Role of *Cecr2* and Candidate Modifier Genes of *Cecr2* in the Neural Tube Defect Exencephaly

by

Renee Yvonne Marie Leduc

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

Doctor of Philosophy

in

Molecular Biology and Genetics

Department of Biological Sciences
University of Alberta

Abstract

Neurulation, the early developmental process that forms the rudiment of the brain and spinal cord, relies upon the intricate interplay of hundreds of genes in multiple genetic pathways within the appropriate environmental conditions. Neural tube defects (NTDs), common congenital birth defects in humans, arise when the process of neurulation is disrupted. Human NTDs are multifactorial disorders, which means that combinations of many genetic, epigenetic, and environmental factors interact in order for disease to manifest. Our mouse model of NTDs develops the perinatal lethal cranial NTD exencephaly (the murine equivalent of anencephaly in humans) when homozygous for a loss-of-function mutation in the ATP-dependent chromatin remodeling gene *Cecr2*. Much like in humans, manifestation of the exencephaly phenotype in *Cecr2* mutant mice is dependent on multiple factors. Work in this thesis focused on identifying and characterizing the multifactorial nature of exencephaly in both *Cecr2* mutant mice and in humans. A previously established incomplete penetrance of exencephaly in *Cecr2* homozygous mutant mouse embryos is indicative of genetic and/or environmental changes that are contributing resistance to exencephaly. An updated penetrance analysis that I performed revealed a reduction in exencephaly penetrance, which demonstrated that genetic and/or environmental factors are changing over time. Previous work in our lab has shown that the penetrance of exencephaly in *Cecr2* mutant mice is dependent on mouse strain, as *Cecr2* mutant BALB/cCrl mice are susceptible to developing exencephaly but *Cecr2* mutant FVB/N mice are resistant. This inter-strain variability suggests the presence of modifier genes, where the BALB/cCrl genetic background harbors susceptibility alleles and FVB/N harbors resistance alleles. Prior studies from the McDermid lab identified a modifier region in mouse chromosome 19 that contains at

least two modifier loci, which contribute to the difference in exencephaly penetrance seen between *Cecr2* mutant BALB/cCrl and FVB/N. I have further characterized the chromosome 19 modifier region and demonstrated that the two modifier loci are not additive, suggesting that the modifiers are involved in the same pathway or process. I then used whole exome sequencing of the two mouse strains to identify candidate modifier genes containing protein-coding variants that differed between the two strains. With this analysis, combined with previously generated data from whole genome microarrays, I produced a list of 26 candidate modifier genes that differ in expression and/or protein code between the two mouse strains. I showed via genetic analysis in the mouse that the top candidate gene, *Arhgap19*, is most likely not a *Cecr2* modifier. The human homologue of *CECR2* and the human homologues of the remaining candidate modifier genes were then sequenced in a cranial NTD cohort consisting of 156 probands. This study in humans identified protein-coding variants that were predicted to affect protein function in *CECR2* and in 17 of the candidate modifier genes, as well as established *DNMBP*, *MMS19*, and *TJP2* as top candidate NTD modifier genes.

As an independent but related project, I also characterized a gene-environment interaction that resulted in circling behavior, a phenotype unrelated to NTDs, in male mice of a specific genetic cohort in our mouse colony. This study is the first to show that environmental enrichment in the form of running-wheels can induce abnormal behaviors in genetically susceptible mice. Additionally, I sought to characterize phenotypes due to homozygous mutation or knockdown of the *Drosophila melanogaster* homologue of *Cecr2* (*dikar*) in an effort to produce a more tractable genetic model to study the molecular function of *Cecr2*. Results indicated that *dikar* is dispensible for normal fly development with no obvious phenotype due to loss of *dikar*.

Overall, I have established the *Cecr2* mutant mouse as a valuable model for studying the multifactorial nature of NTDs and have produced several novel candidate NTD genes in mice and humans. Important future work will be directed towards the functional characterization of protein-coding variants identified in the human cranial NTD cohort. In the event that a variant is shown to have a functional impact, expanded sequencing efforts of this variant or the gene it affects in additional NTD patients will aid in determining the relevance of such a gene in human NTD etiology.

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.

The research project involving human samples received research ethics approval from the University of Alberta Research Ethics Board, project name "Sequencing genes in a cohort of humans with a neural tube defect" (MS3_Pro00042452). This project was part of the larger research project at Duke University, which received research ethics approval from the Duke University Health System Institutional Review Board for Clinical Investigations, project name "Hereditary Basis of Neural Tube Defects" (CR5_Pro00016517).

Some of the research conducted for this thesis has previously been published and/or has been completed in collaboration. Specific details regarding publication and collaborations for each results chapter are given in the appropriate chapter preface.

Acknowledgements

First and foremost, I wish to express my sincere gratitude to my supervisor, Dr. Heather McDermid, for her unrelenting support, expertise, and patience. Her guidance and mentorship provided the foundation for all future successes in my science career. Her enthusiasm is truly contagious; it kept my passion for science alive. I could not have imagined a better mentor or lab environment for my PhD study.

I would also like to extend my thanks to the rest of my thesis committee, Dr. Kirst King-Jones and Dr. Fred Berry, for their insightfulness, guidance, and encouragement. They always were quick and gracious with offering their time and expertise, which has been invaluable for my research.

Much of the work in this thesis would not have been possible without the many collaborators I have had the privilege to work with. In the four short months I spent at Duke University, Dr. Allison Ashley-Koch integrated me into her lab as though I was one of her own. She was a superior mentor, and I want to thank her for sharing her immense knowledge, support and laboratory space. Dr. Simon Gregory also provided me with his unrelenting support and expertise, for which I am extremely grateful. Drs. Nicholas Katsanis and Erica Davis were so quick to offer their expertise and resources for my research project, which has led to a very fruitful collaboration for both this project and future projects. At the University of Alberta, Dr. John Locke has offered me his expertise, resources, and laboratory space since I was in undergraduate student. I have always felt like a member of his lab. I am eternally grateful for his advice, support, and mentorship. I also extend my thanks to Dr. Glen Baker, who graciously offered his time, expertise and resources as part of a collaboration.

I give a special thanks to my labmates, who are also my friends. They offered much advice for my research and immeasurable support for all aspects of my life. My PhD experience would have been a shadow without them. I also extend a very heartfelt thank-you to the rest of my friends and family, who never once stopped offering their support in all my life's endeavors.

I wish to acknowledge the funding provided by the National Science and Engineering Research Council (NSERC), who supported me throughout my PhD, and with their Micheal Smith Foreign Study Scholarship (MSFSS), made my international research possible. Lastly, I offer my sincerest thanks to the patients and their families that have graciously enrolled in the Duke NTD study, without whom much of this research would not have been possible.

Table of Contents **Page #**

1: General Introduction 1

1.1: Introduction 2

1.2: Neurulation 2

1.3: Open neural tube defects 8

1.4: Mice as models of neurulation and neural tube defects 17

1.5: Other models of neurulation and neural tube defects 19

1.6: ATP dependent chromatin remodeling 21

1.7: *Cecr2* 23

1.8: Objectives 26

2: Materials and methods 28

2.1: Mouse husbandry 29

2.2: Embryo collection and exencephaly penetrance analysis 29

2.3: Small inner ear penetrance analysis 30

2.4: DNA extraction 30

2.5: $Cecr2^{GT45bic}$ genotyping 31

2.6: $Cecr2^{tm1.1Hemc}$ genotyping 31

2.7: Sanger sequencing 32

2.8: Whole exome sequencing of BALB/cCrl and FVB/N 33

2.9: Exonic sequencing of 25 candidate NTD genes in 156 human anencephaly samples 35

2.10: LacZ staining 39

2.11: Paraffin embedding and histology of tissue samples 39

2.12: Immunofluorescence of neurulating embryos 40

2.13: Cloning of human *CECR2* 41

2.14: Site-directed mutagenesis of human *CECR2* plasmid constructs 42

2.15: Sub-cloning human *CECR2* plasmid constructs 44

2.16: HEK293 cell culture maintenance, harvest, and transfection 45

2.17: Protein extraction and quantification 47

Table of Contents **Page #**

2.18: CECR2 immunoprecipitation 47
2.19: Protein electrophoresis 48
2.20: Western blot analysis 49
2.21: MOD 5/31, MOD 5/+, and MOD 31/+ crosses 50
2.22: Congenic *Cecr2*^{GT45bic} BALB/cCrl x BALB/cJ crossing scheme 52
2.23: Total RNA extraction and cDNA amplification for quantitative real-time PCR (qRT-PCR) 54
2.24: Quantitative real-time PCR (qRT-PCR) of candidate modifier genes 54
2.25: Generation of FVB/129P2 (*Arhgap19*$^{GT(YHD020)Byg}$) circling mice 56
2.26: *Arhgap19*$^{GT(YHD020)Byg}$ genotyping 56
2.27: Phenotyping of circling behavior in FVB/129P2 mice 57
2.28: Phenotyping pipeline to determine inability to sense gravity due to inner ear defects 58
2.29: High performance liquid chromatography (HPLC) analysis of biogenic amines and amino acids in left and right brain hemispheres 59
2.30: *Drosophila* stocks and husbandry 60
2.31: RNAi in *Drosophila* to knock-down *dikar* mRNA and characterize phenotype 66
2.32: Gamma irradiation (IR) of *Drosophila* embryos/larvae to test *dikar* function in DNA double-strand break repair 67
2.33: DNA sequence analysis of *Drosophila* genes *dikar* and *Gen* to detect possible second-site mutations 69

3: *Cecr2* and candidate modifier genes of *Cecr2*-associated exencephaly in mice and humans 71
3.1: Introduction 72

