

**The role of the CECR2 chromatin remodeling complex in
mouse developmental gene regulation**

by

Alaina Nicole Terpstra

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Molecular Biology and Genetics

Department of Biological Sciences
University of Alberta

© Alaina Nicole Terpstra, 2018

Abstract

Early mammalian development cannot progress without targeted temporal and spatial expression of genes. Changing the accessibility of DNA to transcriptional machinery is one critical way gene expression is controlled. This process, known as chromatin remodeling, is vital for formation of the embryonic neural tube and for spermatogenesis. *Cecr2* is strongly expressed in the embryo, testicular germ-line stem cells, and embryonic stem (ES) cells. Mutations in mouse *Cecr2* can lead to either the lethal neural tube defect exencephaly, analogous to human anencephaly, or a non-lethal subfertility phenotype.

The CECR2 protein forms a chromatin remodeling complex named CERF (CECR2-containing Remodeling Factor) and immunoprecipitation analysis has revealed tissue-specific binding partners. In both ES cells and testis CECR2 binds SNF2H, a protein that drives remodeling. CECR2 interacts with LUZP1, a protein required for neurulation, in ES cells but not in the testis. CCAR2 was a recently identified member of CERF in ES cells, but its presence in the testis CERF complex was equivocal. To confirm and clarify the newest CERF complex members, LUZP1 and CCAR2, I used immunofluorescence to investigate spatial localization in testis and ES cells with CECR2. While results for LUZP1 were mainly inconclusive due to assay difficulties, I showed that CCAR2 is not a part of the CERF protein complex in testis as it is not present in the same cell type as CECR2.

To understand the transcriptional effects of CECR2 loss, I used RNA sequencing (RNA-Seq). This identifies any misregulated genes, potentially revealing candidate genes or signaling pathways in *Cecr2* mutant mice which may be involved in the exencephalic

phenotype. It also allows us to compare to a previously performed chromatin immunoprecipitation sequencing (ChIP-Seq) experiment which identified overlapping genomic binding sites for CECR2, SNF2H, and/or LUZP1 in ES cells. We hypothesize that CERF may be regulating gene expression as a chromatin remodeller. Together, these results may be used as a powerful comparative approach to determine if CECR2 acts as a direct transcriptional regulator.

RNA-Seq on neurulating wildtype and mutant embryo heads revealed 143 misregulated genes, 102 of which were upregulated. I validated using qRT-PCR *Aldh1a2*, *Cdh7*, and *Cyp26c1* as significantly downregulated, and *Dbx1* and *Kcna6* as significantly upregulated. *Cdh7*, *Dbx1*, and *Kcna6* all had presumptive binding sites for CECR2 in the ChIP-Seq experiment. *Aldh1a2* and *Cyp26c1* are both involved in retinoic acid signaling, and mutations in *Aldh1a2* cause exencephaly. Perhaps CERF plays a role in this important signaling pathway.

I also looked at expression of ChIP-Seq candidate genes across many time points using qRT-PCR. My analysis found that *Alx1*, *Lrp6*, and *Gli2* are significantly downregulated, and *Nf1* is significantly upregulated in the neurulating embryo head. Mutations in all these genes leads to exencephaly, though whether this misregulation is real or an artifact since it was not seen in my RNA-Seq requires further analysis. I showed that the two transcription factors analyzed, *Alx1* and *Dbx1*, were the only genes misregulated at all time points – from 12-14 somites, to 21+ somites. This may suggest that CECR2 is involved in the regulation of transcription factors, or factors which influence their transcription.

I also found that *ApoB* is significantly upregulated 3.5-fold in the 15-16 somite mutant embryo head, right after neurulation. This gene was also upregulated 2-fold in my RNA-Seq in the neurulating embryo head. *ApoB* is an interesting candidate gene as homozygous mutants develop exencephaly. APOB is critical for clearing low density lipoprotein (LDL)/very low-density lipoprotein (VLDL) from the blood plasma, and I suggest that perhaps this leads to a decrease in LRP6 receptors available to bind the WNT ligand and initiate PCP signaling. *Lrp6* knockout mutants develop exencephaly, so it would be interesting to investigate the levels of LDL/VLDL in *Cecr2* mutants.

Overall, my research contributes to the overarching theme of cranial neurulation: it is a complex, dynamic process requiring incredible spatiotemporal coordination. Loss of *CECR2* has a considerable impact in the mouse embryo, resulting in failure of the neural tube to close. I have provided some interesting candidate genes for further investigation, with our goal to elucidate how disruption of this process leads to lethal neural tube defects.

Preface

The research project involving mice, of which this thesis is a part, received research ethics approval from the Animal Care and Use Committee of the University of Alberta, University of Alberta AUP 00000094.

Some of this research was done in collaboration:

Section 2.9.5, 3.4, and 3.5: Analysis of RNA-Sequencing data and generation of heat maps was performed by Arun Kommandath in collaboration with Dr. Paul Stothard's Lab.

Section 3.3: An undergraduate student, Jennifer Just, helped with the selection of testis ChIP-Seq candidate genes.

Figure B1: Figure was generated by Marc Parsons with data I provided. I also made some minor edits after the graph was created.

Acknowledgments

I would first like to thank Dr. Heather McDermid for being an exceptional supervisor and mentor to me throughout my time in her lab. She is incredibly kind, empathetic, ambitious, and patient. She fosters a working environment conducive to learning, exploration, curiosity, and empowerment. I without a doubt would not be where I am today without her, as she has supported me in all aspects of my life and career. I am forever grateful for her unrelenting encouragement, seemingly endless reference letters, late-night edits, and emails, feminist and political discussions, and morning crossword answers.

I would also like to extend my thanks to my committee, Dr. Frank Nargang and Dr. Paul Stothard, for their expertise, guidance and patience. Their support and suggestions have been invaluable in my project. I would also not have been able to navigate qRT-PCR without the help of Troy Locke. He answered unrelenting questions with patience and clarity, and provided comfort in an otherwise challenging task. My level of comprehension and skills obtained are a direct result of his mentorship and I am incredibly grateful.

This degree would not have been possible without my labmates and colleagues who fostered an environment that made every day enjoyable. The network of support they provided for both successes and failures made graduate school an overwhelmingly positive experience. I've learned so much from them and gained incredible friendships. I'd like to specially recognize Kacie Norton, who shared a bench with me and many bottles of wine. Farshad Niri also provided mentorship and late-night conversations over western blots. From neighbouring labs, Francesca Jean and Sonya Widen offered unrelenting support and pictures of cats. Finally, thanks goes to my family for their love, care, and genuine interest in my research. I truly would not have survived or flourished without these people.

Last, I would like to recognize the National Science and Engineering Research Council (NSERC), Alberta Innovates Technology Futures (AITF), and the Province of Alberta's QEII Graduate Scholarship for funding me throughout my degree. I'd also like to thank the University for the opportunity to be a Teaching Assistant, and to Mark Wolansky for his enthusiasm and dedication as a facilitator.

Table of Contents

Chapter 1 – Introduction	1
1.1 Neurulation	1
1.1.1 Overview	1
1.1.2 Genetics and mechanisms of primary neurulation	3
1.1.3 Defects in neurulation	4
1.2 Chromatin	9
1.2.1 Chromatin structure	9
1.2.2 Chromatin remodeling and gene regulation	10
1.2.3 ATP-dependent chromatin remodeling.....	12
1.2.4 Imitation switch (ISWI): a sub-family of ATP-dependent chromatin remodellers	13
1.2.5 ISWI chromatin remodellers and mammalian development	15
1.3 Cat eye syndrome critical region candidate 2	17
1.3.1 <i>Cecr2</i> structure	17
1.3.2 <i>Cecr2</i> mutations.....	18
1.3.3 Phenotypes of <i>Cecr2</i> mutants	19
1.3.4 <i>Cecr2</i> expression and localization	21
1.3.5 CERF: CECR2-containing remodeling factor	21
1.3.6 Genomic binding sites of CERF	24
1.4 Purpose and Aims	27
Chapter 2 – Materials & Methods	29
2.1 Maintaining the mouse colony	29
2.1.1 Housing.....	29
2.1.2 Breeding.....	29
2.1.3 Euthanasia	30
2.2 Genotyping.....	30
2.2.1 Genomic DNA extraction	30
2.2.2 PCR reactions.....	30
2.2.3 <i>Cecr2^{Del}</i> genotyping.....	30
2.2.4 <i>Cecr2^{GT}</i> genotyping	31
2.2.5 SRY genotyping	31
2.2.6 Agarose gel electrophoresis	31
2.3 Cell culture of mouse embryonic stem (ES) cells.....	31
2.3.1 Freezing down cells	32
2.4 Immunofluorescence on testis	32
2.4.1 Collecting and preparing testis samples.....	32
2.4.2 Immunofluorescence using a single antibody.....	33

2.4.3 Colocalization immunofluorescence	33
2.5 Immunofluorescence on mouse embryonic stem (ES) cells	34
2.5.1 Preparing cells	34
2.5.2 Colocalization immunofluorescence	34
2.6 Immunofluorescence of <i>Cecr2</i> transfected Caco-2 cells	35
2.7 Vaginal cytology	36
2.7.1 Vaginal swabbing	36
2.7.2 Staining of vaginal swabs	36
2.7.3 Analysis of vaginal swabs	36
2.8 Collection of tissues for RNA extraction	38
2.8.1 Breeding for E9.5 embryos	38
2.8.2 Dissection of E9.5 embryo heads	38
2.8.3 Dissection of testis	39
2.9 Analysis of gene expression	39
2.9.1 Individual embryo head RNA extraction	39
2.9.2 Testis RNA extraction	40
2.9.3 cDNA synthesis	40
2.9.4 qRT-PCR	40
2.9.5 RNA sequencing	41
2.10 Protein analysis of the CERF complex	42
2.10.1 Whole lysate protein extraction	42
2.10.2 Western Blotting	42
Chapter 3 - Results	44
<i>Objective 1: Confirm and clarify the CERF protein complex composition</i>	
3.1 CECR2 is present in the nucleus with CCAR2 and LUZP1 in mouse ES cells	44
3.2 CECR2 does not colocalize with CCAR2 in the testis and LUZP1 colocalization is inconclusive	49
3.3 ChIP-Seq candidate genes	53
<i>Objective 2: Investigate the role of CECR2 as a transcriptional regulator in neurulation</i>	
3.4 RNA-Seq reveals a range of potentially misregulated genes in <i>Cecr2^{Del/Del}</i> mutants	54
3.5 Very few binding sites identified in ChIP-Seq correspond to differentially expressed genes identified in RNA-Seq	67
3.6 Surprising misregulation of ChIP-Seq candidate genes discovered in <i>Cecr2^{Del/Del}</i> embryos	72
3.7 qRT-PCR results validate RNA-Seq and shows temporal gene expression changes	78
Chapter 4 – Discussion	81
4.1 CECR2 interactions with CCAR2 and LUZP1 in the nucleus of ES cells are difficult to demonstrate using immunofluorescence	81
4.2 CECR2 does not appear to interact with LUZP1 or CCAR2 in the testis	83
4.3 RNA-sequencing analysis of the neurulating transcriptome	84
4.4 The function of CECR2 as an ATP-dependent chromatin remodeller is still unclear	89

4.5 qRT-PCR results suggest CECR2 may regulate transcription factors	91
4.6 Some RNA-Seq validated genes may play a small role in the exencephalic phenotype, or are downstream of this defect.....	94
4.7 <i>ApoB</i> and lipid metabolism are interesting candidates for involvement with an exencephalic phenotype.....	96
4.8 Summary and future directions	100
Literature Cited.....	103
Appendices	116
Appendix A – List of primers.....	116
Appendix B – Vaginal cytology analysis	117

List of Tables

Table 1.1.3.1. Epigenetic regulators crucial for neurulation in animal models of NTDs.	8
Table 1.3.5.1. CERF protein complex members based on immunoprecipitation results.	24
Table 1.3.6.2. Genes that have CECR2 binding sites within 5 kb upstream of TSS in both ES cells and testis.	26
Table 3.4.1. RNA sequencing information for submitted samples.	61
Table 3.4.2. Top 10 differentially expressed genes filtered by FDR from RNA-sequencing of five <i>Cecr2</i> ^{+/+} wildtype and five <i>Cecr2</i> ^{Del/Del} mutant embryo heads.	62
Table 3.4.3. Top 10 differentially expressed genes filtered by fold change from RNA-sequencing of five <i>Cecr2</i> ^{+/+} wildtype and five <i>Cecr2</i> ^{Del/Del} mutant embryo heads.	63
Table 3.4.4. Top 10 differentially expressed genes filtered by FDR from RNA-sequencing of five <i>Cecr2</i> ^{+/+} wildtype and four <i>Cecr2</i> ^{Del/Del} mutant embryo heads (outlier removed).	64
Table 3.4.5. Top 10 differentially expressed genes filtered by fold change from RNA-sequencing of five <i>Cecr2</i> ^{+/+} and four <i>Cecr2</i> ^{Del/Del} mutant embryo heads (outlier removed).	65
Table 3.4.6. Enriched gene ontology (GO) biological process terms in RNA sequencing data on <i>Cecr2</i> ^{+/+} and <i>Cecr2</i> ^{Del/Del} embryo heads.	66
Table 3.6.1. Comparison of genes from an RNA-Seq experiment and with qRT-PCR analysis.	77
Table 3.7.1. Genes validated using qRT-PCR from an RNA-sequencing experiment performed on <i>Cecr2</i> ^{+/+} and <i>Cecr2</i> ^{Del/Del} embryo heads.	80
Table 4.3.1. List of genes involved with Hedgehog, Wnt, PCP, BMP Signaling and associated developmental processes.	86
Table B1. Mouse estrous stage at each analysis day.	123
Table B2. Pregnancy analysis of plugged female mice.	124

List of Figures

Figure 1.1.1.1. Morphogenesis of neurulation.	2
Figure 1.1.1.2. Sites of initiation and closure during vertebrate neurulation.	2
Figure 1.1.3.3. Failure of closure site 2 to fuse in mice results in exencephaly.	7
Figure 1.3.1.1. <i>Cecr2</i> gene diagram.	18
Figure 2.7.3.1. Vaginal cytology and identification tools of the mouse estrous cycle.	37
Figure 2.8.2.1. Cut site of E9.5 mouse embryo.....	39
Figure 3.1.1. Immunofluorescence of mouse embryonic stem cells.....	47
Figure 3.1.2. Immunofluorescence of transiently transfected CaCo2 epithelial cells with human <i>CECR2</i> shows antibody specificity.	48
Figure 3.2.1. Immunofluorescence on paraffin sections of mouse testis.....	51
Figure 3.2.2. Localization of LUZP1 in mouse testis using immunofluorescence.	52
Figure 3.4.1. Visualization of E9.5 embryonic neural tube at the cranial region.....	57
Figure 3.4.2. Multi-dimensional scaling (MDS) plot of RNA-sequencing samples.....	58
Figure 3.4.3. Heat map of RNA-sequencing samples.....	59
Figure 3.4.4. Heat map of RNA-sequencing samples with outlier MT_02 removed.	60
Figure 3.5.1. Heat map of Top 104 genes from ChIP-Seq clusters wildtype and mutant samples together indicating a mild expression trend.	69
Figure 3.5.2. Heat map of all genes with a <i>CECR2</i> binding site clusters wildtype and mutant RNA-Seq samples.....	70
Figure 3.6.1. qRT-PCR analysis of ChIP-Seq candidates genes during and after neurulation in <i>Cecr2</i> ^{+/+} and <i>Cecr2</i> ^{Del/Del} mouse embryo heads.	74
Figure 3.6.2. ChIP-Seq and microarray candidate gene expression throughout early mouse embryonic development.	75
Figure 3.6.3. A possible sex difference in the expression of <i>Nf1</i> and <i>Lrp6</i> is observed in <i>Cecr2</i> ^{Del/Del} embryo heads when compared to wildtype with qRT-PCR analysis.	76
Figure 3.7.1. RNA-Seq candidate gene expression throughout early mouse embryonic development. ...	80
Figure 4.5.1. Gene expression throughout early mouse embryonic development.....	93
Figure 4.5.2. <i>CECR2</i> may occupy the region upstream of <i>Alx1</i> in ES cells.	94
Figure B1. Estrous cycling of 6 representative mice.....	121
Figure B2. Representative vaginal swab analysis for mouse pregnancy prediction.....	122

List of Abbreviations

°C	Degrees Celsius
aa	Amino acid
ACF	ATP-utilizing chromatin assembly and remodeling factor
ADP	Adenosine diphosphate
Aldh1a2	Aldehyde dehydrogenase family 1, subfamily A2
Alx1	ALX homeobox 1 (aka Cart1)
Apob	Apolipoprotein B
AT-hook	Adenosine thymine DNA binding hook
ATAC	Ada-two-a-containing
ATAC-Seq	Assay for transposase-accessible chromatin using sequencing
ATP	Adenosine triphosphate
BAF	BRG1-associated factor
Baf155	BRG1-associated factor 155
bam	Binary format for storing sequence data
BMP	Bone morphogenetic protein
bp	Base pair
Brg1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily A, member 4 (aka Smarca4)
BSA	Bovine serum albumin
BVC	Biological coefficient of variation
Caco-2	Caucasian colon adenocarcinoma cell line (human)
Cart1	Cartilage paired-class homeoprotein 1 (aka Alx1)
Casp9	Caspase 9
Cdh7	Cadherin 7, type 2
cDNA	Complementary DNA
Cecr2	Cat eye syndrome chromosome region, candidate 2
CERF	CECR2-containing remodeling factor
CESCR	Cat eye syndrome critical region
CHD	Chromodomain helicase DNA-binding
ChIP-Seq	Chromatin immunoprecipitation with massively parallel DNA sequencing
CHRAC	Chromatin accessibility complex
co-IP	Co-immunoprecipitation (reciprocal)
CPM	Counts per million
CT	Cycle threshold
CT45	Cancer/testis antigen 45
CUT&RUN	Cleavage under targets and release using nuclease
Cyp26a1	Cytochrome P450 family 26 subfamily A polypeptide 1
Cyp26c1	Cytochrome P450 family 26 subfamily C polypeptide 1

D. melanogaster	Drosophila melanogaster
DAPI	4',6-diamidino-2-phenylindole
Dbx1	Developing brain homeobox 1
DDT	DNA binding homeobox and different transcription factors
DEPC	Diethylpyrocarbonate
DLHP	Dorso-lateral hinge point
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
E	Embryonic day
EDTA	Ethylenediaminetetraacetic acid
Elmo2	Engulfment and cell motility 2
ES	Embryonic stem
FAB	Antigen-binding fragment
FC	Fold change
FDR	False discovery rate
FGF	Fibroblast growth factor
Fgl2	Fibrinogen-like protein 2
FH β	Familial hypobetalipoproteinemia
Foxg1	Forkhead box G1
FRET	Förster/fluorescence resonance energy transfer
g	Gram
Gli2	GLI-Kruppel family member GLI2
Gli3	GLI-Kruppel family member GLI3
GO	Gene ontology
GT	Genetrap
H2A	Histone H2A
H2A.Z	Histone H2A.Z (variant of histone H2A)
H2B	Histone H2B
H3	Histone H3
H4	Histone H4
HCl	Hydrogen chloride
HDAC1	Histone deacetylase 1
HDL	High density lipoprotein
HEK293	Human embryonic kidney 293 cell line
HRP	Horseradish peroxidase
HSA	Helicase-SANT-associated domain
HSS	HAND-SANT-SLIDE
IF	Immunofluorescence

IgG	Immunoglobulin G
IGV	Integrative genomics viewer
INDEL	Insertion/deletion
IP	Immunoprecipitation
ISWI	Imitation switch
kb	Kilobase
Kcna6	Potassium voltage-gated channel, shaker-related, subfamily, member 6
kDa	Kilodalton
KEGG	Kyoto encyclopedia of genes and genomes
L	Litre
LDL	Low density lipoprotein
LDS	Lithium dodecyl sulfate
LoxP-Cre	Locus of X-over P1-causes-recombination
Lpar1	Lysophosphatidic acid receptor 1
Lrp6	LDL receptor related protein 6
Luzp1	Leucine zipper protein 1
M	Molar
M	Merge
MBSU	Molecular biology services unit
MDa	Megadalton
MDS	Multi-dimensional scaling
MECO	Meta-coactivator complex
MED	Mediator coactivator
MENA	Mammalian enabled (enah)
mg	Milligrams
MgCl ₂	Magnesium chloride
MGI	Mouse genome informatics
MHP	Medial hinge point
mL	Millilitre
mM	Millimolar
mm9	UCSC mouse genome assembly build 9
Mnase	Micrococcal nuclease
mRNA	Messenger ribonucleic acid
MT	Mutant
n	Sample number
NaCl	Sodium chloride
NaOAc	Sodium acetate
NCoR	Nuclear receptor co-repressor
Nf1	Neurofibromatosis type 1

ng	Nanograms
NGS	Next generation sequencing
nm	Nanometers
NoRC	Nucleolar remodeling complex
NTD	Neural tube defect
NURD	Nucleosome remodeling and deacetylase complex
NURF	Nucleosome remodeling factor
P	P-value
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PBS-T	Phosphate buffered saline with tween-20
PCP	Planar cell polarity
PCR	Polymerase chain reaction
Pfn1	Profilin 1
pGT1	Dual-tagging gene trap vector pGT1
Phactr4	Phosphatase and actin regulator 4
PTM	Post translational modification
PVDF	Polyvinylidene fluoride
qRT-PCR	Quantitative real-time PCR
RefSeq	Reference sequence database
RhoA	Ras homolog family member A
rhPCR	RNase H-dependent PCR
RIN	RNA integrity number
RNA	Ribonucleic acid
RNA Pol II	RNA polymerase II
RNA-Seq	RNA extraction with massively parallel DNA sequencing
RNAi	RNA interference
RPM	Revolutions per minute
RSF1	Remodeling and spacing factor 1
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SANT	Yeast SWI3, yeast ADA2, human ncor, human TFIIB
SASS	Sciences animal support services
SDS	Sodium dodecyl sulfate
Shh	Sonic hedgehog
Shroom3	Shroom family member 3
SIRT1	Sirtuin 1
SLIDE	SANT-like ISWI domain
SMAD	Sma and Mad related protein

Smarca1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily A, member 1 (aka SNF2L)
Smarca5	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily A, member 5 (aka SNF2H)
SMRT	Silencing mediator for retinoid and thyroid hormone receptors
SNF2	Sucrose non-fermenting 2
Snf2h	Sucrose non-fermenting 2-homolog
Snf2l	Sucrose non-fermenting 2-like
SRY	Sex determining region Y
SWI/SNF	Switching defective/Sucrose non-fermenting
TBP	TATA-binding protein
TBS	Tris-buffered saline
TBST	TBS with 0.1% Tween-20
TE	Tris EDTA
TEMED	Tetramethylethylenediamine
TIF1- α	Transcriptional intermediary factors 1-alpha
Tris	Tris (hydroxymethyl)ethylaminomethane
Tris-Cl	Tris base HCL
TSS	Transcription start site
Twist	Twist basic helix-loop-helix transcription factor
U	Unit
UTR	Untranslated region
V	Volts
VANGL2	VANGL planar cell polarity protein 2
VASP	Vasodilator-stimulated phosphoprotein
VLDL	Very low-density lipoprotein
WICH	WSTF-ISWI chromatin remodeling complex
Wnt	Wingless type
WSTF	Williams syndrome transcription factor (aka BAZ1B)
WT	Wildtype
X-gal	5-bromo-4-chloro-indolyl-%-D-galactopyranoside
β -Gal	Beta-galactosidase
μ g	Micrograms
μ l	Microliters
μ M	Molar concentration in micromoles/litre
μ m	Micrometer

Chapter 1 – Introduction

1.1 Neurulation

1.1.1 Overview

Early embryonic development is marked by rapid growth and dynamic morphogenesis, transforming the one-cell zygote into a well-defined multicellular organism. Neurulation is a critical and complex part of early embryonic vertebrate development, comprised of multiple coordinated morphogenic steps leading to the formation of the brain and spinal cord (reviewed in Greene & Copp, 2009). During embryonic gastrulation, the ectoderm is induced to form a one cell thick neuroepithelium; this forms the neural plate (Figure 1.1.1.1.) (reviewed in Copp, 2005; Greene & Copp, 2014). The neural plate undergoes more morphogenic changes creating neural folds which rise and eventually fuse to form the neural tube. This is termed primary neurulation and is completed in the mouse by day 10.5 of development (Embryonic or E10.5, where the morning post-conception is considered E0.5). Secondary neurulation is also required to form an epithelial rod at the most caudal/sacral end of the spinal cord, where its lumen connects with the neural tube. The focus of this thesis is on primary neurulation.

In mice and humans, primary neurulation has multiple sites of initiation and closure. This process occurs discontinuously, with three separate closure points in mice and likely only two in humans (Figure 1.1.1.2.)(Greene & Copp, 2009.) Closure 1 occurs at the hindbrain/cervical boundary at around E8.5, closure 2 at the forebrain/midbrain boundary at approximately E9.5, and closure 3 at the rostral forebrain at approximately E9.5. The closing of closure 1 and 2 occur bidirectionally, with closure 3 only closing in the caudal direction. As this process occurs, the regions between the closure sites termed neuropores gradually shorten until closure is complete (Greene & Copp, 2009; 2014). Interestingly the position and presence of closure 2 in mice is strain-dependent and highly polymorphic while the presence of closure site 2 in humans is controversial (reviewed in Copp, 2005; Copp, Greene, & Murdoch, 2003; Greene & Copp, 2009).

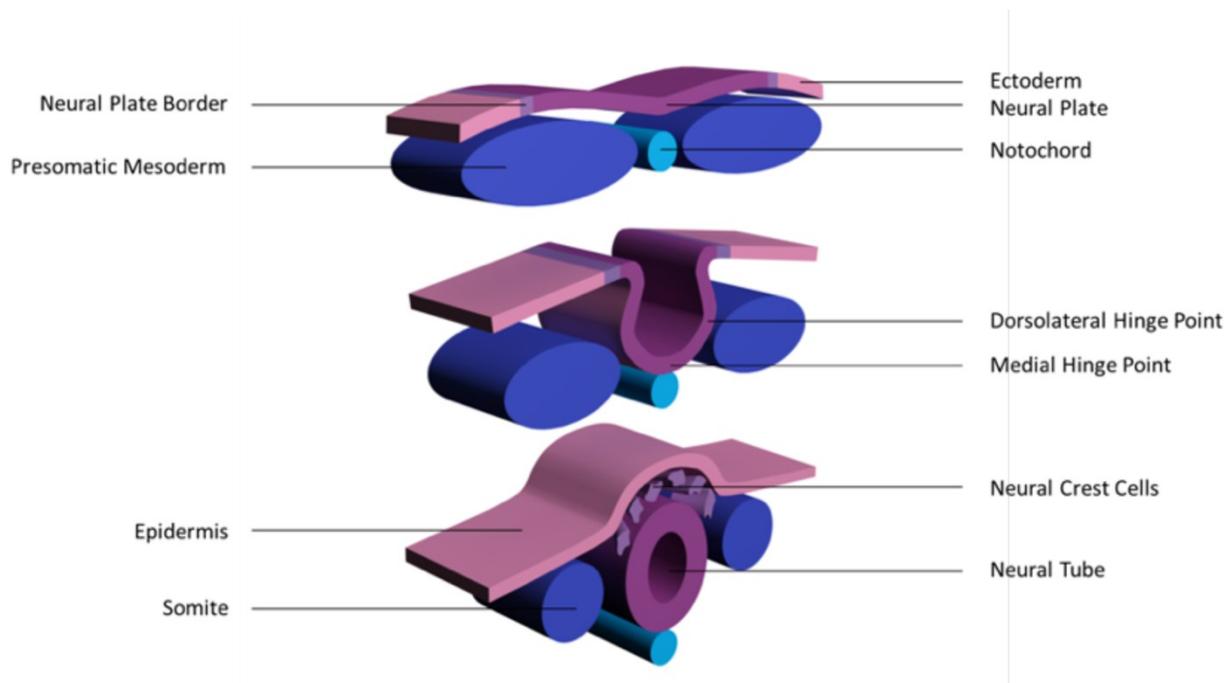


Figure 1.1.1.1. Morphogenesis of neurulation. Vertebrate neurulation is a complex morphogenic process where the neural plate thickens and raises, forming the medial hinge point and the dorsolateral hinge points. These hinge points bend the plate and the edges fuse, forming the neural tube and overlying ectoderm. Figure provided by Singh, 2016.

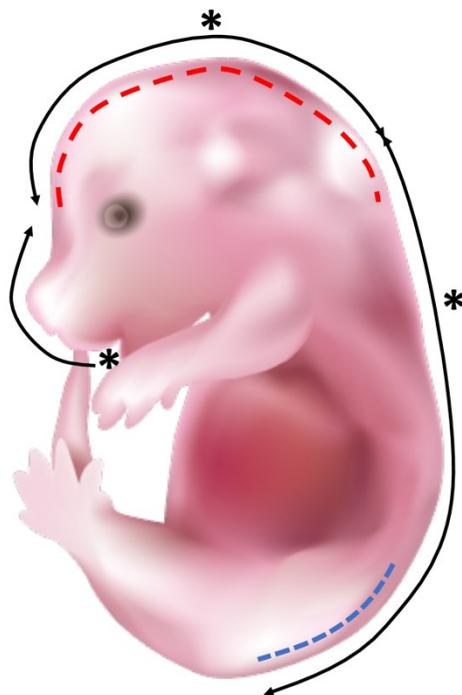


Figure 1.1.1.2. Sites of initiation and closure during vertebrate neurulation. From caudal to rostral (right to left) the asterisks (*) indicate closure sites 1, 2, and 3 respectively. Closure 1 occurs at the hindbrain/cervical boundary, closure 2 at the forebrain/midbrain boundary, and closure 3 at the rostral forebrain. The presence of closure site 2 is controversial in humans. The black line indicates direction of closure once initiation has occurred and indicates the region of craniorachischisis where the neural tube remains open along the body axis (see 1.1.3). The red dotted line indicates the region of cranial neural tube defects, such as exencephaly; the blue dotted line indicates the region of spinal neural tube defects, such as spina bifida. Photo modified from: DBCLS 統合TV [CC BY 4.0 (<https://creativecommons.org/licenses/by/4.0>)], via Wikimedia Commons.

1.1.2 Genetics and mechanisms of primary neurulation

Primary neurulation can be categorized into three step-wise processes: Initiation and induction, apposition of neural folds, and fusion of neural folds (reviewed in Copp et al., 2003). The first step is the well-conserved induction of the embryonic ectoderm to diverge into either neural or non-neural tissue (reviewed in Greene & Copp, 2009; H.K. Lee, Lee, & Moody, 2014). In the presence of bone morphogenic proteins (BMPs), and wingless type (Wnt) signalling, the ectoderm has a non-neural epidermal fate; in their absence and along with fibroblast growth factor (FGF) signalling, it forms the neuroepithelium (neural plate). Neural plate induction continues by a process termed convergent-extension, whereby there is a medial-lateral displacement of cells towards the midline of the neural plate (convergence), and a rostro-caudal elongation of the neural plate (extension). This process is largely governed by non-canonical Wnt signalling, also known as the planar cell polarity (PCP) pathway. Disruptions of an array of PCP signalling molecules in mice produce multiple NTDs discussed in 1.1.3, demonstrating its importance in neurulation.

As the neural plate concomitantly narrows and lengthens via convergent-extension, several other processes work in conjunction to form and raise the neural folds at the lateral edges of the neural plate (Copp, 2005; Greene & Copp, 2009). Notably, there is an increase in extracellular matrix molecules and space that accompanies cranial mesenchymal proliferation and expansion, pushing the folds upwards. The future lumen of the neural tube is also bordered with actin, which may actively constrict the folds together or be a stabilizing force by providing structure for polarized cell motility via the PCP pathway.

Two important structures are also being formed in this region – the medial hinge point (MHP) and two dorsal-lateral hinge points (DLHPs) which bend the epithelium (Figure 1.1.1.1.). This process is governed by factors secreted from the notochord and neuroepithelium itself (reviewed in Copp et al., 2003; Greene & Copp, 2009). Principally, sonic hedgehog (Shh) and BMP signalling are critical for establishing the MHP and DLHPs. SHH is secreted from the notochord and promotes MHP formation at the central region of the neural plate to which it is in close proximity. SHH inhibits Noggin, which in turn inhibits BMP signalling. In the presence of SHH, there is a decrease of Noggin which allows a BMP

effector (BMP2) to inhibit DLHP formation. At the tips of the neural folds most distal to the notochord, there is an increase in Noggin expression which overcomes Shh repression, thereby repressing BMP2 which allows for DLHP formation. With the MHP and two DLHPs properly formed, this brings the neural folds in parallel and in close proximity (apposition), promoting the final step of primary neurulation – fusion – to occur.

Fusion is a critical step for the formation of the neural tube. After fusion, a final remodeling occurs to create the surface ectoderm (the future epidermis), and the inner neural tube (reviewed in Copp, 2005; Greene & Copp, 2009). Unfortunately fusion is difficult to study separate from apposition of the neural folds because unfused folds splay apart quickly, mimicking defects seen when the neural folds fail to appose (Copp et al., 2003). Nevertheless, there are a few theories on the mechanism of fusion. Cellular protrusions have been observed on the cells at the tips of the neural folds, resembling filopodia, and cell-adhesion molecules are present in this region (Copp et al., 2003; Greene & Copp, 2014). Both of these would contribute to epithelial fusion. Alternatively, the rate of apoptosis in that region or the emigration of neural crest cells to the region between the newly formed epidermis and neural tube may be involved. Most likely, it is a combination of factors that leads to the successful fusion of the neural fold to complete primary neurulation.

1.1.3 Defects in neurulation

In both humans and mice, disruption of either induction and morphogenesis of the neural plate or elevation and fusion of the neural folds may lead to failure of any or all sites to close. Consequently, this leads to neural tube defects (NTDs). Though neurulation is similar between mice and humans, NTD inheritance does differ. In humans, neurulation has incredibly complex genetics and is also heavily influenced by environmental factors (reviewed in Wilde, Petersen, & Niswander, 2014). In mice there are over 300 genetic mutations that cause NTDs in inbred strains, however differences in genetic backgrounds and environmental factors still play a role in inheritance (reviewed in Leduc, Singh, & McDermid, 2017). The mouse therefore provides an excellent model to study neurulation and NTDs (reviewed in Zohn, 2012). Not only is the process of neurulation well conserved, allowing genetic and mechanistic studies, but it also provides a system to study in a

controlled setting the multifactorial nature of NTDs which includes but is not limited to modifier genes, maternal nutrition, diabetes, obesity, and sex.

Genetic mutations in the mouse can affect all points of neurulation – from initiation to fusion. Disruptions at any point in the process can cause a myriad of defects (Greene & Copp, 2009)(Figure 1.1.1.2). When closure fails caudally at the posterior neuropore, spina bifida occurs. In mice, this mutation is lethal. While this NTD is compatible with life in humans, those affected have permanent disabilities. When closure site 1 fails to fuse, both mice and humans develop craniorachischisis. This defect is severe, leading to an open neural tube along the axis of the body (Greene & Copp, 2009). Interestingly, most mouse mutants known to date that develop craniorachischisis have genetic defects in the PCP pathway.

When the mouse cranial closure site 2 fails to fuse, exencephaly develops as the skull vault fails to form leaving the neuroepithelium exposed to amniotic fluid (Greene & Copp, 2009). Not only does the brain develop abnormally, but the amniotic fluid also eventually damages and degrades the forming neural tissue. Interestingly, neurulation is not coupled to neurogenesis (Copp, 2005). Neuronal differentiation and synaptic connections still occur but are halted by amniotic degeneration. In humans, exencephaly is known as anencephaly, as fetuses are born without a developed brain or skull cap (Greene & Copp, 2014). Anencephaly represents more than half of NTDs observed and is lethal; fetuses rarely live more than a few hours.

Cranial NTDs such as exencephaly can arise from defects in a variety of molecular mechanisms. Knock-outs of *Twist* and *Cart1* (also known as *Alx1*) cause mice to develop cranial NTDs due to abnormal mesenchymal proliferation and expansion (Chen & Behringer, 1995; Zhao, Behringer, & deCrombrugge, 1996). Exencephaly is also observed in mutants lacking actin-binding proteins such as SHROOM3 (Hildebrand & Soriano, 1999) and vinculin (Xu, Baribault, & Adamson, 1998), along with cytoskeletal double mutants in *Mena* and *Pfn1* (Lanier et al., 1999). Disrupting the actin cytoskeleton chemically can also cause cranial NTDs (Copp, 2005). Interestingly, although spina bifida can be comorbid with exencephaly, spinal closure is normal in the mutants described

above. Disrupting Shh signaling also can result in NTDs. *Shh* null mutants do not form a MHP and instead all DLHPs, however the neural tube still closes (Ybot-Gonzalez, Cogran, Gerrelli, & Copp, 2002). Cranial NTDs do arise though when *Shh* is overexpressed, and with mutations in *Gli3* a negative regulator of Shh signalling (Echelard et al., 1993). Although teasing apart the mechanism of fusion from elevation is difficult, mice with null mutations of a cell-surface ephrin-A5 ligand develop exencephaly as do mice with mutations in the ephrin receptor Eph7A (Holmberg, Clarke, & Frisén, 2000). Other mutations such as a *Casp9* and *ApoB* knockouts cause a respective decrease and increase in apoptosis which both result in cranial NTDs (Homanics et al., 1995; Kuida et al., 1998). Exencephaly is clearly genetically heterogeneous – these are only a few of the hundreds of genetic mutations that can lead to exencephaly (Copp, 2005; Copp et al., 2003). Intriguingly, cranial closure is also susceptible to teratogens and/or the environment. Overall, this highlights the complexity and multitude of processes that contribute to cranial neural tube closure.

Along with disruptions in essential signaling pathways and mechanisms of neurulation, epigenetic disruptions can also cause NTDs (Wilde et al., 2014). For example, mutations in the ATP-dependent chromatin remodeling BAF complex members BRG1 and BAF155 result in NTDs. When either *Brg1* or *Baf155* are heterozygous for a loss of function or hypomorphic allele, respectively, it results in exencephaly (Bultman et al., 2008; Harmacek et al., 2014). Chromatin remodellers like the BAF complex are crucial for repositioning nucleosomes, thereby regulating genomic access for transcription (see 1.2.2 for more details). Epigenetic regulators can also function together, creating either functional redundancy or more likely initiating a cascade of disruption when one regulator is mutated. This may be particularly essential in cranial closure, as functional loss of an epigenetic modifier often results in exencephaly (Table 1.1.3.1.) (Wilde et al., 2014). Moreover, epigenetic regulators are responsible for creating a transcriptional profile, and profile variability can affect penetrance of phenotypes such as NTDs. Therefore, tight regulation of the chromatin architecture is crucial for neurulation to occur.

