klumpp SA, Liang H, Tailor RC, Raymond AC, Estrov Z, Brandt SJ, et al. 2014. Requirement for ssbp2 in hematopoietic stem cell maintenance and stress response. *J Immunol* **193**: 4654–4662.

Li L, Freudenberg J, Cui K, Dale R, Song SH, Dean A, Zhao K, Jothi R, Love PE. 2013. Ldb1-nucleated transcription complexes function as primary mediators of global erythroid gene activation. *Blood* **121**: 4575–4585.

Liao Y, Smyth GK, Shi W. 2013a. featureCounts: An efficient general-purpose program for assigning sequence reads to genomic features. *arXiv* **q-bio.GN**:

923–930.

Liao Y, Smyth GK, Shi W. 2013b. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res* **41**: e108.

Liu G, Dean A. 2019. Enhancer long-range contacts: The multi-adaptor protein LDB1 is the tie that binds. *Biochim Biophys Acta Gene Regul Mech* **1862**: 625–633.

Liu S, Zhang L, Quan H, Tian H, Meng L, Yang L, Feng H, Gao YQ. 2018. From 1D sequence to 3D chromatin dynamics and cellular functions: a phase separation perspective. *bioRxiv* 255174.

Liu Z, Legant WR, Chen B-C, Li L, Grimm JB, Lavis LD, Betzig E, Tjian R. 2014. 3D imaging of Sox2 enhancer clusters in embryonic stem cells. *Elife* **3**: e04236.

Love MI, Huber W, Anders S. 2014a. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**:550.

Love PE, Warzecha C, Li L. 2014b. Ldb1 complexes: the new master regulators of erythroid gene transcription. *Trends Genet* **30**: 1–9.

Loyd C, Liu Y, KIM T, Holleman C, Galloway J, Bethea M, Ediger BN, Swain TA, Tang Y, Stoffers DA, et al. 2017. LDB1 regulates energy homeostasis during diet-induced obesity. *Endocrinology* **158**: 1289–1297.

Lumsden A. 1995. Neural Development: A "LIM code" for motor neurons? *Current Biology* **5**: 491–495.

Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, et al. 2015. Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* **161**:1012-1025.

Ma W, Noble WS, Bailey TL. 2014. Motif-based analysis of large nucleotide data sets using MEME-ChIP. *Nat Protoc* **9**: 1428–1450.

Marti-Renom MA, Almouzni G, Bickmore WA, Bystricky K, Cavalli G, Fraser P, Gasser SM, Giorgetti L, Heard E, Nicodemi M, et al. 2018. Challenges and guidelines toward 4D nucleome data and model standards. *Nat Genet* **50**: 1352–1358.

Marti-Renom MA, Mirny LA. 2011. Bridging the resolution gap in structural modeling of 3D genome organization. *PLoS Computational Biology* **7**: e1002125.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetJournal* **17**: 10–12.

Mateo LJ, Murphy SE, Hafner A, Cinquini IS, Walker CA, Boettiger AN. 2019. Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568**: 49–54.

Matthews JM, Bhati M, Craig VJ, Deane JE, Jeffries C, Lee C, Nancarrow AL, Ryan DP, Sunde M. 2008. Competition between LIM-binding domains. *Biochemical Society transactions* **36**: 1393–1397.

Matthews JM, Visvader JE. 2003. LIM-domain-binding protein 1: a multifunctional cofactor that interacts with diverse proteins. *EMBO Rep* **4**: 1132–1137.

Mazzoni EO, Mahony S, Closser M, Morrison CA, Nedelec S, Williams DJ, An D, Gifford DK, Wichterle H. 2013. Synergistic binding of transcription factors to cell-specific enhancers programs motor neuron identity. *Nat Neurosci* **16**: 1219–1227.