Table of Contents **Page #**

RESULTS

3.2: Exencephaly penetrance in congenic *Cecr2*^{GT45bic} BALB/cCrl mice is lower than the penetrance in the earlier N6 generation 81

3.3: Small inner ears are secondary to the exencephaly phenotype 83

3.4: Exencephaly penetrance of congenic *Cecr2*$^{tm1.1Hemc}$ FVB/N mice is lower than the penetrance in the earlier N3 generation 85

3.5: Modifier genes within the chromosome 19 modifier region are not additive 86

3.6: qRT-PCR validation of microarray data yields a list of 9 candidate modifier genes differentially expressed between BALB/cCrl and FVB/N 92

3.7: Whole exome sequencing of BALB/cCrl and FVB/N identifies coding variants of interest in 19 genes within the chromosome 19 candidate modifier region 94

3.8: *Arhgap19* is not a modifier of *Cecr2*-associated exencephaly 98

3.9: Exonic sequencing of CECR2 in 156 human cranial NTD samples identifies protein-coding variants of interest 106

3.10: Exonic sequencing of 24 *CECR2* candidate modifier genes in 156 human cranial NTD samples identifies protein-coding variants of interest, particularly in *DNMBP*, *MMS19*, and *TJP2* 110

3.11: Discussion 130

4: *dikar*, the *Drosophila melanogaster* homologue of *Cecr2*, is dispensable for normal development 152

4.1: Introduction 153

RESULTS

4.2: RNAi mediated targeted gene silencing of *dikar* does not result in a planar cell polarity or other overt phenotype 158

Table of Contents **Page #**

4.3: Irradiation sensitivity of homozygous *dikar* mutants appears to be due to a second-site mutation 161

4.4: The locus conferring irradiation sensitivity is likely within the chromosomal region encompassed by the *Df(3L)ZN47* deficiency: *Gen* and *dikar* are candidate genes 165

4.5: Discussion 177

5: Circling behavior and stereotypic route-tracing in a genetically distinct cohort of male mice 180

5.1: Introduction 181

RESULTS

5.2: Circling behavior in FVB/129P2 male mice 183

5.3: Circling behavior is not associated with the *Arhgap19*$^{GT(YHD020)Byg}$ mutation 189

5.4: Circling behavior in female mice 190

5.5: Circling mice do not have obvious behaviors typical of inner ear defects 191

5.6: Dopamine and serotonin biochemistry is altered in circling FVB/129P2 male mice 193

5.7: Discussion 195

6: General discussion and future directions 202

6.1: Analyzing exencephaly penetrance changes over time in *Cecr2* mutant mice 204

6.2: Identifying NTD modifier genes in susceptible genetic backgrounds in mice 206

6.3: Gene-environment interactions contributing to abnormal mouse behavior 211

6.4: Characterizing *dikar*, the *Drosophila* homologue of *Cecr2* 212

Table of Contents **Page #**

6.5: Concluding remarks 212

Bibliography 214

Appendices 234
Appendix A: List of primers 235
Appendix B: Cranial NTD phenotypes for 156 probands 242
Appendix C: Co-immunoprecipitation assay comparing human wildtype CECR2 to variant CECR2 (E32K and P632L) in the ability to complex with SNF2L in transfected HEK293 cells 246
Appendix D: *CECR2* ORF constructs in the pENTR^{TM}11 plasmid 250
Appendix E: Biogenic amine and amino acid concentrations in half brain samples 260
Appendix F: *Drosophila* progeny counts for RNAi crosses 262

List of tables **Page #**

Table 1.3.1: Genes containing SNVs associated with NTDs in humans 10

Table 1.3.2: Genes containing rare, novel, or *de novo* variants that alter the protein code (nonsense, missense, frameshift) identified in human sequencing studies 13

Table 2.8.1: Summary of whole exome sequencing quality metrics 35

Table 2.9.1: Coding sequence that was not covered by QIAGEN® GeneRead DNAseq custom primers 37

Table 2.14.1: Four site-directed mutagenesis reactions that vary in primer concentration and presence/absence of DMSO used for optimal generation of mutagenized *CECR2* plasmid constructs 44

Table 2.26.1: PCR programs for $Arhgap19^{Gt(YHD020)Byg}$ genotyping 57

Table 2.30.1: *Drosophila* stocks 62

Table 2.32.1: Fly crosses set up for IR experiments 69

Table 3.1.1: Genes within the chromosome 19 modifier region that differ in expression by at least 1.5 fold using GC-RMA/Student's t-test analysis 79

Table 3.1.2: Genes within the chromosome 19 modifier region that differ in expression by at least 1.5 fold using RMA/ANOVA analysis 80

Table 3.2.1: Congenic $Cecr2^{GT45bic}$ BALB/cCrl homozygous mutant exencephaly penetrance analysis 83

Table 3.3.1: Statistical analysis of inner ear size comparing E17.5 congenic $Cecr2^{GT45bic}$ BALB/cCrl embryos wildtype for *Cecr2* (WT), homozygous mutant for the $Cecr2^{GT45bic}$ allele but not penetrant for exencephaly (NP), and homozygous mutant for the $Cecr2^{GT45bic}$ allele with exencephaly (E) 84

Table 3.4.1: Congenic $Cecr2^{tm1.1Hemc}$ FVB/N homozygous mutant exencephaly penetrance analysis 86

Table 3.5.1: χ^2 test-of-independence comparing exencephaly penetrance between congenic $Cecr2^{GT45bic}$ BALB/cCrl, MOD 5, MOD 31, MOD 5/31, MOD 5/+, and MOD 31/+ mouse embryos homozygous for the $Cecr2^{GT45bic}$ allele 89

List of tables **Page #**

Table 3.5.2: Summary of exencephaly penetrance of males, females, and combined sexes for congenic *Cecr2*^{GT45bic} BALB/cCrl, MOD 5, MOD 31, MOD 5/31, MOD 5/+, and MOD 31/+ 90

Table 3.5.3: Statistical analysis of inner ear size comparing E17.5 embryos wildtype for *Cecr2* (WT), homozygous mutant for the *Cecr2*^{GT45bic} allele but not penetrant for exencephaly (NP), and homozygous mutant for the *Cecr2*^{GT45bic} allele with exencephaly (E), between congenic *Cecr2*^{GT45bic} BALB/cCrl, MOD 5, and MOD 31 mouse lines 92

Table 3.6.1: qRT-PCR validation of candidate modifier genes identified by whole genome microarray 93

Table 3.7.1: Summary of SNVs and INDELs identified by WES in the chromosome 19 modifier region 95

Table 3.7.2: Candidate modifier genes containing single nucleotide variants (SNVs) that alter the protein code that differ between BALB/cCrl and FVB/N wildtype mice 96

Table 3.7.3: Candidate modifier genes containing an insertion or deletion (INDEL) that alter the protein code that differ between BALB/cCrl and FVB/N wildtype mice 96

Table 3.7.3: Refined list of candidate modifier genes that differ in expression and/or protein code between BALB/cCrl and FVB/N 97

Table 3.8.1: SNVs that differ between BALB/cCrl and BALB/cJ within the chromosome 19 modifier region 103

Table 3.8.2: Summary of exencephaly penetrance in BALB/c (50% Charles River substrain, 50% Jackson substrain) *Cecr2*^{GT45bic} homozygous mutant embryos that are homozygous mutant, heterozygous, or homozygous wildtype for the *Arhgap19*^{ex6non} nonsense mutation and compared to exencephaly penetrance of congenic *Cecr2*^{GT45bic} BALB/cCrl mutant embryos 106

List of tables **Page #**

Table 3.9.1: List of 9 variants of interest identified in CECR2 in the human cranial NTD cohort 108
Table 3.9.2: *CECR2* protein-coding sequence variants of interest 109
Table 3.10.1: List of 53 variants of interest identified in candidate NTD modifier genes in the human cranial NTD cohort 111
Table 3.10.2: Protein-coding variants of interest in cranial NTD probands 113
Table 3.10.3: List of candidate modifier genes based on variants identified in a human cranial NTD cohort, protein function, and gene expression 116
Table 3.10.4: *DNMBP* protein-coding sequence variants of interest 120
Table 3.10.5: *MMS19* protein-coding sequence variants of interest 124
Table 3.10.6: *TJP2* protein-coding sequence variants of interest 127
Table 3.11.1: Exencephaly penetrance associated with homozygous mutation in *Cecr2* dropped over time in four different mouse lines 132
Table 4.2.1: GAL4 driver expression patterns 159
Table 4.2.2: External phenotypes seen in *Drosophila melanogaster* as a result of RNAi mediated targeted gene silencing of *dikar*, *fz*, and *Vang* 160
Table 4.4.1: Gene ontology (GO) analysis of 176 genes within the *Df(3L)ZN47* region identified 13 DNA binding genes 167
Table 4.4.2: Protein predictions of deleteriousness of the 11 variants in the Gen Protein 169
Table 4.4.3: Secondary structure predictions for the 11 amino acids that differ in Gen between the $dikar^{WT}$ chromosome and the *TM3* chromosome 170
Table 4.4.4: Secondary structure predictions for the 3 amino acids that differ in Dikar between the $dikar^{WT}$ chromosome and the *TM3* chromosome 176
Table 5.3.1: Circling behavior is not associated with the $Arhgap19^{Gt(YHD020)Byg}$ mutation 190
Table 5.5.1: Times for individual mice spent engaging in each behavior scored in the behavioral tests for inner ear function 192

List of tables **Page #**

Table 5.6.1: Analyses of concentrations of biogenic amines, their metabolites, and their precursor amino acids in male FVB/129P2 comparing circling mice housed with wheels, non-circling mice housed with wheels, and non-circling mice housed with cardboard housing (standard) 195

Table 5.6.2: Analyses of amino acid concentrations in male FVB/129P2 comparing circling mice housed with wheels, non-circling mice housed with wheels, and non-circling mice housed with cardboard housing (standard) 195