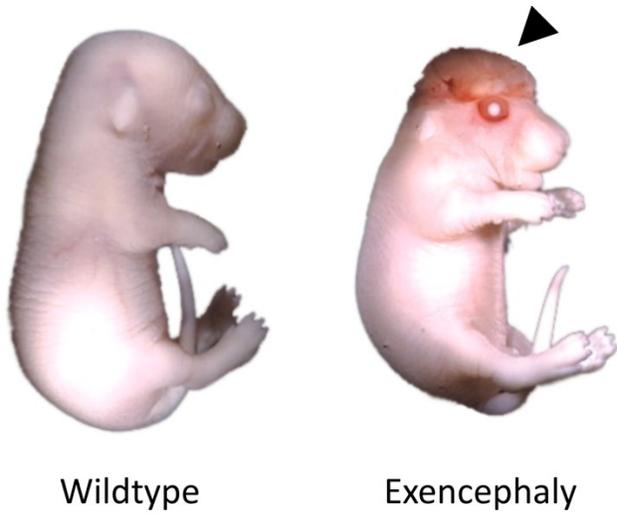


Figure 1.1.3.3. Failure of closure site 2 to fuse in mice results in exencephaly. When neurulation proceeds and all closure points fuse, mice develop normally (wildtype). Mice that do not undergo complete neurulation, and have closure site 2 remaining open, develop exencephaly, a lethal condition where the skull vault fails to form leaving the neuroepithelium exposed to amniotic fluid (indicated by black arrow). These mice were dissected at E18.5, and the exencephalic mouse is homozygous for a mutation in *Cecr2*. Figured modified from Banting et al., 2005.

Table 1.1.3.1. Epigenetic regulators crucial for neurulation in animal models of NTDs.
(modified from Wilde et. al 2014).

Gene	Protein	Function	Observed NTD
<i>Ppm1g</i>	PPM1G	Chromatin remodeling	Exencephaly
<i>Dnmt3a</i>	DNMT3A	DNA methylation	Exencephaly
<i>Dnmt3b</i>	DNMT3B	DNA methylation	Exencephaly
<i>Dnmt3L</i>	DNMT3L	DNA methylation	Exencephaly
<i>CBP</i>	CBP	Histone acetylation	Exencephaly
<i>Kat2a</i>	GCN5	Histone acetylation	Exencephaly
<i>Ep300</i>	p300	Histone acetylation	Exencephaly
<i>Hdac4</i>	HDAC4	Histone deacetylation	Exencephaly
<i>Sirt1</i>	SIRT1	Histone deacetylation	Exencephaly
<i>Kdm2b</i>	FBXL10	Histone demethylation	Exencephaly
<i>Kdm6a</i>	UTX	Histone demethylation	Exencephaly
<i>Uty</i>	UTY	Histone demethylation	Exencephaly
<i>Alkbh1</i>	ALKBH1	Histone methylation	Exencephaly
<i>Jmj</i>	JARID2/Jumonji	Histone methylation	Exencephaly
<i>Cited2</i>	CITED2	Co-regulator of CBP/p300	Exencephaly
<i>Smarcc1</i>	BAF155	Nucleosome remodeling	Exencephaly
<i>Smarca4</i>	BRG1	Nucleosome remodeling	Exencephaly
<i>Cecr2</i>	CECR2	Nucleosome remodeling	Exencephaly
<i>mIR-124a</i>	N/A	Nucleosome remodeling	Spina bifida
<i>mIR-9*</i>	N/A	Nucleosome remodeling	Spina bifida
<i>Nap1L2</i>	NAP1L2	Nucleosome assembly	Exencephaly

1.2 Chromatin

1.2.1 Chromatin structure

In order for a diploid genome to occupy only the nucleus of a cell, a 10 to 20,000X compaction is required (reviewed in Wright, Fernandez-Fuentes, Oliva, & Beato, 2016; Zentner & Henikoff, 2013). Chromatin is the dynamic structure where DNA associates with histone proteins for compaction. This structure also allows for other functions such as gene regulation discussed in 1.2.2 (reviewed in Venkatesh & Workman, 2015; Wright et al., 2016). The structural unit of chromatin is the nucleosome, which is present approximately every 200 base pair (bp) (Wright et al., 2016). The nucleosome is comprised of a core histone octamer, with a ~147 bp segment of DNA wrapped around it (reviewed in Luger, Dechassa, & Tremethick, 2012). Nucleosomes are joined together by short pieces of linker DNA to form a nucleosomal array. These arrays interact to form a more compact 30 nm chromatin fibre, and this can further compact to form a condensed chromosome. The nucleosomal array, chromatin fibre, and chromosome are respectively the primary, secondary, and tertiary structural levels of chromatin compaction.

The primary level of compaction – the nucleosomal array – is arguably one of the most well studied and understood. The protein core of the nucleosome is a histone octamer typically composed of four canonical proteins as dimers: H3/H4 and H2A/H2B (Venkatesh & Workman, 2015). These histone proteins are incorporated into the nucleosome in a replication-dependent manner, and are responsible for the interaction with DNA through a histone fold domain (Luger et al., 2012). The assembly of the histone octamer is a highly ordered process, where there is heterodimerization of H3/H4 and H2A/H2B, formation of tetramers, and finally the octamer (Venkatesh & Workman, 2015). Surprisingly, all four histone proteins have a highly conserved structure, especially in the histone fold domain, despite little sequence homology (Cutter & Hayes, 2015). At the N-terminal domain of the histone protein, variant or canonical, is a flexible tail domain. This string of peptides containing many arginines and lysines exists unstructured until chromatin-bound. It can interact with neighbouring nucleosomes and has functions in epigenetic signalling and genome compaction.

Variants of these histone proteins, which range from a change in a few amino acids to additional domains, can affect the biochemical properties of the nucleosome – from protein-protein interaction to post-translational modifications and higher-order chromatin structure (Venkatesh & Workman, 2015). Similarly, post-translational modifications (PTMs) of histones can also affect the dynamic structure of chromatin (Venkatesh & Workman, 2015). Through acetylation, methylation, phosphorylation, ubiquitination, and sumoylation of the histone tail or fold domains, PTMs can affect the chemical interactions of the nucleosome or even the physical structure by acting as a scaffold or hindrance for binding of proteins.

1.2.2 Chromatin remodeling and gene regulation

Chromatin clearly has a more complex role than simply the packaging of DNA into the nucleus. The compaction method has functional consequences affecting accessibility of binding factors important for many cellular processes that requires modulation. Given the octamer structure of the nucleosomal histone core, the numerous variants of the canonical histone proteins, and the many types of PTMs, the possible variations in primary structure are numerous (Luger et al., 2012). Interestingly, these changes in primary structure seem subtle – the nucleosomal architecture is indifferent – until assessing the stability and dynamics of high-order chromatin structure.

Even control of transcription may not be as affected by sequence and PTMs as originally believed; condensed chromatin does not necessarily correlate to no transcription. In fact, the secondary structure of chromatin is heavily debated and access for transcription may be directly from the tertiary structure. Furthermore, DNA spontaneously wraps and unwraps from the nucleosome (Cutter & Hayes, 2015). This transient and rapid action allows high-affinity factors to bind, demonstrating that even the canonical nucleosomal structure only acts as a surmountable thermodynamic barrier. Therefore, how does the cell create directed transcription and regulate genetic pathways? A major way this is achieved is through chromatin remodeling.

Chromatin remodeling occurs via two major routes: covalent PTMs and ATP-dependent chromatin remodeling (Zentner & Henikoff, 2013). Covalent PTMs are made by enzymes on the lysine and arginine's of the histone tails. The most common covalent

modifications are acetylation and methylation. Acetylation of lysine generally results in activation, or an increase in accessibility to the chromatin landscape. It achieves this through neutralization of a positive charge that therefore weakens any dependent interactions, creating important implications for transcription, replication, and DNA repair. Interestingly, covalent PTMs implemented by chromatin remodellers can in turn affect remodeling. For example, acetylated lysines are recognized by the bromodomain of some chromatin remodeling proteins such as *D. melanogaster* ISWI (Tamkun et al., 1992). H4K16ac in particular was shown to inhibit *D. melanogaster* ISWI activity *in vitro* (Shogren-Knaak et al., 2006). Both lysine and arginine can be mono, di, or tri-methylated, however less is known about the dynamics of arginine modifications (Zentner & Henikoff, 2013). In all cases of methylation, it does not affect the charge of the histone tail and seems to have a lesser transcriptional effect than acetylation.

A lesser discussed covalent histone PTM is phosphorylation. While widely known for its role in signal transduction, it imparts a negative charge to the histone tail which can increase DNA accessibility (Zentner & Henikoff, 2013). It is largely tied to DNA repair and an increase in acetylation and chromatin-binding affinity. The biological and developmental consequences of these modulations are not well understood, but some findings such as the large increase of phosphorylation of sperm histones immediately after fertilization in sea urchins suggest an importance. Many other types of covalent histone PTMs exist such as ADP ribosylation, glycosylation, and the larger scale (76-100 aa) ubiquitylation and sumoylation, where the functions are less understood and are implicated in either transcriptional activation, repression, or both (Luger et al., 2012; Venkatesh & Workman, 2015; Zentner & Henikoff, 2013).

Considering all these possible modifications, along with the increasingly more complex PTMs and possible combinations not discussed, it is interesting to note that ChIP-chip and ChIP-Seq have only identified a small number of covalent histone PTM patterns that correspond to transcriptional changes (Zentner & Henikoff, 2013). Therefore, it seems that perhaps it is not the specific histone modifications themselves that have specific transcriptional roles, but rather they act together to promote general outcomes like nucleosomal positioning, composition, and occupancy which in turn results in

transcriptional changes. Overall, they are a small part of chromatin dynamics that is mediated by other processes such as transcription, remodeling, and targeted action of non-coding RNAs.

1.2.3 ATP-dependent chromatin remodeling

ATP-dependent chromatin remodellers use the energy from ATP hydrolysis to allow DNA translocase to slide or evict nucleosomes (Venkatesh & Workman, 2015). Remodellers can work alone or collaborate with transcription factors and/or enzymes to affect nucleosome density, spacing, and composition (Clapier, Iwasa, Cairns, & Peterson, 2017). This alters the chromatin landscape and will affect accessibility of DNA to binding factors that influence biological processes. More commonly, ATP-dependent chromatin remodellers work in part of a larger, multiprotein complex to achieve changes in transcription, replication, DNA repair, and chromatin assembly (Clapier et al., 2017; Hargreaves & Crabtree, 2011). In fact, subunit composition in these multiprotein complexes determines the stoichiometric complex:nucleosome ratio which is one factor, in addition to many others, that can affect the specific remodeling function.

ATP-dependent chromatin remodellers all share a central mechanism of ATP-dependent DNA translocation and are within the RNA/DNA helicase superfamily 2 (Clapier et al., 2017). They all have a single catalytic subunit, containing an ATPase domain split into 2 RecA-like lobes, as well as domains and/or proteins that regulate this catalytic subunit and can interact with other chromatin proteins. All remodellers affinity for binding to the nucleosome is also greater than for DNA alone. Differences do exist however, and therefore they can be divided into four sub-families: imitation switch (ISWI), chromodomain helicase DNA-binding (CHD), switch/sucrose non-fermentable (SWI/SNF), and INO80. Along with specific differences in structure and composition, their biological functions also differ. Remodellers can be classified to function either in nucleosome assembly and organization, chromatin access, or nucleosome editing (change in histone composition). The focus of this thesis is on the ISWI family.

1.2.4 Imitation switch (ISWI): a sub-family of ATP-dependent chromatin remodellers

The first identified ATP-dependent chromatin remodeller was SWI/SNF, isolated in *S. cerevisiae* (Hargreaves & Crabtree, 2011). It is a large and multiprotein complex that is well conserved and a general activator of transcription. ISWI – imitation SWI – was discovered *in vitro* in extracts from *Drosophila* embryos, and is also a very conserved family (Erdel & Rippe, 2011; Hargreaves & Crabtree, 2011). All ISWI proteins contain the catalytic ATPase domain, as previously described, and in addition have a characteristic C-terminal HAND-SANT-SLIDE (HSS) domain (Clapier et al., 2017). This HSS domain is responsible for substrate recognition, and the SANT-SLIDE in particular is critical for coupling protein binding to ATP catalysis (Erdel & Rippe, 2011). This unique C-terminal portion of the ISWI structure is necessary to bind unmodified histone H3 tails, as well as flanking linker DNA (Clapier et al., 2017). There are also many other regulatory domains and associated proteins that are outside the scope of this thesis. The ISWI protein family is part of larger multiprotein complexes, ranging from a simple composition in yeast to a more complex arrangement in mammals. In mammals, SNF2L (SNF2-like) and SNF2H (SNF2-homolog) were first identified in human cells and are orthologs of the yeast ISW1 and ISW2 (Erdel & Rippe, 2011; Lazzaro & Picketts, 2001). SNF2L and SNF2H are also known as SMARCA1 or SMARCA5 respectively. Both of these proteins are found in mice and are developmentally important in different ways, despite their structural similarity (Hargreaves & Crabtree, 2011; Lazzaro & Picketts, 2001) (see 1.2.5 for more on remodellers and development).

The ISWI family of ATP-dependent chromatin remodellers is primarily involved in nucleosome assembly and spacing (Clapier et al., 2017). They are able to space nucleosomes through nucleosomal sliding, generally limiting chromatin accessibility and promoting heterochromatin formation (Clapier et al., 2017; Hargreaves & Crabtree, 2011). The mechanism by which this is achieved can be examined from two perspectives: physically, and biologically. Physically, nucleosomal sliding is achieved through the HSS domain binding linker DNA. This acts as a ‘molecular ruler’, which allows the ATPase domain to bind the proximal side of the nucleosomal dyad (Clapier et al., 2017). In the presence of ATP, the two lobes of the ATPase domain come together in a unidirectional fashion to slide the nucleosome by 1-2 bp (termed ‘inchworming mechanism’). Exactly how

this occurs is still not well understood – one large reason being discrepancies in *in vitro* versus *in vivo* experiments (Clapier et al., 2017; Hargreaves & Crabtree, 2011). Biologically, targeted nucleosomal sliding occurs via a ‘continuous sampling’ mechanism whereby ISWI complexes continuously interact with nucleosomes, albeit transiently, until they encounter a signal that increases their affinity for binding. These signals such as DNA sequence, PTMs, histone variants, and accessory proteins, in combination with continuous sampling, allow for quick and targeted chromatin remodeling.

ISWI remodeling activity is regulated by autoinhibition at the level of the ATPase and coupling (Clapier et al., 2017). ISWI autoinhibition of ATPase occurs by the N-terminal mock H4 tail basic patch. This is antagonized by the actual H4 tail basic patch, which allows activity. The binding of HSS to linker DNA also relieves autoinhibition which uncouples ATP hydrolysis from translocation. Histone variants and modifications can also regulate ISWI remodeling activity.

Nucleosomal sliding has many functional consequences; ISWI has been associated with roles in transcription, heterochromatin formation, DNA replication, DNA repair, and ES cell pluripotency (Clapier et al., 2017; Erdel & Rippe, 2011; Hargreaves & Crabtree, 2011). The role of ISWI remodellers in transcription is vast and includes both transcriptional activation and repression though it is more heavily associated with repression. A large factor influencing the transcriptional effects of an ISWI remodeling complex is its composition. SNF2H and SNF2L work in separate complexes to achieve different transcriptional outcomes (Erdel & Rippe, 2011). Some proteins, such as RSF1 (remodeling and spacing factor 1) and the later discussed CECR2, can bind either SNF2H or SNF2L to exert different effects. This is further enhanced by the presence of cell-type-specific splice variants, reported for *Snf2l*, which have unique functions and localizations (Barak, Lazzaro, Cooch, Picketts, & Shiekhattar, 2004; Lazzaro et al., 2008).

Since chromatin remodellers also need an appropriate signal to slide nucleosomes, these signals can influence transcription (Erdel & Rippe, 2011). For example, the histone variant H2A.Z has been shown to increase the activity of SNF2H and SNF2L (Goldman, Garlick, & Kingston, 2010). The H2A family has a high divergence in the C-terminus tail, a

region shown to regulate SNF2H and SNF2L activity, so it is unsurprising that histone variants such as H2A.Z show translocation differences (Vogler et al., 2010). If this histone variant was incorporated into the nucleosome of a promoter or enhancer, where it is often found, it would increase the chance of nucleosomal sliding making it accessible to transcriptional machinery (Erdel & Rippe, 2011; Venkatesh & Workman, 2015). Transcription factors are also key players in the functioning of ISWI complexes (Bowman & McKnight, 2016). In humans, transcription factors surround nucleosomes in an organized and patterned way that often requires SNF2L and/or SNF2H. Together they work to establish nucleosome patterning and therefore genomic access. This overarching theme of chromatin patterning is also seen with ISWI's role in DNA replication. Both before and after the replication fork, ISWI works in a complex-dependent manner to define accessibility whether that be deconstructing or recapitulating the chromatin architecture.

1.2.5 ISWI chromatin remodellers and mammalian development

Although they share many similarities, SNF2H and SNF2L have distinct roles in development (reviewed in Hota & Bruneau, 2016; Lazzaro et al., 2006). While SNF2H is required for survival, SNF2L is not required (reviewed in Hargreaves & Crabtree, 2011; Hota & Bruneau, 2016). *Snf2h* is expressed ubiquitously, and particularly critical in the developing early embryo as homozygous mutations are lethal (Lazzaro & Picketts, 2001; Stopka & Skoultchi, 2003). Embryos without *Snf2h* are phenotypically normal at E3.5, but die and resorb by E7.5 (Stopka & Skoultchi, 2003). Furthermore, by day 3 in culture increasing apoptosis is observed, and derivation of an ES cell line is impossible. *Snf2h* clearly plays a role in cellular proliferation, and is required for inner mass cell survival and in later progenitor cells (Hargreaves & Crabtree, 2011; Hota & Bruneau, 2016).

Contrarily, *Snf2l* has specific expression in post-natal gametes and the brain, and has important roles in cellular differentiation and maturation (Hota & Bruneau, 2016; Lazzaro & Picketts, 2001). When the remodeling activity of SNF2L is removed, mice are healthy and fertile (Yip et al., 2012). Their only apparent defect is a 1.4X increase in their brain:body ratio. Brain analysis revealed an increase in cortical thickness, and hypercellularity caused by increased proliferation. Although no structural or morphological defects were observed, there appears to be some change in cell fate. This is likely because SNF2L inhibits *Foxg1*

expression, thereby allowing cellular differentiation. In the absence of SNF2L, differentiation is repressed allowing for continued progenitor proliferation.

The ability of SNF2H and SNF2L to regulate chromatin accessibility has direct links to their associated functions. SNF2H is associated with maintaining open chromatin (Hota & Bruneau, 2016). As previously described, SNF2H is crucial for cellular proliferation which would require an open chromatin state for gene transcription (Hota & Bruneau, 2016; Stopka & Skoultchi, 2003). It has clear roles in murine zygotic transcription regulation, as inhibition of *Snf2h* expression misregulates select TIF1- α (transcription intermediary factor 1 α) genes that mediate initial one-cell zygote transcription (Torres-Padilla & Zernicka-Goetz, 2006). SNF2H also plays a role in the proliferation of neural progenitor cells during nervous system development, as this process is impaired in conditional mutants and leads to abnormal cerebellar morphogenesis and neural maturation (Alvarez-Saavedra et al., 2014). Interestingly, SNF2L antagonizes this process to regulate brain size through repression of *Foxg1* and is believed to achieve this by binding its promoter rendering the chromatin inaccessible to transcriptional machinery (Yip et al., 2012). SNF2L is also important during folliculogenesis – without it, *Fgl2* is misregulated and females have decreased egg count due to a failure of both proliferation and differentiation of the granulosa cells (Pépin, Paradis, Perez-Iratxeta, Picketts, & Vanderhyden, 2013).

Taken together, ISWI remodellers SNF2H and SNF2L work to increase transcriptional specificity, regulate chromatin structure, and participate in DNA replication to exert developmental outcomes (reviewed in Ho & Crabtree, 2010; Hota & Bruneau, 2016). They do this by working with tissue-specific subunits and varying their combinatorial assembly which ultimately alters their targets and role in moderating the chromatin landscape. This results in SNF2H and SNF2L working as a part of many different remodeling complexes. For example, SNF2L works as part of the complex NURF (nucleosome remodeling factor) which is involved in a variety of processes such as embryo implantation, thymocytopoiesis, and erythropoiesis (Landry et al., 2011; 2008; Stopka & Skoultchi, 2003). SNF2H is also present in many complexes, including NoRC (nucleolar remodeling complex), WICH (WSTF ISWI chromatin remodeling), ACF (chromatin-assembly

factor), and human CHRAC (chromatin accessibility complex) (Ho & Crabtree, 2010; Hota & Bruneau, 2016). While NURF and NoRC act as transcriptional activators or repressors, WICH, ACF, and CHRAC are involved in nucleosomal assembly, spacing, and DNA replication. ISWI remodellers also increase specificity by collaborating with transcription factors. For example, NURF is shown to be a cofactor for SMAD transcription factors (Landry et al., 2008).

There is also a myriad of other factors to consider when examining the role of ISWI in development such as non-coding RNAs, the physical chromatin architecture, along with ones less understood such as transcriptional memory and *in vivo* effects (reviewed in Hota & Bruneau, 2016). Long non-coding RNAs have been implicated with targeting of ISWI complexes, as well as antagonizing them via sequestration or binding of catalytic domains (Han & Chang, 2015). The chromatin landscape can have important implications for ISWI complexes since some histone marks, such as at regulatory sites for transcription, can influence remodeling ability (Shogren-Knaak et al., 2006). The amount and impact of inherited epigenetic changes on transcription is not well understood, nor is the variability of assays performed *in vivo* compared to its *in vitro* counterparts (Hargreaves & Crabtree, 2011; Hota & Bruneau, 2016). It is likely that structural and mechanistic factors play important roles that have not yet been identified.

1.3 Cat eye syndrome critical region candidate 2

1.3.1 *Cecr2* structure

The focus of this thesis will be on CECR2, which is part of an ISWI chromatin remodeling complex involved in neurulation. *Cecr2* was originally identified as one of 14 genes in the cat eye syndrome critical region (CESCR) in humans at 22q11 (Footz et al., 2001). Cat eye syndrome in humans is the result of CESCR trisomy or tetrasomy and is characterized by various defects including ocular coloboma, mental retardation, and kidney, heart, and anal defects (Schinzel, Schmid, Fraccaro, genetics, 1981, n.d.). A homologous region is shared with mice, where 10 of the 14 CESCR genes are found on chromosome 6, including *Cecr2*. To date, the role of *Cecr2* in cat eye syndrome is unknown.

Mouse *Cecr2* is comprised of 1453 amino acids, 19 exons, and nuclear localization sequence (NLS); it makes a 180 kDa chromatin remodeling protein with an AT hook,

bromodomain, DDT domain (Banting et al., 2005)(Figure 1.3.1.1.). AT hooks and bromodomains are important for recognition of DNA sequence and acetylated lysines on histone tails respectively, and are predicted to affect the outcome of chromatin remodeling (Längst & Manelyte, 2015). DDT domains are important for protein-protein interaction in the ISWI family of chromatin remodellers, as is shown in *Drosophila* and *Arabidopsis* (Dong et al., 2013; Fyodorov & Kadonaga, 2002). These domains should all contribute to CECR2 acting in a chromatin remodeling protein complex, termed CERF (CECR2 containing remodeling factor). This protein complex will be elaborated on in 1.3.5.

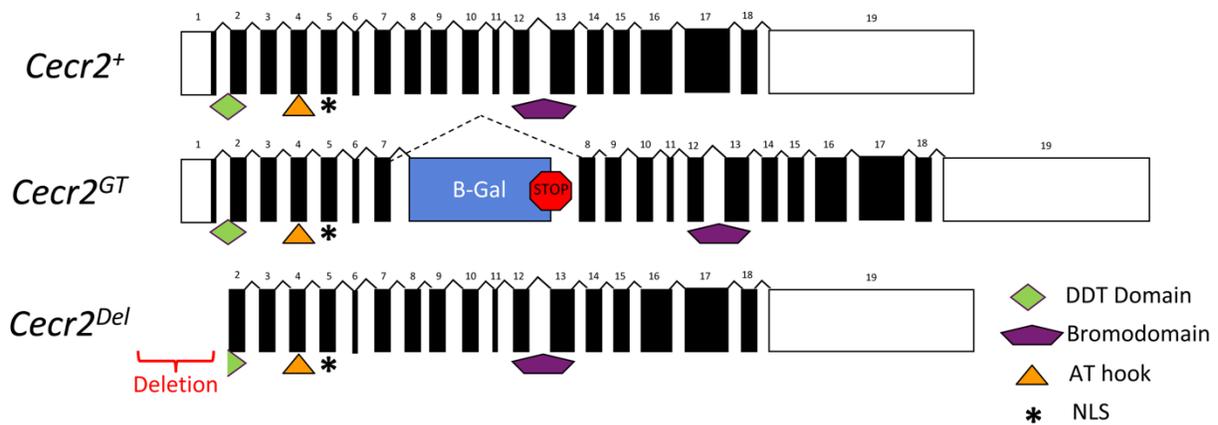


Figure 1.3.1.1. *Cecr2* gene diagram. Wildtype *Cecr2*⁺ is shown with 19 exons, a DDT domain, an AT hook, a nuclear localization sequence (NLS), and a bromodomain. Domain symbols that span an intron do not include the intron. *Cecr2*^{GT} has a β -galactosidase gene inserted between the seventh and eighth exon, causing a truncated protein. Some splicing around the insert is indicated by the dotted line. *Cecr2*^{Del} has a deletion of exon 1 and 1 kb upstream, disrupting the DDT domain and transcription start site, thereby creating a null mutation.

1.3.2 *Cecr2* mutations

To study the function of CECR2, two mutations in mice were created (Figure 1.3.1.1.). A pGT1 gene trap vector inserted between exons 7 and 8 was used to create the *Cecr2*^{Gt45Bic} mutation (also annotated as *Cecr2*^{Gt(pGT1)1Hemc} and referred to as *Cecr2*^{GT} throughout this thesis)(Banting et al., 2005). This mutation creates a truncated CECR2- β -galactosidase fusion protein that can be detected with X-gal staining and was

used for the expression analysis described in 1.3.4. This mutation eliminates the conserved bromodomain of CECR2, and therefore should eliminate the chromatin-binding functionality of CECR2 and targeting of the CERF complex (Längst & Manelyte, 2015). However, the DDT domain, NLS, and AT hook still remain so this fusion protein still localizes to the nucleus and can interact other proteins, such as SNF2H and SNF2L (Banting et al., 2005; Thompson, Norton, Niri, Dawe, & McDermid, 2012). This mutation is in fact a hypomorph; in *Cecr^{GT/GT}* mutants there is only a ~14-fold decrease in wildtype transcript as shown by a microarray and qRT-PCR on embryos (Fairbridge, Dawe, Niri, Kooistra, King-Jones, & McDermid, 2010). This can be explained by splicing around the gene trap mutation, which produces some amount of wildtype transcript (Leduc et al., 2017). A small amount of wildtype CECR2 can also be detected on a western blot using a CECR2 specific antibody (Niri, in prep). All things considered, this mutation disrupts CECR2 sufficiently to observe severe defects in homozygous mutants (see 1.3.3).

Another mutation, *Cecr2^{tm.1.1Hemc}* (referred to as *Cecr2^{Del}* throughout this thesis) was created by deleting exon 1 and 1 kb upstream using LoxP-Cre recombination (Fairbridge et al., 2010). This deletion removes the start codon, promoter, and disrupts the DDT domain of *Cecr2* which should eliminate any *Cecr2* transcript and protein. A ~200-fold decrease in *Cecr2* transcript was observed in the heads of homozygous mutants using qRT-PCR. Western blot analysis using a CECR2-specific antibody detected no CECR2 protein in homozygous mutants, confirming disruption by the deletion (Niri, in prep). This mutation is considered to be more severe than the gene trap, as an increased penetrance of defects are observed in homozygous mutants.

1.3.3 Phenotypes of *Cecr2* mutants

The main phenotype observed with *Cecr2* mutations is the NTD exencephaly – a lethal disorder where the brain develops abnormally and eventually degrades (Figure 1.1.3.3.)(see 1.1.3 for more details)(Banting et al., 2005). The exencephaly penetrance for *Cecr2^{GT/GT}* mutants in a congenic Balb/c line is ~54%, compared to ~100% penetrance for *Cecr2^{Del/Del}* mutants on the same background. In compound heterozygous mice, with an allele of each mutation, exencephaly penetrance is ~75% (Norton, unpublished).

Mutations in *Cecr2* can also cause other concurrent phenotypes with exencephaly, including open eyelids and inner ear defects (Banting et al., 2005; Dawe, Kooistra, Fairbridge, Pizio, & McDermid, 2011). The smaller cochlea and stereocilia disorganization in *Cecr2^{GT/GT}* mutants appears indicative of PCP defects (see 1.1.2 on PCP), however qRT-PCR analysis of 13 PCP-associated genes, including *Vangl2*, showed no differential expression (Dawe et al., 2011). Interestingly, there is a genetic interaction between *Cecr2* and *Vangl2*. Mice who are heterozygotes for mutations in both genes show a significant increase in spina bifida penetrance, yet the degree of inner ear defects of the double heterozygotes was unchanged. Therefore, the role of *Cecr2* in the PCP pathway remains unclear.

The surviving non-penetrant *Cecr2^{GT/GT}* mutants were found to have fertility defects. Interestingly, both males and females were found to be subfertile (Thompson et al., 2012, Norton, unpublished). Recent studies in compound heterozygous mice (*Cecr2^{GT/Del}*) have shown that males have smaller testis, and exhibit defects in the seminiferous tubules (Norton, unpublished).

Since CECR2 works in a chromatin remodeling complex, a microarray gene expression analysis was done in 11-14 somite (time of neurulation) *Cecr2^{GT/GT}* mutant embryos compared to wildtype (Fairbridge et al., 2010). The results were confirmed by performing qRT-PCR on embryo heads. In embryos around the time of neurulation, three genes (*Alx1*, *Dlx5*, *Six1*) were found to be downregulated. In addition to those, three other genes (*Epha7*, *Eya1*, *Lix1*) were downregulated in 18-20 somite embryos which is after neural tube closure. *Epha7*, *Alx1*, and *Dlx5* all can cause exencephaly at varying penetrance when mutated. Interestingly, *Alx1* and *Dlx5* are mesenchymal/ectodermal transcription factors which suggests that mutations in *Cecr2* can affect downstream targets and pathways that are regulated by other transcription factors. Either *Cecr2* is directly regulating these transcription factors or is regulating another element affecting their expression.

While all information described above was collected using Balb/c mice, the *Cecr2*^{GT} mutation was also put onto an FVB/N background. When homozygous for the mutated allele, these mice do not exhibit exencephaly; they appear healthy and reproduce contrasting the 54% exencephaly penetrance observed in the Balb/c strain (Banting et al., 2005; Kooistra et al., 2012). This suggests the presence of modifier genes which can confer resistance or affect susceptibility to developing exencephaly.

1.3.4 *Cecr2* expression and localization

Taking advantage of the gene trap and X-gal staining, an expression analysis showed that in E10.5-14.5 embryos *Cecr2* is strongly expressed in the central nervous system (CNS), in the limb and intercostal mesenchyme, as well as in the nasal epithelium and the eye (Banting et al., 2005). As the embryo develops, expression in the brain decreases but remains present until shortly after birth when there is a sharp decrease (Elliot, in prep). At E18.5, *Cecr2* expression is retained in the hippocampus, cerebellum, olfactory bulbs, and cortex of the brain, whereas in adulthood only expression in some cells of the hippocampus and cerebellum remains.

Furthermore, *Cecr2* expression in adulthood remains strong in the testis, specifically in the spermatogonia type A cells as was established using immunofluorescence and immunohistochemistry (Norton, in prep). The presence of CECR2 has been confirmed by western blotting in the embryo, testis, ovary, and cerebellum (Niri, in prep). It is also present in mouse embryonic stem (ES) cells and neural stem cells (neurospheres). Nuclear localization of CECR2 in testis, ES cells, and embryonic kidney has also been confirmed by immunofluorescence.

1.3.5 CERF: CECR2-containing remodeling factor

As mentioned, CECR2 functions as part of a chromatin remodeling complex termed CERF. It was originally isolated from HEK293 cells as a 0.6 MDa protein complex that contains not only CECR2, but also the well-known ISWI chromatin remodeller SNF2L (Banting et al., 2005). An *in vitro* assay demonstrated the ability of CERF to shift nucleosomes using ATP hydrolysis and established CERF as an ATP-dependent chromatin remodeller.

CERF localizes to the nucleus as shown by western blot following cell fractionation and by immunofluorescence (Niri, unpublished). It has also been isolated from mouse ES cells and testis for molecular weight determination. At physiological salt conditions (150mM), the protein complex size is ~2 MDa in both cell types – a size much larger than what was observed in HEK293 cells. At higher salt concentrations the nuclear extraction of CECR2 is more successful, however the protein-protein interactions of complex are disrupted – a characteristic common for protein heterocomplexes (Z. Zhang, Witham, & Alexov, 2011). CERF protein complex analysis was done in both ES cells and testis, using a combination of gel filtration, mass spectrometry, and co-immunoprecipitation (co-IP) (Niri, in prep). Some preliminary analysis in neurospheres has been done using co-IP.

In ES cells, confirmed members of the CERF protein complex are SNF2H, SNF2L, LUZP1, and CCAR2 (see Table 1.3.5.1.)(Niri and Lim, in prep). These proteins were identified by mass spectrometry and confirmed with reciprocal co-IP except in the case of SNF2L, as the antibody was not suitable for IP. Furthermore, an interaction between SNF2H and LUZP1 was also shown with co-IP. However, LUZP1 does not act as a mediator between CECR2 and SNF2H since they still associate and co-IP in ES cells without any LUZP1 protein. It does though likely act as a bridging protein within the overall complex, since smaller CECR2-containing protein complexes are eluted in these LUZP1-deficient cells.

In testis, confirmed members of the CERF protein complex are SNF2H and SNF2L (see Table 1.3.5.1.)(Niri, in prep). CCAR2 may also be part of the testis complex, however the co-IP showing interaction was not reliably reproduced (Lim, in prep). LUZP1 was not identified as part of the protein complex, as co-IPs were negative. In neurospheres, CECR2 interacts with SNF2H and LUZP1 (Niri, in prep).

LUZP1 and CCAR2 are interesting members of the CERF protein complexes. LUZP1 (Leucine zipper protein 1) contains leucine zipper motifs and nuclear localization signals (M. W. Lee, Chang, Sun, Hsu, & Chang, 2001). It is present in mouse ES cells, and predominantly in the brain. Interestingly, LUZP1 is present in the hippocampus of the mouse adult brain where *Cecr2* is also expressed (Banting et al., 2005). Although the function of LUZP1 remain unclear, *Luzp1*^{-/-} embryos can develop exencephaly much like

Cecr2 mutants (Hsu et al., 2008). Although both embryos fail to close their neural tube at the cranial region and are strikingly phenotypically similar, the penetrance of *Luzp1*^{-/-} mutants is lower at 42% compared to 100% of *Cecr2*^{Del/Del} mutants (Hsu et al., 2008). However, the *Luzp1* mutation is on a C57BL/6J background which may affect the penetrance comparison. *Luzp1*^{-/-} mutants also display ectopic sonic hedgehog signalling and increased apoptosis in the cranial region, whereas *Cecr2* mutants do not (Niri, unpublished). *Luzp1* mutants also develop omphalocele and complex cardiac defects (Hsu et al., 2008). Omphalocele is not seen in *Cecr2* mutants and heart defects have not been investigated. Taken together, this suggests that LUZP1 may have functions in other protein complexes aside from CERF. There is no evidence in the literature that LUZP1 plays a role in fertility or is found in the testis which supports the CECR2 IP data. However, fertility defects are often missed or simply not investigated.

CCAR2 (cell cycle and apoptosis regulator 2), commonly known as DBC1 (deleted in breast cancer 1), also has a nuclear localization signal and a leucine zipper motif (Joshi, Quach, Giguere, & Cristea, 2013). It is well documented to have an array of roles including cellular proliferation, apoptosis, and survival in response to DNA damage or heat shock. Most notable is CCAR2's indirect control of epigenetic modifications through interaction and inhibition of chromatin remodeling enzymes and transcription factors. In particular, it indirectly represses several histone modification enzymes which in turn modulate various cellular pathways. One example is the inhibition of a histone methyltransferase which is responsible for heterochromatin formation. In the absence of inhibition, more heterochromatin is formed which would result in decreased gene expression. Differential expression of genes is observed in *Cecr2* mutants, and downregulation in particular is a common trend for proteins in the ISWI family (Clapier et al., 2017; Erdel & Rippe, 2011; Fairbridge et al., 2010; Hota & Bruneau, 2016). This may suggest a role for CCAR2 and CECR2 as master regulators of transcriptional processes in the cell. There is also evidence that CCAR2 interacts with estrogen and androgen receptors, which could have implications in fertility (Fu et al., 2009; Koyama et al., 2010).

Table 1.3.5.1. CERF protein complex members based on immunoprecipitation results. Data provided by Niri and Lim, in prep.

Protein	ES Cells	Testis	Neurospheres
SNF2H	Yes	Yes	Yes
SNF2L	Yes	Yes	Unknown
LUZP1	Yes	No	Unknown
CCAR2	Yes	Inconclusive	Unknown

1.3.6 Genomic binding sites of CERF

The CECR2-containing CERF protein complexes possess an ATP-dependent chromatin remodeling ability. This epigenetic modification is specific to certain regions of the genome (Ho et al., 2009). Gene expression analysis cannot differentiate between the direct targets of CERF, or the downstream dysfunction in cellular pathways – both of which could contribute to *Cecr2* mutant mice developing exencephaly. To elucidate where in the ES cell and testis genome CERF is binding, and therefore which direct genetic targets it may be responsible for activating or repressing, a chromatin immunoprecipitation paired with massively parallel sequencing (ChIP-Seq) experiment was performed previously (Niri, unpublished). This experiment, performed in tandem with antibodies against CECR2, LUZP1, and SNF2H, revealed their respective genomic binding sites. Comparison of these sites between all three proteins can elucidate where the protein complex CERF – which contains two to three of these proteins – may be binding and may be controlling gene expression.