Meier N, Krpic S, Rodriguez P, Strouboulis J, Monti M, Krijgsveld J, Gering M, Patient R, Hostert A, Grosveld F. 2006. Novel binding partners of Ldb1 are required for haematopoietic development. *Development* **133**: 4913–4923.

Milan M, Cohen SM. 1999. Regulation of LIM homeodomain activity in vivo: A tetramer of dLDB and apterous confers activity and capacity for regulation by dLMO. *Mol Cell* **4**: 267–273.

Mir M, Stadler MR, Ortiz SA, Hannon CE, Harrison MM, Darzacq X, Eisen MB. 2018. Dynamic multifactor hubs interact transiently with sites of active transcription in Drosophila embryos. *eLife* **7**: e40497.

Miron E, Oldenkamp R, Pinto D, Brown JM, Faria AR, Shaban HA, Rhodes JDP, Innocent C, de Ornellas S, Buckle VJ, et al. 2019. Chromatin arranges in chains of mesoscale domains with nanoscale functional topography independent of cohesin. *bioRxiv*: 566638.

Misteli T. 2007. Beyond the sequence: cellular organization of genome function. *Cell* **128**: 787–800.

Misteli T. 2010. Higher-order genome organization in human disease. *Cold Spring Harbor Perspectives in Biology* **2**: a000794.

Mizunuma H, Miyazawa J, Sanada K, Imai K. 2003. The LIM-only protein, LMO4, and the LIM domain-binding protein, LDB1, expression in squamous cell carcinomas of the oral cavity. *British Journal of Cancer* **88**: 1543–1548.

Monahan K, Horta A, Lomvardas S. 2019. LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. *Nature* **565**: 448–453.

Morcillo P, Rosen C, Baylies MK, Dorsett D. 1997. Chip, a widely expressed chromosomal protein required for segmentation and activity of a remote wing margin enhancer in Drosophila. *Genes & Development* **11**: 2729–2740.

Morcillo P, Rosen C, Dorsett D. 1996. Genes regulating the remote wing margin enhancer in the Drosophila cut locus. *Genetics* **144**: 1143–1154.

Mukhopadhyay M, Teufel A, Yamashita T, Agulnick AD, Chen L, Downs KM, Schindler A, Grinberg A, Huang S-P, Dorward D, et al. 2003. Functional ablation of the mouse Ldb1 gene results in severe patterning defects during gastrulation. *Development* **130**: 495–505.

Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, Greenleaf WJ, Chang HY. 2016. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods* **13**: 919–922.

Nagashima R, Hibino K, Ashwin SS, Babokhov M, Fujishiro S, Imai R, Nozaki T, Tamura S, Tani T, Kimura H, et al. 2019. Single nucleosome imaging reveals loose genome chromatin networks via active RNA polymerase II. *J Cell Biol*. **218**: 1511-1530.

Nam H, Jeon S, An H, Yoo J, Lee H-J, Lee S-K, Lee S. 2019. Critical roles of ARHGAP36 as a signal transduction mediator of Shh pathway in lateral motor columnar specification. *Elife* **8**: 1122.

Narayanan A, Meriin A, Andrews JO, Spille J-H, Sherman MY, Cissé II. 2019. A first order phase transition mechanism underlies protein aggregation in mammalian cells. *Elife* **8**:e39695.

Naumova N, Imakaev M, Fudenberg G, Zhan Y, Lajoie BR, Mirny LA, Dekker J. 2013. Organization of the Mitotic Chromosome. *Science* **342**: 948–953.

Niekamp S, Sung J, Huynh W, Bhabha G, Vale RD, Stuurman N. 2019. Nanometer-accuracy distance measurements between fluorophores at the single-molecule level. *Proceedings of the National Academy of Sciences*. **116**: 4275-4284.

Nishioka N, Nagano S, Nakayama R, Kiyonari H, Ijiri T, Taniguchi K, Shawlot W, Hayashizaki Y, Westphal H, Behringer RR, et al. 2005. Ssdp1 regulates head morphogenesis of mouse embryos by activating the Lim1-Ldb1 complex. *Development* **132**: 2535–2546.