Table E.1: Individual sample concentrations of biogenic amines, their metabolites, and their precursor amino acids in male FVB/129P2 in circling mice housed with wheels, non-circling mice housed with wheels, and non-circling mice housed with cardboard housing (standard) 260

Table E.2: Individual sample amino acid concentrations in male FVB/129P2 in circling mice housed with wheels, non-circling mice housed with wheels, and non-circling mice housed with cardboard housing (standard) 261

Table F.1: *Drosophila* progeny counts for RNAi CG32394 2478/GD (*dikar*) crosses 262

Table F.2: *Drosophila* progeny counts for RNAi CG32394 100383/KK (*dikar*) crosses 263

Table F.3: *Drosophila* progeny counts for RNAi CG32393 107312/KK (*dikar*) crosses 264

Table F.4: *Drosophila* progeny counts for RNAi CG17697 43077/GD (*fz*) crosses 265

Table F.5: *Drosophila* progeny counts for RNAi CG17697 105697/KK (*fz*) crosses 266

Table F.6: *Drosophila* progeny counts for RNAi CG8075 100819/KK (*Vang*) crosses 267

Table F.7: *Drosophila* progeny counts for RNAi CG8075 7376/GD (*Vang*) crosses 268

List of figures **Page #**

Figure 1.2.1: Neurulation 6

Figure 1.2.2: Neural plate induction 7

Figure 1.2.3: Planar cell polarity (PCP) pathway in mammals 8

Figure 1.4.1: Mouse as a genetic model for neural tube defects 18

Figure 1.7.1: *Cecr2* and the *Cecr2*^{GT45bic} allele 25

Figure 2.21.1: MOD 5/31, MOD 5/+, and MOD 31/+ crossing scheme 51

Figure 2.21.2: MOD 5, MOD 31, MOD 5/31, MOD 5/+, MOD 31/+ chromosome 19 modifier region genetic background 52

Figure 2.22.1: Crossing scheme to test if the *Arhgap19* nonsense mutation contributes to exencephaly susceptibility 53

Figure 2.30.1: Crossing scheme to generate the homozygous viable *P{SUPor-P}dikar*^{KG00884} stock to act as a control 64

Figure 2.30.2: Crossing scheme to precisely excise the *P{SUPor-P}dikar*^{KG00884} P element in order to generate the *dikar*^{WT} stock 65

Figure 2.31.1: Crossing scheme and summary of RNAi mediated gene silencing in *Drosophila* 67

Figure 3.1.1: The threshold model of multifactorial disease inheritance 72

Figure 3.1.2: Crossing scheme to generate sub-interval congenic mouse lines MOD 4, MOD 5, and MOD 31 77

Figure 3.1.3: Exencephaly penetrance analyses of sub-interval congenic mouse lines demonstrate more than one modifier locus within the chromosome 19 modifier region 78

Figure 3.3.1: Small inner ears are secondary to the exencephaly phenotype in congenic *Cecr2*^{GT45bic} BALB/cCrl mutant embryos 84

Figure 3.5.1: Exencephaly penetrance analyses in MOD 5, MOD 31, MOD 5/31, MOD 5/+, and MOD 31/+ 88

Figure 3.5.2: Inner ears are significantly smaller in embryos with exencephaly 91

Figure 3.6.1: *Foxd4* expression is significantly down-regulated in BALB/cCrl and 129S2, but not C57BL/6, relative to FVB/N 94

List of figures **Page #**

Figure 3.8.1: Wildtype and mutant alleles of *Arhgap19* 99

Figure 3.8.2: *Arhgap19* expression in whole embryos 100

Figure 3.8.3: *Arhgap19* expression in E13.5 and E14.5 embryo brains 101

Figure 3.9.1: Location of all variants of interest within the CECR2 protein 110

Figure 3.10.1: DNMBP is present in the neural epithelium at the time of neurulation 118

Figure 3.10.2: Variants identified in the candidate modifier gene *DNMBP* in the human cranial NTD cohort 121

Figure 3.10.3: MMS19 is present in the neural epithelium at the time of neurulation 122

Figure 3.10.4: Variants identified in the candidate modifier gene *MMS19* in the human cranial NTD cohort 125

Figure 3.10.5: Variants identified in the candidate modifier gene *TJP2* in the human cranial NTD cohort 127

Figure 3.10.6: Variants identified in the remaining 14 candidate modifier genes in the human cranial NTD cohort 129

Figure 4.1.1: Three dikar deletion alleles generated by imprecise excision of the P element $P\{SUPor\text{-}P\}dikar^{KG00884}$ 155

Figure 4.1.2: Distribution of the six core PCP proteins in *drosophila* wing cells 156

Figure 4.3.1: Flies homozygous for the $dikar^{5}$ allele are sensitive to IR doses of 20 grey, 30 grey, and 40 grey 162

Figure 4.3.2: Irradiation sensitivity for $dikar^{3}$ and $dikar^{5}$ flies on three ranges of embryo/larva ages (0-2 days, 2-4 days, 4-6 days) dosed with 0 grey, 15 grey, and 30 grey 163

Figure 4.3.3: Irradiation sensitivity in flies homozygous for the $dikar^{3}$ and $dikar^{5}$ alleles, heterozygous for a *dikar* deletion over a deficiency, as well as for homozygous flies produced in control crosses ($P\{SUPor\text{-}P\}dikar^{KG00884}$ and $dikar^{WT}$) at 22.5 grey 165

Figure 4.4.1: Genomic region encompassed by the Df(3L)ZN47 deficiency 166

List of figures **Page #**

Figure 4.4.2: Gen protein sequence alignment shows 11 Gen amino acids that differ between the *Gen* allele on the *dikar*^{WT} chromosome and the *Gen* allele on the *TM3* chromosome 168

Figure 4.4.3: The most comparable Gen protein prediction models of tertiary structure between the two protein variants 171

Figure 4.4.4: Dikar protein sequence alignment shows 4 Dikar amino acids that differ between the *dikar* allele on the *dikar*^{WT} chromosome and the *dikar* allele on the *TM3* chromosome 172

Figure 4.4.5: The most comparable Dikar protein prediction models of tertiary structure between the two protein variants 176

Figure 5.2.1: Mouse pedigree showing the ancestry of the FVB/129P2 line 185

Figure 5.2.2: Circling in FVB/129P2 male mice 187

Figure 5.2.3: Socially housed mice can develop circling behavior prior to separation as well as after separation due to fighting 188

Figure 5.6.1: Male FVB/129P2 mice that circle have altered brain biochemistry 194

Figure 6.1.1: The *Cecr2* mutant mouse is a multifactorial model of exencephaly 204

Figure 6.1.2: Encephalocele in a homozygous mutant congenic *Cecr2*^{GT45bic} FVB/N mouse pup (generation N10) 205

Figure 6.2.1: The cholesterol synthesis pathway appears to be down-regulated in FVB/N relative to BALB/cCrl neurulating embryos 210

Figure C.1: HEK293 cells have low endogenous levels of CECR2 when compared to the level of CECR2 in mouse TT2 ES cells 248

Figure C.2: α-CECR2 antibody is capable of specifically pulling down SNF2L in HEK293 cells 249

List of Abbreviations

°C	degrees celsius
1KG	1000 Genomes database
2700078E11Rik	CDK2 associated, cullin domain 1
3-MT	3-methoxytyramine
5-HIAA	5-hydroxyindoleacetic acid
5-HT	serotonin
53BP1	tumor protein P53 binding protein 1
AA	amino acid
ACF1	bromodomain adjacent to zinc finger domain, 1A
ADHD	attention deficit hyperactivity disorder
aka	also known as
ALA	alanine
Alx1	ALX homeobox 1
Amp	ampicillin
ANKRD32	ankyrin repeat domain 32
Ankrd6	ankyrin repeat domain 6
ANOVA	analysis of variance
Ap2	activator protein 2
APEX1	APEX nuclease (multifunctional DNA repair enzyme) 1
ARG	arginine
Arhgap19	Rho GTPase activating protein 19
ASP	aspartate
ATAD2	ATPase family, AAA domain containing 2
ATM	ataxia telangiectasia mutated
ATP	adenosine triphosphate
BAF	BRG1- or HRBM-associated factors
BAF155	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily C, member 1
BAG1	BCL2-associated athanogene
BAM	binary version of a SAM file
BAR	bin-amphiphysin-rvs
BAZ	bromodomain adjacent to zinc finger
BMP	bone morphogenetic protein
bp	base pair
BPTF	bromodomain PHD finger transcription factor
Brca1	breast cancer 1
Brg1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily A, member 4 (aka Smarca4)
BSA	bovine serum albumin
BTAF1	BTAF1 RNA polymerase II, B-TFIID transcription factor associated, 170 kDa
BWA	Burrows-Wheeler Aligner
c-Myc	V-myc avian myelocytomatosis viral oncogene homolog
C17orf80	chromosome 17 open reading frame 80
Caco-2	human epithelial colorectal adenocarcinoma cell line
caspase-3	caspase 3, apoptosis-related cysteine peptidase
CBS	cystathionine-beta-synthase