In ES cells, ChIP-Seq was performed for CECR2, SNF2H, and LUZP1 as all three are confirmed members of CERF. CCAR2 was not yet identified at the time of this experiment. Gene ontology revealed that CECR2 binds cis-regulatory elements involved in the development of the brain, heart and kidney (Niri, unpublished). When looking at CECR2 and SNF2H together, it highlighted their roles in brain development and reproduction. When LUZP1 was included in the analysis, it was largely genes associated with kidney development.

The overlapping DNA binding sites between CECR2 and SNF2H, when focusing on regions 5 kb upstream of the transcription start site (TSS) of genes, gave rise to 103 coding and 33 non-coding RNA genes (Niri, unpublished). Adding in LUZP1 for comparison narrowed the list to 23 coding and 7 non-coding RNA genes. The binding sites for CECR2 alone were also investigated, at a more stringent parameter of 1 kb upstream of the TSS, and revealed 298 coding and 31 non-coding RNA genes.

In testis, CHIP-Seq was performed for CECR2 and SNF2H only, as IP evidence suggested that LUZP1 is not a part of CERF and CCAR2 was not yet identified at the time of this experiment. The gene ontology term with the highest number of hits suggests that CECR2 binds cis-regulatory regions for genes involved in the acrosome reaction in sperm (Niri, unpublished). There was insufficient data to perform gene ontology for both CECR2 and SNF2H together. The overlapping DNA binding sites between CECR2 and SNF2H, when focusing on regions 5 kb upstream of the TSS of genes, gave rise to 73 coding and 13 non-coding RNA genes. Broadened to CECR2 alone, there were 106 coding and 22 non-coding RNA genes with binding sites. Interestingly, there are only 14 genes that share binding sites for CECR2 in both ES cells and testis (Table 1.3.5.1.). This suggests that CERF may function in a tissue-dependent manner.

Table 1.3.6.1. Genes that have CECR2 binding sites within 5 kb upstream of TSS in both ES cells and testis. Binding sites were evaluated at a p-value < 10⁻⁵. Table is modified from Niri, 2016.

Coding genes		
Gene ID	Gene name	Mutant phenotypes
<i>Aqp7</i>	aquaporin 7	abnormal sperm physiology, abnormal kidney physiology
<i>C8b</i>	complement component 8	abnormal hematopoietic system physiology
<i>Caly</i>	calcyon neuron-specific vesicular protein	abnormal synaptic vesicle recycling
<i>Cmah</i>	cytidine monophospho-N-acetylneuraminic acid hydroxylase	cochlear outer hair cell degeneration, abnormal inner ear morphology,
<i>Galnt6</i>	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 6	
<i>Galnt9</i>	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 9	
<i>Nedd9</i>	neural precursor cell expressed, developmentally down-regulated gene 9	
<i>Nefh</i>	neurofilament, heavy polypeptide	abnormal neuron morphology
<i>Nfia</i>	nuclear factor I/A	male-sterility, low female fertility, lack of corpus callosum, hydrocephalus.
<i>Olig1</i>	oligodendrocyte transcription factor 1	adult cortical interneuron numbers, Fewer Numbers of Oligodendrocytes,
<i>Prdm14</i>	PR domain containing 14	female and male sterility
<i>Slc25a19</i>	solute carrier family 25 (mitochondrial thiamine pyrophosphate carrier), member 19	exencephaly
<i>Slc46a1</i>	solute carrier family 46, member 1	cell division cycle 20 homolog
<i>Tomm40</i>	translocase of outer mitochondrial membrane 40 homolog	

1.4 Purpose and Aims

The ChIP-Seq experiment provided potential information on where the CERF complex interacts with chromatin. This experiment, performed in ES cells, allowed for quick protocol optimization, ample sample, and comparative studies with other previously performed assays such as mass spectrometry and co-IP. ES cells also serve as a model to compare with other ChIP-Seq analyses, such as chromatin modifications and transcription factors, done in ES cells and found in public databases. Furthermore, many of these techniques are not feasible to do *in vivo*, for example on mouse neuroulating tissue, as the amount of tissue is insufficient and collection is lengthy and difficult.

This experiment showed that much like other chromatin remodellers, CECR2 has many binding sites in the genome (Niri, unpublished). SNF2H and LUZP1 both interact with thousands of genomic sites in the context of many different remodeling complexes, therefore comparing the binding sites that overlap with CECR2 enhanced the analysis specificity of CERF genomic interactions. Genes with overlapping DNA-binding signatures in their regulatory sequences in ES cells and testis may provide some evidence for CERF functioning as a transcriptional regulator. However, this experiment needs to be validated. Whether any of these targets also directly controls gene expression remains unclear.

Although the microarray and qRT-PCR data previously discussed show misregulation of specific genes in *Cecr2^{Del/Del}* mutant mice, an RNA sequencing experiment would provide a more complete analysis as the genechip used did not have complete coverage of the transcriptome (Fairbridge et al., 2010). Using mutant and wildtype embryo heads at the time of neurulation targets the most biologically relevant tissue at a very specific time point. Together, RNA-Seq and ChIP-Seq could synergistically reveal if the transcriptional changes in mutant mice are due to direct regulation by CERF. This experiment might also elucidate which genetic misregulations may be contributing to the exencephalic phenotype of *Cecr2^{Del/Del}* mice.

Furthermore, I propose to confirm and clarify the CERF complex composition. CCAR2 and LUZP1 have both been identified in ES cells and testis and assayed for possible interaction with CECR2 through co-IP (Niri, unpublished). Although the IP was robust for

LUZP1, whether CCAR2 may interact with CECR2 is still very unclear. Immunoprecipitation is only one way to assay for interaction, and confirmation using another method especially in the case of any uncertainty is valuable. Therefore, immunofluorescence could help resolve this ambiguity by demonstrating the spatial localization of these proteins. If LUZP1 and CCAR2 are interacting with CECR2 in the CERF complex, they should localize together in the nucleus of ES cells and in testis.

Hypothesis: The CERF complex directly regulates specific genes involved in embryogenesis, specifically during neurulation.

Objectives:

1. Confirm and clarify the CERF protein complex composition
2. Investigate the role of CECR2 as a transcriptional regulator in neurulation

Specific Aims:

- A)** Perform immunofluorescence colocalization experiments in ES cells and testis for newly identified protein complex members CCAR2 and LUZP1 to confirm their tissue-specific interaction or lack of interaction
- B)** Perform an RNA sequencing experiment on neurulating embryo heads to investigate transcriptional differences between wild-type and mutant *Cecr2* mice. Identify candidate gene targets from ChIP-Seq analysis and compare to RNA-Seq to look for potential direct targets regulated by the CERF complex.

Chapter 2 – Materials & Methods

2.1 Maintaining the mouse colony

All experimental protocols and procedure were reviewed and approved by the Animal Care and Use Committee of the University of Alberta (AUP 00000094). *Cecr2* mutations were bred onto a Balb/c sub-strain originating from Charles River Laboratories. In 1988 the BALB/c strain from Charles River Laboratories was isolated in a University of Alberta colony, and bred with some selection rendering it approximately 20 substrains removed from the founder.

2.1.1 Housing

Colony mice were housed in one room in the University of Alberta Biological Sciences Animal Services facility (SASS) and maintained daily by technicians. The colony was kept on a 14:10 light:dark cycle at 22±2°C. Mice were kept in filter-top individually ventilated cages (31.8 cm × 16.5 cm × 12.7 cm; IVC Blue Line, Tecniplast, Buguggiate, Italy), with no more than five per cage. The colony was fed 5053-PicoLab Mouse Diet 20 formula containing 4% fat (LabDiet 9F 5020) except for breeder cages and plugged female cages which were fed 5058-PicoLab Mouse Diet 20 formula containing 9% fat (LabDiet 5001), and weanlings which were fed a 1:1 mixture of the two diets described above.

2.1.2 Breeding

Mice were sexed, and ear notched for identification just prior to weaning at 3 weeks. Ear notches were collected for genotyping, and mice were weaned by SASS technicians. Stock breeder cages contained permanent trios, consisting of one male and two females. For experiment-specific breeding, males were caged alone until introducing one or two females. Mating was determined by visual observation of a post-copulatory vaginal plug. If males were set up with females during the afternoon or evening, plug testing would occur before 10:00 am the next morning. For timed matings, males were set up with females in the morning around 6:00 am and were plug-tested every hour until 10:00 am to check for mating. Once a plug was found, females were placed into a separate cage and housed as described in 2.1.1. Pregnant mice were considered at embryonic day 0.5 into gestation

from the morning their plug was found, with the exact hour of mating known for timed matings.

2.1.3 Euthanasia

Most mice were euthanized in the lab using cervical dislocation, although some were culled using the SASS HiRoad Euthanasia Chamber, which anesthetizes mice with isoflurane, before CO₂ euthanasia. Death was confirmed with a toe pinch test, and carcasses were stored at -20 until incineration.

2.2 Genotyping

2.2.1 Genomic DNA extraction

Ear notches or E9.5 decapitated embryo bodies were collected in 1.5 mL microcentrifuge tubes for genotyping. The DNA extraction method was based on Lopez, 2012 protocol. Samples were submerged in 50mM NaOH and incubated at 95°C for 30 minutes. Tubes were vortex and incubated another 15 minutes at 95°C. Extractions were cooled and kept at 4°C for genotyping. Extractions were later stored at -20°C.

2.2.2 PCR reactions

Generally, 22 µl reactions were prepared for genotyping PCR. All PCR reactions used 2 µl of extracted genomic DNA added to 20 µl of master mix. The master mix consisted of 0.75 mM dNTPs, 1 µM of each primer, 1.1X DreamTaq Buffer, and 1.5U of Dream Taq DNA Polymerase (Thermo Scientific, #EP0705). Nuclease-free water purchased from Integrated DNA Technologies (IDT) was used as a diluent in all master mixes. PCR reactions were amplified using the Bio-Rad T100™ Thermal Cycler with an initial denaturation step at 94°C for 3 minutes, followed by a repetitive cycle (37X) of denaturation at 94°C for 15 seconds, annealing at 60°C for 20 seconds, and extension at 68°C for 40 seconds. A final extension step followed at 68°C for 5 minutes and then PCR reactions were kept at 4°C until electrophoresis.

2.2.3 *Cecr2*^{Del} genotyping

The presumptive null mutation, *Cecr2*^{Del}, is a deletion of exon 1 and 1kb upstream of the *Cecr2* gene; generation of this mutation is described in Fairbridge et al. (2010). The IngeniousLox1, Ingenious SDL2, and LoxCECR2_DEL3R primers were used to amplify either

the *Cecr2* wild type allele (220bp amplicon) or the *Cecr2^{Del}* mutant allele (~450bp amplicon). See Appendix A – List of primers for sequences of these primers.

2.2.4 *Cecr2^{GT}* genotyping

The *Cecr2^{GT}* gene trap mutation has a β -galactosidase gene spliced in between the 7th and 8th exon creating a truncated fusion protein; generation of this mutation is described in Banting et al. (2005). The Mmu Intron7 F4, Mmu Intron7 R4, and pGT1R4 primers were used to amplify either the *Cecr2* wild type allele (376bp amplicon) or the *Cecr2^{GT}* mutant allele (573bp amplicon). This PCR was multiplexed to also genotype for sex by detecting the male specific SRY gene (266bp amplicon) using the SRY FOR and SRY REV primers. See Appendix A for sequences of these primers.

2.2.5 SRY genotyping

Embryos collected for RNA sequencing and qPCR analysis carrying only the wildtype *Cecr2* allele and/or the *Cecr2^{Del}* allele were genotyped separately for sex by detecting the male specific SRY gene (266bp amplicon). Primers used were SRY FOR and SRY REV; see Appendix A for sequences of these primers.

2.2.6 Agarose gel electrophoresis

PCR amplifications were visualized on a 2% agarose gel with 0.1 μ g/mL of ethidium bromide. Gels were run in 1X TAE Buffer (40 mM Tris-acetate, 1 mM EDTA) at 130V for 50 minutes. Band size was compared to a nucleic acid marker (GeneRuler™ 1kb DNA Ladder) and detected using a UV fluorescence gel imager (Alpha Innotech).

2.3 Cell culture of mouse embryonic stem (ES) cells

TT2 (from Matt Lorincz's lab, University of British Columbia) and *Luzp1^{+/+}* and *Luzp1^{GT/GT}* E14 (from Dr. Laszlo Tora, Institut de Génétique et de Biologie Moléculaire et Cellulaire) mouse ES cells were revived from liquid nitrogen by transferring the frozen vial to a 37°C water bath until almost completely thawed. Cells were then transferred into 6 mL of warmed medium [Dulbecco's Modified Eagle high glucose (Sigma, #D6429-1L) supplemented with 15% embryonic stem cell-qualified Fetal Bovine Serum (FBS) (Life technologies, #10439-024), 2 mM L-glutamine (Life Technologies, #25030-081), 100 mM β -mercaptoethanol, 0.1 mM MEM non-essential amino acids solution (Life technologies, #

11140-050), 100 Units/mL Penicillin-Streptomycin (Life technologies, #15070-063) and 1000 U/mL recombinant leukemia inhibitory factor (Sigma-Aldrich #L5158-5UG)]. After pelleting the cells at 90 g for 5 minutes, they were resuspended in 6 mL of warmed media and plated. The cells were grown on 100 mm Cell Culture Dishes (Thermo Scientific, #NC0479278) coated with 0.1% gelatin and incubated at 37°C in a 5% CO₂ atmosphere. Cultures were fed every day and split at a 1:5 ratio into 100 mm Cell Culture Dishes, or into a 24 well dish (see 2.5.1) before reaching ≥80% confluency using 0.25% trypsin-EDTA solution (Invitrogen, #25200-072).

2.3.1 Freezing down cells

Cells were cultured as is described in 2.3. When cells reached 70-90% confluency they were harvested, and each plate was suspended in 1mL of media. Drop-by-drop, 1 mL of ice-cold freezing media (80% embryonic stem cell-qualified FBS (Life technologies, #10439-024), 20% DMSO) was added, and then cells were aliquoted into ice-cold freezing vials. Cells were packed tightly into an insulated freezing container and kept at -80°C for at least 24 hours. Cells were then moved to liquid nitrogen for storage.

2.4 Immunofluorescence on testis

2.4.1 Collecting and preparing testis samples

Adult male (42-150 days old) Balb/c *Cecr2*^{+/+} and *Cecr2*^{Gt45Bic/tm.1.1Hemc} mice, henceforth referred to as *Cecr2*^{GT/Del}, were culled (see 2.1.3) and testis were dissected using sterilized tools. Each testis was fixed in ≥10X volume of 4% formaldehyde diluted in PBS for up to 48 hours at room temperature. Testis were rinsed three time in 1X PBS, then 25% ethanol and 50% ethanol for a minimum of 30 minutes each, before being stored in 70% ethanol at either room temperature or 4°C. Testis were submitted to the Biological Sciences microscopy for paraffin processing using the Fisher Histomatic Tissue Processor (Model 166). Samples were then placed in embedding molds and filled with melted paraffin wax. After cooling at RT for a minimum for 3 hours, the testis paraffin blocks were sectioned using a microtome (LEICA RM2235) at a thickness of 7 μm. Paraffin sections were mounted onto Superfrost Plus Microscope Slides (Fisher Scientific, #12-550-15) using a 42°C water bath and left to dry overnight (O/N) at 37°C, or until used for immunofluorescence.

2.4.2 Immunofluorescence using a single antibody

Paraffin sections of testis were collected and mounted on slides as described. Sections were deparaffinized by washing three times in toluene for 5 minutes each, and then rehydrated in 100% ethanol with two 10 minutes washes, followed by two 10 minute washes in 95% ethanol, and two 5 minute washes in distilled water. Slides were put into an Antigen Retrieval Buffer (10 mM Tris Base, 1 mM EDTA, 0.05% Tween 20) and heated until boiling and then incubated for 10 minutes. This Antigen Retrieval Step was repeated four times sequentially. Sections were blocked in 10% normal goat serum (Sigma, #G9023) and 0.6% Triton-X (Sigma, #T8787) diluted in PBS for 1 hour at room temperature, before being placed in diluted primary antibody (Antibody Dilution Buffer: 0.1% bovine serum albumin, 0.3% Triton-X, diluted in PBS) overnight at 4°C in a humidifying chamber. Primary antibody concentrations used were: purified rabbit polyclonal anti-CECR2 (Niri et al., in prep) at 1:10 000, rabbit polyclonal anti-LUZP1 (Protein tech #17483-1-AP) at 1:1000, and rabbit polyclonal anti-CCAR2 (Protein tech #22638-1-AP) at 1:1000. Sections were washed in PBS three times for 5 minutes each, then incubated in secondary antibody (AlexaFluor 488 goat anti-rabbit, Life Technologies, # A11008) at a concentration of 1:200 in Antibody Dilution Buffer in a humidifying chamber for 2 hours at room temperature in the dark. Sections were then stained with 0.1% DAPI in PBS for 5 minutes before being washed in PBS three times for 5 minutes each. Finally, a coverslip was mounted using Fluoromount G (SouthernBiotech, #0100-01) and sealed before being imaged with the Nikon Eclipse 80i confocal microscope (CVI Melles Griot Ion Laser, NIS-Elements v4.0 software).

2.4.3 Colocalization immunofluorescence

Sections were put through the same protocol as described in 2.4.2 until the addition of the secondary antibody. Since all the primary antibodies were made in rabbit, the use of fluorophore-conjugated FAB fragments to label and cover the epitopes of the first rabbit primary antibody was necessary to use a second rabbit primary antibody. See 3.1 for more explanation of FAB fragments. After being incubated in the first rabbit primary antibody overnight and washed in PBS three times for 5 minutes each, sections were incubated in Alexa Fluor® 488 AffiniPure Fab Fragment Goat Anti-Rabbit IgG (H+L) (Jackson

ImmunoResearch, #111-547-003) reconstituted in nuclease-free water and used at a 1X concentration for 2 hours at room temperature in the dark. All following steps were done in the dark. Sections were then washed in PBS three times for 5 minutes each, and blocked in 20% normal goat serum (Sigma, #G9023) and 0.6% Triton-X (Sigma, #T8787) diluted in PBS for 1 hour at room temperature. The sections were then incubated in the second primary antibody diluted in Antibody Dilution Buffer in a humidifying chamber overnight at 4°C. This step was also done for 2 hours at room temperature, although the antibody staining was less successful for CECR2 and LUZP1. Sections were washed three times in PBS for 5 minutes each, and then incubated in secondary antibody (Goat anti-Rabbit IgG (H+L) Secondary Antibody-Alexa Fluor 555, Invitrogen #A-21429) at a concentration of 1:200 in Antibody Dilution Buffer in a humidifying chamber for 2 hours at room temperature. Sections were then stained with 0.1% DAPI in PBS for 5 minutes before being washed in PBS three times for 5 minutes each. Finally, a coverslip was mounted using Fluoromount G (SouthernBiotech, #0100-01) and sealed before being imaged with the Nikon Eclipse 80i confocal microscope (CVI Melles Griot Ion Laser, NIS-Elements v4.0 software).

2.5 Immunofluorescence on mouse embryonic stem (ES) cells

2.5.1 Preparing cells

Cells were cultured as is described in 2.3. For immunofluorescence, Round German Glass Cover Slips (Chemglass, #89167-106) were autoclaved and sterilized with 95% ethanol before being placed into a 24 well cell culture plate (Costar, #3524) to let dry completely. The coverslips were coated in 0.1% gelatin and then cells were harvested and split into the 24 well plate at a concentration of approximately 75 000 cells/mL. ES cells were incubated for up to 3 days to allow for colonies to grow.

2.5.2 Colocalization immunofluorescence

Cells were fixed in ice-cold methanol for 20 minutes at -20°C and then washed three time in PBS for 5 minutes each. Cells were then permeabilized with 0.1% Triton X-100 diluted in PBS for 15 minutes at room temperature. All subsequent steps were also performed at room temperature. After being washed three time in PBS for 5 minutes each, cells were blocked in 5% non-fat milk diluted in TBS. Cells were then incubated in primary

antibody diluted in 5% milk/TBS for 1 hour. Antibody concentrations were as follows: purified rabbit polyclonal anti-CECR2 (Niri et al., in prep) at 1:10,000, rabbit polyclonal anti-LUZP1 (Protein tech #17483-1-AP) at 1:1000, and rabbit polyclonal anti-CCAR2 (Protein tech #22638-1-AP) at 1:1000. After being washed three time in PBS for 5 minutes each, cells were incubated in Alexa Fluor® 488 AffiniPure Fab Fragment Goat Anti-Rabbit IgG (H+L) (Jackson ImmunoResearch, #111-547-003) reconstituted in PBS and used at a 1X concentration for 1 hour in the dark. Since all the primary antibodies were made in rabbit, the use of FAB fragments was necessary (see 2.4.3). All subsequent steps were performed in the dark. Cells were washed three time in PBS for 5 minutes each, blocked in 5% milk/TBS for 30 minutes, and then put into the second primary antibody diluted in 5% milk/TBS for 1 hour. After being washed three time in PBS for 5 minutes each, cells were incubated in secondary antibody (Goat anti-Rabbit IgG (H+L) Secondary Antibody-Alexa Fluor 555, Invitrogen #A-21429) diluted in PBS for 30 minutes. Cells were washed 3 times in PBS for 10 minutes each, stained with DAPI at 1:500 diluted in PBS for 5 minutes, and washed again twice for 10 minutes each. Coverslips were mounted onto slides using Fluoromount G (SouthernBiotech, #0100-01) and sealed before being imaged with the Nikon Eclipse 80i confocal microscope (CVI Melles Griot Ion Laser, NIS-Elements v4.0 software).

2.6 Immunofluorescence of Cccr2 transfected Caco-2 cells

Caco-2 epithelial cells were cultured by Parmveer Singh according to methods listed in Singh, 2016. TranIT-LT1 Transfection Reagent (Mirus Bio LLC, #MIR2304) was mixed with 50 µl of Opti-MEM I Reduced-Serum Medium (Gibco, # 31985062) and 1 ug of human CECR2 pENTRTM11 plasmid (Leduc, 2015) in a microcentrifuge tube. This was incubated at room temperature for 20 minutes, and then the tube was mixed again. The solution was added dropwise onto cells, and then swirled gently to mix into media. Cells were incubated at 37°C for 48 hours. Immunofluorescence and imaging was performed as described in 2.5.2 except anti-CECR2 was used at a concentration of 1:500.

2.7 Vaginal cytology

2.7.1 Vaginal swabbing

Since the cycling of females seemed abnormal, due to large proportions of unsuccessful matings and pregnancies, vaginal cytology was used. This method allowed me to determine which females were ready for mating and aimed to determine if mice were pregnant 9 days after a vaginal plug was detected. Vaginal swabs were prepared by wrapping a small piece of cotton around the tip of a wooden toothpick and autoclaved prior to use. The sterilized vaginal swab was inserted into the mouse vagina at a depth of approximately 2 mm, and gently pressed against the vaginal wall with a circular rolling motion to collect cells (as described by Byers, Wiles, Dunn, & Taft, 2012). The swab was immediately rolled onto a glass slide (Technologist Choice, #LAB-034) and left to air dry for a few minutes. Slides were stored at room temperature until staining for analysis.

2.7.2 Staining of vaginal swabs

Slides were briefly dipped in 0.025% methylene blue, ensuring the total smear was submerged. Any excess liquid was removed with a Kimwipe without disrupting the smear. The slides were placed vertically and allowed to dry completely before being stored at room temperature.

2.7.3 Analysis of vaginal swabs

Using a compound microscope, stained slides were analyzed to visualize the relative proportions of leukocytes, cornified epithelial cells, and nucleated epithelial cells. The presence/absence and relative proportions of these cells types determine the stage of the estrous cycle. The estrous cycle stage identification tool from Byers et al., 2012 was used in the analysis (see Figure 2.7.3.1.)

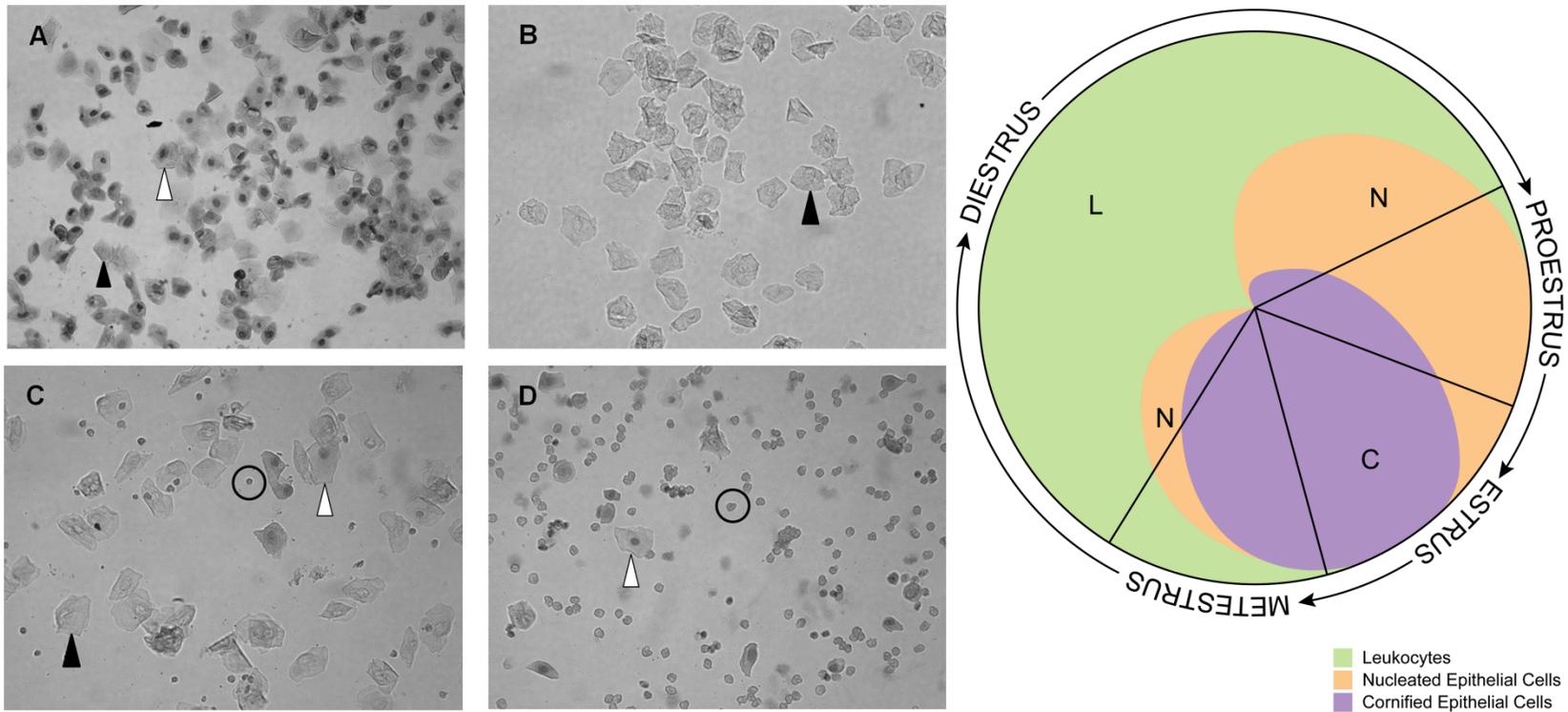


Figure 2.7.3.1. Vaginal cytology and identification tools of the mouse estrous cycle. Proestrus (A), estrus (B), metestrus (C), and diestrus (D) are represented demonstrating various cell types. The identified cells are leukocytes (circle), cornified epithelial cells (black arrow), and nucleated epithelial cells (white arrow). A visual representation of the relative amounts of cell types and the corresponding estrous cycle are seen in E. Each stage is divided by a solid black line with an entire cycle taking 4-5 days. Each quadrant shows the relative time spent in that cycle. Figure assembled from Byers et al. (2012).

2.8 Collection of tissues for RNA extraction

2.8.1 Breeding for E9.5 embryos

The vaginal cytology of Balb/c *Cecr2^{+/-Del}* and Balb/c *Cecr2^{+/+}* mice was analyzed as described in 2.6. All mice were vaginally swabbed daily from Sunday to Wednesday, no earlier than 4:00 pm. Mice staged in either late diestrus or proestrus, determined by the overall presence of nucleated epithelial cells, were set up for timed matings the next morning as described in 2.1.2.

2.8.2 Dissection of E9.5 embryo heads

Mice were vaginally swabbed 9 days after a plug was observed during timed matings and analyzed to determine if pregnancy could accurately be predicted (see 2.1, 2.6). The mice were then euthanized by cervical dislocation regardless of prediction. Embryos were dissected in DEPC treated PBS with RNase free tools and using a M5 Stereo Microscope with trans illuminator stand (Wild Heerbrugg). Somites were counted to stage the developmental progress of each embryo, and the degree of neural tube closure was visually observed and recorded. Embryos were selected based on the required criteria for the experiment, and the heads were always separated from the body to focus on neural tissue only. Forceps were used to steady the body of the embryo under 50X magnification, while a disposable carbon steel scalpel blade (Integra™ Miltex®, #21909-618) was used to precisely cut just above the brachial arch and otic vesicle to separate the head from the body (Figure 2.8.2). The body was collected for genotyping, and the head was carefully placed in a 1.5 mL tube and snap frozen in liquid nitrogen for storage at -80°C.

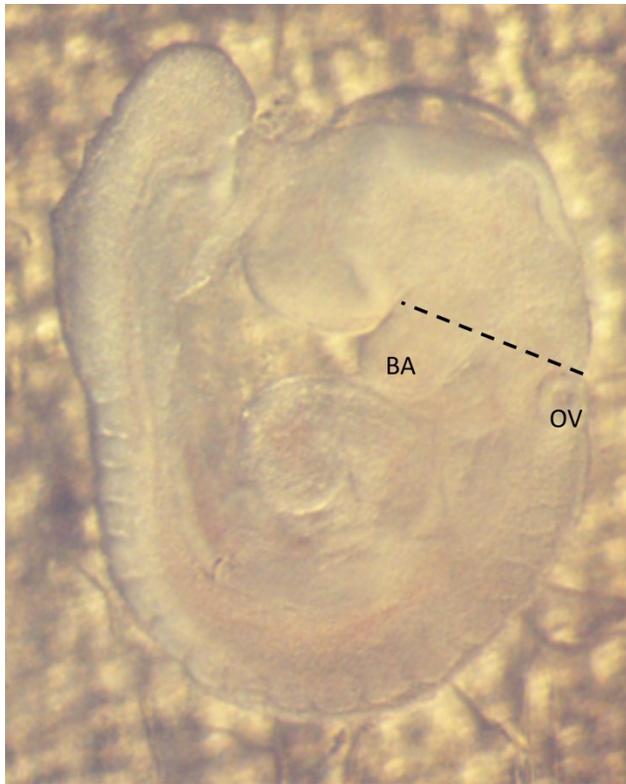


Figure 2.8.2.1. Cut site of E9.5 mouse embryo. Black dotted line represents where the embryonic head is separated from the body. OV is the otic vesicle and BA is the brachial arch.

2.8.3 Dissection of testis

Mice were euthanized by cervical dislocation and testis were dissected using RNase free tools. Each testis was placed into a 1.5 mL tube and snap frozen in liquid nitrogen for storage at -80°C.

2.9 Analysis of gene expression

2.9.1 Individual embryo head RNA extraction

RNA was isolated using the E.Z.N.A.® MicroElute Total RNA Kit (Omega Biotek, #R6831-01) in an RNase free workspace, with slight modifications to the recommended protocol. Embryo heads were removed from storage at -80°C and immediately submerged in 350 µl of TRK Lysis Buffer. Each lysate was passed through a 20 ½ gauge needle 15-20 times. Tubes were vortexed for 30 seconds and then 350 µl of 70% ethanol was added and mixed thoroughly by tube inversion. This entire lysate was transferred to the MicroElute LE RNA Column. The tube was spun at 14 000 rpm for 20 seconds, and then the column

was washed with 500 μ l of RWF Wash Buffer and centrifuged again for 30 seconds. The column was then washed twice with 500 μ l of RNA Wash Buffer II, spinning for 30 seconds after each wash. Finally, the column was spun for 2 minutes to completely dry it, before eluting and re-eluting with 20 μ l each of nuclease free water. RNA was then treated with the DNA-free™ DNA Removal Kit (Invitrogen, #AM1906) as per recommended instructions. Quality of the RNA was assayed with the NanoDrop ND-100 Spectrophotometer and quantified using the Qubit fluorometer at the Molecular Biology Service Unit (MBSU).

2.9.2 Testis RNA extraction

RNA from half of one previously collected testis (see 2.7.3) was isolated using the RNeasy Lipid Tissue MiniKit (Qiagene #74804) as per recommended instructions. RNA was then treated with the DNA-free™ DNA Removal Kit (Invitrogen, #AM1906) and quality was assayed with the NanoDrop Spectrophotometer and quantified using the Qubit fluorometer at the MBSU.

2.9.3 cDNA synthesis

cDNA was synthesized using qScript cDNA SuperMix (Quanta Biosciences, 95048-100) with 100 ng - 1 μ g of RNA. Typically, 320 ng of embryo head RNA and 1 μ g of testis RNA was used for cDNA synthesis. The manufacturer's included protocol was used.

2.9.4 qRT-PCR

Primers were designed to span one exon-exon junction near the 3' end of the gene using PrimerQuest and ordered through IDT. Some primers were ordered to be compatible with RNase-H dependent PCR (rhPCR). These primers have an RNA base and blocking moiety at the 3' end to increase primer specificity (Dobosy et al., 2011). Each reaction occurred in a MicroAmp® Optical 384-Well Reaction Plate (Applied Biosystems by Life Technologies, #4309849) and consisted 1.6 μ M of each forward and reverse primer, 5 μ l of SYBR master mix (made in house at MBSU), 2.5 μ l of cDNA template at a 1/16 dilution, and 0.5 μ l of RNase-H2 enzyme (IDT) if applicable. Reactions were set up using the Biomek® 3000 Laboratory Automation Workstation (Beckman Coulter). The plate was then sealed using MicroAmp® Optical Adhesive Film (Applied Biosystems by Life Technologies, #4311971). The plate was shaken for 30 seconds at speed setting 6.5, and then centrifuged for 2 minutes at 2000 rpm using the Allegra X-22 Centrifuge (Beckman Coulter). qRT-PCR

was performed using the QuantStudio 6 Flex Real-Time PCR System and software (Applied Biosystems by Life Technologies). For primer validation, a 5-point ¼ dilution series was run using the following cycling conditions: 2 minutes at 50°C, 10 minutes at 95°C, 15 seconds at 95°C and 1 minute at 60°C for 40 cycles, 15 seconds at 95°C, 1 minute at 60°C, and then 15 seconds at 95°C. For relative quantification, cycling conditions were: 2 minutes at 95°C, 15 seconds at 95°C and 1 minute at 60°C for 40 cycles, 15 seconds at 95°C, 1 minute at 60°C, and then 15 seconds at 95°C. For relative quantification analysis, samples were normalized to the endogenous control *Gapdh* and the delta-delta CT ($\Delta\Delta CT$) method was used with a minimum of 3 biological replicates with 3 technical replicates each.

2.9.5 RNA sequencing

RNA was extracted as described in 2.8.1 for 10 samples - 5 *Cecr2*^{+/+} and 5 *Cecr2*^{Del/Del} embryo heads. The integrity of the RNA was also measured using the Agilent 2100 Bioanalyzer, following the manufacturers protocol. All RNA submitted for library prep had an RNA Integrity Number of >8. Barcoded cDNA libraries were prepared with the TruSeq RNA Sample Prep Kit v2 through Delta Genomics and in conjunction with the Agricultural, Food & Nutritional Genomics and Proteomics Lab at the University of Alberta. Libraries were sent to Delta Genomics where they were processed together in the same lane on the Illumina HiSeq 4000. The data analysis was performed by Arun Kommadath in Dr. Paul Stothard's Lab at the University of Alberta using the following software: Read trimming – Trimmomatic v0.36 with the parameters phred33 ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36 (Bolger, Lohse, & Usadel, 2014); Quality control – FastQC v0.11.5, QualiMap v.2.2, and MultiQC v0.8 (Okonechnikov, Conesa, & García-Alcalde, 2016); Read mapping – STAR v2.5.2b with default parameters (Dobin et al., 2013); Read counting – featureCounts v1.5.0-p3 with parameter -s 0 -p -t exon -g gene_id -a MmusculusNCBI37.gtf (Liao, Smyth, & Shi, 2014); Sorting/Indexing alignment (bam) files – samtools v1.3.1 software (Li, 2012); Read mapping visualization – IGV v2.3.91 (Thorvaldsdóttir, Robinson, & Mesirov, 2013); Differential expression analysis and associated tests – R (R version 3.3.1 (2016-06-21)) statistical programming language. In R, the specific software packed used from CRAN and Bioconductor version (3.3) were: Differential expression analysis – edgeR version 3.14.0 (complete citations found in Chen,

McCarthy, Robinson, & Smyth, 2014) using the calcNormFactors function for data normalization (applies to data a trimmed mean of M-values (TMM) normalization to adjust for RNA library compositional differences (Robinson & Oshlack, 2010); Gene ontology database – GO.db version 3.14.0; Mouse gene annotation database – org.Mm.eg.db version 3.14.0.