Norton HK, Phillips-Cremins JE. 2017. Crossed wires: 3D genome misfolding in human disease. *J Cell Biol* **62**: jcb.201611001.

Omari El K, Hoosdally SJ, Tuladhar K, Karia D, Vyas P, Patient R, Porcher C, Mancini EJ. 2011. Structure of the leukemia oncogene LMO2: Implications for the assembly of a hematopoietic transcription factor complex. *Blood* **117**: 2146–2156.

Ong C-T, Corces VG. 2014. CTCF: an architectural protein bridging genome topology and function. *Nature Reviews Genetics* **15**: 234–246.

Oudelaar AM, Davies JOJ, Downes DJ, Higgs DR, Hughes JR. 2017. Robust

detection of chromosomal interactions from small numbers of cells using low-input Capture-C. *Nucleic Acids Res* **45**: e184–e184.

Oudelaar AM, Davies JOJ, Hanssen LLP, Telenius JM, Schwessinger R, Liu Y, Brown JM, Downes DJ, Chiariello AM, Bianco S, et al. 2018. Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains. *Nat Genet* **489**: 109.

O'Geen H, Frietze S, Farnham PJ. 2010. Using ChIP-seq technology to identify targets of zinc finger transcription factors. *Methods Mol Biol* **649**: 437–455.

Paliou C, Guckelberger P, Schöpflin R, Heinrich V, Esposito A, Chiariello AM, Bianco S, Annunziatella C, Helmuth J, Haas S, et al. 2019. Preformed chromatin topology assists transcriptional robustness of Shh during limb development. *Proceedings of the National Academy of Sciences* **116**: 12390–12399.

Papantonis A, Cook PR. 2013. Transcription Factories: Genome Organization and Gene Regulation. *Chem Rev* 130418061411007.

Parmar JJ, Woringer M, Zimmer C. 2019. How the Genome Folds: The Biophysics of Four-Dimensional Chromatin Organization. *Annu Rev Biophys* **48**: 231–253.

Petrovic J, Zhou Y, Fasolino M, Goldman N, Schwartz GW, Mumbach MR, Nguyen SC, Rome KS, Sela Y, Zapataro Z, et al. 2019. Oncogenic Notch Promotes Long-Range Regulatory Interactions within Hyperconnected 3D Cliques. *Mol Cell* **73**: 1174–1190.e12.

Pfaff SL, Macfarlan T. 2009. Transcriptional Networks and the Spinal Cord. In *Encyclopedia of Neuroscience*, pp. 1093–1098, Elsevier.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

Quinodoz SA, Ollikainen N, Tabak B, Palla A, Schmidt JM, Detmar E, Lai MM, Shishkin AA, Bhat P, Takei Y, et al. 2018. Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* **174**: 744–757.e24.

Rabl C. 1885. *Uber zelltheilung*.

Rada Iglesias A, Grosveld FG, Papantonis A. 2018. Forces driving the three-dimensional folding of eukaryotic genomes. *Mol Syst Biol* **14**: e8214.

Ramirez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**:W160-5.

Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT,

Sanborn AL, Machol I, Omer AD, Lander ES, et al. 2014. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*. **159**:1665-1680.

Rétaux S, Bachy I. 2002. A short history of LIM domains (1993-2002): from protein interaction to degradation. *Mol Neurobiol* **26**: 269–281.

Rhee HS, Closser M, Guo Y, Bashkirova EV, Tan GC, Gifford DK, Wichterle H. 2016. Expression of Terminal Effector Genes in Mammalian Neurons Is Maintained by a Dynamic Relay of Transient Enhancers. *Neuron* **92**: 1252–1265.

Rivella S, Breda L, Dong A, Blobel GA, Deng W, Deng W. 2018. Viral vectors for prophylaxis and therapy of hemoglobinopathies. *US Patent Office*.