CCW	counterclockwise
Cdc20	cell division cycle 20
Cdc42	cell division cycle 42
cDNA	complementary DNA
Cecr2	cat eye syndrome chromosome region, candidate 2
CELSR1	cadherin, EGF LAG seven-pass G-type receptor 1
CERF	CECR2-containing remodeling factor
CES	cat eye syndrome
CHD	chromodomain, helicase, DNA binding
CHKA	choline kinase alpha
CNV	copy number variant
CO_2	carbon dioxide
CPEB3	cytoplasmic polyadenylation element binding protein 3
CSTF2T	cleavage stimulation factor, 3' pre-RNA, subunit 2, 64kDa, Tau variant
CT	cycle threshold
CTH	cystathionine gamma-lyase
CUBN	cubulin
Cutc	cutC copper transporter homolog
CW	clockwise
D-SER	D-serine
D19Ertd737e	family with sequence similarity 204, member A
D2 receptor	dopamine receptor D2
DA	dopamine
DACT1	dishevelled-binding antagonist of beta-catenin 1
DAPI	4',6-diamidino-2-phenylindole
DAT	dopamine transporter
dbSNP137	SNP database build 137
DDT	DNA binding homeobox and different transcription factor
DDX3X	DEAD (Asp-Glu-Ala-Asp) box helicase 3, X-linked
DENND6A	DENN/MADD domain containing 6A
DEPC	diethylpyrocarbonate
DExx	Asp-Glu-xx(two indeterminate amino acids) box helicase domain
DH	Dbl homolgy
DHFR	dihydrofolate reductase
DLHP	dorso-lateral hinge point
Dlx5	distal-less homeobox 5
Dlx6	distal-less homeobox 6
Dmp1	dentin matrix acidic phosphoprotein 1
Dmrt2	doublesex and mab-3 related transcription factor 2
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNApol-ε58	DNA polymerase ε 58kD subunit
DNMBP	dynamin binding protein
dNTP	deoxyribonucleotide triphosphate
DOPAC	3,4-dihydroxyphenylacetic acid
DPX	dibutyl phthalate, xylene
DSB	double-strand break

Dvl1	dishevelled segment polarity protein 1
DVL2	dishevelled segment polarity protein 2
Dvl3	dishevelled segment polarity protein 3
E	embryonic day
E. coli	Escherichia coli
EDTA	ethylenediaminetetraacetic acid
Elmo2	engulfment and cell motility 2
en	engrailed
EphA	ephrin receptor A
ES	embryonic stem
esp6500	exome sequencing project (6503 samples sequenced)
Exo	exonuclease I
EXOSC1	exosome component 1
ey	eyeless
Eya1	eyes absent (drosophila) homolog 1
Eya2	eyes absent (drosophila) homolog 2
FAM129B	family with sequence similarity 129, member B
FAM160B1	family with sequence similarity 160, member B1
FAM45A	family with sequence similarity 45, member A
FAS	fas cell surface death receptor
FGF	fibroblast growth factor
Foxd3	forkhead box D3
FOXD4	forkhead box D4
FOXF2	forkhead box F2
Foxi1	forkhead box L1
Foxi3	forkhead box L3
FUZ	fuzzy planar cell polarity protein
FW	formula weight
Fz3	frizzled class receptor 3
Fz6	frizzled class receptor 6
Fzd3	frizzled class receptor 3
Fzd6	frizzled class receptor 6
g	gram
GABA	gamma-aminobutyric acid
Gadd45a	growth arrest and DNA-damage-inducible 45 alpha
GAP	GTPase activating protein
Gata2	GATA binding protein 2
Gata3	GATA binding protein 3
GATK	Genome Analysis Toolkit
Gbx2	gastrulation brain homeobox 2
GD library	P element library (Drosophila RNAi)
Gen	XPG-like endonuclease
GERP	genomic evolutionary rate profiling
GIGYF1	GRB10 interacting GYF protein 1
GLIS3	GLIS family zinc finger 3
GLN	glutamine
GLU	glutamate

GLUT1	solute carrier family 2 (facilitated glucose transporter), member 1
GLY	glycine
GRCh37	UCSC human genome assembly (synonymous with hg19)
Grhl2	grainyhead-like 2
Grhl3	grainyhead-like 3
GT	genetrap
GWAS	genome wide association study
H1	histone H1
H2A	histone H2A
H2A.Z	histone H2A.Z (variant of histone H2A)
H2Av	histone H2Av (variant of histone H2A)
H2AX	histone H2AX (variant of histone H2A)
H2B	histone H2B
H3	histone H3
H4	histone H4
H5	histone H5
HCl	Hydrogen chloride
HEAT	Huntingtin, elongation factor 3, protein phosphatase 2A, kinase TOR1
HEK	human embryonic kidney
HEK293	human embryonic kidney 293 cell line
HEK293T	HEK293 cell line containing SV40 large T antigen
HELICc	helicase superfamily c-terminal domain
hg19	UCSC human genome assembly (synonymous with GRCh37)
Hh	hedgehog
HIF1AN	hypoxia inducible factor, alpha subunit inhibitor
HIP1R	huntingtin interacting protein 1 related
HK1	hexokinase 1
HNRNPL	heterogeneous nuclear ribonucleoprotein L
HPLC	high performance liquid chromatography
HPS6	Hermansky-Pudlak syndrome 6
HR	homologous recombination
HRP	horseradish peroxidase
HS	heat-shock
HSA	helicase-SANT-associated domain
hScrib	scribbled planar cell polarity protein (aka SCRIB, Scrb1)
HSLAS	Health Sciences Laboratory Animal Services
HUWE1	HECT, UBA and WWE domain containing 1, E3 ubiquitin protein ligase
HVA	homovanillic acid
ICMT	isoprenylcysteine carboxyl methyltransferase
IgG	immunoglobulin
IGV	Integrative Genomics Viewer
INDEL	insertion/deletion
INF2	inverted formin, FH2 and WH2 domain containing
INO80	inositol requiring 80
IP6K1	inositol hexakisphosphate kinase 1
IR	gamma irradiation
ISWI	imitation switch

ITGAX	integrin, alpha X (complement component 3 receptor 4 subunit)
JNK	c-Jun NH2-terminal kinase
Kan	kanamycin
KANK1	KN motif and ankyrin repeat domains 1
kb	kilobase
kDa	kilo Dalton
KK library	ΦC31 library (Drosophila RNAi)
L	litre
L-SER	L-serine
LB	lysogeny broth
LCMT2	leucine carboxyl methyltransferase 2
LEPR	leptin receptor
Lipa	lysosomal acid lipase A
LIPJ	lipase, family member J
Lipo1	lipase member, O1
LRFN1	leucine rich repeat and fibronectin type III domain containing 1
LRP6	low density lipoprotein receptor-related protein 6
M	molar concentration in moles/litre
MAF	minor allele frequency
March5	membrane-associated ring finger (C3HC4) 5
Mb	megabase
MBSU	Molecular Biology Services Unit
MDC1	mediator of DNA-damage checkpoint 1
MDCK	Madin-Darby canine kidney epithelial cell line
MDHFD2	methylenetetrahydrofolate dehydrogenase 2
MFAP1	microfibrillar-associated protein 1
$MgCl_2$	magnesium chloride
MHP	medial hinge point
mL	millilitre
mM	molar concentration in millimoles/litre
mm9	UCSC mouse genome assembly build 9
MMS19	MMS19 nucleotide excision repair homolog
MRN	Mre11, Rad50, and Nbs1
mRNA	messenger RNA
Msx1	msh homeobox 1
MTHFD1	methylenetetrahydrofolate dehydrogenase 1
MTHFR	methylenetetrahydrofolate reductase
MTRR	5-methyltetrahydrofolate-homocysteine methyltransferase reductase
n	sample number
N	generation number
N-Myc	V-myc avian myelocytomatosis viral oncogene neuroblastomahderived Homolog
NA	noradrenaline
NaCl	sodium chloride
NANOS1	Nanos homolog 1
NaOAc	sodium acetate
NAT1	N-acetyltransferase 1
NCAM1	neural cell adhesion molecule 1

NGS	next generation sequencing
NHEJ	non-homologous end-joining
NTD	neural tube defect
ORF	open reading frame
Otx2	orthodenticle homeobox 2
P	P-value
PAGE	polyacrylamide gel electrophoresis
PAR	partitioning defective
PAR-3	partitioning defective 3 homolog
PAX1	paired box 1
PAX3	paired box 3
PAX7	paired box 7
PAX8	paired box 8
PBS	phosphate buffered saline
PCMT1	protein-L-isoaspartate (D-aspartate) O-methyltransferase
PCP	planar cell polarity
PCR	polymerase chain reaction
PDZ	post synaptic density protein, Drosophila disc large tumor suppressor, zona occludens-1
PHD	plant homeodomain
PHF14	PHD finger protein 14
Pk1	prickle 1
Pk2	prickle 2
PNLIPRP2	pancreatic lipase-related protein 2
PP2 (HDIV)	polyphen2 (HumDiv-trained model)
prd	paired
PREPL	prolyl endopeptidase-like
PRMT1	protein arginine methyltransferase 1
PTBP1	polypyrimidine tract binding protein 1
Pten	phosphatase and tensin homolog
PTPRS	protein tyrosine phosphatase, receptor type, S
PVDF	polyvinylidene fluoride
qRT-PCR	quantitative real-time PCR
Rac1	ras-related C3 botulinum toxin substrate 1
Rcc1	regulator of chromosome condensation 1 ortholog
RefSeq	Reference Sequence database
RhoA	ras homolog family member A
RISC	RNA-induced silencing complex
RMA	robust multi-array
RNA	ribonucleic acid
RNAi	RNA interference
RNLS	renalase, FAD-dependent amine oxidase
Rock	rho-associated, coiled-coil containing domain protein kinase
RP1L1	retinitis pigmentosa 1-like 1
RSF-1	remodeling and spacing factor 1
Rxn	reaction
SAM	sequence alignment data (file)
SANT	Swi3, Ada2, N-cor, and TFIIIB domain