2.10 Protein analysis of the CERF complex

2.10.1 Whole lysate protein extraction

Tissue was dissected and kept on ice in cold PBS until ready for use. In the case of cells, they were washed 3 times with cold sterile PBS before use. For testis, the tunica was removed and 1 mL of freshly-made cold non-denaturing lysis buffer (20 mM Tris-HCl pH 8.0, 420 mM NaCl, 10% glycerol, 1% IGEPAL CA-630, 2 mM EDTA, in HPLC grade water) + 1% protease inhibitor cocktail (Sigma-Aldrich, #P8340) was added to a 15 mL dounce homogenizer. Twenty slow strokes were applied to create a uniform homogenate which was then transferred to a cooled 1.5 mL microcentrifuge tube. For cells, 1 mL of freshly-made cold non-denaturing lysis buffer + 1% of proteinase inhibitor as previously described were added to the plate, and cells were then dislodged using a clean plastic scraper and the cell suspension was transferred to a cooled 1.5 mL microcentrifuge tube. For brain tissue, 2X volume of freshly-made cold non-denaturing lysis buffer + 1% of proteinase inhibitor was added to a 1.5 mL tube. Homogenization for both cells and brain tissue was performed by passing the lysate through a 23 $\frac{3}{4}$ g needle until no longer viscous. In all cases, the homogenate was shaken for 30 minutes to 1 hour at 4°C, and then spun at 12000 rpm for 20 minutes at 4°C. The pellet of cell debris was discarded, the supernatant was collected and quantified using the Bio-Rad DC Protein Assay. Samples were then equalized when possible and aliquots were snap frozen in liquid nitrogen for storage at -80°C.

2.10.2 Western Blotting

SDS-polyacrylamide gel electrophoresis was performed using the Mini Protean III™ electrophoresis system (Bio-Rad) according to manufacturer instructions. A 4% stacking gel (4% acrylamide, 125 mM Tris pH 6.8, 0.1% SDS, 0.1% APS, 0.1% TEMED) and 7.5% separating gel (7.5% acrylamide, 375 mM Tris pH 8.8, 0.1% SDS, 0.08% TEMED) were made with HPLC-grade water to separate proteins. Protein samples were prepared for

western blot (1X NuPAGE LDS sample buffer, 0.1 M DTT) either fresh after extraction, or from a frozen aliquot. In either case, protein was denatured by boiling for 5 minutes and then kept on ice or stored at 4°C until use. After loading, protein was electrophoresed through the stacking gel at 120V, and through the separating gel at 175V in a Tris-glycine running buffer (25 mM Tris, 190 mM glycine, 0.1 % SDS) until desired separation was achieved. Protein migration was monitored using a pre-stained 170 kDa protein ladder (Thermo Scientific, # 26616). Proteins were transferred to a PVDF membrane (Millipore, cat. no. IPVH00010) in a Tris-glycine transfer buffer (25 mM Tris, 190 mM glycine, 10 % methanol) for 30 minutes at 350 mA using the Mini Protean IIITM submerged tank wet transfer unit (Bio-Rad). The membrane was washed with 0.05% TBST for a few minutes before blocking in 5% non-fat milk diluted in TBST on a shaker for 1 hour at room temperature. The membrane was then incubated in primary antibody diluted in 5% milk/TBST at the following concentrations: Affinity purified rabbit polyclonal anti-CECR2 antibody (Niri, in prep) 1:10,000, rabbit polyclonal anti-LUZP1 (Protein tech, #17483-1-AP) 1:1000, mouse monoclonal anti-tubulin (Sigma-Aldrich, cat. no. T6199) 1:10,000, rabbit polyclonal anti-CCAR2 (Protein tech, #22638-1-AP) 1:1000.

Chapter 3 - Results

Objective 1: Confirm and clarify the CERF protein complex composition

3.1 CECR2 is present in the nucleus with CCAR2 and LUZP1 in mouse ES cells

Co-immunoprecipitation assays on mouse ES cells previously showed an interaction between CECR2 and CCAR2, and CECR2 and LUZP1 (Niri, unpublished). Previous studies suggested that CCAR2 and LUZP1 would reside inside the nucleus of the cell (C. C. S. Chini, Escande, Nin, & Chini, 2010; Lee et al., 2001). Therefore, to assay the spatial localization of recently identified CERF protein complex members LUZP1 and CCAR2, I performed immunofluorescence in mouse ES cells and co-stained for CECR2. DAPI was used in all experiments to mark the DNA. Unfortunately, all of the IF -validated antibodies against these proteins were made in rabbit. Therefore FAB (fragment antigen-binding) fragments were utilized in order to achieve colocalization. FAB fragments act as a conjugated secondary antibody, except they are monovalent and therefore able to cover the rabbit antigen binding sites of the primary antibody (Lewis Carl, Gillete-Ferguson, & Ferguson, 1993). This allows for the application of another same-host primary antibody on the same sample, with a subsequent secondary antibody conjugated to a different fluorophore, without the second fluorophore labelling the first applied primary antibody.

Previous experiments demonstrated that *Cecr2^{GT}* is localized to the nucleus of CT45 ES cells, and CECR2 was detected in western blots using a nuclear protein extract from both ES cells and embryo (Thompson et al., 2012, Niri, unpublished). Immunofluorescence in P19 cells also showed nuclear localization (Niri, unpublished). Based on previous immunoprecipitation experiments which showed protein-protein interactions, I was expecting to see CECR2 colocalize with both CCAR2 and LUZP1 in the nucleus of mouse ES cells.

I performed this experiment primarily using TT2 ES cells, which were used for previous IP experiments. In LUZP1 colocalizations I also used the E14 ES cells from Dr. Laszlo Tora since he had provided both a wildtype (*Luzp1^{+/+}*) and LUZP1-deficient (*Luzp1^{GT/GT}*) line. As Figure 3.1.1. demonstrates, CECR2, CCAR2, and LUZP1 are all present in the nucleus in TT2 ES cells. CECR2 and CCAR2 were both almost exclusively located in

the nucleus, but LUZP1 was very strongly present in the cytoplasm with faint nuclear staining. This result is analogous to staining in other cell lines using the HPA028506 antibody as shown on The Human Protein Atlas (www.proteinatlas.org). The strong cytoplasmic staining makes confocal visualization of the nuclear staining difficult.

Since LUZP1 is predominantly not in the nucleus of ES cells, visualizing the colocalization with CECR2 is difficult. While some LUZP1 puncta are seen in the nucleus, the staining is not bright enough to produce a yellow overlap when the channel is merged with CECR2. However, as indicated by the white arrows in Figure 3.1.1.B, there are areas within the nucleus where both CECR2 and LUZP1 are present. Interestingly, this cytoplasmic localization is contrary to the nuclear localization reported in more differentiated cells (Hsu et al., 2008; M. W. Lee et al., 2001). I also noticed in the *Luzp1^{+/+}* E14 ES cell lines that sometimes there were differentiated cells present. In these cells, the LUZP1 staining was more predominant in the nucleus (Figure 3.1.1.C). Interestingly, the non-nuclear LUZP1 staining in these differentiated cells does not perfectly align with actin or tubulin (data not shown) and it is therefore unknown which cellular component is being labelled.

Colocalization of puncta was difficult in all colocalization assays. The use of FAB fragments restricted imaging settings as even with good epitope coverage, some amount of bleed-through from the primary antibody is inevitable. To control for this, I put one coverslip of cells for each antibody coated with FAB fragments in buffer instead of a second primary antibody, and continued the protocol as described in 2.5.2, adding the secondary anti-rabbit antibody. This allowed me to ensure the FAB fragments sufficiently covered the first rabbit primary antibody epitopes, which therefore did not bind to the secondary antibody. This control was used to determine the appropriate imaging settings by imaging at the highest setting that showed no fluorescence/bleed-through.

All colocalizations were assayed reciprocally however even at a concentration of 1:50 FAB fragments could not sufficiently cover anti-LUZP1 to be able to image the second channel without large amounts of bleed-through. The LUZP1 staining was very intense in the cytoplasm and along the border of the cell colonies, which could explain the difficulty in

covering the epitopes using FAB. This is not surprising as studies show this technique is most effective when labelling the least abundant epitope first (Lewis Carl et al., 1993), making our particular experiment difficult as LUZP1 is the least abundant protein assayed in the nucleus, but the most abundant overall. There is also diffuse, ubiquitous, and faint staining in the LUZP1-deficient cell line which restricted the settings for imaging LUZP1. I used this control to define the maximum settings to image the wildtype cell lines. Under these settings, little-to-no staining was visualized in the LUZP1-deficient cell line, but nuclear staining was seen in the wildtype cell lines. This allowed me to attribute all seen staining to LUZP1 and not background, but likely also reduced the amount of nuclear LUZP1 able to be visualized.

Colocalization of CCAR2 with CECR2 was performed reciprocally and showed similar results to each other, presumably due to a lack of imbalance in abundance. Although both proteins are located exclusively in the nucleus, their patterning is quite different (Figure 3.4.1). There is minimal, albeit present overlap between these two proteins in the nucleus of ES cells. A yellow overlap is difficult to observe but is indicated in Figure 3.1.1.A. Colocalization with LUZP1 showed similar results and problems as described with CECR2. Both CCAR2 and LUZP1 are in the nucleus and seen in overlapping spatial locations as indicated by the white arrow heads and small amounts of yellow overlap.

Since a negative control for CECR2 does not exist in ES cells, I checked the specificity of the antibody for immunofluorescence by transiently transfecting Caco-2 cells which do not express *Cecr2* endogenously with a vector containing human *Cecr2*. I performed immunofluorescence with our *Cecr2* antibody and DAPI to mark the DNA. Only select transformed cells had CECR2 staining, which was primarily in the nucleus (Figure 3.1.2). There was some staining in the cytoplasm, seen as isolated puncta. This is likely a result of overexpression of CECR2 in the cell, and has been seen in other transfection assays performed with this cell line and others, along with transfection of other proteins (Singh, 2016).

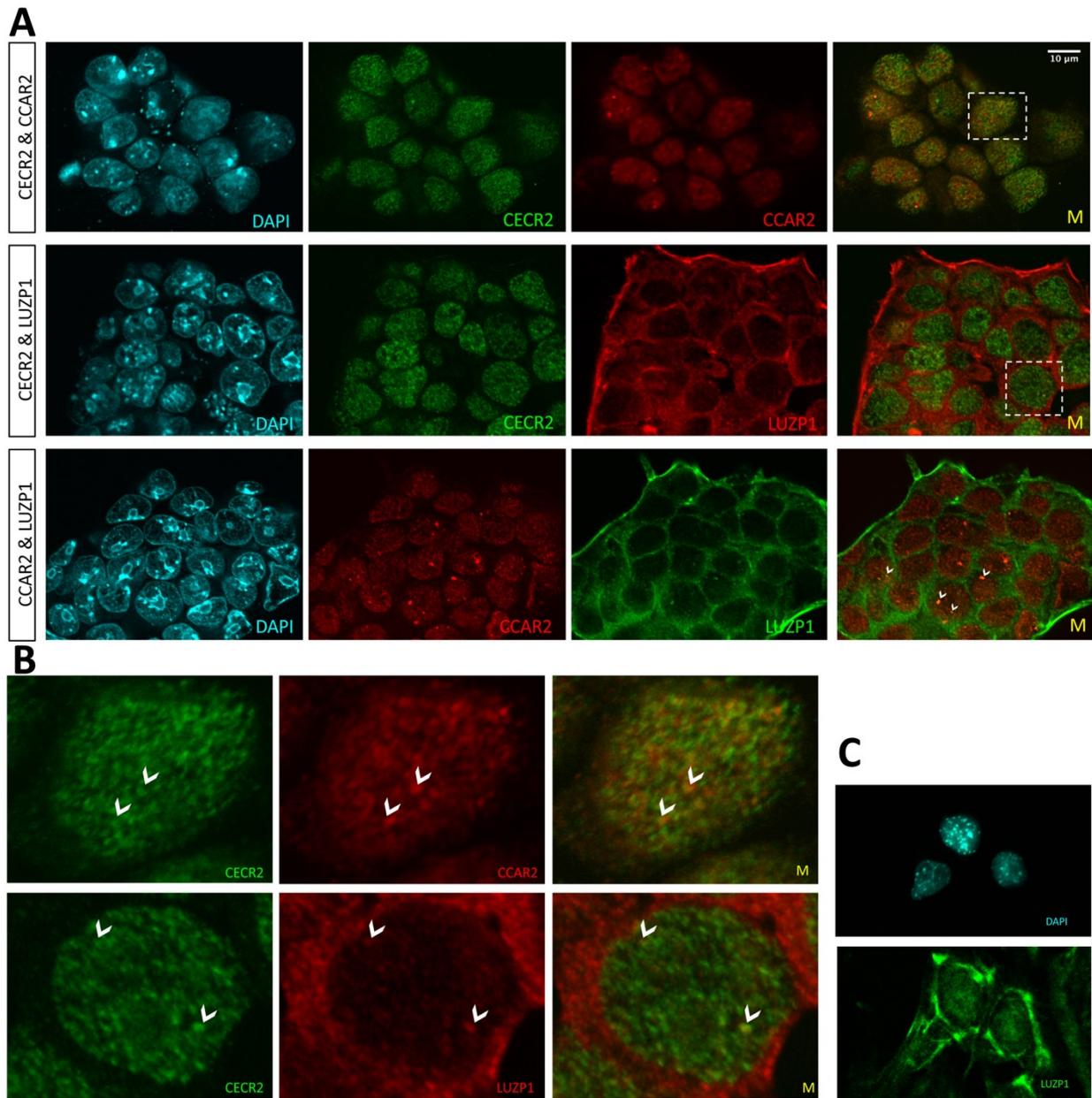


Figure 3.1.1. Immunofluorescence of mouse embryonic stem cells. (A) Colocalization immunofluorescence was performed to localize CECR2, CCAR2, and LUZP1 in TT2 ES cells. DAPI was used to mark the DNA. Panels are arranged in the order staining was performed using FAB fragments. The right panel M indicates the merge of the red and green channels. White dotted rectangles are enlarged in panel (B), with white arrow heads indicating regions of colocalization. (C) In differentiating *Luzp1*^{+/+} E14 ES cells, LUZP1 is more obviously present in the nucleus.

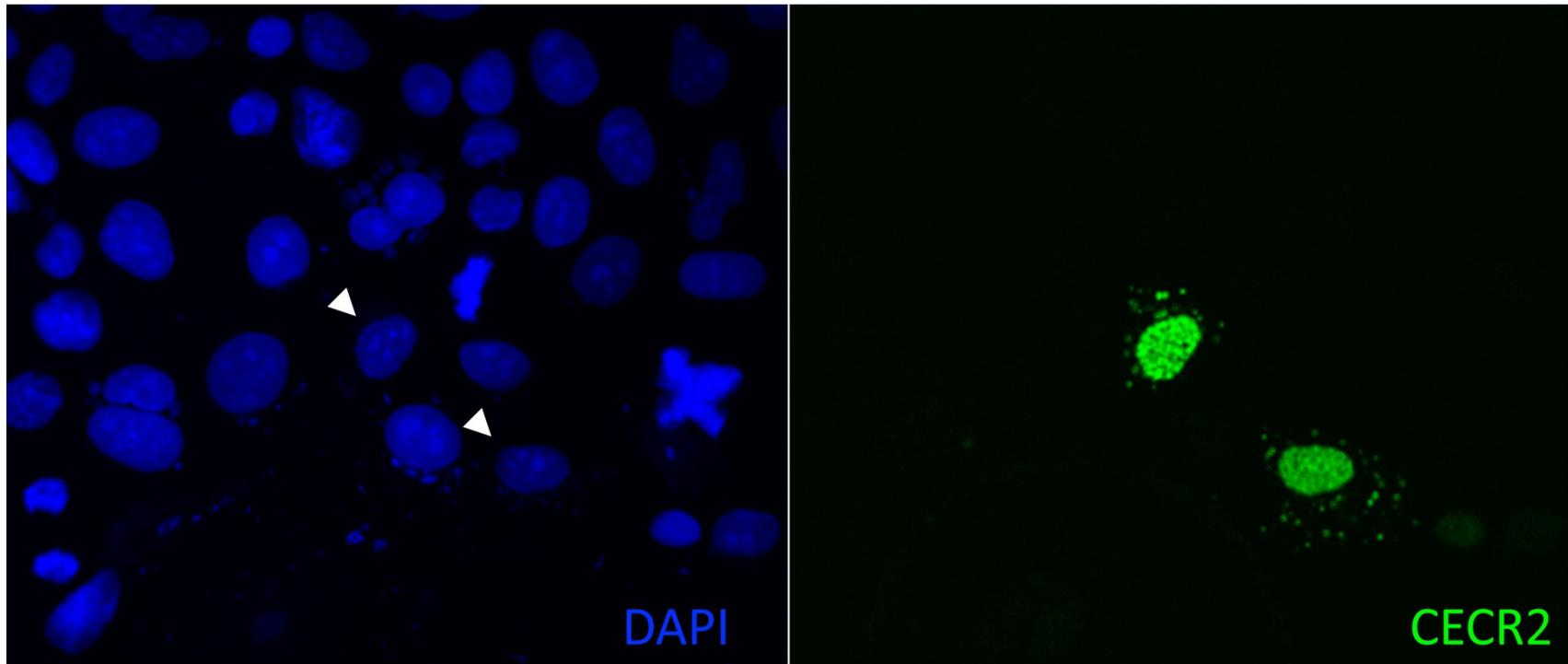


Figure 3.1.2. Immunofluorescence of transiently transfected CaCo2 epithelial cells with human *CECR2* shows antibody specificity. Cells were stained with a *Cecr2* antibody (green) and DAPI (blue) after transfection. The two green cells in the right panel represent two transfected cells expressing *CECR2*, since CaCo2 cells do not express *Cecr2*. White arrow heads in the left DAPI panel indicate the transfected cells.

3.2 CECR2 does not colocalize with CCAR2 in the testis and LUZP1 colocalization is inconclusive

Previous immunofluorescence experiments demonstrated that CECR2 localizes to the spermatogonia A cells of the testis; these cells are spaced intermittently in the outermost layer of the seminiferous tubule (Norton, unpublished). Considering the immunoprecipitation of CECR2 and LUZP1 robustly showed no interaction, I was expecting to see CECR2 fail to colocalize with LUZP1 in the testis. I was unsure if I would see CECR2 colocalize with CCAR2, since the immunoprecipitation showing interaction with CCAR2 was less robust and only successful in 1/9 attempts. CCAR2 and LUZP1 were both previously uncharacterized in the testis.

To assay the spatial localization of recently identified CERF protein complex members LUZP1 and CCAR2 I performed immunofluorescence on testis paraffin sections and co-stained for CECR2. DAPI was used in all experiments to mark the DNA and visualize the cells of the testis. These are the same antibodies used in ES cell colocalization, therefore FAB fragments were once again used.

CCAR2 immunofluorescence showed nuclear staining throughout the cells of the seminiferous tubules – from the inner more differentiated cells to the outermost stem-like cells (Figure 3.1.2). Interestingly, it was not in all of the outer spermatogonia cells. When I co-stained with CECR2, it was striking that the two proteins were not present in the same cell type. CECR2 is present in spermatogonia A cells, and CCAR2 is not. CCAR2 is present in the other outer cells of the tubules – presumably spermatogonia B cells – however immunohistochemistry would be needed to conclusively determine the identity of these CECR2-/CCAR2+ cells. These results support the 8/9 immunoprecipitations that did not show CECR2 and CCAR2 interaction (Niri, in prep). The one co-IP that was successful was likely an artifact. It should be noted that all diffuse CECR2 staining seen in the inner tubule is also an artifact, as this is also present in the mutant testis that lacks CECR2. Therefore, any overlapping localization showing yellow in the central area of the testis in the merge panels of Figure 3.2.1. should be disregarded.

Colocalization of CECR2 and CCAR2 was assayed in both directions with a corresponding FAB control (data not shown). Unfortunately, even at high FAB

concentrations the CCAR2 epitopes were never sufficiently covered to image the lack of colocalization with CECR2. Background CCAR2 bleed-through was always detected when imaging CECR2, or the settings were too dim to detect CECR2. Therefore, the only successful direction was using anti-CECR2 first. This is presumably because CECR2 is less ubiquitous within the testis, and so epitope coverage is more easily achieved. This is also the recommended method to assay the least abundant protein first (Lewis Carl et al., 1993).

Staining LUZP1 in the testis was problematic, similar to ES cells. LUZP1 appears primarily in the inner, more differentiated cells of the seminiferous tubule with a primarily cytoplasmic localization (Figure 3.2.1. and Figure 3.2.2.). However, there does appear to be a small amount of LUZP1 in the nucleus of testis cells, specifically spermatogonia A cells where CECR2 is present (Figure 3.2.2. white arrow heads). Unfortunately, we do not have LUZP1-deficient testis so it was difficult to determine appropriate settings for imaging. Moreover, the confocal used for imaging does not have the power or resolution to image at a higher magnification to look for nuclear localization. Therefore, it is not possible to conclude whether or not CECR2 and LUZP1 are spatially present together.

When I colocalized LUZP1 and CECR2 there were problems in both directions. With LUZP1 first, the FAB fragments did not completely cover the epitopes so there was a lot of bleed-through. LUZP1 staining also diminished in quality when done prior to CECR2, becoming fainter and more diffuse-looking. This is perhaps unsurprising as the downstream protocol is longer and includes more washes. When CECR2 was stained for first (Figure 3.2.2), it appears that anti-LUZP1 is binding to the FAB fragments as there is nuclear staining when viewing LUZP1. This nuclear staining in the cells of the tubule periphery is not seen when immunofluorescence for LUZP1 is assayed first, or alone. This is a facet of this experiment I noticed with all the antibodies, and is a noted caveat of using FAB fragments (Lewis Carl et al., 1993). However, I was able to control for this with CECR2 and CCAR2 together because the immunofluorescence staining is brighter and less diffuse, and the second blocking after FAB fragments proved more effective likely due to lower epitope abundance. I also assayed for CCAR2 and LUZP1 colocalization but encountered the same difficulties as CCAR2 is also nuclear (Figure 3.2.2). Therefore, colocalization cannot be conclusively determined for LUZP1 and CECR2 or CCAR2.

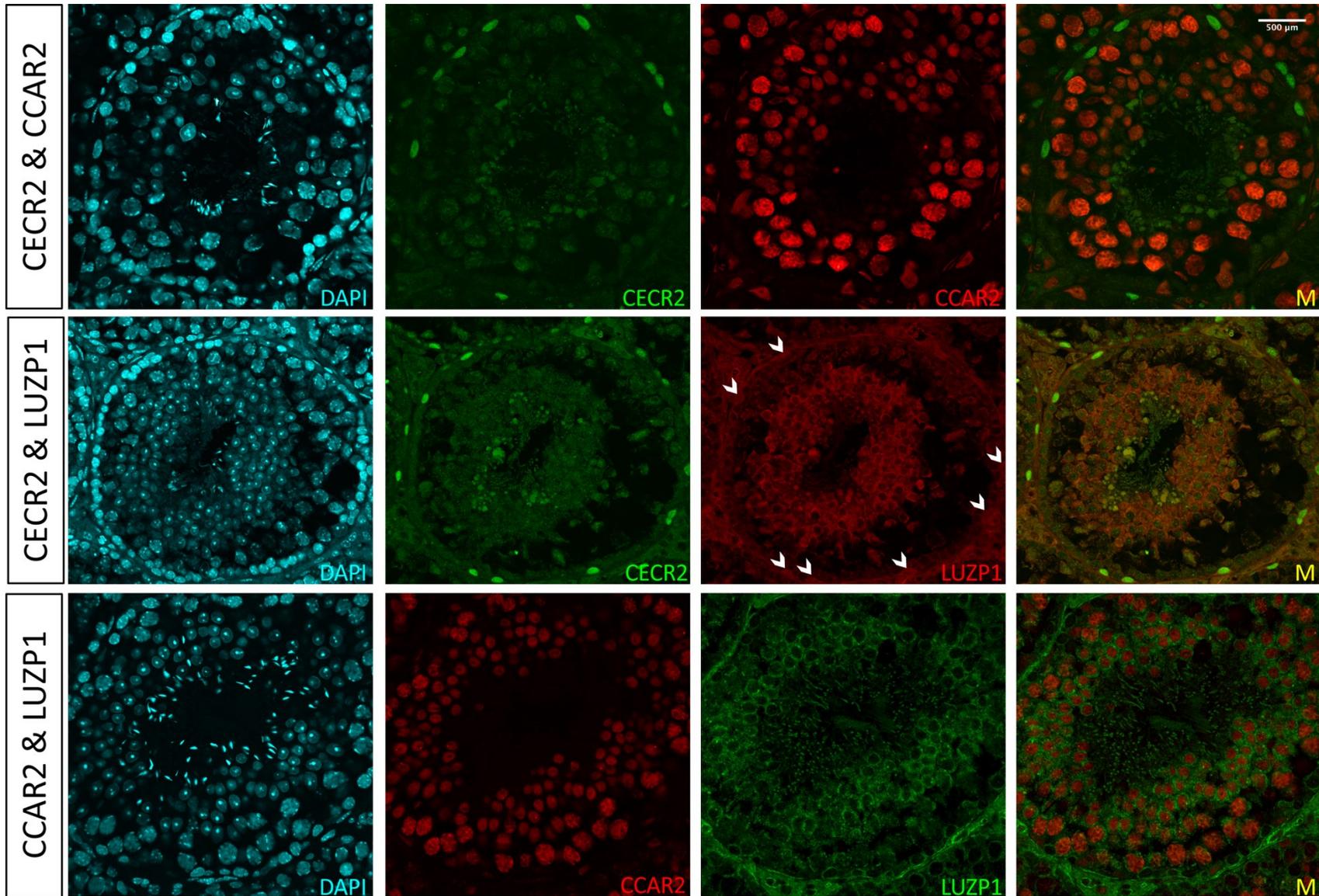


Figure 3.2.1. Immunofluorescence on paraffin sections of mouse testis.

Colocalization immunofluorescence was performed to localize CECR2, CCAR2, and LUZP1 in paraffin sections of mouse testis. DAPI was used to mark the DNA. The right panel M indicates the merge of the red and green channels. CECR2 staining in the lumen of the seminiferous tubules is an artifact also seen in *Cecr2* mutant testis and should be disregarded when shown as yellow in the merge. All panels are arranged in the order staining was performed using FAB fragments. White arrow heads in the red LUZP1 panel indicate cells that have CECR2 staining.

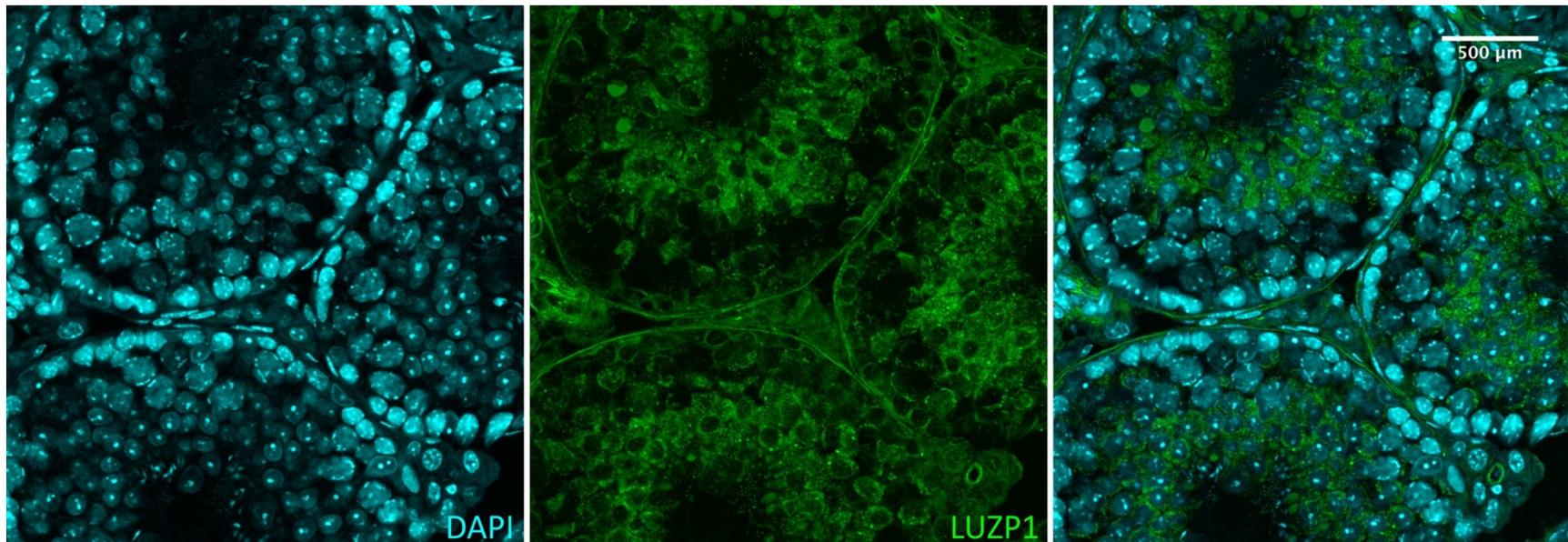


Figure 3.2.2. Localization of LUZP1 in mouse testis using immunofluorescence. LUZP1 localizes primarily to the cytoplasm of the inner testis cells. DAPI (cyan) marks the DNA in the left panel, LUZP1 (green) is shown in the middle panel, and a merge of the two channels in the right panel.

Objective 2: Investigate the role of CECR2 as a transcriptional regulator in neurulation

3.3 ChIP-Seq candidate genes

Since the previously performed ChIP-Seq experiment in ES cells and wildtype testis revealed many potential CECR2 binding sites within the promoter region of genes (Niri, unpublished), our goal was to identify CECR2 targets that may have a role in either neurulation or fertility. Since CECR2 could be regulating certain genes in both the testis and ES cells, I also investigated testis binding sites. I focused on those genes involved in neurulation while another graduate student focused on those involved in reproduction. Farshad and I started by investigating genes bearing an overlapping potential binding site within 1 kb of the TSS between CECR2, SNF2H, and LUZP1 in ES cells, and CECR2 and SNF2H in testis. This was decided to be too stringent, as there were not many identified binding sites, so we looked within 5 kb of the TSS. We first considered a p-value of 10^{-5} and then also a less stringent p-value of 10^{-3} . Since the ChIP-Seq assay is not perfect and some binding sites may not have been detected for SNF2H and/or LUZP1, we also considered genes with only a binding site for CECR2.

In testis, we determined 11 candidate genes to be (in alphabetical order) *Elmo1*, *Fgfr4*, *Ggt1*, *Insr*, *Itgb3*, *Lrp6*, *Nfia*, *Pcsk1*, *Phactr4*, *Schip1*, and *Styx*. All of these genes have overlapping ChIP binding sites for CECR2 and SNF2H within 5 kb of the TSS using a p-value of 10^{-5} . My undergraduate student, Jennifer Just, assisted in choosing the testis candidate genes under my supervision. *Lrp6* and *Phactr4* were both selected as knock-out mouse mutants develop neural tube defects. They also have similar respective ChIP binding sites in both ES cells and in testis which would support our hypothesis of some overlapping gene control in ES cells and testis. The other 9 listed genes have mutations that are associated with fertility defects and are good candidates to be explored in the reproduction project.

In ES cells, we determined six candidate genes to be (in alphabetical order) *Hsd17b2*, *Lpar1*, *Lrp6*, *Nf1*, and *Phactr4*. *Hsd17b2*, *Lpar1*, and *Nf1* all have overlapping ChIP binding sites between CECR2 and SNF2H within 5 kb of the TSS using a p-value of 10^{-5} . When the p-value is raised to 10^{-3} , LUZP1 also has an overlapping ChIP binding site for *Hsd17b2*, and *Lpar1*. Using that less stringent p-value, binding sites for *Lrp6* and *Phactr4*

are also detected for CECR2 and SNF2H. These genes have mutations that cause defects in central nervous system development and were candidates for an RNA-Seq comparison and explored for misregulation using qRT-PCR (elaborated on in the subsequent sections).

3.4 RNA-Seq reveals a range of potentially misregulated genes in *Cecr2*^{Del/Del} mutants

I used RNA-Seq to investigate the transcriptome changes in *Cecr2*^{Del/Del} mutant embryos to better understand the neurulation defect and possibly identify direct gene targets of CECR2 when comparing to the ChIP-Seq. To enrich for genes involved in neurulation, I isolated only the neurulating heads of E9.5 12-14 somite embryos as described in 2.8.2. Embryos of any genotype at this stage, had elevated, sometimes bent at the DLHPs, but always non-fused neural folds. It was not possible to distinguish genotype from the appearance of the cranial region. It was however possible to distinguish a closed, or nearly fused neural tube, from one that was open, especially when abnormal (Figure 3.4.1.). After collection, RNA extraction, and library preparation, I submitted 5 female somite-matched embryo head samples of each *Cecr2*^{+/+} and *Cecr2*^{Del/Del} for RNA sequencing. Data analysis was performed by Arun Kommadath in collaboration with Dr. Paul Stothard's lab.

RNA sequencing was successful for all submitted samples. Raw read depths were over 30 million for each sample; the minimum and maximum read count respectively was 30.34 and 46.53, with a mean of 37.66 million reads. A full list of the read depths for each sample can be found in Table 3.4.1. This read depth meets the standards for RNA sequencing (Chen et al., 2014; Conesa et al., 2016). There were no quality control issues for the samples. Data was filtered based on the library sizes post read mapping – this removes any low-level gene expression that may interfere with downstream analysis. The cut-off for gene expression was determined to be counts per million (CPM) greater than 0.21 in at least 5 samples. Overall, there were 16,339 genes expressed in the samples. The biological coefficient of variation (BVC) for this experiment is 0.07739 which is considered normal for inbred mice (Chen et al., 2014). A power analysis determined what fold-change could be reliably detected for differential analysis considering sample read depth and an alpha of 0.05.

Before differential analysis, I requested a cluster analysis of the samples to identify any potential outlier samples. A multi-dimensional scaling (MDS) plot (Figure 3.4.2.) shows the level of similarity between individual samples in the data set. Here, the top 500 genes with the largest absolute log-fold-changes between wildtype and mutant samples is considered. This MDS plot clearly shows an outlier – MT_02 – however this does not determine whether or not this is within a normal range of biological variability.

Differential analysis proceeded initially with the outlier sample and then without. Significance was determined using three parameters: False discovery rate (FDR) < 0.05, fold change (FC) > 1.5 (logFC > 0.5829625), and logCPM > 0.21. When comparing the wildtype and mutant samples, 170 genes were identified as being differentially expressed. 124 genes (72.9%) were upregulated in the mutant, and 46 genes (27.1%) were downregulated. The top 10 upregulated and downregulated genes based on FDR are found in Table 3.4.2., while the top 10 upregulated and downregulated genes based on FC are found in Table 3.4.3. A heat map was generated to observe differences between all 10 samples for all the differentially expressed genes identified (Figure 3.4.3.). All samples are fairly consistent for the set of up and downregulated genes, except MT_02 which was previously identified as an outlier by the MDS plot. This sample does not match its mutant counterparts, showing 21 highly induced genes (shown by intense red bands in heat map). Furthermore, this sample was one of the first embryos collected and had the lowest RIN value. It is likely a technical reason, such as the exact location of the cut during dissection, that accounts for this variability rather than it being within a normal biological variability range. Therefore, another analysis of differentially expressed genes was done with the outlier removed.

Using the same criteria as previously described, this analysis revealed 143 significant differentially expressed genes; 102 genes (71.3%) are upregulated and 41 genes (28.7%) are downregulated. The top 10 upregulated and downregulated genes based on FDR are found in Table 3.4.4., while the top 10 upregulated and downregulated genes based on FC are found in Table 3.4.5. The heat map (Figure 3.4.4.) is now more consistent among the wildtype and mutant samples with this outlier removed.

To further characterize the differentially expressed genes, the data underwent a Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis. Only one GO biological process – nervous system development – was identified as enriched in upregulated genes. In downregulated genes, 10 GO biological processes were enriched, including the regulation of transcription, multicellular organism development, and proximal/distal pattern formation. These results are shown in Table 3.4.6. Enriched gene ontology (GO) biological process terms in RNA sequencing data on *Cecr2*^{+/+} and *Cecr2*^{Del/Del} embryo heads. KEGG analysis only found retinol metabolism to be enriched in downregulated genes.

One interesting additional result from the RNA-Seq read mapping is the lack of any read mapping to a region >1 kb from the first exon of *Cecr2*. Previous data in *Cecr2*^{GT/GT} mice suggested that an alternative start site, exon 0, may result in a *Cecr2* isoform in testis (Norton, unpublished). In *Cecr2*^{Del/Del} mouse embryo heads however, there does not seem to be any *Cecr2* transcript resulting from this alternative start site.

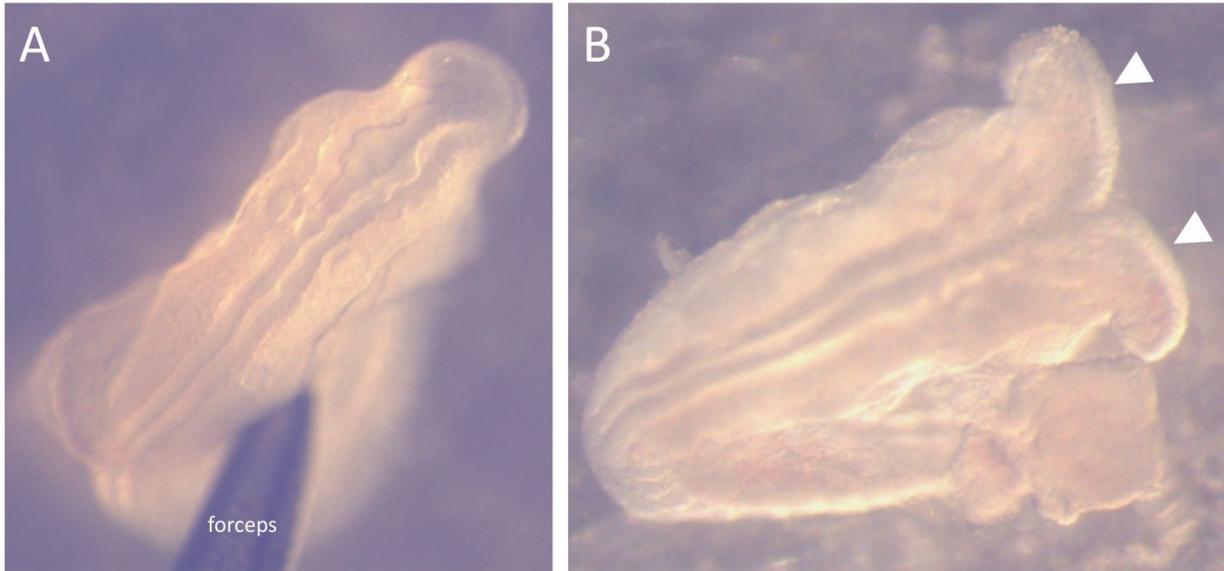


Figure 3.4.1. Visualization of E9.5 embryonic neural tube at the cranial region. Embryos were dissected and the degree of neural tube closure was assessed. (A) Cranial neurulation is normal, and nearly completely fused. (B) Cranial neurulation is abnormal, as indicated by the open and splayed apart neural folds (indicated by white arrows).