Robertson NO, Smith NC, Manakas A, Mahjoub M, McDonald G, Kwan AH, Matthews JM. 2018. Disparate binding kinetics by an intrinsically disordered domain enables temporal regulation of transcriptional complex formation. *Proc Natl Acad Sci USA* **84**: 201714646.

Rocha PP, Raviram R, Bonneau R, Skok JA. 2015. Breaking TADs: insights into hierarchical genome organization. *Epigenomics* **7**: 523–526.

Rowley MJ, Nichols MH, Lyu X, Ando-Kuri M, Rivera ISM, Hermetz K, Wang P, Ruan Y, Corces VG. 2017. Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol Cell* **67**: 837–852.e7.

Ryan DP, Duncan JL, Lee C, Kuchel PW, Matthews JM. 2007. Assembly of the oncogenic DNA-binding complex LMO2-Ldb1-TAL1-E12. *Proteins* **70**: 1461–1474.

Schermelleh L, Carlton PM, Haase S, Shao L, Winoto L, Kner P, Burke B, Cardoso MC, Agard DA, Gustafsson MGL, et al. 2008. Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy. *Science* **320**: 1332–1336.

Schermelleh L, Ferrand A, Huser T, Eggeling C, Sauer M, Biehlmaier O, Gregor P. C. Drummen. 2019. Super-resolution microscopy demystified. *Nature Cell Biology* **21**: 72–84.

Schermelleh L, Heintzmann R, Leonhardt H. 2010. A guide to super-resolution fluorescence microscopy. *J Cell Biol* **190**: 165–175.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. 2012. Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**: 676–682.

Schlesinger S, Meshorer E. 2019. Open Chromatin, Epigenetic Plasticity, and Nuclear Organization in Pluripotency. *Dev Cell* **48**: 135–150.

Schuetzmann D, Walter C, van Riel B, Kruse S, König T, Erdmann T, Tönges A, Bindels E, Weilemann A, Gebhard C, et al. 2018. Temporal autoregulation during human PU.1 locus SubTAD formation. *Blood* **132**: 2643–2655.

Semina EV, Altherr MR, Murray JC. 1998. Cloning and chromosomal localization of two novel human genes encoding LIM-domain binding factors CLIM1 and CLIM2/LDB1/NLI. *Mamm Genome* **9**: 921–924.

Seo SY, Lee B, Lee S. 2015. Critical Roles of the LIM Domains of Lhx3 in Recruiting Coactivators to the Motor Neuron-Specifying Isl1-Lhx3 Complex. *Mol Cell Biol* **35**: 3579–3589.

Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, Vert J-P, Heard E, Dekker J, Barillot E. 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16**: 11.

Shachar S, Misteli T. 2017. Causes and consequences of nuclear gene positioning. *bioRxiv* 118711.

Shapiro JA. 1997. Genome organization, natural genetic engineering and adaptive mutation. *Trends Genet* **13**: 98–104.

Shapiro JA. 2009. Revisiting the central dogma in the 21st century. *Ann N Y Acad Sci* **1178**: 6–28.

Shirasaki R, Pfaff SL. 2002. Transcriptional codes and the control of neuronal identity. *Annu Rev Neurosci* **25**: 251–281.

Shrinivas K, Sabari BR, Coffey EL, Klein IA, Boija A, Zamudio AV, Schuijers J, Hannett NM, Sharp PA, Young RA, et al. 2019. Enhancer Features that Drive Formation of Transcriptional Condensates. *Mol Cell* **75**: 549–561.e7.

Simonik EA, Cai Y, Kimmelshue KN, Brantley-Sieders DM, Loomans HA, Andl CD, Westlake GM, Youngblood VM, Chen J, Yarbrough WG, et al. 2016. LIM-Only Protein 4 (LMO4) and LIM Domain Binding Protein 1 (LDB1) Promote Growth and Metastasis of Human Head and Neck Cancer (LMO4 and LDB1 in Head and Neck Cancer) ed. G. Sethi. *PLoS ONE* **11**: e0164804.