SAP	shrimp alkaline phosphatase
SARDH	sarcosine dehydrogenase
SASS	Sciences Animal Support Services
SCD	stearoyl-coA desaturase 5
Scd1	stearoyl-Coenzyme A desaturase 1
Scd2	stearoyl-Coenzyme A desaturase 2
Scrb1	scribbled planar cell polarity protein (aka hScrib, SCRIB)
SCRIB	scribbled planar cell polarity protein (aka hScrib, Scrb1)
SDS	sodium dodecyl sulfate
sec	seconds
Sfrp5	secreted frizzled-related sequence protein 5
SFXN2	sideroflexin 2
SH3	SRC homology 3
shep	alan shepard
Shh	sonic hedgehog
SHROOM3	shroom family member 3
SIFT	sorting intolerant from tolerant
Six1	SIX homoebox 1
Six4	SIX homeobox 4
SLAIN2	SLAIN motif family, member 2
SLC19A1	solute carrier family 19 (folate transporter), member 1
SLC22A13	solute carrier family 22 (organic anion/urate transporter), member 13
SLC26A4	solute carrier family 26 (anion exchanger), member 4
SLC2A4RG	SLC2A4 regulator
SLIDE	SANT-like ISWI domain
Smarca1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily A, member 1 (aka SNF2L)
Smarca4	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily A, member 4 (aka Brg1)
Smarca5	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily A, member 5 (aka SNF2H)
Smurf1	SMAD specific E3 ubiquitin protein ligase 1
Smurf2	SMAD specific E3 ubiquitin protein ligase 2
SNF2H	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily A, member 5 (aka Smarca5)
SNF2L	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily A, member 1 (aka Smarca1)
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
Sox2	SRY (sex determining region Y) box-2
Sox3	SRY (sex determining region Y) box-3
Sry	sex determining region Y
ssDNA	single-stranded DNA
STK31	serine/threonine kinase 31
SWI/SNF	switching defective/sucrose non-fermenting
T. brachyury	T, brachyury homolog
T279	maternal tubulin αTub67C
TAUR	taurine
TBS	tris-buffered saline

TBST	TBS with 0.1% Tween-20
TCTN3	tectonic family member 3
TE	Tris EDTA
TEMED	tetramethylethylenediamine
TIP5	bromodomain adjacent to zinc finger domain, 2A
TJP2	tight junction protein 2
TLR4	toll-like receptor 4
TMEM180	transmembrane protein 180
TRDMT1	tRNA aspartic acid methyltransferase 1
TRIM10	tripartite motif containing 10
Tris-Cl	Tris base HCl
TRUB1	TruB pseudouridine (Psi) synthase family member 1
TRYP	tryptophan
TTLL11	tubulin tyrosine ligase-like family member 11
TXN2	thioredoxin 2
TYMS	thymidylate synthetase
TYR	tyrosine
U2SURP	U2 snRNP-associated SURP domain containing
UAS	upstream activating sequence
UCP2	uncoupling protein 2 (mitochondrial, proton carrier)
UCSC	University of California, Santa Cruz
VANGL1	VANGL planar cell polarity protein 1
VANGL2	VANGL planar cell polarity protein 2
Vldlr	very low density lipoprotein receptor
vvl	ventral veins lacking
WBSCR28	Williams-Beuren syndrome chromosome region 28
WES	whole exome sequencing
WHIM	WSTF, HB1, Itc1p, MBD9
Wnt	wingless type
Wnt5a	wingless-type MMTV integration site family, member 5A
WSTF	bromodomain adjacent to zinc finger domain, 1B
WT	wildtype
X-gal	5-bromo-4-chloro-indolyl-β-D-galactopyranoside
XPD	excision repair cross-complementation group 2
Xpnpep1	X-prolyl aminopeptidase (aminopeptidase P) 1, soluble
Xrcc2	X-ray repair complementing defective repair in Chinese hamster cells 2
Zic2	zinc finger protein of the cerebellum 2
ZO-1	zona occludens 1 (tight junction protein 1)
ZO-2	zona occludens 2 (tight junction protein 2)
β-geo	β-galactosidase gene and neomycin resistance gene fusion
μM	molar concentration in micromoles/litre
γH2Av	phosphorylated histone H2Av variant histone
γH2AX	phosphorylated histone H2AX variant histone

Chapter 1

General Introduction

1: General Introduction

1.1: Introduction

Early embryonic development is characterized by rapid growth accompanying considerable morphogenetic changes, which transforms the single celled zygote into a multicellular organism with defined tissues and organs. One early embryonic process is neurulation, which results in the formation of the neural tube; the earliest rudiment of what will eventually become the brain and spinal cord. This process is highly dynamic, involving complex morphological movements along the entire rostral-caudal axis of the embryo. Neurulation occurs during the first 3-4 weeks of gestation in humans, a time at which the embryo is undergoing rapid growth (reviewed in [1]). Neural tube defects (NTDs) result from a disruption in the process of neurulation, and are one of the most common and devastating birth defects in humans. NTDs have a complex, multifactorial etiology that includes both genetic and environmental contributions. Mice are often used to model NTDs because of all commonly used animal models, mouse neurulation most resembles human neurulation. Mice are an attractive model for studying the complex multifactorial nature of NTDs because mice display a similar NTD etiology as humans, with gene-gene interactions, genetic background influences, and environmental factors all contributing to the development of NTDs in mice. A mouse gestation period is 19 days, with the process of neurulation occurring during embryonic days 7.5 to 10 (E7.5-E10). In this thesis, the mouse containing a homozygous mutation in *Cecr2* served as a genetic model of human NTDs. Research performed using *Cecr2* mutant mice included studies of gene-gene interactions in these mice, and led to studies of the genetic contributions to NTDs in a human cranial NTD cohort and the exploration of the function of the *Cecr2* homologue, *dikar*, in another genetic model *Drosophila melanogaster*.

1.2: Neurulation

Neurulation can be broken up into two distinct processes, primary neurulation and secondary neurulation. Primary neurulation involves the brain and the majority of the spinal cord, excluding only the portion of the neural tube caudal to the midsacral region, which is formed by

secondary neurulation. Primary neurulation can be further broken down into four distinct consecutive events, which are (1) induction of the neural plate, (2) convergent extension, (3) folding, and (4) fusion (Figure 1.2.1).

The first event is the induction of the neural plate, which is achieved by intricate morphogenetic signaling that is conserved in vertebrates (Figure 1.2.1A, reviewed in [2]). The embryonic ectoderm diverges into neural ectoderm (neural plate) and non-neural ectoderm, which is accomplished by fibroblast growth factor (FGF), wingless type (Wnt), and bone morphogenetic protein (BMP) signaling [2]. In order to initiate neural ectoderm formation, induction of the early neural transcription factors, *Sox3* and *Otx2*, is accomplished by FGF signaling in combination with Wnt and BMP antagonists, while the non-neural ectoderm transcription factors, *Ap2*, *Dlx5, Dlx6*, *Gata2, Gata3*, *Foxi1, Foxi3*, and *Msx1*, are induced by Wnt and BMP signaling (Figure 1.2.2) [2]. As neural plate induction continues, the FGF/Wnt/BMP signaling interplay shifts in order to maintain neural and non-neural fates, along with establishing the border pre-neural crest, characterized by *Msx1*, *Pax3*, *Pax7*, *Foxd3*, and *c-Myc* transcription factor expression, and pre-placodal regions, characterized by *Six1*, *Six4*, *Eya1*, and *Eya2* transcription factor expression (Figure 1.2.2) [2]. In mouse, mutations in several genes involved in FGF, Wnt, and BMP signaling as well as the transcription factors they induce, result in NTDs. These include the FGF receptor *Fgfr1* [3], the *Wnt3a* [4] and *Wnt5a* [5,6] secreted signaling factors, the *Bmp2* [7] and *Bmp5/Bmp7* (digenic) [8] secreted signaling factors, as well as the induced transcription factors *Ap2* [9], *Dlx5* [10], *Pax3* [11], and *Msx1/Msx2* (digenic) [12,13].

The next event in neurulation is convergent extension, a process by which cells within the neural plate migrate towards the midline and intercalate in order to narrow and lengthen the neural plate (Figure 1.2.1B) [14]. The non-canonical Wnt signaling planar cell polarity (PCP) pathway largely governs this process, which involves polarizing cells in a planar orientation in order to allow for and direct morphological movements required for convergent extension. A depiction of the core mammalian PCP pathway is provided in Figure 1.2.3. There are mutations in 26 different PCP genes that result in NTDs in mice (listed in http://ntdwiki.wikispaces.com [15]), which exemplifies the size and complexity of the PCP pathway and the important role PCP plays in the process of neurulation.

The third event, which begins while convergent extension is still in progress, is the bending and folding of the neural plate. This event involves the formation of a medial hinge point

(MHP) and two dorso-lateral hinge points (DHLPs) (Figure 1.2.1C). Along with planar cell polarity, another type of cell polarity called apicobasal polarity exists within the neural epithelium. The apical side of the neural epithelial cells initially face outwards, or dorsally, in the neural plate and will eventually face towards the inner lumen of the closed neural tube. The basal side of the neural epithelial cells initially face ventrally and will eventually face away from the inner lumen of the closed neural tube. Within the cell, apical compartments are characterized by the presence of apical proteins, including the PAR3-PAR6-atypical protein kinase C (aPKC) complex, located within the tight and adherens junctions, and basal compartments are defined by basolateral proteins such as the Lethal giant larva (LGL)-Scribble-Discs large (Dlg) complex [16]. During the process of hinge point formation, cells within the MHP and DLHPs must undergo apical constriction, which involves nuclear migration to the basal region of the cell and remodeling of apical epithelial junctions. Cellular apical constriction is important for the bending of the neural plate at these regions. Using the neurulating chick embryo as a model to study apicobasal changes within the MHP, Eom et al. (2011) demonstrated that a blockade of BMP signaling, which can be achieved by the BMP antagonist noggin, is necessary and sufficient for MHP formation [17]. They went on to show that a BMP blockade was capable of inducing both apical constriction and basal nuclear migration [17]. They also established that the cells at the MHP display a partial loss of polarity induced by a BMP blockade. This results in targeting LGL to the apical compartment that leads to endocytosis of apical membrane proteins as shown by immunofluorescent labeling of endosomes containing PAR3 [17]. The DLHPs form under a complex interplay of morphogenetic signals, where Shh signaling from the notochord along with BMP signaling from the ectoderm inhibits DLHP formation. However, these signals are counteracted by the BMP antagonist noggin emanating from the dorsal neural folds, which is sufficient for DLHP formation [18,19]. Taken together, these findings indicate that inhibition of BMP signaling by noggin plays a role in both MHP and DLHP formation. An essential regulator of apical constriction at the neural tube hinge points is *Shroom*, which encodes for an actin binding protein that results in impairment of cranial neurulation when mutated [20]. The medial and dorso-lateral hinge points do not form uniformly along the neural tube. The cranial neural tube contains the MHP and is unique in that the adjacent neural folds are initially biconvex, and then switch to a biconcave shape as neurulation progresses [1]. As for the spinal region, the upper spine

only has the MHP, the midspine has both the MHP and the DLHPs, and the caudal spine only has DLHPs [1].