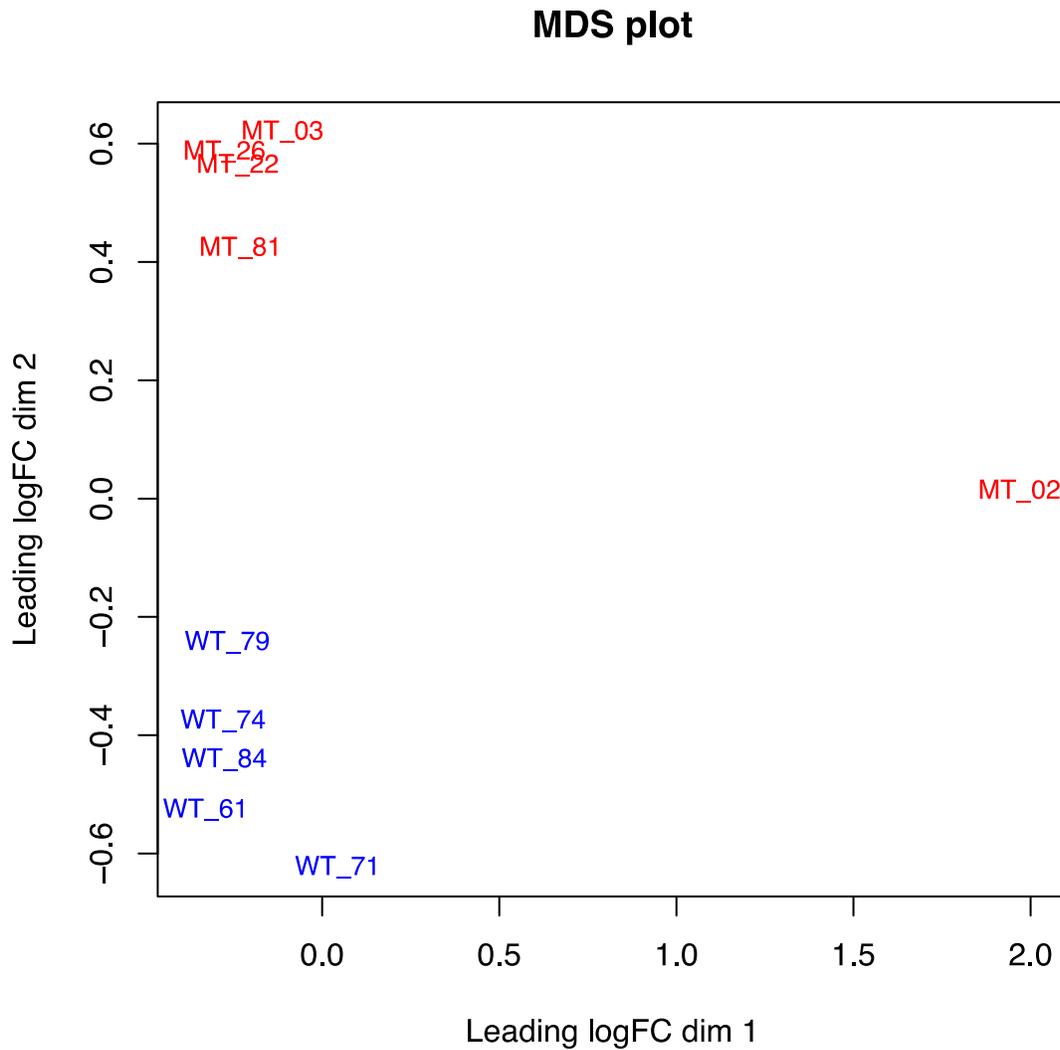


Figure 3.4.2. Multi-dimensional scaling (MDS) plot of RNA-sequencing samples. Five *Cecr2^{Del/Del}* mutant (labelled MT_02, MT_03, MT_22, MT_26, MT_81) and five *Cecr2^{+/+}* wildtype (labelled WT_61, WT_71, WT_74, WT_79, WT_84) samples were clustered to evaluate sample similarity (Arun Kommadath, Stothard Lab).

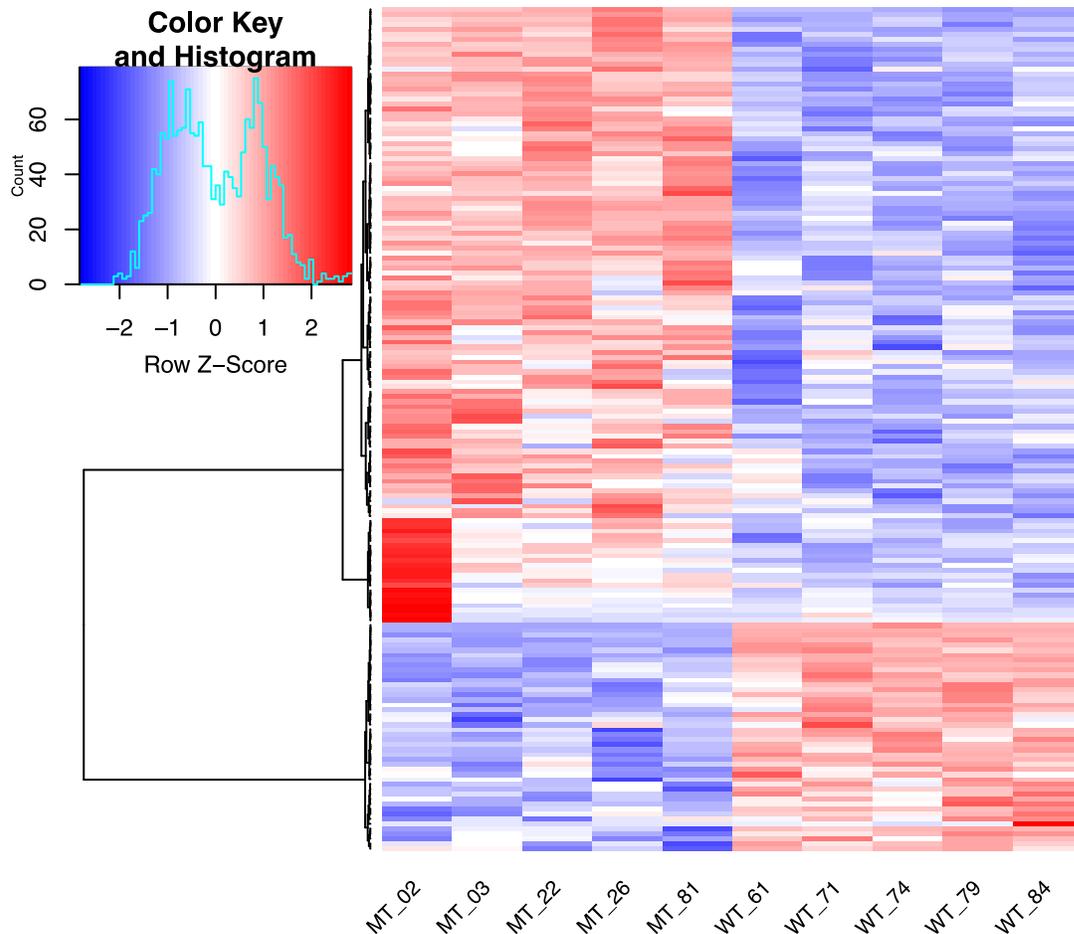


Figure 3.4.3. Heat map of RNA-sequencing samples. Five *Cecr2^{Del/Del}* mutant (labelled MT_02, MT_03, MT_22, MT_26, MT_81) and five *Cecr2^{+/+}* wildtype (labelled WT_61, WT_71, WT_74, WT_79, WT_84) samples had the differentially expressed genes clustered to compared relative expression. The redder the colour, the more highly expressed a gene is; the bluer the colour is, the less expressed that gene is.

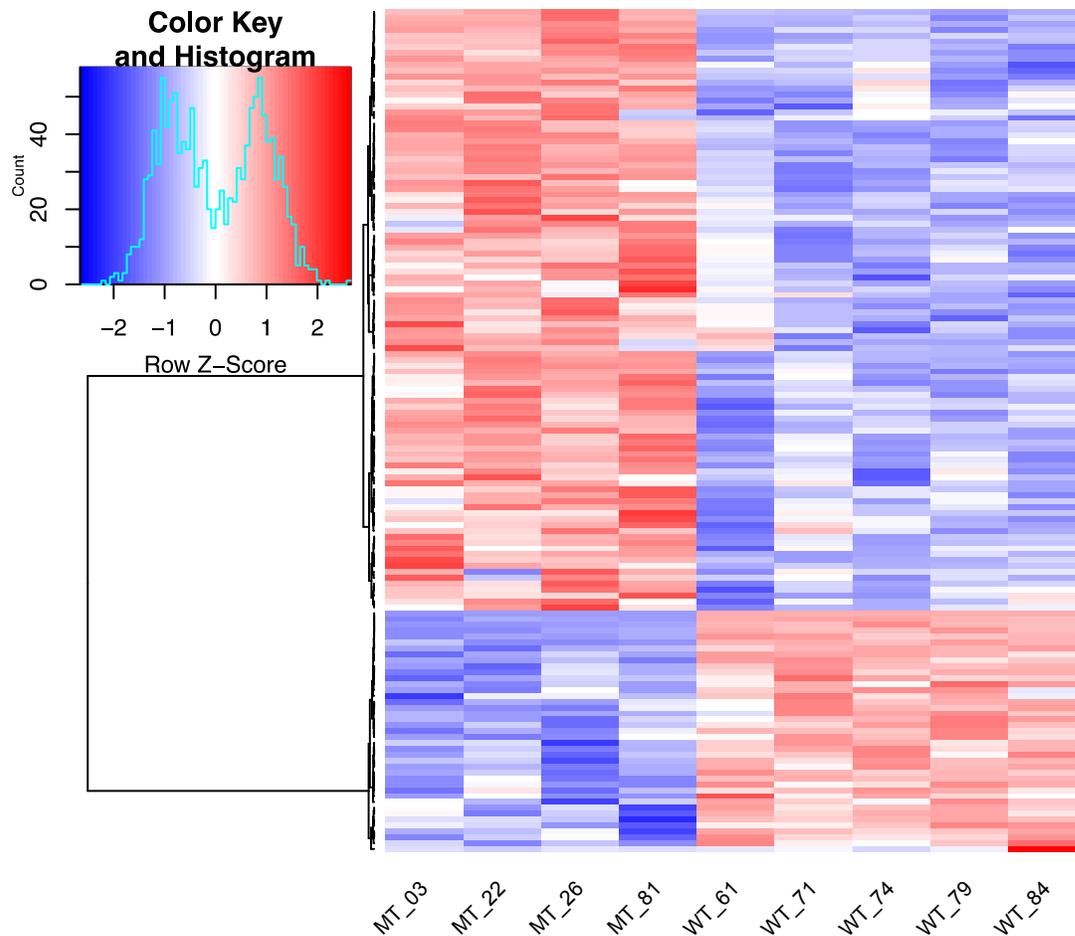


Figure 3.4.4. Heat map of RNA-seq samples with outlier MI_UZ removed. Four *Cecr2^{Del/Del}* mutant (labelled MT_03, MT_22, MT_26, MT_81) and five *Cecr2^{+/+}* wildtype (labelled WT_61, WT_71, WT_74, WT_79, WT_84) samples had the differentially expressed genes clustered to compared relative expression. The redder the colour, the more highly expressed a gene is; the bluer the colour is, the less expressed that gene is.

Table 3.4.1. RNA sequencing information for submitted samples. Information provided by Genome Quebec Nanuq server. MT corresponds to a mutant sample and WT corresponds to a wildtype sample, followed by the sample ID. Quality refers to an array of measures including read quality, composition, mapping, etc. with a score >30 considered very good.

Sample	Number of reads (count per million)	Number of bases (million)	Average Quality
MT_02	36.39	7278.53	38
MT_03	43.40	6880.31	38
MT_22	32.99	6597.52	38
MT_26	34.71	6942.42	38
MT_81	43.56	8712.62	38
WT_61	30.34	6067.26	36
WT_71	35.88	7175.95	38
WT_74	46.53	9305.82	38
WT_79	43.98	8796.20	38
WT_84	37.81	7561.29	38

Table 3.4.2. Top 10 differentially expressed genes filtered by FDR from RNA-sequencing of five *Cecr2*^{+/+} wildtype and five *Cecr2*^{Del/Del} mutant embryo heads.

ensemblID	Gene name	FDR	Fold change
<u>Upregulated genes</u>			
ENSMUSG00000091243	<i>Vgll3</i>	0.000000e+00	2.37765699
ENSMUSG00000001270	<i>Ckb</i>	3.358218e-28	1.71030924
ENSMUSG00000033965	<i>Slc16a2</i>	2.213400e-19	1.70878111
ENSMUSG00000057777	<i>Mab2112</i>	1.568975e-18	1.83491216
ENSMUSG00000023336	<i>Wfdc1</i>	6.899662e-18	2.25612844
ENSMUSG00000032085	<i>Tagln</i>	5.915080e-15	1.79663569
ENSMUSG00000047976	<i>Kcna1</i>	6.640576e-15	1.7624397
ENSMUSG00000024063	<i>Lbh</i>	7.749091e-15	1.54649664
ENSMUSG00000035783	<i>Acta2</i>	7.188202e-14	2.20828105
ENSMUSG00000030507	<i>Dbx1</i>	1.695785e-13	1.8742309
<u>Downregulated genes</u>			
ENSMUSG00000071226	<i>Cecr2</i>	0.000000e+00	-497.58794
ENSMUSG00000024565	<i>Sall3</i>	0.000000e+00	-2.7535893
ENSMUSG00000025927	<i>Tfap2b</i>	7.116900e-46	-2.0246898
ENSMUSG00000068859	<i>Sp9</i>	2.274469e-35	-4.6178823
ENSMUSG00000001622	<i>Csn3</i>	1.776406e-27	-4.6140585
ENSMUSG00000020160	<i>Meis1</i>	2.755676e-24	-1.5203216
ENSMUSG00000048349	<i>Pou4f1</i>	6.317758e-19	-2.7663423
ENSMUSG00000026739	<i>Bmi1</i>	5.645942e-18	-1.6542662
ENSMUSG00000026586	<i>Prrx1</i>	7.736153e-16	-1.9233877
ENSMUSG00000013584	<i>Aldh1a2</i>	1.177102e-10	-2.8909781

Table 3.4.3. Top 10 differentially expressed genes filtered by fold change from RNA-sequencing of five *Cecr2*^{+/+} wildtype and five *Cecr2*^{Del/Del} mutant embryo heads.

ensemblID	Gene name	Fold change	FDR
<u>Upregulated genes</u>			
ENSMUSG00000032083	<i>Apoa1</i>	14.7035956	0.00834585
ENSMUSG00000020609	<i>Apob</i>	14.0988757	3.18E-06
ENSMUSG00000026726	<i>Cubn</i>	8.87957976	0.00906179
ENSMUSG00000024990	<i>Rbp4</i>	4.70162729	0.01846891
ENSMUSG00000024391	<i>Apom</i>	4.40838016	0.02399364
ENSMUSG00000023886	<i>Smoc2</i>	2.81733284	0.00893843
ENSMUSG00000005124	<i>Wisp1</i>	2.69551492	0.01846891
ENSMUSG00000001506	<i>Col1a1</i>	2.61363443	0.00013369
ENSMUSG000000091243	<i>Vgll3</i>	2.37765697	4.92E-76
ENSMUSG000000015484	<i>Fam163a</i>	2.35085425	1.89E-08
<u>Downregulated genes</u>			
ENSMUSG000000071226	<i>Cecr2</i>	-497.58794	0
ENSMUSG000000061462	<i>Obscn</i>	-25.404118	0.00139187
ENSMUSG000000068859	<i>Sp9</i>	-4.6178824	2.27E-35
ENSMUSG000000001622	<i>Csn3</i>	-4.6140585	1.78E-27
ENSMUSG000000013584	<i>Aldh1a2</i>	-2.8909782	1.18E-10
ENSMUSG000000048349	<i>Pou4f1</i>	-2.7663423	6.32E-19
ENSMUSG000000024565	<i>Sall3</i>	-2.7535893	7.75E-86
ENSMUSG000000035456	<i>Prdm8</i>	-2.4891817	1.04E-08
ENSMUSG000000066224	<i>Arid3c</i>	-2.4485829	8.41E-09
ENSMUSG000000024619	<i>Cdx1</i>	-2.3290457	7.56E-06

Table 3.4.4. Top 10 differentially expressed genes filtered by FDR from RNA-sequencing of five *Cecr2*^{+/+} wildtype and four *Cecr2*^{Del/Del} mutant embryo heads (outlier removed).

ensemblID	Gene name	FDR	Fold change
<u>Upregulated genes</u>			
ENSMUSG00000091243	<i>Vgll3</i>	0.000000e+00	2.41398615
ENSMUSG00000001270	<i>Ckb</i>	6.726711e-27	1.73667994
ENSMUSG00000001506	<i>Col1a1</i>	9.272179e-20	1.67501382
ENSMUSG00000033965	<i>Slc16a2</i>	1.618927e-19	1.72057574
ENSMUSG00000023336	<i>Wfdc1</i>	6.483088e-19	2.30333076
ENSMUSG00000035783	<i>Acta2</i>	7.462009e-16	1.99176729
ENSMUSG00000057777	<i>Mab2112</i>	1.312712e-15	1.82714636
ENSMUSG00000047976	<i>Kcna1</i>	1.869802e-15	1.79301766
ENSMUSG00000030507	<i>Dbx1</i>	2.326800e-13	1.90708071
ENSMUSG00000024063	<i>Lbh</i>	1.257200e-12	1.53750788
<u>Downregulated genes</u>			
ENSMUSG00000071226	<i>Cecr2</i>	0.000000e+00	-533.82786
ENSMUSG00000024565	<i>Sall3</i>	0.000000e+00	-2.7718623
ENSMUSG00000025927	<i>Tfap2b</i>	4.788198e-42	-1.9669318
ENSMUSG00000068859	<i>Sp9</i>	1.195983e-33	-4.2617214
ENSMUSG00000001622	<i>Csn3</i>	9.433540e-27	-4.8111911
ENSMUSG00000020160	<i>Meis1</i>	6.132718e-25	-1.5511824
ENSMUSG00000048349	<i>Pou4f1</i>	5.249887e-18	-2.6817278
ENSMUSG00000026739	<i>Bmi1</i>	1.908443e-15	-1.6088424
ENSMUSG00000026586	<i>Prrx1</i>	4.665125e-13	-1.868517
ENSMUSG00000013584	<i>Aldh1a2</i>	1.292425e-10	-3.0403873

Table 3.4.5. Top 10 differentially expressed genes filtered by fold change from RNA-sequencing of five *Cecr2*^{+/+} and four *Cecr2*^{Del/Del} mutant embryo heads (outlier removed).

ensemblID	Gene name	Fold change	FDR
<u>Upregulated genes</u>			
ENSMUSG00000064348	<i>mt-Tn</i>	2.42865298	0.00579101
ENSMUSG00000091243	<i>Vgll3</i>	2.41398623	1.42E-74
ENSMUSG00000023336	<i>Wfdc1</i>	2.30333077	6.48E-19
ENSMUSG00000015484	<i>Fam163a</i>	2.29464988	7.59E-08
ENSMUSG00000056569	<i>Mpz</i>	2.17317426	7.02E-06
ENSMUSG00000047586	<i>Nccrp1</i>	2.15943495	4.70E-06
ENSMUSG00000036766	<i>Dner</i>	2.12312191	5.34E-06
ENSMUSG00000022018	<i>1190002H23Rik</i>	2.08942292	6.17E-05
ENSMUSG00000061718	<i>Ppp1r1b</i>	2.06181495	0.01578207
ENSMUSG00000024366	<i>Gfra3</i>	2.04021533	3.58E-05
<u>Downregulated genes</u>			
ENSMUSG00000071226	<i>Cecr2</i>	-533.82785	0
ENSMUSG00000061462	<i>Obscn</i>	-26.105822	0.00062288
ENSMUSG00000001622	<i>Csn3</i>	-4.811191	9.43E-27
ENSMUSG00000068859	<i>Sp9</i>	-4.2617215	1.20E-33
ENSMUSG00000013584	<i>Aldh1a2</i>	-3.0403872	1.29E-10
ENSMUSG00000024565	<i>Sall3</i>	-2.7718623	6.54E-72
ENSMUSG00000048349	<i>Pou4f1</i>	-2.6817278	5.25E-18
ENSMUSG00000035456	<i>Prdm8</i>	-2.5141411	3.36E-08
ENSMUSG00000066224	<i>Arid3c</i>	-2.4968427	2.30E-08
ENSMUSG00000041911	<i>Dlx1</i>	-2.400037	0.0007381

Table 3.4.6. Enriched gene ontology (GO) biological process terms in RNA sequencing data on *Cecr2*^{+/+} and *Cecr2*^{Del/Del} embryo heads.

GO term	Count	List Total	P-value	Genes
<u>Upregulated genes</u>				
Nervous system development	13	90	0.00561	Apob; Ascl1; Camk1d; Dll3; Erbb4; Gfra1; Grik1; Mdga1; Ndrgr2; Ntf3; Ntrk2; Sim2; Slc1a2
<u>Downregulated genes</u>				
Regulation of transcription, DNA-templated	17	35	0.00682	Arid3c; Bmi1; Cdx1; Dlx1; Dlx2; Gsc; Meis1; Pitx1; Pou3f3; Pou4f1; Prdm8; Prrx1; Rel; Sall3; Smad9; Sp9; Tfp2b
Negative regulation of transcription from RNA polymerase II promoter	10	35	0.00929	Bmi1; Cd36; Dlx1; Dlx2; Gsc; Pou3f3; Pou4f1; Prrx1; Rel; Tfp2b
Retinal metabolic process	3	35	0.0228	Aldh1a2; Aldh1a3; Cyp1b1
Retinoic acid metabolic process	3	35	0.0291	Aldh1a2; Aldh1a3; Aldh1a7
Multicellular organism development	10	35	0.0249	Cdx1; Dlx1; Dlx2; Gsc; Meis1; Obscn; Pitx1; Pou3f3; Pou4f1; Prrx1
Positive regulation of transcription from RNA polymerase II promoter	10	35	0.0251	Arid3c; Cdx1; Dlx2; Meis1; Pitx1; Pou3f3; Pou4f1; Prrx1; Rel; Tfp2b
Hindlimb morphogenesis	3	35	0.0254	Pitx1; Sall3; Tfp2b
Retinol metabolic process	3	35	0.0313	Aldh1a2; Aldh1a3; Cyp1b1
Cartilage development	4	35	0.0469	Dlx2; Pitx1; Prrx1; Smad9
Transcription, DNA-templated	13	35	0.0469	Arid3c; Bmi1; Cdx1; Meis1; Pitx1; Pou3f3; Pou4f1; Prdm8; Rel; Sall3; Smad9; Sp9; Tfp2b
Proximal/distal pattern formation	3	35	0.044	Aldh1a2; Dlx1; Dlx2

3.5 Very few binding sites identified in ChIP-Seq correspond to differentially expressed genes identified in RNA-Seq

Our goal in doing a ChIP-Seq:RNA-Seq comparison was to attempt to understand which genes *Cecr2* may be directly regulating, and which misregulated genes are downstream of direct CECR2 targeting. Unfortunately, when I compared the differentially expressed genes from my RNA-Seq to the list of genes from Farshad's ChIP-Seq experiment there was little overlap. Only one gene, *Kcna6*, had a ChIP binding site for CECR2, LUZP1, and SNF2H in ES cells. *Kcna6* is a voltage-gated potassium channel and loss of function mutation in mice is implicated with increased blood triglycerides and thermal nociceptive threshold (Lexicon Genetics Inc, 2005). It is upregulated 1.8-fold in my RNA-Seq. I also searched for any other differentially expressed genes that appeared in the ChIP analysis regardless of overlapping binding sites. Only four other genes have CECR2 ChIP binding sites: *Cdh7*, *Dbx1*, *Pde4a*, and *Slc6a1*. In my RNA-Seq, these genes are upregulated by 0.5, 1.9, 1.5, and 1.6-fold respectively.

Some additional comparative analysis was performed to try and identify any existing trends between the RNA-Seq and ChIP-Seq. In collaboration with Farshad, we determined the "Top 100" genes from his ChIP-Seq that had an overlapping binding site for CECR2 and SNF2H. This would therefore also include all the genes that also had overlapping binding sites with LUZP1. This gave 104 genes for comparison. A heat map was generated to compare expression profiles of these genes across my samples (Figure 3.5.1.). Mutant and wild-type embryos clustered together (horizontal axis) based on their expression profiles of these genes (vertical axis) indicating a trend in gene expression, with most genes being slightly more expressed in mutant embryos than they are in their wildtype counterparts. In this heat map, and the subsequent ones to be discussed, all genes with no variation across samples have been removed from the heat map.

Another comparison was done to a broader list of all genes that had a CECR2 binding site (Figure 3.5.2.). Although wildtype and mutant embryos cluster together in a similar way to Figure 3.5.1., it is difficult to make conclusions about gene expression profiles. The row side bars (lines between the gene tree and heat map) on the left of Figure

3.5.2. collapse to a few indicating profile similarities. However, whether those show a trend within the data or simply collapse together because genes within the same family demonstrate similar expression profiles cannot be elucidated. To compare the genes with DNA binding sites to those without, a more rigorous resampling-based multiple testing should be performed.

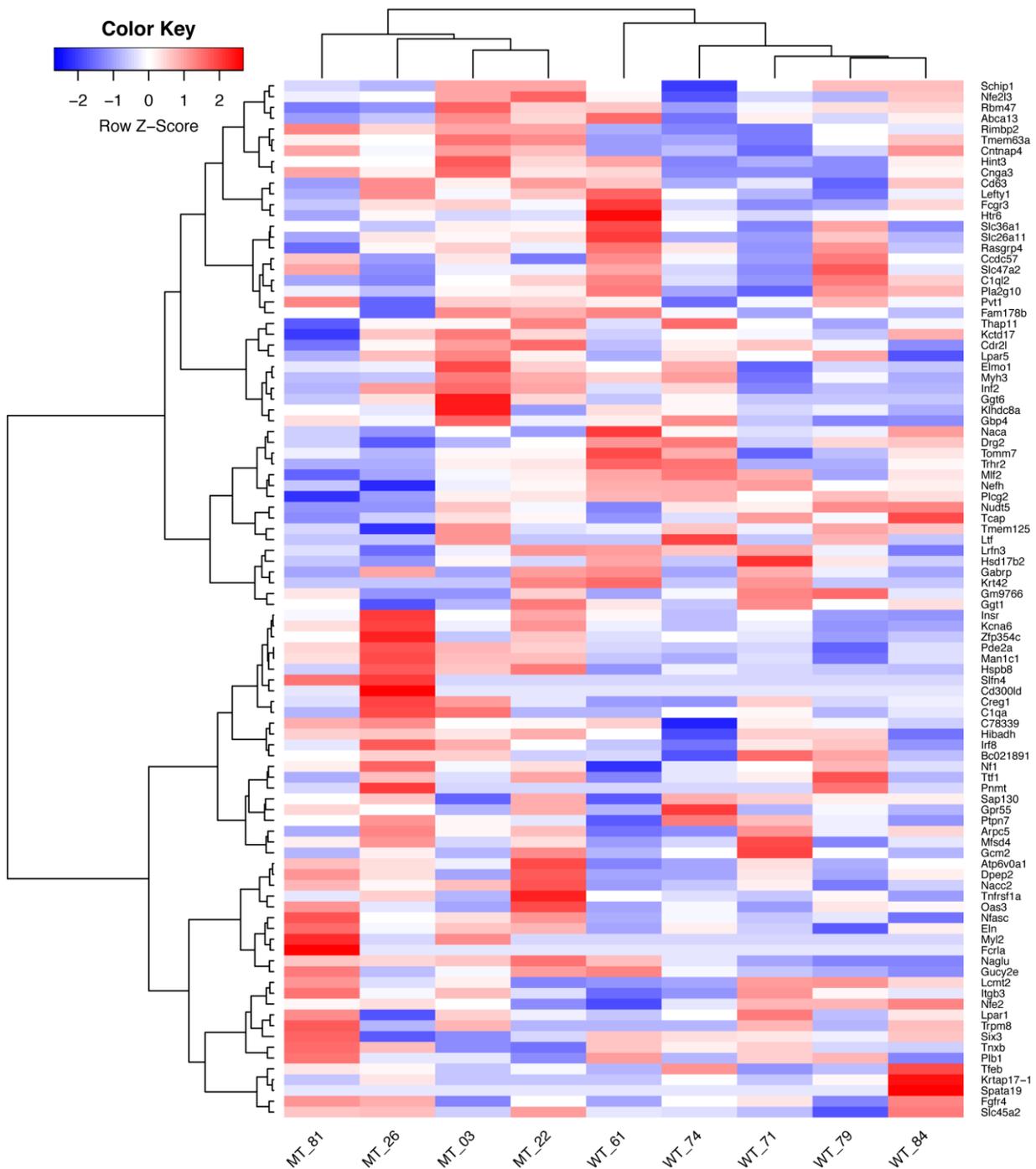


Figure 3.5.1. Heat map of Top 104 genes from CHIP-Seq clusters wildtype and mutant samples together indicating a mild expression trend. The Top 104 genes represent genes from a ChIP-Sequencing experiment in embryonic stem cells that had overlapping binding sites for CECR2 and SNF2H. Four *Cecr2^{Del/Del}* mutant (labelled MT_03, MT_22, MT_26, MT_81) and five *Cecr2^{+/+}* wildtype (labelled WT_61, WT_71, WT_74, WT_79, WT_84) samples cluster based on gene expression. The redder the colour, the more highly expressed a gene is; the bluer the colour is, the less expressed that gene is.

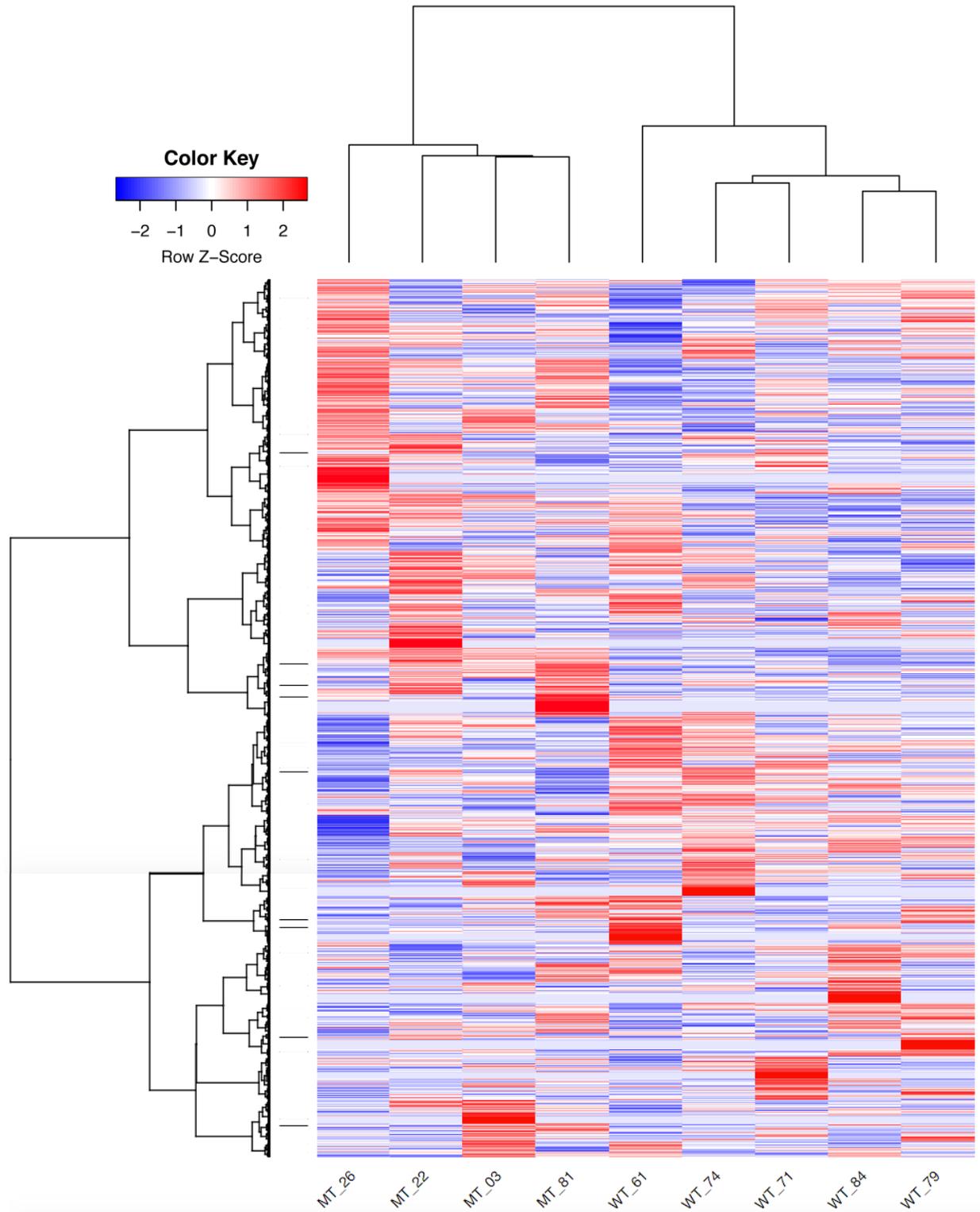


Figure 3.5.2. Heat map of all genes with a CECR2 binding site clusters wildtype and mutant RNA-Seq samples. DNA binding sites were identified in a ChIP-Seq experiment in ES cells. Four *Cecr2^{Del/Del}* mutant (MT_03, MT_22, MT_26, MT_81) and five *Cecr2^{+/+}* wildtype (WT_61, WT_71, WT_74, WT_79, WT_84) samples cluster based on gene expression. The redder the

colour, the more highly expressed a gene is; the bluer the colour is, the less expressed that gene is. A single line on the left-hand side represents a collapsed subset of genes due to expression profile similarities. Figure shown is representative; original file is large and collapsed regions change when viewing depth changes.

3.6 Surprising misregulation of ChIP-Seq candidate genes discovered in *Cecr2*^{Del/Del} embryos

Before the RNA-Seq experiment and analysis was complete, I started looking at the expression of the ChIP-Seq candidate genes in 15-16 somite embryos and continued to collect embryos. We hypothesized that if a gene was misregulated at 12-14 somites (time of neurulation), it may continue to be differentially expressed in 15-16 somite embryos (just after neurulation). This also served as a method to start optimizing qRT-PCR for RNA-Seq validation, and pre-emptively start expression analysis. Interestingly, a few of these genes (*Lrp6*, *Nf1*, *Phactr4*) seemed to be upregulated at 15-16 somites, although the n value was too low for significance (Figure 3.6.1.). Since these three genes are associated with NTDs, I also decided to look at expression in the 12-14 somite male embryos which were not going to be used for RNA-Seq. Surprisingly, none showed any noticeable deviations from wildtype expression. It was interesting that there seemed to be an upregulation after the time of neurulation that was not seen in the neurulating embryos.

Based on this preliminary data, we decided to expand our time course and I collected all 11+ somite E9.5 embryo heads for qRT-PCR analysis. I looked at 5 timepoints in total: 12-14 somites, 15-16 somites, 17-18 somites, 19-20 somites, and 21+ somites. Each group had a minimum of 3 biological somite-matched replicates. Only the two oldest time points had male embryos (~50%), but all other timepoints used female embryos. I was diligent to look for sex-based expression differences. I examined all the candidate ChIP-Seq genes at all time points using qRT-PCR. I also assayed for *Alx1*, which was a qRT-PCR verified differentially expressed gene from the microarray which was not differentially expressed in my RNA-Seq. *Alx1* is a strong candidate for association with the NTD phenotype.

Alx1 was differentially regulated at all time points (Figure 3.6.2.). In the neurulating 12-14 somite embryos, it showed a >2-fold decrease in expression when compared to wildtype. In the older embryos, it remained significantly downregulated. Although the expression between timepoints is not significantly different, it does appear to be trending towards wildtype levels as the embryo develops. In my RNA-Seq however, *Alx1* was not determined to be expressed presumably due to low counts that were below the cut-off.

Transcripts that mapped to *Alx1* can be seen in the raw bam files of my RNA-Seq (files that contain sequence alignment data).

I also found that *Nf1*, *Lrp6*, and *Gli2* expression at 12-14 somites was differentially expressed. This is not seen in my RNA-Seq data, and a comparison is summarized in Table 3.6.1. In the qRT-PCR analysis in Figure 3.6.2., *Nf1* is upregulated at ~1.5-fold, *Lrp6* is downregulated ~2.5-fold, and *Gli2* is downregulated ~1.2-fold. *Lpar1* expression remained at wildtype levels throughout all 5 timepoints, as did *Phactr4*. In all cases, the expression levels in non-neurulating (15+ somite) embryos was not significantly different from wildtype. Interestingly though, in the 19-20 somite stage for *Nf1* and *Lrp6* the fold change and error deviates quite a bit from 1. When I looked more closely, this correlated to a male/female difference. When I analyzed males and females separately there was a clear distinction (n=2) (Figure 3.6.3.). However, whether this is within the normal range of biological variability or indeed a sex-specific phenotype cannot be concluded as more biological replicates would be needed. In both cases, the male expression levels are close to wildtype levels, whereas the females are misregulated. This is interesting since females are known to develop NTDs at a higher penetrance than males (Copp, 2005). This is one of the reasons the RNA-Seq was performed using only female embryos.