Simonis M, Kooren J, de Laat W. 2007. An evaluation of 3C-based methods to capture DNA interactions. *Nat Methods* **4**: 895–901.

Sizemore AE, Phillips-Cremins J, Ghrist R, Bassett DS. 2018. The importance of the whole: topological data analysis for the network neuroscientist. *Network Neuroscience.* **3**:656-673.

Smeets D, Neumann J, Schermelleh L. 2014. Application of Three-Dimensional Structured Illumination Microscopy in Cell Biology: Pitfalls and Practical Considerations. In *Super-Resolution Microscopy Techniques in the Neurosciences*, Vol. 86 of, pp. 167–188, Neuromethods.

Smith NC, Matthews JM. 2016. Mechanisms of DNA-binding specificity and functional gene regulation by transcription factors. *Current Opinion in Structural Biology* **38**: 68–74.

Soler E, Andrieu-Soler C, de Boer E, Bryne JC, Thongjuea S, Stadhouders R, Palstra RJ, Stevens M, Kockx C, van IJcken W, et al. 2010. The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. *Genes & Development* **24**: 277–289.

Son EY, Ichida JK, Wainger BJ, Toma JS, Rafuse VF, Woolf CJ, Eggan K. 2011. Conversion of Mouse and Human Fibroblasts into Functional Spinal Motor Neurons. *Cell Stem Cell* **9**: 205–218.

Song M-R, Sun Y, Bryson A, Gill GN, Evans SM, Pfaff SL. 2009. Islet-to-LMO stoichiometries control the function of transcription complexes that specify motor neuron and V2a interneuron identity. *Development* **136**: 2923–2932.

Song S-H, Hou C, Dean A. 2007. A positive role for NLI/Ldb1 in long-range beta-globin locus control region function. *Mol Cell* **28**: 810–822.

Song S-H, Kim A, Dale R, Dean A. 2012. Ldb1 regulates carbonic anhydrase 1 during erythroid differentiation. *Biochim Biophys Acta* **1819**: 885–891.

Stevens TJ, Lando D, Basu S, Atkinson LP, Cao Y, Lee SF, Leeb M, Wohlfahrt KJ, Boucher W, O'Shaughnessy-Kirwan A, et al. 2017. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**: 59–64.

Stifani N. 2014. Motor neurons and the generation of spinal motor neuron diversity. *Front Cell Neurosci* **8**: 293.

Storbeck CJ, Wagner S, O'Reilly P, McKay M, Parks RJ, Westphal H, Sabourin LA. 2009. The Ldb1 and Ldb2 transcriptional cofactors interact with the Ste20-like kinase SLK and regulate cell migration. *Mol Biol Cell* **20**: 4174–4182.

Strickfaden H, Sharma AK, Hendzel M. 2019. A charge-dependent phase transition determines interphase chromatin organization. *bioRxiv* 541086.

Strickfaden H, Zunhammer A, van Koningsbruggen S, Köhler D, Cremer T. 2010. 4D chromatin dynamics in cycling cells: Theodor Boveri's hypotheses revisited. *nucleus* **1**: 284–297.

Suleiman H, Heudobler D, Raschta A-S, Zhao Y, Zhao Q, Hertting I, Vitzthum H, Moeller MJ, Holzman LB, Rachel R, et al. 2007. The podocyte-specific inactivation of Lmx1b, Ldb1 and E2a yields new insight into a transcriptional network in podocytes. *Dev Biol* **304**: 701–712.

Susa T, Sato T, Ono T, Kato T, Kato Y. 2006. Cofactor CLIM2 promotes the repressive action of LIM homeodomain transcription factor Lhx2 in the

expression of porcine pituitary glycoprotein hormone alpha subunit gene. *Biochim Biophys Acta* **1759**: 403–409.