The final event is fusion and adhesion of the neural fold tips resulting in the formation of a tube separate from the overlying ectoderm (Figure 1.2.1D). Fusion occurs by the formation of cell membrane protrusions resembling lamellipodia and filopodia [21,22], which are regulated by the small GTPases, Rac1 and Cdc42 [23]. Adhesion is assisted by ephrin-A/EphA interactions [24,25]. In humans and mice, fusion does not happen simultaneously along the neural tube. Rather, there are discrete sites of initial contact. In humans, the initial site of closure occurs at the hindbrain-cervical boundary, with a second closure site at the most rostral end of the neural tube [26]. Mice share these two closure sites, which are termed closure 1 and closure 3 respectively. However, mice also appear to have a unique closure site around the fore-mid brain boundary, called closure 2, that forms after the initial closure site (closure 1) but before the rostral closure site (closure 3) (Figure 1.4.1) [26]. In mice, the position of closure 2 is generally located at the fore-mid brain boundary, but is highly variable [27]. For example, closure 2 has a more caudal position in the DBA/2 mouse strain, a more rostral position in the NZW strain, and the complete absence of closure 2 in the SELH strain [28]. Although the SELH mouse strain is susceptible to the cranial NTD exencephaly, the fact that 83% of SELH embryos go on to develop normally demonstrates that closure 2 is dispensable for successful neural tube closure in mice [29]. The presence or absence of closure 2 is still controversial in humans (reviewed in [29]). Originally, it was proposed that humans likely possessed this closure site because of the similarities of NTDs between mice and humans [30]. This hypothesis was corroborated by Nakatsu et al (2000), who observed fifteen human embryos at different stages of neurulation, with a second closure site at the mid-hind brain boundary seen in three embryos [31]. However, studies of neurulating human embryos performed by Sulik et al (1998) and O'Rahilly & Muller (2002) failed to identify this mid-hind brain closure site [26,32]. A review by Copp (2005) attempted to reconcile these contradictory findings by suggesting that the presence or absence (and possibly the location) of closure 2 is variable, much like what has been seen between various mouse strains [29]. He also suggested that closure 2 may be absent in humans due to its dispensability for neural tube closure, as was seen in the SELH mouse strain, as well as the observation that human embryos have proportionately smaller heads than mouse embryos at the time of neurulation [29]. It is important to understand the differences observed between humans and animal models in order to best interpret human neurulation when

utilizing animal models. Although mice and humans may differ in the number of initial neural tube closure sites, the underlying mechanisms of fusion of the neural folds are likely conserved at the molecular level. Once initial contacts have been made, the neural tube proceeds to "zipper-up" away from the initial closure sites and undergo remodeling events that lead to separation of the neural epithelium from the overlying non-neural ectoderm. The last regions of the neural tube to close are called neuropores, which are found approximately midway between the initial closure sites as well as at the most caudal portion of the neural tube. The molecular mechanism for separation of the neural tube from the overlying epithelium remains unclear (reviewed in [33]); however, it has been shown that apoptosis is not required for these remodeling events [34]. Disruptions in any of the four major neurulation events (neural plate induction, convergent extension, folding, and fusion) result in abnormal neurulation and ultimately in open neural tube defects, which are discussed in more detail in section 1.3.

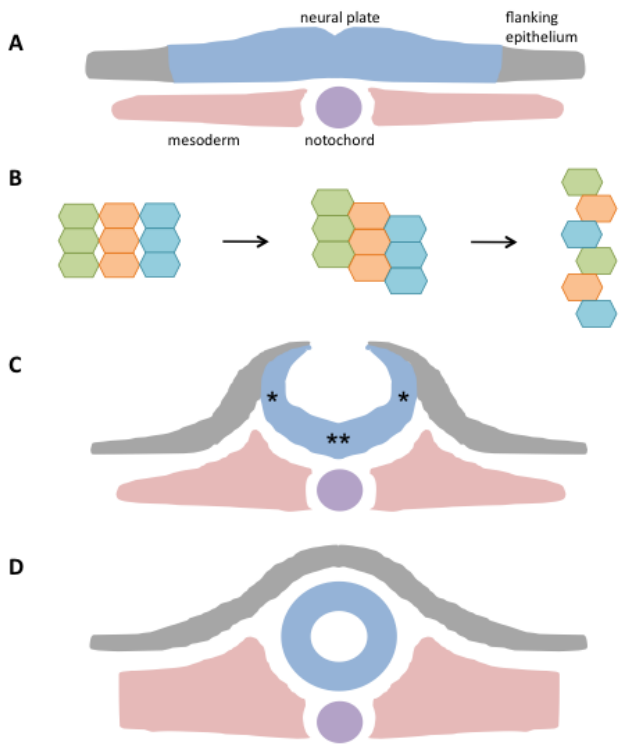


Figure 1.2.1: Neurulation. The process of neurulation can be broken down into four events. The first event is induction of the neural plate (**A**), followed by convergent extension as illustrated by individual cell movement towards the midline (**B**). Thirdly, folding occurs by forming the MHP** and the DLHPs* (**C**). Lastly, fusion of the neural fold tips results in a closed neural tube separate from the overlying epithelium (**D**).

Secondary neurulation is the process by which the most caudal portion of the neural tube forms [35,36]. A population of mesenchymal tail-bud cells coalesces and undergoes a mesenchymal-to-epithelial transition mediated by Rac1 and Cdc42. This forms an epithelial rod, which undergoes cavitation in order to form a lumen continuous with the primary neural tube [37]. When

this process is disrupted, the resulting neural tube defects are closed (covered with skin) since secondary neurulation occurs under an already formed non-neural ectoderm. These types of NTDs often include a tethered spinal cord, which involves attachments between the spinal cord and surrounding tissues within the spinal column, resulting in limited movement [38]. As studies contained herein focus on the genetics of primary neurulation, secondary neurulation will not be discussed any further.

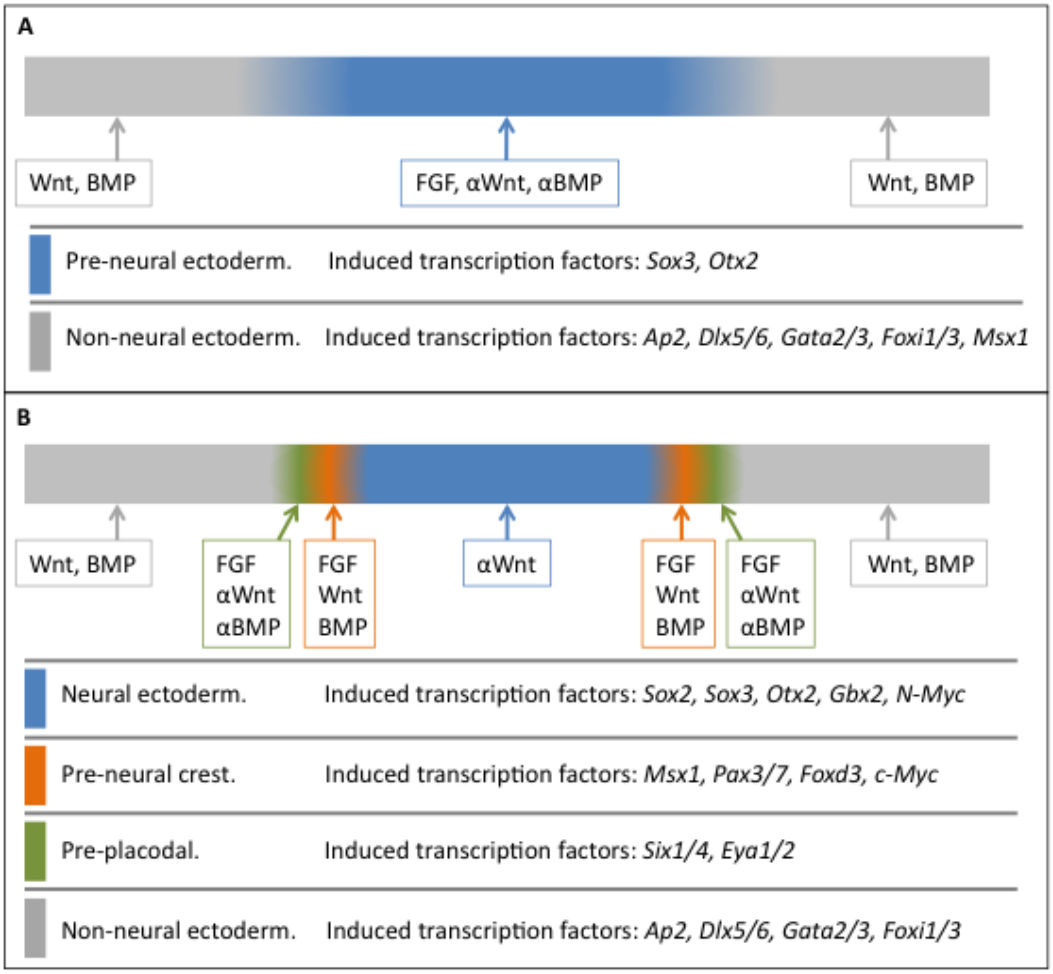


Figure 1.2.2: Neural plate induction. In the earliest stage of neural plate induction, inducing signals FGF, Wnt antagonists (αWnt) and BMP antagonists (αBMP) trigger expression of pre-neural ectoderm transcription factors, whereas Wnt and BMP signaling induce expression of non-neural transcription factors (**A**). FGF, Wnt, and BMP signaling changes within the neural ectoderm region in order to establish border pre-neural crest and pre-placodal regions, with induced transcription factors for each region displayed (**B**). All information for this image was acquired from Groves and Labonne (2014) [2].