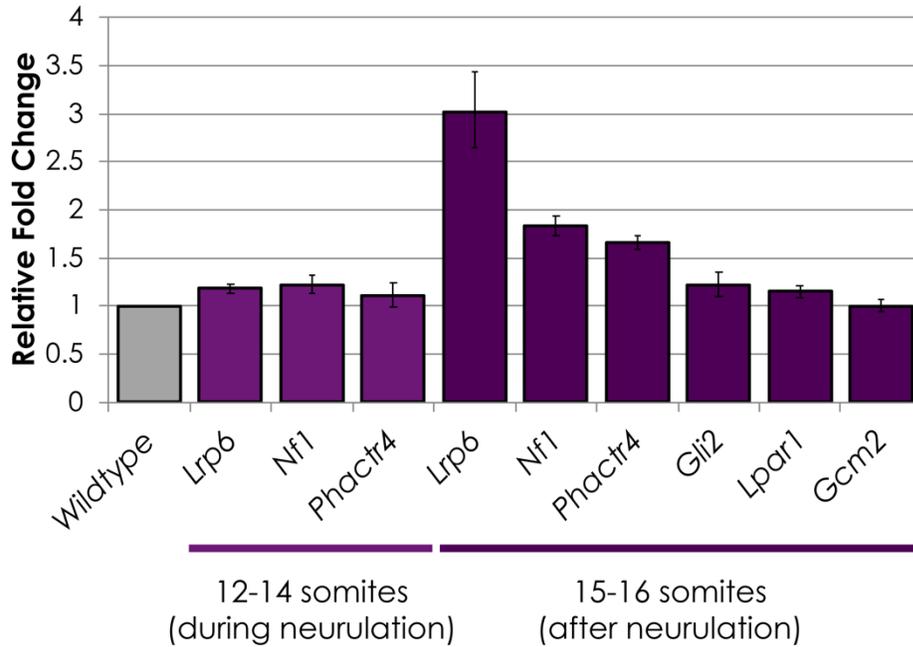


Figure 3.6.1. qRT-PCR analysis of ChIP-Seq candidates genes during and after neurulation in *Cecr2*^{+/+} and *Cecr2*^{Del/Del} mouse embryo heads. Embryos were somite-matched; however, the 12-14 somite embryos were male and the 15-16 somite embryos were female. The n-value of 2 is too low to draw conclusions. Data was analyzed using the $\Delta\Delta\text{CT}$ method, and primer and biological replicates were kept consistent throughout this experiment.

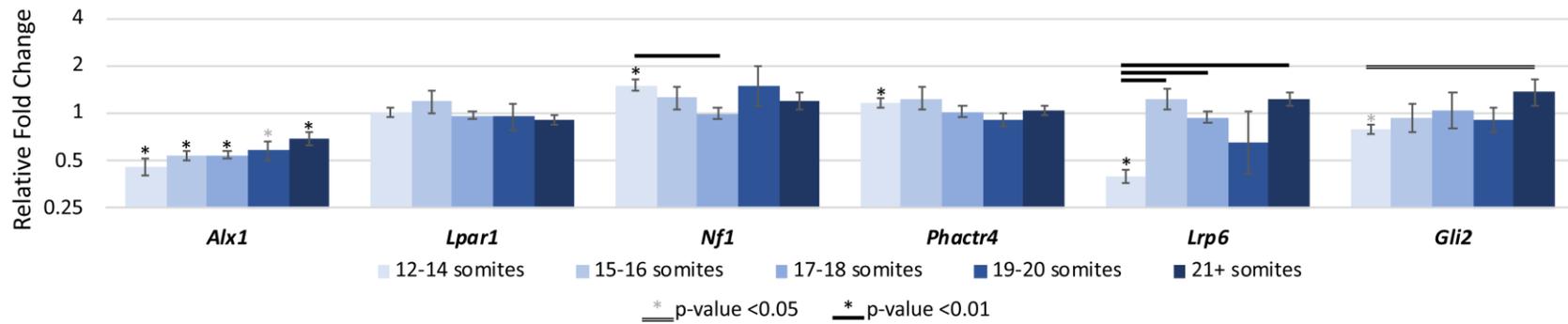


Figure 3.6.2. ChIP-Seq and microarray candidate gene expression throughout early mouse embryonic development. Candidate genes from a microarray (*Alx1*) and ChIP-Seq (*Lpar1*, *Nf1*, *Phactr4*, *Lrp6*, *Gli2*) experiment were chosen to analyze using qRT-PCR across several developmental time points in the heads of *Cecr2^{+/+}* and *Cecr2^{Del/Del}* mouse embryos. Data was analyzed using the $\Delta\Delta$ CT method, and primer and biological replicates were kept consistent throughout this experiment. All data shown are relative to the wildtype level (Fold Change of 1). The y-axis is represented on a logarithmic base 2 scale.

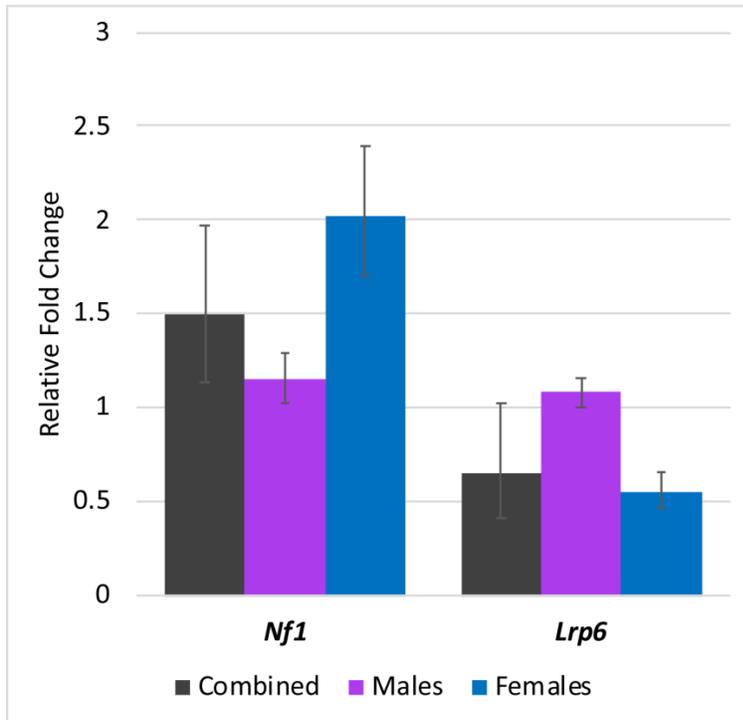


Figure 3.6.3. A possible sex difference in the expression of *Nf1* and *Lrp6* is observed in *Cecr2^{Del/Del}* embryo heads when compared to wildtype with qRT-PCR analysis. Data was analyzed using the $\Delta\Delta\text{CT}$ method, and primer and biological replicates were kept consistent throughout this experiment. All data shown are relative to the wildtype level (Fold Change of 1). The y-axis is represented on a logarithmic base 2 scale. Statistics could not be performed with n=2 biological replicates for each male and female bar shown.

Table 3.6.1. Comparison of genes from an RNA-Seq experiment and with qRT-PCR analysis. Genes were selected from as candidates from a ChIP-Seq experiment and followed up with qRT-PCR comparing *Cecr2^{+/+}* and *Cecr2^{Del/Del}* embryo heads at 12-14 somites. *denotes a significant (p -value <0.05) change from wildtype levels in qRT-PCR.

Gene	Full gene name	Fold Change (RNA-Seq)	Fold Change (qRT-PCR)
<i>Alx1</i>	ALX homeobox 1	Not expressed	0.46*
<i>Lpar1</i>	lysophosphatidic acid receptor 1	0.98	1.01
<i>Nf1</i>	neurofibromin 1	1.05	1.51*
<i>Phactr4</i>	phosphatase and actin regulator 4	1.02	1.16
<i>Lrp6</i>	low density lipoprotein receptor-related protein 6	1.09	0.40*
<i>Gli2</i>	GLI-Kruppel family member GLI2	0.93	0.79*

3.7 qRT-PCR results validate RNA-Seq and shows temporal gene expression changes

Of the 142 significantly misregulated genes identified in my RNA-Seq experiment, I validated 5 of 6 assayed using qRT-PCR. These genes were selected based on overlap with the ChIP-Seq experiment (*Cdh7*, *Dbx1*, *Kcna6*) and then genes which when mutated had a neurulation or nervous system-specific defect (*Aldh1a2*, *Cyp26c1*, *Apob*). I used the same RNA submitted for sequencing, with additional biological replicates. A summary of these results is found in Table 3.7.1. Only one gene (*Apob*) failed to validate within significance, although the average fold change is very similar. Unfortunately, a few genes (*Dbx5*, *Hsd17b2*, *Pde4a*, *Prrx1*, *Slc6a1*) had to be dropped from analysis due to extreme difficulties designing a specific and functional primer set.

Using qRT-PCR, I also evaluated the gene expression levels after neurulation (Figure 3.7.1.). Preliminary analysis (n=2) showed an upregulation of *Lrp6*, a gene known to cause NTDs (Zhou et al., 2010), just after neurulation (15-16 somites) but not at the time of neurulation (12-14 somites)(Figure 3.6.1). However, this was later shown to be an artifact.

Cdh7 validated nearly identically to the RNA-Seq at ~1.7-fold downregulation and had lowered expression at the time of neurulation (12-14 somites), but gene expression returned to wildtype levels in all other time points. *Dbx1* also validated nearly identically to the RNA-Seq and was significantly upregulated nearly two-fold at 12-14 somites. Although expression slightly decreases in all the following time points to 1.5-fold and is significantly different from the expression at 12-14 somites, it remains significantly upregulated in the embryo head. *Kcna6* validated nearly identically to the RNA-Seq and was upregulated ~1.7-fold at neurulation, yet at all other timepoints its expression does not differ significantly from wildtype.

Aldh1a2 was downregulated ~3-fold in my RNA-Seq. but qRT-PCR showed only a ~2-fold downregulation at 12-14 somites. In older embryos, expression returns back to wildtype levels. *Cyp26c1* was downregulated 1.7-fold in my RNA-Seq and validated at a 2-fold decrease with qRT-PCR. It's also significantly downregulated in the 17-18 somite embryos, though the trend seems to indicate a gradual return to wildtype levels in the

older time points. *ApoB* was significantly upregulated 2-fold in my RNA-Seq, and although qRT-PCR revealed a similar fold change it was not significantly different from wildtype. Interestingly upregulation continued to 3.5-fold in the 15-16 somite embryos. By 19-20 somites, *ApoB* expression had decreased to wildtype levels and stayed not significantly different in the older embryos. However, at 21+ somites there seems to show an increase in expression levels with a large amount of variability thereby rendering the fold change not statistically different from wildtype levels.

The high variability in the upregulation of *ApoB* may be due to biological variability but is likely a limitation of the qRT-PCR assay. The amount of *ApoB* template for the qPCR reaction is low, indicated by high CT (cycle threshold) values. Unfortunately, obtaining more RNA from a single embryo to increase the amount which goes into cDNA synthesis is not possible in this experiment, nor was decreasing the dilution of cDNA for the qRT-PCR reaction. With template that low, ~1-5 copies of template are going into each reaction resulting in high variability within not only biological replicates, but also technical replicates.

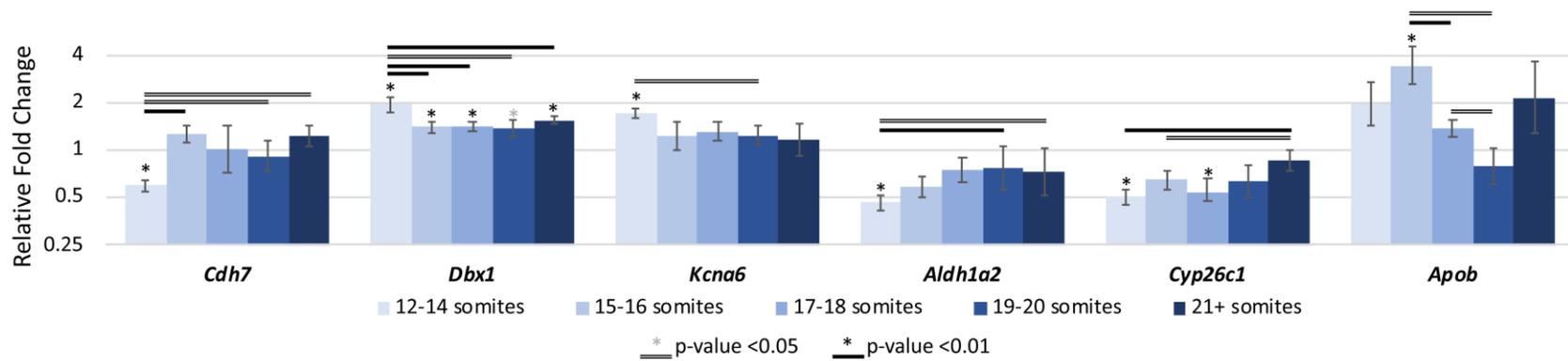


Figure 3.7.1. RNA-Seq candidate gene expression throughout early mouse embryonic development. Candidate genes from an RNA-Seq experiment were chosen to analyze using qRT-PCR across several developmental time points in the heads of *Cecr2^{+/+}* and *Cecr2^{Del/Del}* mouse embryos. Data was analyzed using the $\Delta\Delta$ CT method, and primer and biological replicates were kept consistent throughout this experiment. All data shown are relative to the wildtype level (Fold Change of 1). The y-axis is represented on a logarithmic base 2 scale.

Table 3.7.1. Genes validated using qRT-PCR from an RNA-sequencing experiment performed on *Cecr2^{+/+}* and *Cecr2^{Del/Del}* embryo heads. Embryos for both experiments were 12-14 somites. *denotes a significant (p-value < 0.05) change from wildtype levels using qRT-PCR.

Gene	Full gene name	Fold Change (RNA-Seq)	Fold Change (qRT-PCR)
<i>Cdh7</i>	cadherin 7, type 2	0.57	0.59*
<i>Dbx1</i>	developing brain homeobox 1	1.91	1.96*
<i>Kcna6</i>	potassium voltage-gated channel, shaker-related, subfamily, member 6	1.52	1.72*
<i>Aldh1a2</i>	aldehyde dehydrogenase family 1, subfamily A2	0.33	0.46*
<i>Cyp26c1</i>	cytochrome P450, family 26, subfamily c, polypeptide 1	0.58	0.50*
<i>Apob</i>	apolipoprotein B	2.01	1.96

Chapter 4 – Discussion

In my project I focused on two main aspects of CECR2: its protein complex, and the genetic consequences in its absence. The identified and confirmed members of the CERF protein complex were previously well characterized biochemically. Mass spectrometry identified them, and immunoprecipitation assays confirmed the ability of two proteins to interact. To complement this and resolve any immunoprecipitation uncertainty, immunofluorescence can show the localization of proteins. If two proteins interact they should not only co-immunoprecipitate, but also colocalize.

It was also previously established that CERF is an ATP-dependent chromatin remodeling complex and binds chromatin throughout the genome in both ES cells and testis. Considering the roles of chromatin remodellers, it seemed likely that CERF could be acting as a transcriptional regulator during neurulation, and therefore in its absence mice develop the lethal neural tube defect exencephaly. A few genes had previously been identified as misregulated in mutant neurulating embryos, but this was based on a microarray which has a myriad of caveats (discussed in 4.3).

4.1 CECR2 interactions with CCAR2 and LUZP1 in the nucleus of ES cells are difficult to demonstrate using immunofluorescence

Figure 3.1.1.A clearly shows the presence of CECR2 and CCAR2 in the nucleus of ES cells. It also shows that LUZP1 is present in the ES cell nucleus, albeit much less than CECR2 or CCAR2. When enlarged in panel B, overlapping points of localization are visible for CECR2 and CCAR2, and CECR2 and LUZP1. These points are not numerous, nor is the staining strong making visualization of a yellow overlap colour difficult. However, the co-IP for these interactions was robust, reliably replicating multiple times. Therefore, there is likely an interaction but it is difficult to show using this immunofluorescence protocol. Taken together, the co-IP and immunofluorescence data suggests the proportion of CECR2 that is interacting with CCAR2 and LUZP1 is low as most does not localize together. This corroborates the reciprocal co-IP assay with CECR2 and CCAR2, and CECR2 and LUZP1 where they each pulled-down a small proportion of their binding partner when compared

to the input as indicated by western blot intensity (Niri, in prep). Therefore, both proteins likely have other functions outside their protein complex.

This is not surprising for CCAR2 as it has many other known functions outside the CERF complex (see section 1.3.5). For example, CCAR2 interacts with SIRT1 to promote apoptosis (W. Zhao et al., 2008); there is no evidence to suggest this complex would include CECR2 and apoptosis appears normal in the neurulating *Cecr2*^{-/-} embryo (Niri, unpublished). Perhaps CECR2 is also part of multiple protein complexes in ES cells, most of which do not contain CCAR2. Alternatively, CECR2 may have a role in the cell that does not require or involve binding other proteins. It is also possible CECR2 and CCAR2 interactions are transient and sparse, and therefore difficult to capture in a fixed assay such as immunofluorescence. Likewise, this transient nature could be cell-cycle specific. I did not notice any large differences in the results between islands of ES cells, but an in depth comparison was never performed.

LUZP1 and CECR2 likely have different but overlapping roles during neurulation. LUZP1 deficient mice exhibit phenotypes not seen in CECR2 deficient mice such as ectopic SHH and increased apoptosis, though this has only preliminarily been investigated. This suggests LUZP1 may be involved in more complexes and roles during neural tube development than CECR2. Perhaps a lack of CECR2 impairs the LUZP1 complex involved in neural tube closure at the cranial region, whereas the LUZP1 complexes involved in other neurulation processes remains undisturbed and therefore the other aberrant phenotypes are not observed. Like CCAR2, LUZP1 is also known to be a part of other complexes. For example, it acts as a molecular bridge facilitating binding between the Ada-Two-A-containing (ATAC) histone acetyltransferase and the Mediator coactivator (MED) complexes (Krebs et al., 2010). Together, this forms a meta-coactivator complex (MECO) shown to regulate transcription of some non-coding RNA genes.

Interestingly, although LUZP1 has a nuclear localization sequence, immunofluorescence shows the majority is located outside the nucleus in ES cells. Even more notable is the increased localization of LUZP1 to the nucleus in more differentiated cells. This supports the idea of LUZP1 having a variety of roles outside of CECR2.

Specifically, it may play a more crucial role in gene regulation in differentiated cells regulating the SHH or apoptosis pathways (Hsu et al., 2008). LUZP1 deficient mice also show cardiovascular defects, which could be the result of LUZP1 absence in these more differentiated cells.

Overall, this colocalization experiment could be more easily interpreted without the use of FAB fragments. This method introduces many confounds and the length generally results in a lower quality staining and images. Therefore, a CCAR2 and LUZP1 antibody suited for immunofluorescence which is not made in rabbit would be valuable. Moreover, it seems as though measuring colocalization with a yellow overlap of staining may not be the best method. As seen, differences in intensity can affect the ability to show yellow and therefore conclusions are difficult to draw. A better method would be to mathematically look at the pixel intensity for each channel and compare the percent overlap with another channel. There are several programs available, such as the open source Fiji Coloc 2 (https://imagej.net/Coloc_2) among others. Other assays such as FRET (fluorescence resonance energy transfer) might be better suited to answer the question of protein-protein interactions in cases like this.

4.2 CECR2 does not appear to interact with LUZP1 or CCAR2 in the testis

The localization of CCAR2 and LUZP1 in the tubules of the testis had previously been uncharacterized. I have shown that CCAR2 is expressed in select cells of the outer tubule, and fairly ubiquitously throughout the more differentiated inner cells. Additionally, its expression is restricted to the nucleus of the cells. LUZP1 on the other hand is primarily found in the cytoplasm of cells, but is also shown ubiquitously throughout the tubule with perhaps more localized to the inner differentiated cells of the testis. Neither CCAR2 nor LUZP1 seem to localize to mature sperm. Considering the diverse roles of these two proteins, their ubiquitous expression is unsurprising. It is interesting to note that although LUZP1 seems to be more present in the differentiated cells of the tubules, its nuclear localization does not appear to increase as it does in differentiating ES cells (Figure 3.1.1). This may be an unfair comparison however since these are two very different systems.

I have convincingly shown that although CCAR2 and CECR2 have the potential to interact, as indicated by colocalization in ES cells and co-immunoprecipitation, they do not interact in testis as they are not physically present in the same cells. The conflicting immunoprecipitation that showed interaction could be attributed to a false *in vitro* interaction; it was only successful in 1/9 attempts highlighting the inauthenticity of this single interaction. Crosslinking before protein extraction and immunoprecipitation could rectify this problem but was never successfully performed. Though attempted, it resulted in a messy blot with bands in the IgG control among other difficulties. This evidence confirms the tissue-specific nature of the CERF complex, where CCAR2 interacts in ES cells but not in testis and raises the question of the role of CCAR2 interaction with CECR2 and the CERF complex. CCAR2 binds many different proteins and can regulate transcription (see 1.3.5) – perhaps CCAR2 regulates the transcriptional function of CERF for pathways that are relevant during embryogenesis, but not spermatogenesis.

The colocalization assay with CECR2 and LUZP1 has no firm conclusions. The microscopy resolution and assay quality are insufficient to determine whether LUZP1 is present in the nucleus of testis cells with CECR2. Here a LUZP1-deficient testis is crucial, and a non-rabbit antibody suitable for immunofluorescence would be valuable. However, the biochemical data for a lack of LUZP1 interaction is strong (Niri, unpublished). In all five mass spectrometry experiments in the testis, LUZP1 was not identified. Comparatively, it was identified in three of the mass spectrometry experiments in ES cells. Furthermore, LUZP1 never immunoprecipitated with CECR2 in the testis, even though IP data in ES cells shows they can physically interact. Taken together, this suggests that even if LUZP1 is expressed in the nucleus of cells in testis extracts, it does not interact with CECR2. It is unclear what the role of LUZP1 in the testis may be, but would be interesting to evaluate the fertility and testis morphology of LUZP1-deficient mice.

4.3 RNA-sequencing analysis of the neurulating transcriptome

CECR2 is required for neurulation and loss of function causes exencephaly. As an ATP-dependent chromatin remodeller, we hypothesize it could have a critical role in gene regulation. RNA sequencing is an effective way to assay gene expression levels with spatial-temporal precision. When comparing embryos with a mutation to wildtype, RNA-Seq can

reveal which transcripts are differentially expressed allowing for downstream analysis. Ideally, follow-up experiments based on this analysis may elucidate possible mechanisms to connect the mutation to its phenotype. Although a microarray was previously performed in *Cecr2^{Del/Del}* embryo heads (Fairbridge et al., 2010), this assay has many caveats making RNA-Seq more suitable. Microarrays use only one set of manufacturer-generated probes, and these probes tend to have a considerable bias to the 3' end of transcripts (Jaksik, Iwanaszko, Rzeszowska-Wolny, & Kimmel, 2015). In particular, the Affymetrix MOE 420 2.0 genechip used did not have a complete coverage of the mouse transcriptome and all probes were within the 3' untranslated region of transcripts. Therefore, important genes may have been missed and only changes within that specific region would be seen.

My RNA-Seq data shows 102 significantly upregulated genes from a total of 142. This bias towards upregulated genes suggests CECR2 has a primarily repressive role in the neurulating embryo head, whether it be directly or indirectly. The fact that CECR2 binds to both SNF2H and SNF2L supplements the idea that it can act as an activator or repressor as SNF2H and SNF2L are associated respectively with open and closed chromatin. The chromatin binding motifs of CECR2 suggest it is regulated by co-factors which could also explain either an activating or repressive role (Niri, unpublished).

Interestingly, gene ontology of the upregulated genes only revealed the term 'Nervous System Development' suggesting the repressive role of CECR2 is specifically important during the time of neurulation. In contrast, a variety of gene ontology terms were identified for the downregulated genes (Table 3.4.6.). Perhaps CECR2 plays a principle role in nervous system development and a minor role in a multitude of other functions explaining why mutations in CECR2 cause the highly penetrant exencephaly phenotype, and a low penetrance or absence of other defects.

Although CECR2 plays a clear role in nervous system development, GO and KEGG analysis failed to identify any particular pathways it may be involved in that could elucidate the exact role of CECR2 during neurulation. I hypothesized CECR2 absence may have a mild effect on all genes within a pathway or within a family, and therefore wouldn't show up as significantly differentially expressed. For example, multiple genes within pathways that are

misregulated by 30% could have a significant biological effect, even though statistically each gene is not differentially expressed. Unfortunately, no clear pathway or family was affected as is seen by Table 4.3.1.

The most promising implicated signaling pathway would be the 4/7 BMPs (*Bmp2*, *Bmp4*, *Bmp5*, *Bmp6*) that are slightly misregulated. BMPs are important for dorsal lateral patterning during primary neurulation; when BMP2 is knocked-out mice are nonviable and have an open neural tube, believed to be a secondary effect from cardiovascular defects (H. B. Zhang & Bradley, 1996). Mice without CECR2 do not appear to have patterning problems in the forming neural tube, as indicated by normal SHH localization (Niri, unpublished). However BMP signaling also has complex and diverse crosstalk with Wnt signaling (Itasaki & Hoppler, 2010). Perhaps the combination of slight misregulation of BMPs and WNTs in the absence of CECR2 could lead to exencephaly.

Table 4.3.1. List of genes involved with Hedgehog, Wnt, PCP, BMP signaling and associated developmental processes. List was acquired from Quiagen based on their R² Profiler PCR arrays, and genes were compared to differentially expressed genes from the RNA-Seq data set. Genes in **magenta** were differentially expressed by 20-30%, and genes in **red** were differentially expressed by >30%.

Hedgehog Signaling	
Hedgehog Ligands & Regulators	Boc, Cdon, Dhh, Gas1, Hhip , Ihh, Shh
Hedgehog Receptors & Cofactors	Lrp2, Ptch1, Ptch2, Ptchd2, Ptchd3, Rab23, Smo.
Transcription Factors & Regulators	Btrc (b-TrCP), Csnk1a1, Csnk1e, Fbxw11, Gli1, Gli2, Gli3, Gsk3b, Prkaca, Prkacb, Stk36, Sufu, Zic1 , Zic2 (HPE5)
Other Hedgehog Signaling Genes	Disp1, Disp2 , Fgf9, Fkbp8, Hhat, Kctd11, Otx2, Npc1, Shox2
Hedgehog Signaling Target Genes	Bcl2, Bmp2 , Bmp4 , Bmp5 , Bmp6 , Bmp7, Bmp8a, Bmp8b, Mtss1, Ptch1, Ptch2, Wnt1, Wnt10a , Wnt10b , Wnt11, Wnt16, Wnt2, Wnt2b, Wnt3, Wnt3a, Wnt4 , Wnt5a, Wnt5b, Wnt6, Wnt7a , Wnt7b, Wnt8a, Wnt8b, Wnt9a , Wnt9b , Vegfa.
Pathways Crosstalking with Hedgehog Signaling	
TGFβ / BMP Signaling	Bmp2 , Bmp4 , Bmp5 , Bmp6 , Bmp7, Bmp8b, Grem1 , Sfrp1

WNT Signaling	Btrc (b-TrCP), Ctnnb1 (Catnb), Fbxw11, Fgf9, Gsk3b, Lats1, Lats2, Wif1 , Wnt1, Wnt10a , Wnt10b , Wnt11, Wnt16, Wnt2, Wnt2b, Wnt3, Wnt3a, Wnt4 , Wnt5a, Wnt5b, Wnt6, Wnt7a , Wnt7b, Wnt8a, Wnt8b, Wnt9a, Wnt9b
Hippo Signaling	Fat4 , Frmd6, Lats1, Lats2, Mob1b, Nf2, Stk3
Other Hedgehog Signaling Genes	Fgfr3, Runx2, ErbB4 , Foxe1 , Mapk1 (Erk2), Numb, Trp53 (p53)
WNT Signaling Pathways (includes PCP)	
Canonical	Aes (TLE/Groucho), Apc, Axin1, Axin2, Bcl9, Csnk1a1, Csnk2a1, Ctbp1, Ctnnb1, Ctnnbip1 (ICAT), Dixdc1, Dkk1, Dkk3, Dvl1, Dvl2, Ep300, Frat1, Fzd1, Fzd2, Fzd3, Fzd4, Fzd5, Fzd6, Fzd7, Fzd8, Fzd9, Gsk3b, Lef1, Lrp5, Lrp6, Nkd1, Porcn, Pygo1, Ruvbl1, Sfrp1, Sfrp2, Sfrp4, Sox17, Tcf7l1, Tcf7, Wif1, Wnt1, Wnt10a , Wnt16, Wnt2, Wnt2b, Wnt3, Wnt3a, Wnt4 , Wnt6, Wnt7a , Wnt7b, Wnt8a, Wnt8b
Planar Cell Polarity (PCP)	Daam1, Dvl1, Dvl2, Mapk8 (Jnk1), Nkd1 , Prickle1, Rhoa, Rhou, Vangl2, Wnt9a .
Wnt/Ca ²⁺	Fzd2, Nfatc1, Wnt1, Wnt10a , Wnt11, Wnt16, Wnt2, Wnt2b, Wnt3, Wnt3a, Wnt4 , Wnt5a, Wnt5b, Wnt6, Wnt7a , Wnt7b, Wnt8a, Wnt9a
WNT Signaling Negative Regulation	Apc, Axin1, Axin2, Btrc (bTrCP), Ccnd1, Ctbp1, Ctnnbip1 (ICAT), Dkk1, Dkk3, Fbxw11, Fbxw4, Frzb (FRP-3), Kremen1, Lrp6, Nlk, Nkd1, Sfrp1, Sfrp2, Sfrp4, Sox17, Tle1, Wif1
WNT Signaling Target Genes	Axin2, Btrc (bTrCP), Ccnd1, Ccnd2, Dab2, Fosl1 (FRA-1), Jun , Mmp7, Myc, Pitx2 , Ppard, Wisp1 .
Developmental Processes	
Cell Fate	Ctnnb1, Dkk1, Wnt1, Wnt3, Wnt3a
Tissue Polarity	Axin2, Fzd2, Fzd3, Fzd5, Fzd6, Vangl2
Cell Growth & Proliferation	Apc, Ccnd1, Ccnd2, Ctbp1, Ctnnb1, Ctnnbip1 (ICAT), Dab2, Ep300, Fgf4, Fosl1, Foxn1, Fzd3, Jun , Lrp5, Mmp7, Myc, Ppard, Wisp1 , Wnt3a
Cell Migration	Apc, Dkk1, Lrp5, Lrp6, Rhoa, Wnt1
Cell Cycle	Apc, Btrc (bTrCP), Ccnd1, Ccnd2, Ctnnb1, Ep300, Fosl1, Jun, Myc, Rhoa, Rhou, Ruvbl1, Tcf7l1
Cellular Homeostasis	Apc, Fzd2, Jun , Myc
BMP Signaling	
TGFβ Superfamily Cytokines:	
TGF-β	Tgfb1 , Tgfb2, Tgfb3

BMP	Bmp1, Bmp2 , Bmp3, Bmp4 , Bmp5 , Bmp6 , Bmp7
GDF	Amh , Cdc25a, Dlx2 , Gdf1, Gdf2, Gdf3, Gdf5, Gdf6, Gdf7 , Igf1, Igfbp3, Il6, Ltbp1, Ltbp2 , Ltbp4, Myc, Pdgfb, Smurf1
Activin	Inha, Inhba, Inhbb, Lefty1, Nodal
Receptors	Acvr1, Acvr2a, Acvr1, Amhr2 , Bmpr1a, Bmpr1b, Bmpr2, Cd79a, Igfbp3, Itgb5, Itgb7 , Nr0b1, Tgfb1i1 , Tgfbr1, Tgfbr2 , Tgfbr3, Tgfbrap1
SMAD	Smad1, Smad2, Smad3, Smad4, Smad5
Smad Target Genes:	
TGF- β / Activin Responsive	Cd79a, Cdc25a, Cdkn1a, Cdkn2b, Col1a1 , Col1a2, Col3a1, Fos , Gsc , Igf1 , Igfbp3, Il6, Itgb7 , Jun , Junb , Myc, Pdgfb, Plat, Plau, Serpine1 , Tgfb1i1 , Tsc22d1 (Tgfb1i4), Tgfb1
BMP Responsive	Bglap2, Dlx2 , Id1, Id2, Junb, Sox4, Stat1
Molecules regulating signaling of TGF- β superfamily	Bambi, Bmper, Cdkn2b, Chrd, Eng , Mecom , Fkbp1b, Fst, Ltbp4, Nbl1 , Nog, Runx1, Smurf1
Adhesion and Extracellular Molecules	
Adhesion Molecules	Eng , Itgb5, Itgb7 , Tgfb1
Extracellular Matrix Structural Constituents	Col1a1 , Col1a2, Col3a1, Ltbp1, Ltbp4, Tgfb1, Tgfb2, Tgfb1
Other Extracellular Molecules	Acvr2a, Acvr1, Amhr2 , Bglap2, Bmp1, Bmp2 , Bmp4 , Bmp5 , Bmp6 , Bmp7, Bmper, Bmpr1a, Chrd, Eng , Fst, Gdf1, Gdf2, Gdf3, Gdf5, Gdf6, Igf1 , Igfbp3, Il6, Inha, Inhba, Inhbb, Itgb5, Itgb7 , Lefty1, Nbl1 , Nodal, Nog, Pdgfb, Plat, Plau, Serpine1 , Tdgf1, Tgfb3, Tgfbr1, Tgfbr3
Transcription Factors and Regulators	Dlx2 , Fos , Gsc , Id1, Jun , Junb, Myc, Nr0b1, Runx1 , Smad1, Smad2, Smad3, Smad4, Smad5, Sox4, Stat1, Tgfb1i1 , Tsc22d1 (Tgfb1i4)
Genes Involving in Cellular and Developmental Processes	
Apoptosis	Cdkn1a, Igf1 , Igfbp3, Il6, Inha, Myc
Embryonic Development	Acvr1, Bmp2 , Bmp4 , Bmp5 , Bmp7, Bmpr1a, Chrd, Nodal, Nog, Smad2, Smad3, Smad4, Smurf1, Tdgf1, Tgfb3, Tgfbr1
Muscle Development	Smad3, Tgfb1
Neurogenesis	Bmp4 , Chrd, Fos , Igf1 , Inha, Nog, Runx1
Reproduction	Amhr2 , Bmpr1a, Fst, Inha, Inhbb, Smad1
Skeletal Development	Bglap2, Bmp4 , Bmp5 , Bmpr1b, Chrd, Nog, Tgfb1

4.4 The function of CECR2 as an ATP-dependent chromatin remodeller is still unclear

We knew from the microarray data that there are multiple misregulated genes in *Cecr2* mutants – this is not surprising as neurulation is a complex and multifactorial process. What remained unknown was if CECR2 was directly regulating this expression. We hoped to be able to compare the ChIP-Seq experiment to my RNA-Seq experiment to clarify this large unanswered question. Unfortunately, this did not reveal many overlapping genes: Only *Kcna6* which was misregulated 1.8 fold in my RNA-Seq had binding sites within 5kb of its promoter for CECR2, SNF2H, and LUZP1. This reveals three possible explanations: Bioinformatics has many data caveats, CECR2 is not directly regulating the expression of any/many genes, or ES cells and neurulating embryo heads are not comparable for DNA binding sites.

Unlike RNA-sequencing which has many established pipelines for data analysis, ChIP-sequencing is relatively new and so the combination of parameters chosen can drastically alter the results. Even setting biological parameters – such as the binding distance from the promoter region – can return an entirely different list of results. Furthermore, no known binding sites of CECR2 exists therefore that could not be utilized as a control for a successful ChIP-Seq experiment. However, the data analysis of the ChIP-Seq was thorough and I am confident it was analyzed appropriately. The identified binding sites also have not yet been validated. Therefore, these caveats of bioinformatics and this particular experiment should not be completely overlooked.

Considering the ChIP-Seq data as analyzed, it is also possible that the DNA binding sites of ES cells to neurulating embryos does not translate. There are many morphogenic processes occurring in an embryo which are not happening in an ES cell. This would consequently alter the epigenetic landscape and how a chromatin remodeller like CECR2 may interact. Perhaps in the early embryo, CECR2 has a different role than during neurulation. This temporal difference in gene regulation could have some overlap, which would reflect my results, but is primarily different. A better comparison might have been to use neurospheres as they are a primary neural stem cell line derived from our wildtype and mutant. Overall, it seems reasonable to suppose that an *in vitro* model and an *in vivo* model could not be comparable.

Lastly, it is possible that CECR2 is not directly regulating gene expression. This could be true regardless of whether or not the ChIP-Seq and RNA-Seq are truly comparable. The RNA-Seq had very few differentially expressed genes that seemed likely to be involved in the exencephalic phenotype. Every gene was assessed based on its known function, and any known phenotypes when mutated. A large caveat to this method is that it is not understood exactly how exencephaly arises and all the factors at play, nor is the process of neurulation fully understood. Therefore, it is possible that some of these misregulated genes are in fact critical in neurulation and play a large role in the development of exencephaly when misregulated. It is also possible that it is important to look at gene regulation just before the time of the defect, as perhaps misregulations then are what cause the cascade leading to exencephaly.

Furthering the idea of a function outside of gene regulation is a broader look at the ChIP-Seq data. Although there were many binding sites identified in the promoter region of genes, most of the ChIP binding sites were exonic, intronic, or intergenic. It is possible that these sites still result in gene regulation – it could bind cis-regulatory regions that regulate promoters from a distance, or these are unannotated regions which regulate a nearby gene. It is also possible that this suggests CECR2 does not, at least primarily, regulate gene expression. Some of these sites could instead be background noise. The proportion of ChIP binding sites distributed across the genome is variable, even among ATP-dependent chromatin remodellers. However, Farshad determined the results from his analysis were within a normal range (Niri, 2016).

To answer the question of direct gene regulation, ChIP-Seq or ChIP-PCR should be performed for the promoter region of the misregulated RNA-Seq genes. Unfortunately, this is extremely difficult to do in the embryo head, as the amount of template and protein abundance is low. Even if pooling samples, which are difficult to obtain, this would still be a challenge. However, a new technique called CUT&RUN (Cleavage Under Targets and Release Using Nuclease) has shown to be more efficient and effective and could be possible in embryo heads (Skene & Henikoff, 2017). This technique uses an antibody specific to your protein of interest which is bound to DNA/chromatin and a protein A-MNase. Protein A binds the antibody, and with the addition of calcium the regions of DNA adjacent to the

bound protein of interest are cleaved by MNase, releasing this segment of protein-bound DNA into the supernatant. This segment can then be dissociated from the protein and prepared for sequencing or PCR assays. Since this assay increases the specificity and recovery, low template is less problematic. This method could be combined with ATAC-Seq (Assay for Transposase-Accessible Chromatin), which shows whether chromatin is open or closed using a hyperactive Tn5 transposase to cut and ligate adapters for sequencing at regions of chromatin accessibility (Buenrostro, Wu, Chang, & Greenleaf, 2015). Together, this would reveal if CECR2 is binding to the promoter region of misregulated genes, and if this corresponds to accessible (open) or inaccessible (closed) chromatin.