Thaler JP, Koo SJ, Kania A, Lettieri K, Andrews S, Cox C, Jessell TM, Pfaff SL. 2004. A postmitotic role for Isl-class LIM homeodomain proteins in the assignment of visceral spinal motor neuron identity. *Neuron* **41**: 337–350.

Thaler JP, Lee S-K, Jurata LW, Gill GN, Pfaff SL. 2002. LIM factor Lhx3 contributes to the specification of motor neuron and interneuron identity through cell-type-specific protein-protein interactions. *Cell* **110**: 237–249.

Thor S, Andersson SGE, Tomlinson A, Thomas JB. 1999. A LIM-homeodomain combinatorial code for motor-neuron pathway selection. *Nature* **397**: 76–80.

Toyama R, Kobayashi M, Tomita T, Dawid IB. 1998. Expression of LIM-domain binding protein (ldb) genes during zebrafish embryogenesis. *Mechanisms of Development* **71**: 197–200.

Tsuchida T, Ensini M, Morton SB, Baldassare M, Edlund T, Jessell TM, Pfaff SL. 1994. Topographic organization of embryonic motor neurons defined by expression of LIM homeobox genes. *Cell* **79**: 957–970.

Ueki N, Seki N, Yano K, Ohira M, Saito T, Masuho Y, Muramatsu M. 1999. Isolation and chromosomal assignment of human genes encoding cofactor of LIM homeodomain proteins, CLIM1 and CLIM2. *J Hum Genet* **44**: 112–115.

van Meyel DJ, O'Keefe DD, Jurata LW, Thor S, Gill GN, Thomas JB. 1999. Chip and Apterous Physically Interact to Form a Functional Complex during Drosophila Development. *Mol Cell* **4**: 259–265.

van Meyel DJ, O'Keefe DD, Thor S, Jurata LW, Gill GN, Thomas JB. 2000. Chip is an essential cofactor for Apterous in the regulation of axon guidance in Drosophila. *Development* **127**: 1823–1831.

van Meyel DJ, Thomas JB, Agulnick AD. 2003. Ssdp proteins bind to LIM-interacting co-factors and regulate the activity of LIM-homeodomain protein complexes in vivo. *Development* **130**: 1915–1925.

Velasco S, Ibrahim MM, Kakumanu A, Garipler G, Aydin B, Al-Sayegh MA, Hirsekorn A, Abdul-Rahman F, Satija R, Ohler U, et al. 2017. A Multi-step Transcriptional and Chromatin State Cascade Underlies Motor Neuron Programming from Embryonic Stem Cells. *Cell Stem Cell* **20**: 205–217.e8.

Vermunt MW, Zhang D, Blobel GA. 2019. The interdependence of gene-regulatory elements and the 3D genome. *J Cell Biol* **218**: 12–26.

Visvader JE, Mao X, Fujiwara Y, Hahm K, Orkin SH. 1997. The LIM-domain binding protein Ldb1 and its partner LMO2 act as negative regulators of erythroid differentiation. *Proc Natl Acad Sci USA* **94**: 13707–13712.

Waddington CH. 1957. *The Strategy of the Genes*. George Allen and Unwin, London.

Wadman IA, Osada H, Grütz GG, Agulnick AD, Westphal H, Forster A, Rabbitts TH. 1997. The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *The EMBO Journal* **16**: 3145–3157.

Wang Y, Klumpp S, Amin HM, Liang H, Li J, Estrov Z, Zweidler-McKay P, Brandt SJ, Agulnick A, Nagarajan L. 2010. SSBP2 is an in vivo tumor suppressor and regulator of LDB1 stability. *Oncogene* **29**: 3044–3053.