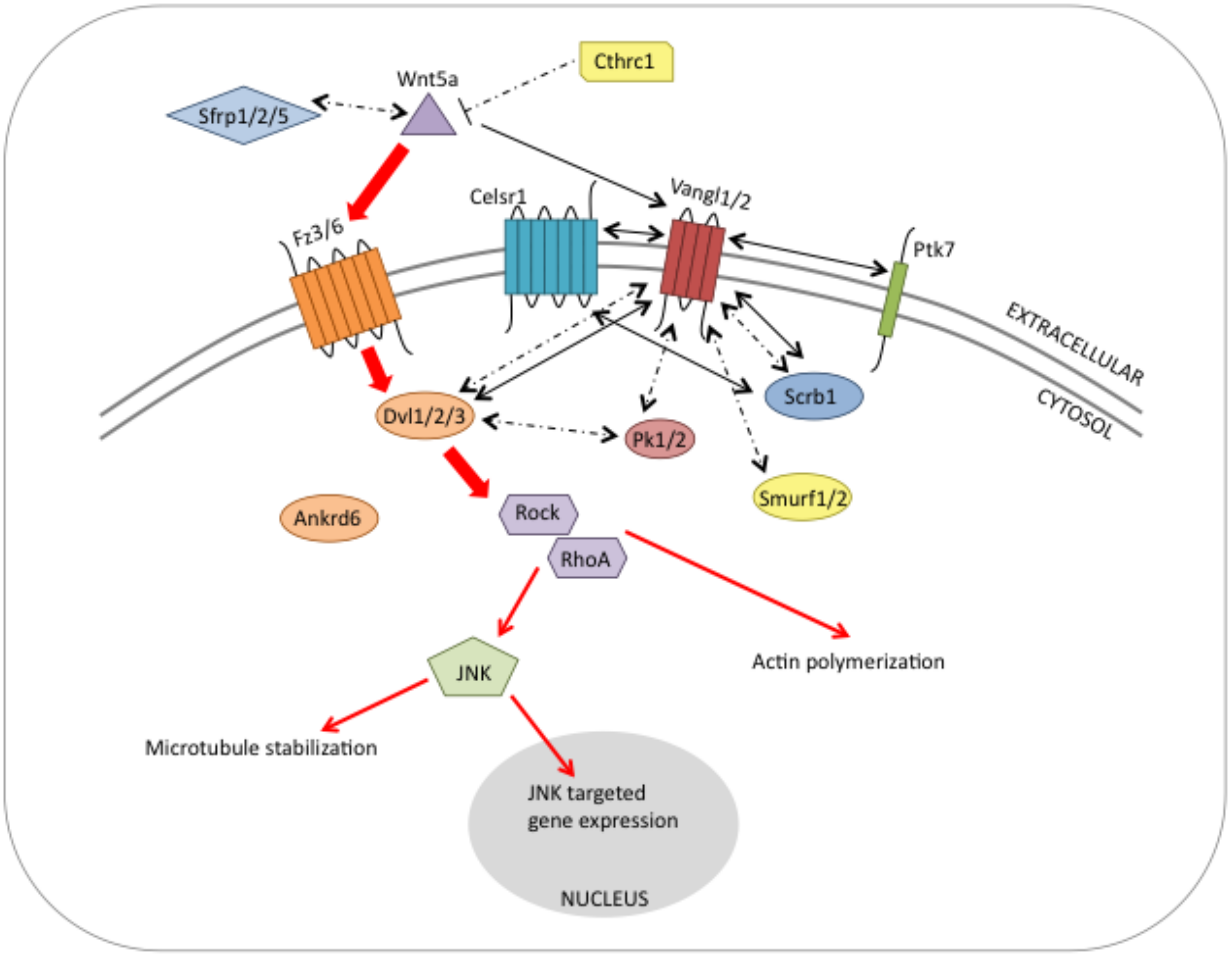


Figure 1.2.3: Planar cell polarity (PCP) pathway in mammals. Red arrows signify the PCP signaling pathway. Known physical interactions are denoted with solid black arrows and known genetic interactions are denoted with dashed black arrows. The *Drosophila* homologue of Ankrd6, Diego, is known to interact with Frizzled, Disheveled, and Vang, but this has not been studied in mammals. Image adapted from Greene et al. 2009 [39].

1.3: Open neural tube defects

Open neural tube defects (NTDs) occur when the process of primary neurulation is disrupted and results in exposure and degeneration of sensitive neural tissue *in utero*. Open NTDs are grouped into three major classifications: spina bifida, anencephaly (or exencephaly in mouse), and craniorachischisis. Spina bifida affects approximately 1-10/1000 human births worldwide [40-

[42]. Spina bifida occurs when the posterior neuropore fails to close, and although spina bifida is the only open NTD compatible with life, it is a debilitating condition that greatly affects quality of life. Anencephaly has a similar incidence as spina bifida, affecting approximately 1-10/1000 human births worldwide [40-42], and is a result of failure of cranial neuropore closure. Craniorachischisis is the most severe form of open NTD, with an incidence ranging from 0.1-11/10,000 births in various populations [43,44]. It is characterized by an entirely open neural tube due to failure of the initial closure site to fuse. In humans, it is well established that NTDs have a complex, multifactorial etiology that likely involves a combination of multiple genetic, environmental, and epigenetic factors that interact with each other.

Neural tube defects have a strong genetic component in humans as evidenced by the 3-5% risk of having an NTD affected pregnancy if there is a first-degree relative with an NTD, or a 1-2% risk if there is a second-degree relative with an NTD [42]. NTD studies of twins are more subjective due to comparisons between "like-sex vs. unlike-sex" twins instead of the more conventional monozygotic vs. dizygotic twins, as well as small sample sizes and limited data. These twin studies indicated rates of NTD concordance in all twins that range from 3.7% to 18% (reviewed in [45]). Also, some ethnic groups have higher risk than others. NTD phenotypes are associated with several syndromes, some of which are monogenic and others that involve large chromosomal abnormalities [42]. A whole genome linkage study utilizing families with high incidence of NTDs identified regions on chromosomes 7 and 10 that are associated with the NTD phenotype, which means that these regions likely contain NTD susceptibility genes [46]. Since 1994, more than 130 studies looking for associations between known single nucleotide variants (SNVs) in specific candidate genes and NTDs were published [42]. To date, no genome-wide association studies (GWAS) have been published on human NTD cohorts. Most of the patient samples utilized for these association studies were obtained from patients diagnosed with spina bifida. These samples were easier to acquire due to spina bifida being compatible with life, although some other NTD types were included in some of the studies. These studies were successful in identifying 31 candidate genes in several biological processes and pathways associated with human NTDs, including but not limited to the folate metabolism pathway, glucose metabolism, DNA repair, and apoptosis (Table 1.3.1) [42]. However, replication studies are recommended in order to further validate many of these reported associations, as it is possible that associations may differ between populations or findings may be spurious due to single

studies with small sample sizes. For example, a meta-analysis of 25 studies with conflicting results regarding the association between the *MTHFR C677T* polymorphism and NTDs found that there was indeed an association; however, this association only appears to be in Asian, Caucasian, and mixed populations but not in African populations [47]. It is possible that the associated SNVs are not directly responsible for the disease, but may be in linkage disequilibrium with the disease-causing sequence variant, which may or may not be located within the gene of interest. Functional analyses (detailed in section 1.5) are required to determine if the associated variants affect protein function, which has been done for one associated SNV in *MTHFR.* This *MTHFR* variant alters an alanine to a valine at position 222 and results in an ~30% reduction in enzymatic activity [48]. Published studies have failed to identify associations between NTDs and SNVs within an additional 89 genes involved in various pathways, including folate metabolism, methionine cycle, methylation, glucose homeostasis, oxidative stress/apoptosis, retinoid metabolism, DNA repair, regulation of gene expression, cell recognition/migration, and planar cell polarity (reviewed in [42]). These 89 genes were also good candidates based on what was known about the genetic pathway they functioned in and/or the fact that they were identified as NTD genes in mouse models. The association studies that did not identify variants within these 89 genes were included in the same or similar studies that identified the associated variants in the 31 genes described above (reviewed in [42]). This may be due to the fact that the study designs only looked for associations of certain SNVs, which themselves may be functionally benign and not linked to the causal mutation. Association study design neglects all other known SNVs and fails to detect novel SNVs or small insertions/deletions.

Table 1.3.1: Genes containing SNVs associated with NTDs in humans (reviewed in [42]).