All of this raises the question then of what the role of CECR2 and the CERF complex is in the neurulating embryo. As an ISWI ATP-dependent chromatin remodeller, if it is not regulating gene expression that suggests it would be involved in DNA replication or repair. Although CECR2 has been reported to be involved in DNA repair in HEK 293 cells (S.-K. Lee, Park, Lee, Lee, & Kwon, 2012), our data suggests this is not the case in mouse neurospheres (Elliot and Norton, in prep). Furthermore, one would speculate that if the CERF complex was primarily involved in DNA repair or replication, we would observe more severe defects in mutant mice as these are ubiquitous and critical cellular mechanisms. One severe example, is the previously discussed lethal *Snf2h* null mutants (see 1.2.5)(Hargreaves & Crabtree, 2011). SNF2H is involved in multiple processes, including the DNA damage response and DNA replication, which helps explain why it is crucial during embryogenesis. Therefore, CERF may have a role in the nucleus interaction with various factors that indirectly regulate gene expression – the mechanisms of which have yet to be elucidated.

4.5 qRT-PCR results suggest CECR2 may regulate transcription factors

A summary of all qRT-PCR results can be seen in Figure 4.5.1. Clearly, gene expression is not static. Significant regulation differences are seen not only between mutant and wildtype embryos, but also between mutants temporally as somite count increases. By 21+ somites, every gene assayed had wildtype expression levels except two – *Alx1* and *Dbx1*. These were also the only two genes that were significantly differentially expressed at all time points. Interestingly, these are both homeobox transcription factors.

Not much is known about *Dbx1* (developing brain homeobox 1), though it is important in neuronal differentiation and axonal projection in the spinal cord (Pierani et al., 2001). *Alx1* (ALX homeobox 1), also known as *Cart1*, is expressed in the embryo mesenchyme and a null mutation when mutated causes severe defects; pups born survive less than 24 hours (Qu, Tucker, Zhao, deCrombrugge, & Wisdom, 1999; Q. Zhao et al., 1996). Notably, all pups develop exencephaly which eventually leads to acrania. Similar to *Cecr2* mutants, their neural folds do elevate but fail to close and penetrance is affected by strain background. It also appears that the NTD is not due to ectopic SHH localization. *Cart1*^{-/-} mice however also exhibit severe craniofacial defects, while *Cecr2* mutants only have craniofacial defects on an FVB/N background. Interestingly, the exencephaly phenotype in *Alx1* mutants is entirely rescued by folic acid supplementation whereas the exencephaly phenotype from *Cecr2* mutants is not.

Perhaps CECR2 has a direct role in regulating the expression of transcription factors which play a role in nervous system development. This regulation is maintained at the time of neurulation and continues afterwards which is why misregulation occurs at all time points in *Cecr2* mutants. CECR2 has a binding site within 5kb upstream of *Dbx1* in the ChIP-Seq (Niri, unpublished), and when looking at *Alx1* using the raw files, CECR2 does have a peak in one of the two biological replicates upstream of *Alx1* as shown by Figure 4.5.2. However, since it is further than 5 kb upstream this was not significant in the ChIP-Seq analysis. More biological replicates are required, but this could be an unannotated TSS. Whether regulation is direct or indirect, perhaps the ~2-fold downregulation seen in *Alx1* expression is sufficient to cause some of the phenotypes previously described but not the others. It is also possible that *Dbx1* plays an unidentified role in neurulation as it is also expressed in the embryo mesenchyme (Diez-Roux et al., 2011). CECR2, as part of CERF, may be regulating transcription factors that then regulate specific genes important for neurulation that when misregulated lead to exencephaly. This may be its primary function, or it may also regulate specific genes important for neurulation.

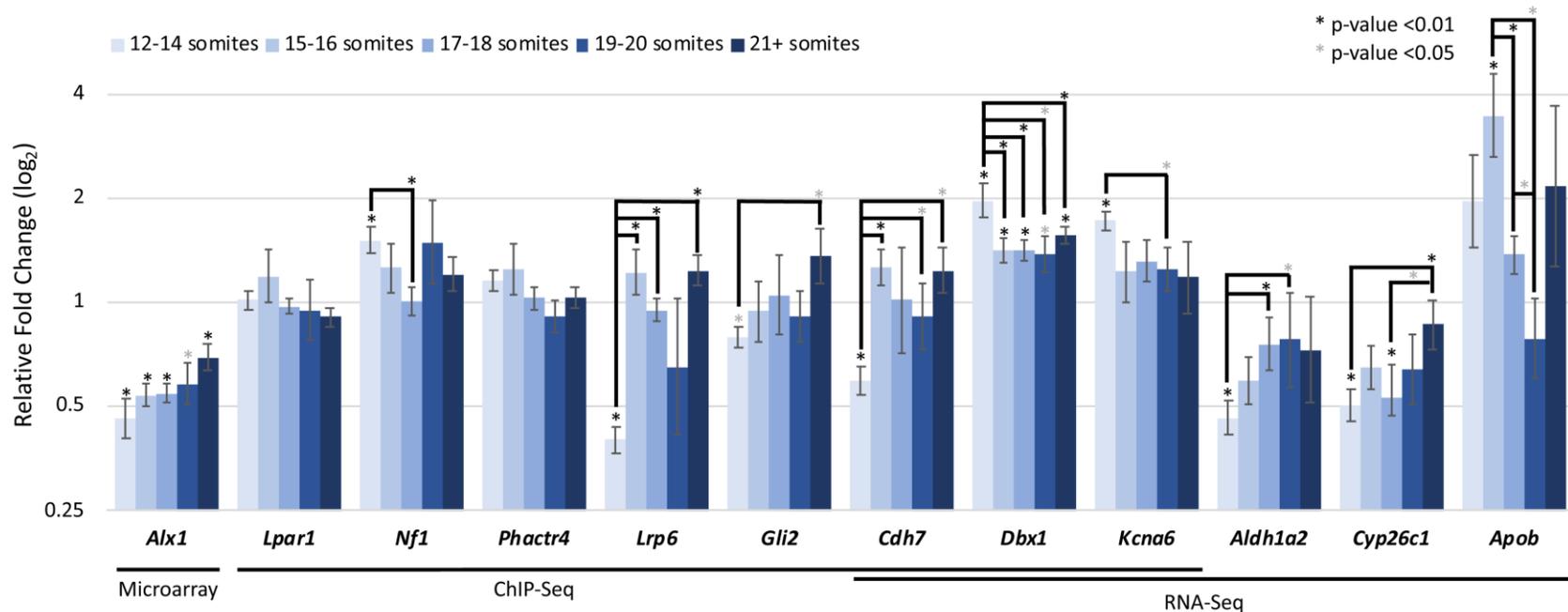


Figure 4.5.1. Gene expression throughout early mouse embryonic development. Candidate genes from a microarray, ChIP-Seq, and RNA-Seq experiment were chosen to analyze using qRT-PCR across several developmental time points in the heads of *Cecr2^{+/+}* and *Cecr2^{Del/Del}* mouse embryos. Data was analyzed using the $\Delta\Delta$ CT method, and primer and biological replicates were kept consistent throughout this experiment. All data shown are relative to the wildtype level (Fold Change of 1). The y-axis is represented on a logarithmic base 2 scale.

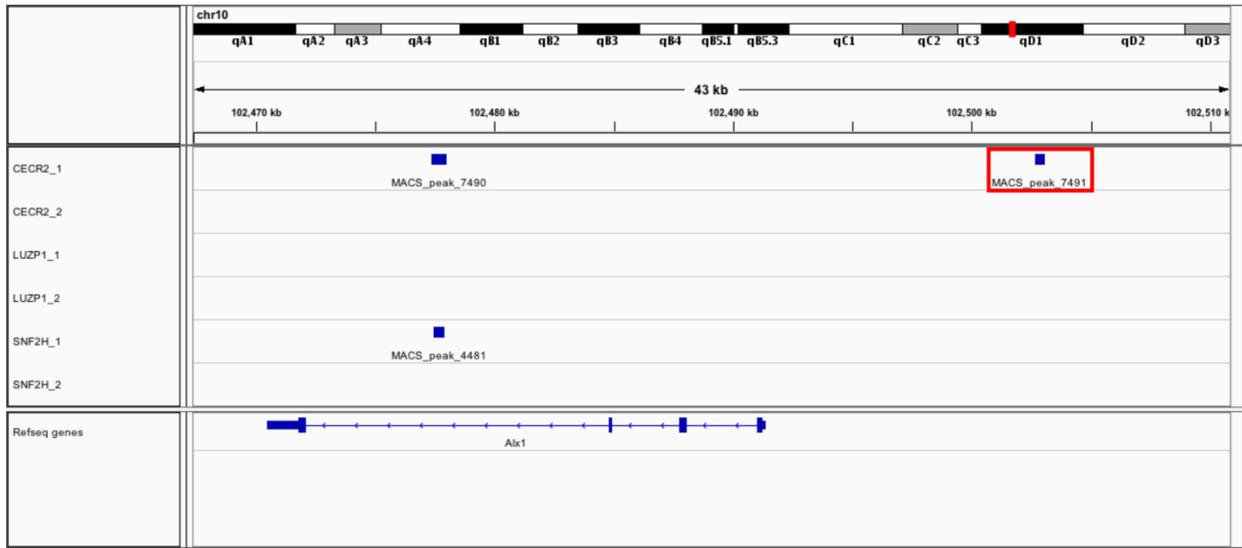


Figure 4.5.2. CECR2 may occupy the region upstream of *Alx1* in ES cells. The Integrative Genomics Viewer (IGV) browser was used to visualize the binding sites of CECR2, LUZP1, and SNF2H upstream of *Alx1*. Shown on the left are the two technical replicates for all three proteins (CECR2_1, CECR2_2, LUZP1_1, LUZP1_2, SNF2H_1, SNF2H_2). *Alx1* is shown on the bottom (Refseq gene), and the top shows the location on chromosome 10. The CECR2 binding site is highlighted by a red box.

4.6 Some RNA-Seq validated genes may play a small role in the exencephalic phenotype, or are downstream of this defect

For some of the genes assayed, little is known about their function. For example, *Cdh7* (cadherin 7 type-2) was selected based on overlap of my RNA-Seq with the ChIP-Seq experiment. It is expressed in the nervous system, but the only current associated phenotype is otocephaly (absence of the lower jaw) and reproductive defects (Juriloff, SULIK, Roderick, & Hogan, 1985) However, more work needs to be done to understand the role of this gene in development.

Another gene selected from overlap with the ChIP-Seq is *Nf1*. The gene when disrupted gives rise to exencephaly, albeit at a low penetrance of ~6% (Brannan et al., 1994). These mice predominantly display heart defects, which have not been examined in *Cecr2* mutants. Interestingly, *Nf1* might act as a negative growth regulator. Although not upregulated in my RNA-Seq, qRT-PCR shows a significant upregulation at the time of neurulation (Figure 4.5.1.). Perhaps the small upregulation of the negative growth regulator could affect whether or not the neural tube closes.

Two downregulated RNA-Seq validated genes, *Aldh1a2* and *Cyp26c1*, are involved in retinoic acid signaling. This pathway was also returned in my GO analysis and is known to be important in patterning in early development (Wilde et al., 2014). When patterning is disrupted, NTDs can occur. *Cyp26c1* is a retinoic acid degrading enzyme, principally expressed at the anterior end of the gastrulating embryo (Uehara et al., 2007). While knockout mice appear phenotypically normal, a double knockout with *Cyp26A1* results in exencephaly along with a host of other CNS defects. The prevalence and severity of the *Cyp26A1/ Cyp26c1* double knockouts is more severe than the *Cyp26A1* knockout alone. Although *Cyp26A1* is not differentially expressed in my RNA-Seq data, perhaps the downregulation of *Cyp26c1* is a contributing susceptibility to developing exencephaly.

Unlike *Cyp26c1*, mutations in *Aldh1a2* alone cause severe NTDs (Niederreither, Subbarayan, Dollé, & Chambon, 1999). *Aldh1a2* is expressed in the mesenchymal cells and the node and is thought to be involved in production of embryonic retinoic acid. In knockout mutants, embryos are small, short, and truncated with an open neural tube. They also display no axial rotation, a critical morphogenic step during embryogenesis, and are resorbed by E10.5. While the posterior of the embryo seems most affected, the anterior neural tube remains open. Since *Aldh1a2* is downregulated at the time of neurulation, a disruption in patterning and retinoic acid signaling could lend to the development of exencephaly.

The discrepancy seen between RNA-Seq and qRT-PCR is confusing. It's possible qRT-PCR is picking up only certain transcripts, possibly due to SNPs in our mice. However, the primers were designed using Ensembl transcript data so there is no transcript-specific bias. The melt curves are also specific suggesting it is not picking up alternative transcripts or making primer dimers. To mitigate any non-specific primers or primer dimers, I used RhPrimers for some genes (see Appendix A). These primers have a RNA base at the 3' end that is only cleaved by RNase-H2 once perfectly annealed to the template – this mitigates any mismatched priming (Dobosy et al., 2011). Furthermore, I looked at the bam files (sequence alignment data) from the RNA-Seq analysis and did not see any exon discrepancies between wildtypes and mutants when looking at the raw counts. The only gene where there seemed to be a notable difference between wildtype and mutants when

looking at the bam files is *Alx1*. Therefore, perhaps since the fold change is not large and counts were low it was not significant. It could also be that there was a large biological variability – except that the qRT-PCR was performed on the exact same RNA which was sent for sequencing. It should not be ignored however that RNA-Seq and qRT-PCR methodologies are very different. With qRT-PCR, there is an additional step to create cDNA and it measures a very specific part of the template whereas RNA-Seq looks at how the fragmented RNA maps to the entire gene.

The discrepancy then between RNA-Seq and qRT-PCR of ChIP-Seq candidates in my experiment is therefore likely is due to RNA breakage/degradation. Normally any small amount of breakage/degradation would not affect a qPCR assay, and I know my RIN was >8 for all samples which indicates intact high-quality RNA. However, if the RNA was breaking at a transcript location more 3' than my primers, and subsequently degrading, with such low template and gene expression it is possible this may be problematic. It is peculiar though that this is seen in only some genes and time points. Therefore, to be certain of these results other primer sets in a different transcript location, or even multiple locations should be assayed. Having one primer set within the UTR would be appropriate here, as it is very large for both *Lrp6* and *Nf1*.

4.7 *Apob* and lipid metabolism are interesting candidates for involvement with an exencephalic phenotype

Apob was ~2-fold upregulated gene in my RNA-Seq, and the only gene in my qRT-PCR assay to be highly upregulated (3.5-fold) shortly after the time of neurulation. However, it was also the only gene which did not validate at 12-14 somites from my RNA-Seq. It is likely that some primer optimization and increased biological replicates would rectify this failure to validate. Since the error is large (>1 fold change), this suggests variability in upregulation (Figure 4.5.1). Increasing biological replicates would give a better understanding of the normal range of variability, and more optimization with primers may help decrease technical variability that becomes apparent with low-expressed genes such as *Apob*.

Mutations in *Apob* cause exencephaly in mice, with a penetrance varying from 28% (Homanics et al., 1993) to 44% (Farese, Ruland, Flynn, Stokowski, & Young, 1995).

However, most homozygotes are resorbed before neurulation occurs, so the penetrance could be higher. *Apob* codes for a structural part of several lipoproteins and is thought to be responsible in part for low density lipoprotein (LDL) clearance. Exencephaly was a surprising finding in the mutants since many studies in mice are trying to model human Familial Hypobetalipoproteinemia (FH β) where NTDs are not observed (reviewed in Linton, Farese, & Young, 1993). In this disorder, humans homozygous for a mutation in *Apob* which creates a truncated protein have pathologically low levels of LDL with an array of phenotypes including neurological defects, but not NTDs. Several mouse mutations in *Apob* have since been characterized, from truncation mutations to nulls, and in all cases exencephaly has been reported in homozygous mutants (Farese et al., 1995; Homanics et al., 1993; Kim, Cham, Veniant, Ambroziak, & Young, 1998) and even in some heterozygous mutants (Huang et al., 1995). Interestingly, one study noted infertility in males heterozygous for a knockout mutation (Huang et al., 1995); *Cecr2* mutant males are subfertile (Norton, unpublished).

A truncated APOB protein is thought to enhance LDL receptor binding as a gain-of-function mutation which would thereby lead to low LDL levels (Kim et al., 1998). Since *Apob* is upregulated in my RNA-Seq, this could mimic a gain-of-function mutation, clearing LDL/VLDL from the plasma of embryos, and may lead to exencephaly. How this might result in exencephaly was hypothesized to be linked to a vitamin E deficiency, since it is a fat soluble vitamin known to be linked to exencephaly in rats (Cheng, Chang, & Bairnson, 1957). Unfortunately not many overexpression studies of *Apob* have been performed, and most focus on the accumulation of amyloid precursor protein which is linked to Alzheimer's (Bjelik et al., 2006).

Lrp6 was shown to be misregulated with qRT-PCR analysis but not in my RNA-Seq. While difficult to make conclusions, it is an excellent candidate for involvement in the exencephalic phenotype as knockout and gain-of-function mutants can develop exencephaly similarly to *Cecr2* mutants (Gray et al., 2013; Zhou et al., 2010). Homozygous nulls also develop other phenotypes seen in *Cecr2* mutants, such as open eyelids. *Lrp6* also had a ChIP binding site for both CECR2 and SNF2h (Niri, unpublished). Furthermore, *Lrp6* is a low density lipoprotein related receptor and therefore involved in lipid endocytosis (Go

& Mani, 2012). However, it is a special member of this family as it is also an important co-receptor in canonical Wnt signaling. While canonical Wnt signaling is not thought to play a role in neurulation, disruption of *Lrp6* clearly affects this process. *Lrp6* may modulate the PCP pathway through RhoA activation (Gray et al., 2013), and RhoA regulation is important for neurulation as demonstrated by the loop-tail mutant mice with NTDs (Ybot-Gonzalez et al., 2007).

Therefore, if *Apob* overexpression during and shortly after the time of neurulation were to indeed increase LDL/VLDL plasma clearance mutant embryos, there may be a lack of LRP6 receptors available for WNT binding and proper signaling in both the canonical and non-canonical pathways. Since lipoprotein clearance is via receptor-mediated endocytosis, these receptors would need to be continuously replaced (Go & Mani, 2012). In mouse and human models where the clearance of LDL is hyperactive, there does not appear to be a feedback loop to regulate the amount of LDL receptors available at the plasma membrane. Perhaps in *Cecr2* mutants, overexpression of *Apob* leads to hyperactive uptake of LDL/VLDL from the blood plasma. Without an upregulation of *Lrp6*, there may not be as many surface receptors available to modulate the PCP pathway, known to be critical during neurulation. This could therefore explain the link between low LDL/VLDL plasma levels and the resulting exencephaly.

It is also likely that a decrease in available LDL/VLDL in the plasma affects other aspects of embryogenesis, explaining other non-exencephalic defects observed in *Apob* mutants such as liver and umbilical hernias (Homanics et al., 1993; Kim et al., 1998). Lipid metabolism and regulation is certainly important for embryogenesis; defects in cholesterol biosynthesis and transport can lead to developmental disorders (reviewed in Farese & Herz, 1998). However, there is not many other genes in lipid metabolism aside from *Apob* which are associated with exencephaly or NTDs. The primary example of cholesterol influence on neurulation involves Shh signaling. As discussed in 1.1.3, overexpression of *Shh* can lead to NTDs, presumably by inhibiting formation of the DLHPs. Though *Shh* is not misregulated in my RNA-Seq, there is evidence it undergoes post-translation modifications whereby it becomes covalently attached to a cholesterol moiety. This modification likely helps SHH to localize properly in the developing embryo (Porter et al., 1996). Therefore,

without this modification SHH may be ectopic and mis-localized leading to defects. This however, is likely not the mechanism behind *Cecr2* exencephaly. First, preliminary analysis shows SHH is not mis-localized in the neurulating embryo (Niri, unpublished). Second, the absence of *Shh* does not cause exencephaly – only overexpression. Nevertheless, this provides evidence of the importance of cholesterol regulation during embryogenesis.

One other lipid-related gene was investigated using qRT-PCR. *Kcna6* was selected as it was the only ChIP-Seq gene in my RNA-Seq with binding sites for CECR2, SNF2H, and LUZP1. It is a potassium voltage-gated channel expressed in the embryo mesenchyme (Lexicon Genetics Inc, 2005). In homozygous knockouts, an increase in triglycerides in females was noted. It is unclear if any other lipids, such as cholesterol, were investigated. Although triglycerides are different than cholesterol, perhaps if *Kcna6* is overexpressed this would further decrease lipid levels in the mouse. However, how this gene is linked to lipid metabolism and if there is any effect on cholesterol remains unclear.

Some preliminary examination of lipids in our mice was done by a previous undergraduate student, Jay Rassmussen, examining if differences in cholesterol could explain the exencephaly penetrance differences between Balb/c (susceptible) mice and FVB/N (resistant) mice (see 1.3.3). He found the plasma cholesterol was lower in Balb/c mice than FVB/N, however the total cholesterol from E9.5 embryos did not differ between the two strains. He also found that a high fat diet seemed to increase the penetrance of exencephaly in the FVB/N mice, however it was not statistically significant. He noted some technical problems with the cholesterol assay used, and that optimization and repetition would be crucial to draw conclusions.

I think moving forward, it would be interesting to examine the levels of LDL/VLDL in neurulating *Cecr2*^{+/+} and *Cecr2*^{Del/Del} mice. The cholesterol assay used previously did not distinguish between HDL and LDL/VLDL, and clearly had some technical issues. Abcam sells a Cholesterol Assay Kit - HDL and LDL/VLDL (ab65390) which would be well suited for this experiment. Not only does it distinguish between the two classes of lipoproteins, it also can be used with tissue homogenate. Due to the nature of collecting E9.5 embryos, a plasma analysis would not be possible. Embryos would have to be pooled and then

homogenized, but a proteomics experiment done on the heads of E9.5 mouse suggest that it would be feasible to obtain sufficient amounts of protein to measure differences in LDL concentration (Hartl et al., 2008). If mice have low levels of LDL, this could be a large contributing reason to the development of exencephaly. If they do not, it must be a lipid-independent mechanism by which the defect arises. Neither outcome would be surprising, as my research and all others demonstrates the complexity and multifactorial nature of NTDs. It would however, provide more clarity as to how the loss of CECR2 results in exencephaly.

4.8 Summary and future directions

My work has clarified the composition of the CERF complex in testis, which was confounded by inconsistent immunoprecipitations. I showed that CCAR2 is not localized in the same cells as CECR2, and therefore the one co-IP showing interaction was an artifact. It also supports other immunoprecipitation data which suggested that CECR2 does not interact with the majority of CCAR2 or LUZP1 present in ES cells (Niri, in prep). However, these conclusions should be quantified using a program such as Fiji Coloc 2 to determine the amount of overlapping protein localization, independent of producing a yellow colour. Assays for LUZP1 in testis likely requires a *Luzp1*^{-/-} testis as a negative control, and should be possible given that not all homozygous mutants develop exencephaly (Hsu et al., 2008). Testis IF would also benefit from a LUZP1 non-rabbit antibody and microscopy with higher magnification and resolution. Overall, I believe these results support the hypothesis that there are still unidentified binding partners of CECR2 in the CERF complexes based on the estimated size of 2 MDa from gel filtration analysis. There remain interesting candidates from mass spectroscopy to investigate, and further analysis of these components will help us understand the role of CECR2 in development.

RNA-Seq analysis revealed an array of misregulated genes in the neurulating embryo head. Moving forward with qRT-PCR, this assay would benefit from re-optimization beginning with RNA extraction, knowing now the caveats and problems in a low-template low-expression experiment. Analysis of some candidate genes from my RNA-Seq and qRT-PCR suggest CECR2 may be involved in the transcription of transcription factors, and perhaps in lipid metabolism. The question of whether CECR2 is directly

regulating these genes remains unknown. Performing ChIP-Seq in neurospheres is an option, and I would predict it translates more closely to embryos than do ES cells. However, the translational caveat from an *in vitro* to an *in vivo* model remains. I believe the best way to answer this question is using CUT&RUN in neurulating embryo heads. Although fairly new, this technique shows promising results making its use in small tissue samples, such as a neurulating mouse embryo head, feasible (Skene & Henikoff, 2017). This would allow comparison to my RNA-Seq and allow us to narrow down a list of candidate genes for further investigation after validation of binding sites using ChIP-qPCR. For example, we could use immunofluorescence to assay for any absent or ectopic expression.

Given that *Apob* gain-of-function mutants can develop exencephaly (Kim et al., 1998), the upregulation of *Apob* shown with RNA-Seq and qRT-PCR in the mutant embryo head suggests it may have a role in this phenotype. APOB is critical for lipoprotein clearance from the blood plasma, specifically LDL/VLDL by mediating binding to receptors. LRP6 is an LDL receptors, however also functions as a receptor for WNT that initiates canonical Wnt signalling (Go & Mani, 2012). *Lrp6* mutants also develop exencephaly, although the mechanism behind this is unknown (Zhou et al., 2010). I hypothesize that perhaps in *Cecr2^{Del/Del}* embryo heads, upregulation of *Apob* causes increased uptake of LDL by receptors such as LRP6. Since lipoproteins are imported by receptor-mediated endocytosis, this increased uptake could result in insufficient LRP6 for Wnt signalling. This loss of LRP6 could be what results in exencephaly in *Apob* mutants.

While this is a complicated cascade to investigate, we could first start by examining the levels of LDL/VLDL in *Cecr2^{Del/Del}* neurulating mouse embryos to see if upregulation of *Apob* does in fact cause excess clearance of lipoproteins. If there are very low levels of LDL/VLDL, we could then investigate if LRP6 is present in the plasma membrane by protein analysis. We could also investigate if there is misregulation of downstream canonical Wnt signalling. There are also several rescue experiments that could be attempted. For example, we could try to rescue the exencephalic phenotype by reducing *Apob* expression or overexpressing *Lrp6*. This can be accomplished by crossing our mice with a mouse with a null *Apob* allele (as used by Farese et al., 1995) or a gain-of-function *Lrp6* mutation (as used by Gray et al., 2013). Another option would be feeding pregnant

females a high-cholesterol diet. However, the dam between mother and fetus is a complex barrier, so an *ex vivo* approach may be more suitable for rescue by increasing the amount of LDL/VLDL in the embryo (Piliszek, Kwon, & Hadjantonakis, 2011).

Overall, it appears that the role of CECR2 in the cranial region of the developing embryo directly or indirectly causes an array of gene misregulation. While its role may be complex, there also must be at least one fundamental function it performs since when disrupted, all *Cecr2^{Del/Del}* embryos develop the lethal NTD exencephaly. Further investigation of CERF composition, function, and the downstream cascade influencing neurulation in the developing embryo is necessary. Not only will this add to our understanding of the role of CERF, but also how disruptions in the complex morphogenic process of neurulation leads to lethal NTDs such as exencephaly in mice, and anencephaly in humans.

Literature Cited

- Alvarez-Saavedra, M., De Repentigny, Y., Lagali, P. S., Raghu Ram, E. V. S., Yan, K., Hashem, E., et al. (2014). Snf2h-mediated chromatin organization and histone H1 dynamics govern cerebellar morphogenesis and neural maturation. *Nature Communications*, *5*, 4181. <https://doi.org/10.1038/ncomms5181>
- Banting, G. S., Barak, O., Ames, T. M., Burnham, A. C., Kardel, M. D., Cooch, N. S., et al. (2005). CECR2, a protein involved in neurulation, forms a novel chromatin remodeling complex with SNF2L. *Human Molecular Genetics*, *14*(4), 513–524. <https://doi.org/10.1093/hmg/ddi048>
- Barak, O., Lazzaro, M. A., Cooch, N. S., Picketts, D. J., & Shiekhattar, R. (2004). A Tissue-specific, Naturally Occurring Human SNF2L Variant Inactivates Chromatin Remodeling. *Journal of Biological Chemistry*, *279*(43), 45130–45138. <https://doi.org/10.1074/jbc.M406212200>
- Bjelik, A., Bereczki, E., Gonda, S., Juhász, A., Rimanóczy, Á., Zana, M., et al. (2006). Human apoB overexpression and a high-cholesterol diet differently modify the brain APP metabolism in the transgenic mouse model of atherosclerosis. *Neurochemistry International*, *49*(4), 393–400. <https://doi.org/10.1016/j.neuint.2006.01.026>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, *30*(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bowman, G. D., & McKnight, J. N. (2016). Sequence-specific targeting of chromatin remodelers organizes precisely positioned nucleosomes throughout the genome. *BioEssays*, *39*(1), e201600183–16. <https://doi.org/10.1002/bies.201600183>
- Brannan, C. I., Perkins, A. S., Vogel, K. S., Ratner, N., Nordlund, M. L., Reid, S. W., et al. (1994). Targeted disruption of the neurofibromatosis type-1 gene leads to developmental abnormalities in heart and various neural crest-derived tissues. *Genes & Development*, *8*(9), 1019–1029. <https://doi.org/10.1101/gad.8.9.1019>
- Buenrostro, J. D., Wu, B., Chang, H. Y., & Greenleaf, W. J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current Protocols in Molecular Biology*, *109*(1), 21.29.1–29.9. <https://doi.org/10.1002/0471142727.mb2129s109>

- Bultman, S. J., Herschkowitz, J. I., Godfrey, V., Gebuhr, T. C., Yaniv, M., Perou, C. M., & Magnuson, T. (2008). Characterization of mammary tumors from Brg1 heterozygous mice. *Oncogene*, *27*(4), 460–468. <https://doi.org/10.1038/sj.onc.1210664>
- Byers, S. L., Wiles, M. V., Dunn, S. L., & Taft, R. A. (2012). Mouse estrous cycle identification tool and images. *PloS One*, *7*(4), e35538–. <https://doi.org/10.1371/journal.pone.0035538>
- Chen, Y., McCarthy, D., Robinson, M., & Smyth, G. K. (2014). edgeR: differential expression analysis of digital gene expression data User's Guide.
- Chen, Z. F., & Behringer, R. R. (1995). Twist Is Required in Head Mesenchyme for Cranial Neural-Tube Morphogenesis. *Genes & Development*, *9*(6), 686–699. <https://doi.org/10.1101/gad.9.6.686>
- Cheng, D. W., Chang, L. F., & Bairnson, T. A. (1957). Gross observations on developing abnormal embryos induced by maternal vitamin E deficiency. *The Anatomical Record*, *129*(2), 167–185. <https://doi.org/10.1002/ar.1091290204>
- Chini, C. C. S., Escande, C., Nin, V., & Chini, E. N. (2010). HDAC3 is negatively regulated by the nuclear protein DBC1. *Journal of Biological Chemistry*, *285*(52), 40830–40837. <https://doi.org/10.1074/jbc.M110.153270>
- Clapier, C. R., Iwasa, J., Cairns, B. R., & Peterson, C. L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nature Publishing Group*, *18*(7), 407–422. <https://doi.org/10.1038/nrm.2017.26>
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., et al. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, *17*(1), 13. <https://doi.org/10.1186/s13059-016-0881-8>
- Copp, A. J. (2005). Neurulation in the cranial region - normal and abnormal. *Journal of Anatomy*, *207*(5), 623–635. <https://doi.org/10.1111/j.1469-7580.2005.00476.x>
- Copp, A. J., Greene, N. D. E., & Murdoch, J. N. (2003). The genetic basis of mammalian neurulation. *Nature Reviews Genetics*, *4*(10), 784–793. <https://doi.org/10.1038/nrg1181>
- Cutter, A. R., & Hayes, J. J. (2015). A brief review of nucleosome structure. *FEBS Letters*, *589*(20PartA), 2914–2922. <https://doi.org/10.1016/j.febslet.2015.05.016>

- Dawe, C. E., Kooistra, M. K., Fairbridge, N. A., Pisio, A. C., & McDermid, H. E. (2011). Role of chromatin remodeling gene *Cecr2* in neurulation and inner ear development. *Developmental Dynamics*, *240*(2), 372–383. <https://doi.org/10.1002/dvdy.22547>
- Diez-Roux, G., Banfi, S., Sultan, M., Geffers, L., Anand, S., Rozado, D., et al. (2011). A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biology*, *9*(1), e1000582. <https://doi.org/10.1371/journal.pbio.1000582>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, *29*(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Dobosy, J. R., Rose, S. D., Beltz, K. R., Rupp, S. M., Powers, K. M., Behlke, M. A., & Walder, J. A. (2011). RNase H-dependent PCR (rhPCR): improved specificity and single nucleotide polymorphism detection using blocked cleavable primers. *BMC Biotechnology*, *11*(1), 80. <https://doi.org/10.1186/1472-6750-11-80>
- Dong, J., Gao, Z., Liu, S., Li, G., Yang, Z., Huang, H., & Xu, L. (2013). SLIDE, the protein interacting domain of Imitation Switch remodelers, binds DDT-domain proteins of different subfamilies in chromatin remodeling complexes. *Journal of Integrative Plant Biology*, *55*(10), 928–937. <https://doi.org/10.1111/jipb.12069>
- Echelard, Y., Epstein, D. J., StJacques, B., Shen, L., Mohler, J., McMahon, J. A., & McMahon, A. P. (1993). Sonic-Hedgehog, a Member of a Family of Putative Signaling Molecules, Is Implicated in the Regulation of Cns Polarity. *Cell*, *75*(7), 1417–1430.
- Erdel, F., & Rippe, K. (2011). Chromatin remodelling in mammalian cells by ISWI-type complexes - where, when and why? *FEBS Journal*, *278*(19), 3608–3618. <https://doi.org/10.1111/j.1742-4658.2011.08282.x>
- Fairbridge, N. A., Dawe, C. E., Niri, F. H., Kooistra, M. K., King-Jones, K., & McDermid, H. E. (2010a). *Cecr2* mutations causing exencephaly trigger misregulation of mesenchymal/ectodermal transcription factors. *Birth Defects Research Part a: Clinical and Molecular Teratology*, *88*(8), 619–625. <https://doi.org/10.1002/bdra.20695>
- Farese, R. V. J., Ruland, S. L., Flynn, L. M., Stokowski, R. P., & Young, S. G. (1995). Knockout of the mouse apolipoprotein B gene results in embryonic lethality in homozygotes and protection against diet-induced hypercholesterolemia in heterozygotes. (Vol. 92, pp.