Werner K, Donow C, Pandur P. 2017. Chip/Ldb1 interacts with Tailup/islet1 to regulate cardiac gene expression in Drosophila. *genesis* **141**: e23030.

Wichterle H, Lieberam I, Porter JA, Jessell TM. 2002. Directed Differentiation of Embryonic Stem Cells into Motor Neurons. *Cell* **110**: 385–397.

Wilkinson White L, Gamsjaeger R, Dastmalchi S, Wienert B, Stokes PH, Crossley M, Mackay JP, Matthews JM. 2011. Structural basis of simultaneous recruitment of the transcriptional regulators LMO2 and FOG1/ZFPM1 by the transcription factor GATA1. *Proc Natl Acad Sci USA* **108**: 14443–14448.

Williamson I, Berlivet S, Eskeland R, Boyle S, Illingworth RS, Paquette D, Dostie J, Bickmore WA. 2014. Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. *Genes & Development* **28**: 2778–2791.

Williamson I, Kane L, Devenney PS, bioRxiv EA, Anderson E, Kilanowski F, Hill RE, Bickmore W, Lettice LA, Lettice. 2019. Developmentally regulated Shh expression is robust to TAD perturbations. *Development.* **146**:dev.179523.

Wolberger C. 1996. Homeodomain interactions. *Current Opinion in Structural Biology* **6**: 62–68.

Xu X, Mannik J, Kudryavtseva E, Lin KK, Flanagan LA, Spencer J, Soto A, Wang N, Lu Z, Yu Z, et al. 2007a. Co-factors of LIM domains (Clims/Ldb/Nli) regulate corneal homeostasis and maintenance of hair follicle stem cells. *Dev Biol* **312**: 484–500.

Xu Z, Meng X, Cai Y, Liang H, Nagarajan L, Brandt SJ. 2007b. Single-stranded DNA-binding proteins regulate the abundance of LIM domain and LIM domain-binding proteins. *Genes & Development* **21**: 942–955.

Yamashita T, Agulnick AD, Copeland NG, Gilbert DJ, Jenkins NA, Westphal H. 1998. Genomic structure and chromosomal localization of the mouse LIM domain-binding protein 1 gene, Ldb1. *Genomics* **48**: 87–92.

Zhang F, Tanasa B, Merkurjev D, Lin C, Song X, Li W, Tan Y, Liu Z, Zhang J, Ohgi KA, et al. 2015. Enhancer-bound LDB1 regulates a corticotrope

promoter-pausing repression program. *Proceedings of the National Academy of Sciences* **112**: 1380–1385.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.

Zhao D, Zheng D. 2018. SMARTcleaner: identify and clean off-target signals in SMART ChIP-seq analysis. *BMC Bioinformatics* **19**: 544.

Zhao Y, Flandin P, Vogt D, Blood A, Hermesz E, Westphal H, Rubenstein JLR. 2014. Ldb1 is essential for development of Nkx2.1 lineage derived GABAergic and cholinergic neurons in the telencephalon. *Dev Biol* **385**: 94–106.

Zhao Y, Kwan K-M, Mailloux CM, Lee W-K, Grinberg A, Wurst W, Behringer RR, Westphal H. 2007. LIM-homeodomain proteins Lhx1 and Lhx5, and their cofactor Ldb1, control Purkinje cell differentiation in the developing cerebellum. *Proc Natl Acad Sci USA* **104**: 13182–13186.

Zheng X, Hu J, Yue S, Kristiani L, Kim M, Sauria M, Taylor J, Kim Y, Zheng Y. 2018. Lamins Organize the Global Three-Dimensional Genome from the Nuclear Periphery. *Mol Cell* **71**: 802–815.e7.

Zhong Z, Ma H, Taniguchi-Ishigaki N, Nagarajan L, Becker CG, Bach I, Becker T. 2011. SSDP cofactors regulate neural patterning and differentiation of specific axonal projections. *Dev Biol* **349**: 213–224.