- 1774–1778). Presented at the Proceedings of the National Academy of Sciences of the United States of America. <https://doi.org/10.1073/pnas.92.5.1774>
- Farese, R. V., & Herz, J. (1998). Cholesterol metabolism and embryogenesis. *Trends in Genetics : TIG*, *14*(3), 115–120.
- Footz, T. K., Brinkman-Mills, P., Banting, G. S., Maier, S. A., Riazi, M. A., Bridgland, L., et al. (2001). Analysis of the Cat Eye Syndrome Critical Region in Humans and the Region of Conserved Synteny in Mice: A Search for Candidate Genes at or near the Human Chromosome 22 Pericentromere. *Genome Research*, *11*(6), 1053–1070. <https://doi.org/10.1101/gr.154901>
- Fu, J., Jiang, J., Li, J., Wang, S., Shi, G., Feng, Q., et al. (2009). Deleted in breast cancer 1, a novel androgen receptor (AR) coactivator that promotes AR DNA-binding activity. *Journal of Biological Chemistry*, *284*(11), 6832–6840. <https://doi.org/10.1074/jbc.M808988200>
- Fyodorov, D. V., & Kadonaga, J. T. (2002). Binding of Acf1 to DNA involves a WAC motif and is important for ACF-mediated chromatin assembly. *Molecular and Cellular Biology*, *22*(18), 6344–6353.
- Go, G.-W., & Mani, A. (2012). Low-density lipoprotein receptor (LDLR) family orchestrates cholesterol homeostasis. *The Yale Journal of Biology and Medicine*, *85*(1), 19–28.
- Goldman, J. A., Garlick, J. D., & Kingston, R. E. (2010). Chromatin remodeling by imitation switch (ISWI) class ATP-dependent remodelers is stimulated by histone variant H2A.Z. *Journal of Biological Chemistry*, *285*(7), 4645–4651. <https://doi.org/10.1074/jbc.M109.072348>
- Gray, J. D., Kholmanskikh, S., Castaldo, B. S., Hansler, A., Chung, H., Klotz, B., et al. (2013). LRP6 exerts non-canonical effects on Wnt signaling during neural tube closure. *Human Molecular Genetics*, *22*(21), 4267–4281. <https://doi.org/10.1093/hmg/ddt277>
- Greene, N. D. E., & Copp, A. J. (2009). Development of the vertebrate central nervous system: formation of the neural tube. *Prenatal Diagnosis*, *29*(4), 303–311. <https://doi.org/10.1002/pd.2206>
- Greene, N. D. E., & Copp, A. J. (2014). Neural tube defects. *Annual Review of Neuroscience*, *37*, 221–242. <https://doi.org/10.1146/annurev-neuro-062012-170354>

- Han, P., & Chang, C.-P. (2015). Long non-coding RNA and chromatin remodeling. *RNA Biology*, 12(10), 1094–1098. <https://doi.org/10.1080/15476286.2015.1063770>
- Hargreaves, D. C., & Crabtree, G. R. (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Nature Publishing Group*, 1–25. <https://doi.org/10.1038/cr.2011.32>
- Harmacek, L., Watkins-Chow, D. E., Chen, J., Jones, K. L., Pavan, W. J., Salbaum, J. M., & Niswander, L. (2014). A unique missense allele of BAF155, a core BAF chromatin remodeling complex protein, causes neural tube closure defects in mice. *Developmental Neurobiology*, 74(5), 483–497. <https://doi.org/10.1002/dneu.22142>
- Hartl, D., Irmeler, M., Römer, I., Mader, M. T., Mao, L., Zabel, C., et al. (2008). Transcriptome and proteome analysis of early embryonic mouse brain development. *Proteomics*, 8(6), 1257–1265. <https://doi.org/10.1002/pmic.200700724>
- Hildebrand, J. D., & Soriano, P. (1999). Shroom, a PDZ domain-containing actin-binding protein, is required for neural tube morphogenesis in mice. *Cell*, 99(5), 485–497.
- Ho, L., & Crabtree, G. R. (2010). Chromatin remodelling during development. *Nature*, 463(7280), 474–484. <https://doi.org/10.1038/nature08911>
- Ho, L., Ronan, J. L., Wu, J., Staahl, B. T., Chen, L., Kuo, A., et al. (2009). An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proceedings of the National Academy of Sciences of the United States of America*, 106(13), 5181–5186. <https://doi.org/10.1073/pnas.0812889106>
- Holmberg, J., Clarke, D. L., & Frisé, J. (2000). Regulation of repulsion versus adhesion by different splice forms of an Eph receptor. *Nature*, 408(6809), 203–206. <https://doi.org/10.1038/35041577>
- Homanics, G. E., Maeda, N., Traber, M. G., Kayden, H. J., Dehart, D. B., & Sulik, K. K. (1995). Exencephaly and Hydrocephaly in Mice with Targeted Modification of the Apolipoprotein-B (Apo-B) Gene. *Teratology*, 51(1), 1–10. <https://doi.org/10.1002/tera.1420510102>
- Homanics, G. E., Smith, T. J., Zhang, S. H., Lee, D., Young, S. G., & Maeda, N. (1993). Targeted modification of the apolipoprotein B gene results in hypobetalipoproteinemia and developmental abnormalities in mice. (Vol. 90, pp. 2389–2393). Presented at the

- Proceedings of the National Academy of Sciences of the United States of America, National Academy of Sciences. <https://doi.org/10.1073/pnas.90.6.2389>
- Hota, S. K., & Bruneau, B. G. (2016). ATP-dependent chromatin remodeling during mammalian development. *Development (Cambridge, England)*, *143*(16), 2882–2897. <https://doi.org/10.1242/dev.128892>
- Hsu, C.-Y., Chang, N.-C., Lee, M. W.-Y., Lee, K.-H., Sun, D.-S., Lai, C., & Chang, A. C. (2008). LUZP deficiency affects neural tube closure during brain development. *Biochemical and Biophysical Research Communications*, *376*(3), 466–471. <https://doi.org/10.1016/j.bbrc.2008.08.170>
- Huang, L. S., Voyiaziakis, E., Markenson, D. F., Sokol, K. A., Hayek, T., & Breslow, J. L. (1995). apo B gene knockout in mice results in embryonic lethality in homozygotes and neural tube defects, male infertility, and reduced HDL cholesterol ester and apo A-I transport rates in heterozygotes. *Journal of Clinical Investigation*, *96*(5), 2152–2161. <https://doi.org/10.1172/jci118269>
- Itasaki, N., & Hoppler, S. (2010). Crosstalk between Wnt and bone morphogenic protein signaling: a turbulent relationship. *Developmental Dynamics*, *239*(1), 16–33. <https://doi.org/10.1002/dvdy.22009>
- Jaksik, R., Iwanaszko, M., Rzeszowska-Wolny, J., & Kimmel, M. (2015). Microarray experiments and factors which affect their reliability. *Biology Direct*, *10*(1), 46. <https://doi.org/10.1186/s13062-015-0077-2>
- Joshi, P., Quach, O. L., Giguere, S. S. B., & Cristea, I. M. (2013). A Functional Proteomics Perspective of DBC1 as a Regulator of Transcription. *Journal of Proteomics & Bioinformatics, Suppl 2*. <https://doi.org/10.4172/jpb.s2-002>
- Juriloff, D. M., SULIK, K. K., Roderick, T. H., & Hogan, B. K. (1985). Genetic and developmental studies of a new mouse mutation that produces otocephaly. *Journal of Craniofacial Genetics and Developmental Biology*, *5*(2), 121–145.
- Kim, E., Cham, C. M., Veniant, M. M., Ambroziak, P., & Young, S. G. (1998). Dual mechanisms for the low plasma levels of truncated apolipoprotein B proteins in familial hypobetalipoproteinemia - Analysis of a new mouse model with a nonsense mutation in the Apob gene. *Journal of Clinical Investigation*, *101*(6), 1468–1477.

- Kooistra, M. K., Leduc, R. Y. M., Dawe, C. E., Fairbridge, N. A., Rasmussen, J., Man, J. H. Y., et al. (2012). Strain-specific modifier genes of *Cecr2*-associated exencephaly in mice: genetic analysis and identification of differentially expressed candidate genes. *Physiological Genomics*, *44*(1), 35–46. <https://doi.org/10.1152/physiolgenomics.00124.2011>
- Koyama, S., Wada-Hiraike, O., Nakagawa, S., Tanikawa, M., Hiraike, H., Miyamoto, Y., et al. (2010). Repression of estrogen receptor beta function by putative tumor suppressor DBC1. *Biochemical and Biophysical Research Communications*, *392*(3), 357–362. <https://doi.org/10.1016/j.bbrc.2010.01.025>
- Krebs, A. R., Demmers, J., Karmodiya, K., Chang, N.-C., Chang, A. C., & Tora, L. (2010). ATAC and Mediator coactivators form a stable complex and regulate a set of non-coding RNA genes. *EMBO Reports*, *11*(7), 541–547. <https://doi.org/10.1038/embor.2010.75>
- Kuida, K., Haydar, T. F., Kuan, C. Y., Gu, Y., Taya, C., Karasuyama, H., et al. (1998). Reduced apoptosis and cytochrome c-mediated caspase activation in mice lacking Caspase 9. *Cell*, *94*(3), 325–337.
- Landry, J. W., Banerjee, S., Taylor, B., Aplan, P. D., Singer, A., & Wu, C. (2011). Chromatin remodeling complex NURF regulates thymocyte maturation. *Genes & Development*, *25*(3), 275–286. <https://doi.org/10.1101/gad.2007311>
- Landry, J., Sharov, A. A., Piao, Y., Sharova, L. V., Xiao, H., Southon, E., et al. (2008). Essential role of chromatin remodeling protein Bptf in early mouse embryos and embryonic stem cells. *PLoS Genetics*, *4*(10), e1000241. <https://doi.org/10.1371/journal.pgen.1000241>
- Lanier, L. M., Gates, M. A., Witke, W., Menzies, A. S., Wehman, A. M., Macklis, J. D., et al. (1999). Mena is required for neurulation and commissure formation. *Neuron*, *22*(2), 313–325.
- Lazzaro, M. A., & Picketts, D. J. (2001). Cloning and characterization of the murine Imitation Switch (ISWI) genes: differential expression patterns suggest distinct developmental roles for *Snf2h* and *Snf2l*. *Journal of Neurochemistry*, *77*(4), 1145–1156.
- Lazzaro, M. A., Pépin, D., Pescador, N., Murphy, B. D., Vanderhyden, B. C., & Picketts, D. J. (2006). The imitation switch protein SNF2L regulates steroidogenic acute regulatory protein expression during terminal differentiation of ovarian granulosa cells. *Molecular*

- Endocrinology (Baltimore, Md.)*, 20(10), 2406–2417. <https://doi.org/10.1210/me.2005-0213>
- Lazzaro, M. A., Todd, M. A., Lavigne, P., Vallee, D., De Maria, A., & Picketts, D. J. (2008). Characterization of novel isoforms and evaluation of SNF2L/SMARCA1 as a candidate gene for X-linked mental retardation in 12 families linked to Xq25-26. *BMC Medical Genetics*, 9(1), 669. <https://doi.org/10.1186/1471-2350-9-11>
- Längst, G., & Manelyte, L. (2015). Chromatin Remodelers: From Function to Dysfunction. *Genes*, 6(2), 299–324. <https://doi.org/10.3390/genes6020299>
- Leduc, R. Y. M. (2015, September 29). *Role of Cecr2 and Candidate Modifier Genes of Cecr2 in the Neural Tube Defect Exencephaly*. Thesis. University of Alberta.
- Leduc, R. Y. M., Singh, P., & McDermid, H. E. (2017). Genetic backgrounds and modifier genes of NTD mouse models: An opportunity for greater understanding of the multifactorial etiology of neural tube defects. *Birth Defects Research*, 109(2), 140–152. <https://doi.org/10.1002/bdra.23554>
- Lee, H.-K., Lee, H.-S., & Moody, S. A. (2014). Neural transcription factors: from embryos to neural stem cells. *Molecules and Cells*, 37(10), 705–712. <https://doi.org/10.14348/molcells.2014.0227>
- Lee, M. W., Chang, A. C., Sun, D. S., Hsu, C. Y., & Chang, N. C. (2001). Restricted expression of LUZP in neural lineage cells: a study in embryonic stem cells. *Journal of Biomedical Science*, 8(6), 504–511.
- Lee, S.-K., Park, E.-J., Lee, H.-S., Lee, Y. S., & Kwon, J. (2012). Genome-wide screen of human bromodomain-containing proteins identifies Cecr2 as a novel DNA damage response protein. *Molecules and Cells*, 34(1), 85–91. <https://doi.org/10.1007/s10059-012-0112-4>
- Lewis Carl, S. A., Gillete-Ferguson, I., & Ferguson, D. G. (1993). An indirect immunofluorescence procedure for staining the same cryosection with two mouse monoclonal primary antibodies. *The Journal of Histochemistry and Cytochemistry : Official Journal of the Histochemistry Society*, 41(8), 1273–1278. <https://doi.org/10.1177/41.8.7687266>
- Lexicon Genetics Inc. (2005). NIH initiative supporting placement of Lexicon Genetics, Inc. mice into public repositories. *MGI Direct Data Submission*.

- Li, H. (2012, March 28). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *arXiv.org*. <https://doi.org/10.1093/bioinformatics/btr509>
- Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*, *30*(7), 923–930. <https://doi.org/10.1093/bioinformatics/btt656>
- Linton, M. F., Farese, R. V., & Young, S. G. (1993). Familial Hypobetalipoproteinemia. *Journal of Lipid Research*, *34*(4), 521–541. <https://doi.org/10.1194/jlr.r300002-jlr200>
- Lopez, M. E. (2012). A Quick, No Frills Approach to Mouse Genotyping. *Bio-Protocol*, *2*(15).
- Luger, K., Dechassa, M. L., & Tremethick, D. J. (2012). New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nature Publishing Group*, *13*(7), 436–447. <https://doi.org/10.1038/nrm3382>
- Nargund, V. H. (2015). Effects of psychological stress on male fertility. *Nature Publishing Group*, *12*(7), 373–382. <https://doi.org/10.1038/nrurol.2015.112>
- Niederreither, K., Subbarayan, V., Dollé, P., & Chambon, P. (1999). Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. *Nature Genetics*, *21*(4), 444–448. <https://doi.org/10.1038/7788>
- Niri, F. H. (2016, November 21). *Identification of CECR2-containing chromatin-remodeling complexes and their Chromatin-binding sites in mice*. Thesis. University of Alberta.
- Okonechnikov, K., Conesa, A., & García-Alcalde, F. (2016). Qualimap 2 - advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics (Oxford, England)*, *32*(2), 292–294. <https://doi.org/10.1093/bioinformatics/btv566>
- Pépin, D., Paradis, F., Perez-Iratxeta, C., Picketts, D. J., & Vanderhyden, B. C. (2013). The imitation switch ATPase Snf2l is required for superovulation and regulates Fgl2 in differentiating mouse granulosa cells. *Biology of Reproduction*, *88*(6), 142. <https://doi.org/10.1095/biolreprod.112.105742>
- Pierani, A., Moran-Rivard, L., Sunshine, M. J., Littman, D. R., Goulding, M., & Jessell, T. M. (2001). Control of interneuron fate in the developing spinal cord by the progenitor homeodomain protein Dbx1. *Neuron*, *29*(2), 367–384.

- Piliszek, A., Kwon, G. S., & Hadjantonakis, A.-K. (2011). Ex utero culture and live imaging of mouse embryos. *Methods in Molecular Biology (Clifton, N.J.)*, 770(Chapter 9), 243–257. https://doi.org/10.1007/978-1-61779-210-6_9
- Porter, J. A., Ekker, S. C., Park, W. J., Kessler, von, D. P., Young, K. E., Chen, C. H., et al. (1996). Hedgehog patterning activity: role of a lipophilic modification mediated by the carboxy-terminal autoprocessing domain. *Cell*, 86(1), 21–34.
- Qu, S., Tucker, S. C., Zhao, Q., deCrombrughe, B., & Wisdom, R. (1999). Physical and genetic interactions between *Alx4* and *Cart1*. *Development (Cambridge, England)*, 126(2), 359–369.
- Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, 11(3), R25. <http://doi.org/10.1186/gb-2010-11-3-r25>
- Schinzel, A., Schmid, W., Fraccaro, M., genetics, L. T. H., 1981. (n.d.). The “cat eye syndrome”: dicentric small marker chromosome probably derived from a no. 22 (tetrasomy 22pter→q11) associated with a characteristic phenotype. *Springer*. <https://doi.org/10.1007/BF00282012>
- Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M. J., Davie, J. R., & Peterson, C. L. (2006). Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions. *Science*, 311(5762), 844–847. <https://doi.org/10.2307/3843617?refreqid=search-gateway:7c7b10475c33e315c9b039fdbcb3c60>
- Singh, P. (2016, October 31). *Analysis of variants of Dnmbp, a Putative Modifier of Cecr2 and Neural Tube Defect Susceptibility Gene*. Thesis. University of Alberta.
- Skene, P. J., & Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife*, 6, e21856. <https://doi.org/10.7554/eLife.21856>
- Stopka, T., & Skoultchi, A. I. (2003). The ISWI ATPase Snf2h is required for early mouse development. *Proceedings of the National Academy of Sciences of the United States of America*, 100(24), 14097–14102. <https://doi.org/10.1073/pnas.2336105100>
- Tamkun, J. W., Deuring, R., Scott, M. P., Kissinger, M., Pattatucci, A. M., Kaufman, T. C., & Kennison, J. A. (1992). *brahma*: A regulator of *Drosophila* homeotic genes structurally related to the yeast transcriptional activator SNF2SWI2. *Cell*, 68(3), 561–572. [https://doi.org/10.1016/0092-8674\(92\)90191-e](https://doi.org/10.1016/0092-8674(92)90191-e)

- Thompson, P. J., Norton, K. A., Niri, F. H., Dawe, C. E., & McDermid, H. E. (2012). CECR2 Is Involved in Spermatogenesis and Forms a Complex with SNF2H in the Testis. *Journal of Molecular Biology*, 415(5), 793–806. <https://doi.org/10.1016/j.jmb.2011.11.041>
- Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14(2), 178–192. <https://doi.org/10.1093/bib/bbs017>
- Torres-Padilla, M. E., & Zernicka-Goetz, M. (2006). Role of TIF1 α as a modulator of embryonic transcription in the mouse zygote. *The Journal of Cell Biology*, 174(3), 329–338. <https://doi.org/10.1083/jcb.200603146>
- Uehara, M., Yashiro, K., Mamiya, S., Nishino, J., Chambon, P., Dolle, P., & Sakai, Y. (2007). CYP26A1 and CYP26C1 cooperatively regulate anterior–posterior patterning of the developing brain and the production of migratory cranial neural crest cells in the mouse. *Developmental Biology*, 302(2), 399–411. <https://doi.org/10.1016/j.ydbio.2006.09.045>
- Venkatesh, S., & Workman, J. L. (2015). Histone exchange, chromatin structure and the regulation of transcription. *Nature Publishing Group*, 16(3), 178–189. <https://doi.org/10.1038/nrm3941>
- Vogler, C., Huber, C., Waldmann, T., Ettig, R., Braun, L., Izzo, A., et al. (2010). Histone H2A C-Terminus Regulates Chromatin Dynamics, Remodeling, and Histone H1 Binding. *PLoS Genetics*, 6(12), e1001234–12. <https://doi.org/10.1371/journal.pgen.1001234>
- Whirledge, S., & Cidlowski, J. A. (2010). Glucocorticoids, stress, and fertility. *Minerva Endocrinologica*, 35(2), 109–125.
- Wilde, J. J., Petersen, J. R., & Niswander, L. (2014). Genetic, epigenetic, and environmental contributions to neural tube closure. *Annual Review of Genetics*, 48, 583–611. <https://doi.org/10.1146/annurev-genet-120213-092208>
- Wright, R. H. G., Fernandez-Fuentes, N., Oliva, B., & Beato, M. (2016). Insight into the machinery that oils chromatin dynamics. *Nucleus*, 7(6), 532–539. <https://doi.org/10.1080/19491034.2016.1255392>
- Xu, W. M., Baribault, H., & Adamson, E. D. (1998). Vinculin knockout results in heart and brain defects during embryonic development. *Development (Cambridge, England)*, 125(2), 327–337.

- Ybot-Gonzalez, P., Cogram, P., Gerrelli, D., & Copp, A. J. (2002). Sonic hedgehog and the molecular regulation of mouse neural tube closure. *Development (Cambridge, England)*, *129*(10), 2507–2517.
- Ybot-Gonzalez, P., Savery, D., Gerrelli, D., Signore, M., Mitchell, C. E., Faux, C. H., et al. (2007). Convergent extension, planar-cell-polarity signalling and initiation of mouse neural tube closure. *Development (Cambridge, England)*, *134*(4), 789–799.
<https://doi.org/10.1242/dev.000380>
- Yip, D. J., Corcoran, C. P., Alvarez-Saavedra, M., DeMaria, A., Rennick, S., Mears, A. J., et al. (2012). Snf2l regulates Foxg1-dependent progenitor cell expansion in the developing brain. *Developmental Cell*, *22*(4), 871–878.
<https://doi.org/10.1016/j.devcel.2012.01.020>
- Zentner, G. E., & Henikoff, S. (2013). Regulation of nucleosome dynamics by histone modifications. *Nature Publishing Group*, *20*(3), 259–266.
<https://doi.org/10.1038/nsmb.2470>
- Zhang, H. B., & Bradley, A. (1996). Mice deficient for BMP2 are nonviable and have defects in amnion chorion and cardiac development. *Development (Cambridge, England)*, *122*(10), 2977–2986.
- Zhang, Z., Witham, S., & Alexov, E. (2011). On the role of electrostatics in protein-protein interactions. *Physical Biology*, *8*(3), 035001. <https://doi.org/10.1088/1478-3975/8/3/035001>
- Zhao, Q., Behringer, R. R., & deCrombrughe, B. (1996). Prenatal folic acid treatment suppresses acrania and meroanencephaly in mice mutant for the Cart1 homeobox gene. *Nature Genetics*, *13*(3), 275–283. <https://doi.org/10.1038/ng0796-275>
- Zhao, W., Kruse, J.-P., Tang, Y., Jung, S. Y., Qin, J., & Gu, W. (2008). Negative regulation of the deacetylase SIRT1 by DBC1. *Nature*, *451*(7178), 587–590.
<https://doi.org/10.1038/nature06515>
- Zhou, C. J., Wang, Y.-Z., Yamagami, T., Zhao, T., Song, L., & Wang, K. (2010). Generation of Lrp6 conditional gene-targeting mouse line for modeling and dissecting multiple birth defects/congenital anomalies. *Developmental Dynamics*, *239*(1), 318–326.
<https://doi.org/10.1002/dvdy.22054>

Zohn, I. E. (2012). Mouse as a model for multifactorial inheritance of neural tube defects.
Birth Defects Research. Part C, Embryo Today : Reviews, 96(2), 193–205.
<https://doi.org/10.1002/bdrc.21011>

Appendices

Appendix A – List of primers. All primers were obtained from Integrated DNA Technologies. All qPCR primers were designed to be spanning an exon-exon junction.

Name	Oligonucleotide sequence 5'-3'	Purpose
IngeniousLox1	TTAGAATAGGTGAGGGAGGAG	<i>Cecr2^{Del}</i> (<i>Cecr2^{tm.1.1Hemc}</i>) genotyping
Ingenious SDL2	GTAGCGCCTATTTGTAATGGTCA	
LoxCECR2_DEL3R	AATGGTGGCGAAATCAACTC	
SRY FOR	GAGAGCATGGAGGGGCAT	SRY genotyping
SRY REV	CCACTCCTCTGTGACACT	
Mmu Intron7 F4	CCCCATTTATTTGCTTGAGCTG	<i>Cecr2^{GT}</i> (<i>Cecr2^{Gt45Bic}</i>) genotyping
Mmu Intron7 R4	CACGAACAATGGAAGGAATGA	
pGT1R4	ACGCCATACAGTCCTCTTCACATC	
Aldh1a2-Fwd	GAGAGAAATGGGTGAGTTTGGCrUTACG/3SpC3/	qRT-PCR analysis
Aldh1a2 New REV	CGGCCTCTTAGGAGTTCTTCTG	
Alx1.7 Set 1 FWD	AGGGTCCAGTTTGGTTTC	
Alx1.7 Set 1 REV	CCGTCCTTGTAAGACTGATATG	
Apob-Fwd	ATCAAAGGAACACTTCAACACTGTGrACTTCT/3SpC3/	
Apob New REV	CCCTGCCAGTCCCAAAGTC	
Cdh7 FWD Set 1	GACACTATTGGTGCTGATCC	
Cdh7 REV Set 1	CCCGCCTTCATCATCATATC	
Cyp26c1-Fwd	GTGATGCCCTGCTCTTGATTATTrAACAGT/3SpC3/	
Cyp26c1 New REV	CAGCCAACCTCCTTCAGCTCTTG	
Dbx1 FWD Set 1	CGCCTAGAAGAGAACTTCG	
Dbx1 REV Set 1	TGGGAAATAGGAGGATCTGATA	
Gapdh_2 001289726.1 -Fwd	GAGAAACCTGCCAAGTA	
Gapdh_2 001289726.1 -Rev	CAGTGTAGCCCAAGATG	
Gli2-Fwd	ATGTGTGTGAACACGAAGGCTGrUAACAT/3SpC3/	
Gli2 New REV	TCTTGCAGATGTAGGGTTTCTCATT	
Kcna6-Fwd	TTGCCTCTGAGGGTTGTGCrUGCACT/3SpC3/	
Kcna6 New REV	GGGTCTCATCCTCCAGAGAAGTT	
Lpar1 FWD Set 1	GTGCTTGGTGCCTTTATTG	
Lpar1 REV Set 1	GAGGAGGAAGAAGTTCATAG	
Lrp6 FWD Set 2	GGTCAGTGCCTTGAAA	
Lrp6 REV Set 2	GCTCCTCAGTTGGATAACAG	
Nf1 FWD Set 1	CCACAGTACCAGACATCTTAC	
Nf1 REV Set 1	GGGATTTGTGTTTGCTTTGA	
Phactr4 FWD Set 1	CTCACTAGAAAGCTCAGTCAA	
Phactr4 REV Set 2	AGCATCAGTTACTTCCACATAC	

Appendix B – Vaginal cytology analysis

The estrous cycle of the mouse has physiological and anatomical consequences, affecting their mating and pregnancies (Byers et al., 2012.) Fortunately, we can use this information to better time pregnancies using a variety of methods including visual observation and vaginal cytology. The vagina of a mouse can be visually observed, as its appearance differs over the four stages of estrous: proestrus, metestrus, estrus, and diestrus. More accurately, vaginal cytology can be performed by using vaginal swabbing to assess the types and proportions of cell types of the mouse vagina, which are unique for each stage (see 2.7 for more details). Mice who are in metestrus or diestrus should not be ovulating and therefore should not have successful matings, whereas mice in proestrus or estrus are receptive to mating and pregnancy (Byers et al., 2012).

While collecting embryos for RNA-Seq, it became apparent that the mice had an abnormal estrous cycle. I would set up 2 female mice and 1 male for mating on Sunday evening and look for copulatory plugs every following morning as evidence of mating. Mice with a copulatory plug would be separated from the male, while any mice without a plug would remain with the male up to a maximum of 4 nights. Overall, the number of plugged females was low, and many mice who plugged were not pregnant upon dissection for embryos at E9.5. Most importantly, their mating time was erratic making it very difficult to time the dissections to obtain embryos at the 11-14 somite stage (time of neurulation).

Therefore, I started monitoring their estrous cycle using vaginal cytology to increase the chances of successful mating. I also used timed matings, where copulatory plugs were looked for every hour once the females were caged with the male, to better predict the developmental stage upon dissection (see 2.1.2 on timed matings). Vaginal swabs were collected in the afternoon prior to the morning timed mating, therefore it was possible the mice may move into the next estrous stage before the morning. Consequently, mainly mice at the end of diestrus or in proestrus were selected for mating.

I recorded the estrous stage of each mouse swabbed until they plugged (Table B1). The estrous cycle should repeat every 4 to 5 days, unless pregnancy occurs (Byers et al., 2012). My results show that *Cecr2* mice display abnormal cycling patterns, with a variety of

anomalies . Figure B1 shows representative abnormal mice (2-6) in comparison to a mouse cycling normally (Mouse 1). Many mice remain in one stage of the estrous cycle for an extended period of time (indicated in green). Others have skipped a stage (indicated in pink) or actually cycled in the reverse direction (Indicated in orange) . Mice should cycle through proestrus, estrus, metestrus, and diestrus every 4-5 days but these mice demonstrate deviations from this norm.

When I used vaginal cytology, the pregnancy success rate of plugged females was an insignificant increase to 67.6%, compared to 56.3% when I selected the most visually ready females (Table B2). Of the 32.4% of staged plugged females who were not pregnant, 54.5% of them had a suspicious plug (i.e. late or less obvious - that is deep and small in the vagina). The pregnancy success rate of females set up in proestrus, estrus, metestrus, and diestrus respectively (with varying n values) is 60.9%, 66.7%, 100%, and 85.7% (Table B2). The low n values for mice in estrus and metestrus make comparisons between these stages difficult, however mice should not have a successful pregnancy from mating in either of these stages (Byers et al., 2012). Mice were only set up if in estrus or metestrus when the number of females available for plug testing was low and there was no better option, or in a couple cases set up erroneously.

Of all the pregnancies, 9.7% had litters with 100% early embryonic death and 12.9% with >50% but <100% early embryonic death (Table B2). When set up in proestrus, 60.9% of females that plugged were pregnant. Of those pregnancies, 14.3% showed early embryonic deaths >50% but <100%. When set up in estrus, 66.7% that plugged were pregnant and had normal litters with <50% early embryonic deaths. When set up in metestrus, 1/1 female was pregnant with a normal litter. When set up in diestrus, 85.7% of females that plugged were pregnant. Of those pregnancies, 33.3% had a litter with 100% early embryonic death and 16.7% had a litter with >50% but <100% early embryonic death. Although n values are too low for statistics and a thorough comparison, there does not appear to be a correlation between the estrus stage when plugged and number of litters with varying proportions of early embryonic deaths.

Before dissection at E9.5, vaginal cytology was used to attempt to predict whether mating had resulted in pregnancy; all mice were dissected regardless of prediction. Predicting pregnancy based on pre-dissection vaginal cytology was 76.6% successful. Of the incorrect predictions, 19.1% were false positives, and 4.3% were false negatives. The vaginal swabs of pregnant mice have a characteristic look to them (Figure B2). Typically, there are many nucleated epithelial cells and a lot of mucous (A), and these cells are often dense and clump together (A, B). Mice who are producing a lot of mucous but are not pregnant have other cell types, such as leukocytes (D). The clumping of cells in non-pregnant mice also differs (E). Even when nucleated epithelial cells are less apparent, pregnant smears show dense and dark coloured clumps (C) compared to their non-pregnant counterparts (F).

The presence of many nucleated epithelial cells can sometimes resemble proestrus, where this is the predominant cell type. While this may help explain the high false positive rate, the quality of the smear prepared also certainly affects the ability to predict pregnancy. Although the false positive rate is high, predicting pregnancy using vaginal cytology was generally successful (76.6%). With more practice, I would anticipate that number could rise. In embryo collection experiments where the number of females available are limited, this appears to be a simple, quick, and effective method to reduce the number of mice culled who mated but are not pregnant. This therefore allows them to be used for the next round of matings, until they successfully become pregnant.

These vaginal cytology results clearly demonstrate that the estrous cycling is abnormal in our mice, although a full cycle analysis is difficult because I only swabbed mice for 4 days a week. Discovering the abnormal cycling was an unexpected result of attempting to increase mating and successful pregnancies. I do not have sufficient data to compare the cycling of heterozygous and wildtype mice, nor did I establish a quantifiable measure of disordered cycling for comparison. Overall, the pregnancy success rate is still quite low considering that nearly all mice set up after being staged were in theory receptive to mating. Using vaginal cytology for mating is an incredible amount of work and it did not increase the successful pregnancies over external vaginal assessment of mating receptivity.

In the future, I would use vaginal cytology to predict pregnancy, but would eliminate its use for mating.

It is possible that missing one wildtype copy of *Cecr2* has a mild effect on the fertility of the heterozygous mice. *Cecr2* is clearly involved in fertility, as mutants demonstrate defects (Norton, unpublished). The fertility of heterozygous mice has not been assessed. While this may be the case, it does not explain why wildtype mice also had abnormal cycling. It is likely that the mice were affected by environmental disruptions, such as unrequested and overt changes in the lighting in the room and the constant loud noise and vibration due to construction occurring in close proximity. How these disruptions could lead to defects in estrous cycling is unclear, however clear links between stressors and hormonal changes have been established (reviewed in Nargund, 2015; Whirledge & Cidlowski, 2010).

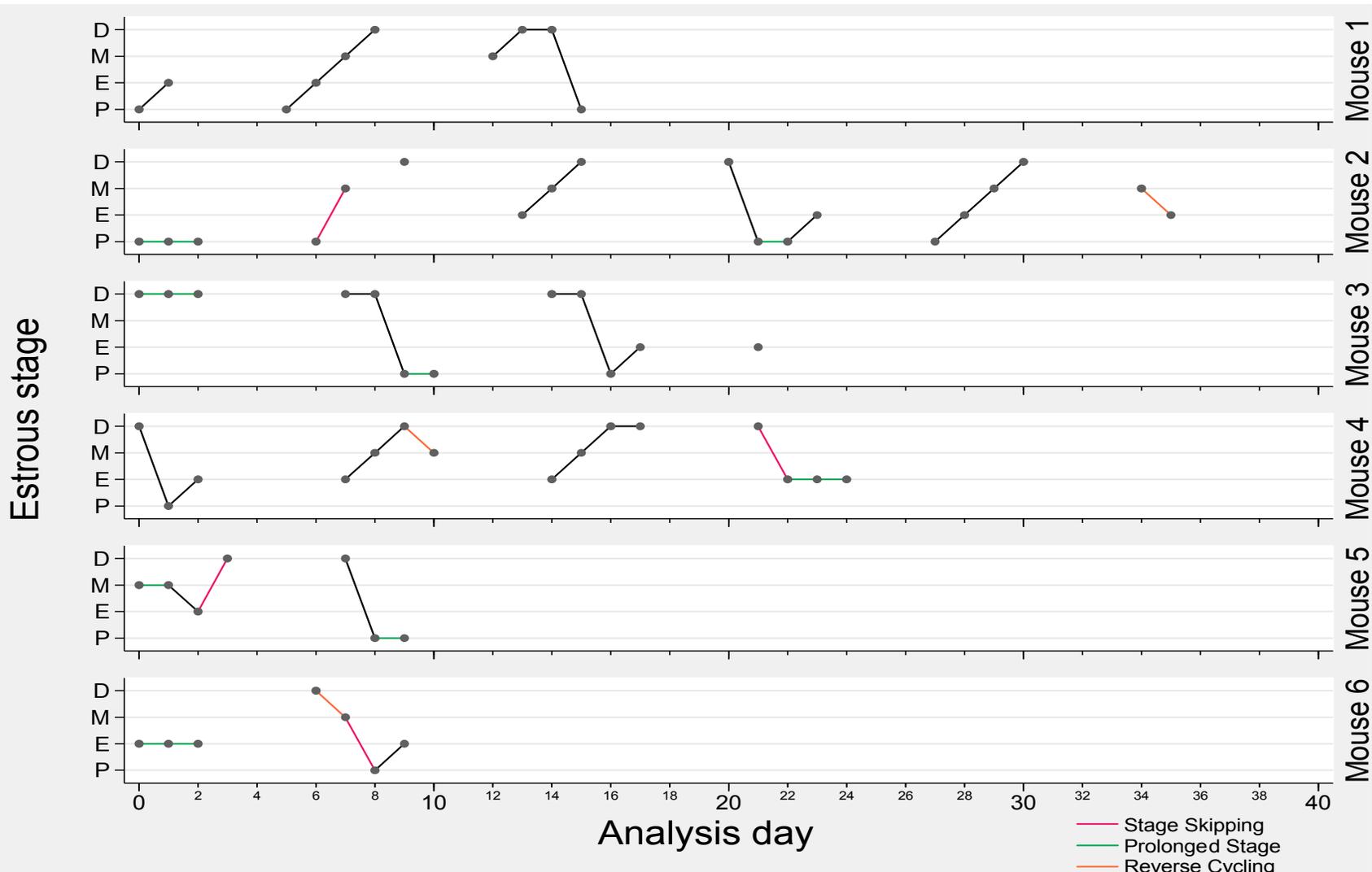


Figure B1. Estrous cycling of 6 representative mice. Mouse 1 represents a normal-cycling mouse that transitions through proestrus (P), estrus (E), metestrus (M), and diestrus (D) with a positive slope. The only normal negative slope is the transition from D to P. Mice 2-6 are cycling abnormally indicated by stage skipping (pink), a prolonged stage (green), and reverse cycling (orange). Graph created by Marc Parsons, and representative data is from Table B1.

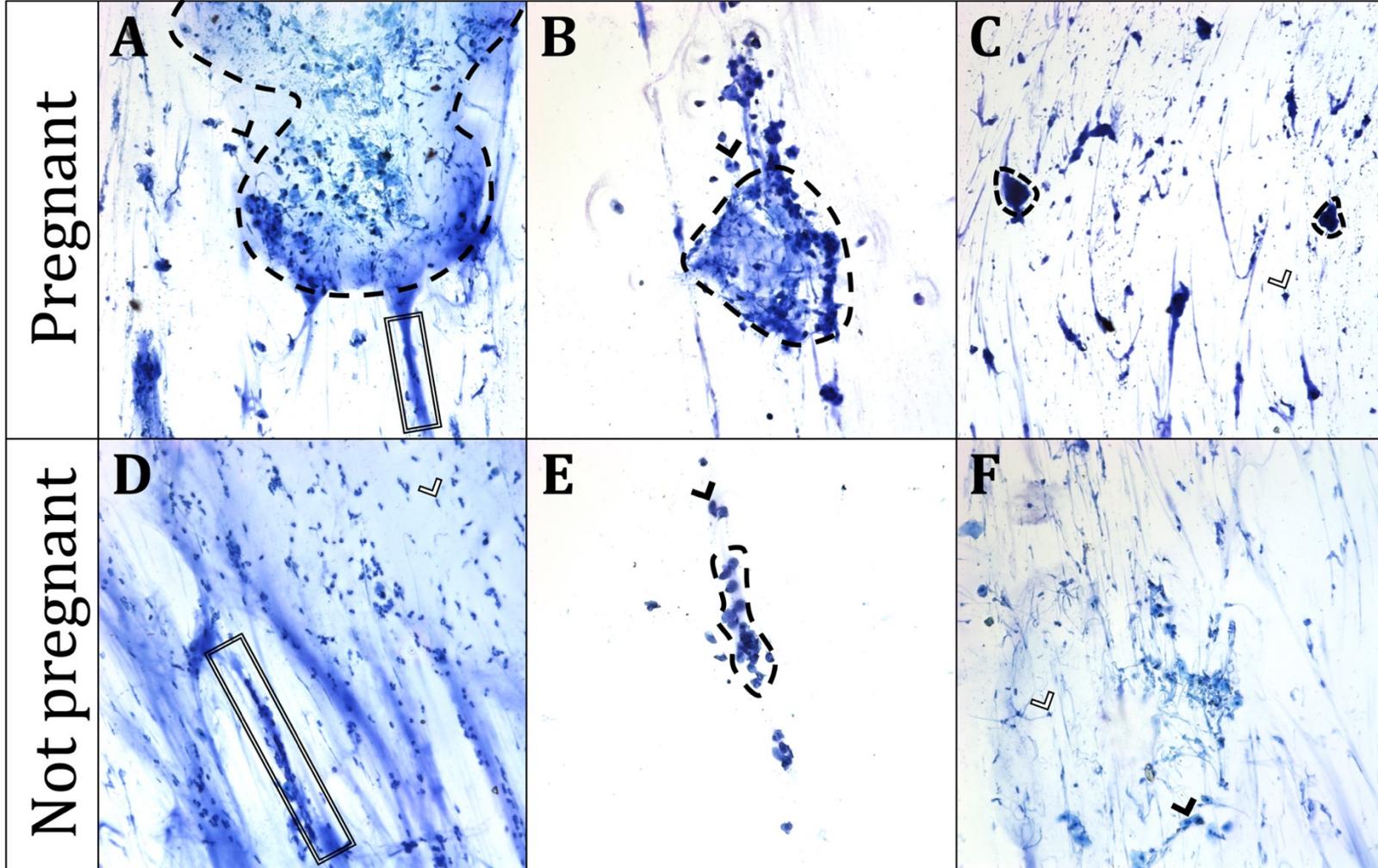


Figure B2. Representative vaginal swab analysis for mouse pregnancy prediction. Vaginal smears stained with methylene blue were viewed under 20X magnification to predict if the mouse was pregnant at E9.5. A-C show representative smears of accurately predicted pregnant mice. D-F show representative smears of accurately predicted not pregnant mice. Nucleated epithelial cells are indicated by the black arrow head, and leukocytes are indicated by the white arrow head. Areas of cell clumping are outlined by a black dashed line, and examples of areas of mucous are boxed with a double black line.

Table B2. Pregnancy analysis of plugged female mice.

Estrous Cycle	Number of pregnant mice	Number of not pregnant mice	Number of pregnant mice with suspicious plug	Number of not pregnant mice with suspicious plug	Number of litters with 100% early embryonic death	Number of litters with >50% but <100% early embryonic death
Proestrus	14	9	0	5	0	2
Estrus	2	1	0	0	0	0
Metestrus	1	0	0	0	0	0
Diestrus	6	1	1	1	2	1
Not staged	8	5	0	0	1	1
Sum	31	16	1	6	3	4