Gene name	**Biological process**	**Reference(s)**
SLC19A1	Folate/vitamin B12 transport	De Marco et al. (2003) [49] Pei et al. (2009) [50]
CUBN	Folate/vitamin B12 transport	Franke et al. (2009) [51]
MTHFR	Folate metabolism	Shaw et al. (2009) [52] Martinez et al. (2009) [53]
TYMS	Folate metabolism	Shaw et al. (2009) [52] Martinez et al. (2009) [53]
DHFR	Folate metabolism	Martinez et al. (2009) [53]
MTHFD1	Folate metabolism	Shaw et al. (2009) [52]
MTHFD2	Folate metabolism	Shaw et al. (2009) [52]
MTHFD1L	Folate metabolism	Parle-McDermott et al. (2009) [54]

Gene name	Biological process	Reference(s)
MTRR	Folate metabolism	Shaw et al. (2009) [52]
CHKA	Methylation	Enaw et al. (2006) [55]
NAT1	Methylation	Jensen et al. (2005) [56]
ICMT	Methylation	Franke et al. (2009) [51]
PCMT1	Methylation	Zhu et al. (2006) [57]
PRMT1	Methylation	Franke et al. (2009) [51]
SARDH	Methylation	Franke et al. (2009) [51]
TRDMT1	Methylation	Franke et al. (2009) [51]
CBS	Transsulfuration	Shaw et al. (2009) [52] Martinez et al. (2009) [53]
CTH	Transsulfuration	Franke et al. (2009) [51]
GLUT1	Glucose transport	Davidson et al. (2008) [58]
HK1	Glucose transport	Davidson et al. (2008) [58]
LEPR	Glucose transport	Davidson et al. (2008) [58]
UCP2	Energy metabolism (glycolysis)	Davidson et al. (2008) [58]
TXN2	Apoptosis/oxidative stress	Wen et al. (2009) [59]
*APEX1**	DNA repair	Olshan et al. (2005) [60]
XPD	DNA repair	Olshan et al. (2005) [60]
PAX1	Transcription	Volcik et al. (2002) [61]
PAX7	Transcription	Volcik et al. (2002) [61]
PAX8	Transcription	Volcik et al. (2002) [61]
PAX3	Transcription	Lu et al. (2007) [62]
T. brachyury	Transcription	Shields et al. (2000) [63]
NCAM1	Cell communication/migration	Deak et al. (2005) [64]

*Associated SNV demonstrated a protective affect for spina bifida.

Hypotheses regarding the frequencies of SNVs contributing to common, multifactorial diseases are currently in flux. There has been a shift away from the concept that common variants (allele frequency >1%) are significant contributors to common disease risk and towards rare variants (allele frequency <1%) with high penetrance as the major culprits for several common diseases (reviewed in [65]). Unlike patients with other common diseases, such as diabetes and heart disease, patients with NTDs most often do not reproduce, meaning that the genetic components contributing to NTDs are under a higher selective pressure. This supports the hypothesis that rare variants with large effect are likely contributing to NTD etiology. Therefore, recent NTD studies have employed a more unbiased approach by utilizing sequence technologies on candidate genes or whole exomes. Sequencing studies can be used to confirm associations identified in previous studies, as well as complement association studies by capturing additional variants, including rare or novel variants. Sanger sequencing is a well-established methodology that can be employed;

however, this method can become quite labor intensive and expensive if used to sequence large or several candidate genes in patient cohorts. High-throughput next generation sequencing (NGS) technologies are now being utilized to overcome the costs and labor of Sanger sequencing. Custom NGS technologies can be employed to capture entire coding and known regulatory regions of a selection of candidate genes, whole exome sequencing (WES) can be performed to capture all known coding and non-coding exons of the entire exome (all known genes within the genome), and whole genome sequencing would capture almost the entire genome. Many NTD studies involving sequencing candidate genes in a patient cohort have been performed since 2007, and have been successful in identifying rare or novel variants (Table 1.3.2). Focus has been directed towards sequencing planar cell polarity genes, which was successful in identifying rare variants of interest in *CELSR1*, *FZD6*, *PRICKLE1*, *VANGL1*, *VANGL2*, *FUZ*, *SCRIB*, *DACT1*, *DVL2*, and *LRP6*. Recently, whole exome sequencing of affected child and unaffected parent trios by Lemay et al. (2015) identified *de novo* (not in either parent of the patient) loss-of-function mutations in 5 genes, 2 of which are known NTD genes (*SHROOM3* and *PAX3*), as well as several *de novo* missense mutations [66]. This study underscored the potential role of severe *de novo* mutations in NTD etiology as well as demonstrated the utility of sequencing approaches for unraveling the genetic architecture of NTDs. Unlike in association studies, where SNVs may simply be linked to the causal mutation, sequencing studies aim to directly identify causal mutations. Therefore, variants identified in sequencing studies are often followed up with bioinformatic analyses (*in silico* predictions of deleteriousness such as SIFT and Polyphen2) and functional analyses. For example, yeast two-hybrid assays identified impaired protein-protein interactions due to protein-coding variants in *VANGL1* [67] and *VANGL2* [68], cell culture assays identified cellular mislocalization or functional impairments due to protein-coding variants identified in *CELSR1* [69], *SCRIB* [69], *FUZ* [70], *DACT1* [71], and *LRP6* [68,72], and *in vivo* zebrafish assays established that protein-coding variants in *PRICKLE1* are capable of antagonizing the wildtype allele of *PRICKLE1* [73]. Interestingly, all of the rare variants identified in the human NTD cases in these studies were heterozygous, meaning that the effect of these variants would have had to act in a dominant fashion.

Table 1.3.2: Genes containing rare, novel, or *de novo* variants that alter the amino acid sequence (nonsense, missense, frameshift) identified in human sequencing studies.

Gene name	Biological process	Reference(s)
CELSR1	Planar cell polarity	Robinson et al. (2012) [69] Allache et al. (2012) [74]
FZD6	Planar cell polarity	De Marco et al. (2012) [75]
PRICKLE1	Planar cell polarity	Besoi et al. (2011) [73]
VANGL1	Planar cell polarity	Kibar et al. (2007) [67] Kibar et al. (2009) [76]
VANGL2	Planar cell polarity	Lei et al. (2010) [68] Kibar et al. (2011) [77]
FUZ	Planar cell polarity	Seo et al. (2011) [70]
SCRIB	Planar cell polarity	Robinson et al. (2012) [69]
DACT1	Planar cell polarity	Shi et al. (2012) [71]
DVL2	Planar cell polarity	De Marco et al. (2013) [78]
LRP6	Planar cell polarity	Allache et al. (2014) [72] Lei et al. (2015) [79]
SHROOM3	Cell morphology, cytoskeleton	Lemay et al. (2015) [66]
PAX3	Transcription	Lemay et al. (2015) [66]
MFAP1	Pre-mRNA processing	Lemay et al. (2015) [66]
DDX3X	RNA helicase	Lemay et al. (2015) [66]
WBSCR28	Unknown	Lemay et al. (2015) [66]
GRHL3	Transcription	Lemay et al. (2015) [66]
LRFN1	Unknown	Lemay et al. (2015) [66]
TTLL11	Microtubule dynamics	Lemay et al. (2015) [66]
PTPRS	Cell signaling	Lemay et al. (2015) [66]
HIP1R	Endocytosis	Lemay et al. (2015) [66]
C17orf80	Unknown	Lemay et al. (2015) [66]
HUWE1	Ubiquitination/apoptosis	Lemay et al. (2015) [66]
INF2	Actin dynamics	Lemay et al. (2015) [66]
RP1L1	Microtubule dynamics	Lemay et al. (2015) [66]
SLC2A4RG	Transcription	Lemay et al. (2015) [66]
SLC22A13	Small molecule transport	Lemay et al. (2015) [66]
ITGAX	Immunity/complement	Lemay et al. (2015) [66]
TRIM10	Unknown	Lemay et al. (2015) [66]
STK31	Kinase activity	Lemay et al. (2015) [66]
FOXF2	Transcription	Lemay et al. (2015) [66]
ATAD2	Chaperone/chromatin remodeling	Lemay et al. (2015) [66]
PTBP1	Pre-mRNA processing, RNA metabolism, RNA transport	Lemay et al. (2015) [66]
GIGYF1	Unknown	Lemay et al. (2015) [66]
IP6K1	Inositol metabolism	Lemay et al. (2015) [66]
SLAIN2	Microtubule dynamics	Lemay et al. (2015) [66]
DENND6A	Endocytosis, cell junctions	Lemay et al. (2015) [66]

Gene name	Biological process	Reference(s)
FAM129B	Apoptosis	Lemay et al. (2015) [66]
HNRNPL	Pre-mRNA processing	Lemay et al. (2015) [66]
U2SURP	Unknown	Lemay et al. (2015) [66]
ANKRD32	Unknown	Lemay et al. (2015) [66]
PREPL	Unknown	Lemay et al. (2015) [66]
BAG1	Apoptosis	Lemay et al. (2015) [66]
LCMT2	Methyltransferase activity	Lemay et al. (2015) [66]
PHF14	Proliferation	Lemay et al. (2015) [66]
TLR4	Immunity	Lemay et al. (2015) [66]

Sequencing technologies are able to capture SNVs as well as small insertions and deletions. However, due to size limitations of sequence reads, these methodologies are not very robust at capturing large duplications or deletions, herein referred to as copy number variants (CNVs), that involve one or more genes. Chen et al. (2013) utilized a different methodology, comparative genome hybridization, in order to identify novel CNVs greater than 30 kb in size in NTD patients [80]. This study identified an enrichment of novel CNVs in cases compared to controls, with 55 novel CNVs (40 genic CNVs) in cases and 26 novel CNVs (19 genic CNVs) in controls. The authors also found that cilia genes were particularly affected by novel CNVs in cases (41 cilia genes) compared to controls (12 cilia genes). This study further demonstrated the highly complex genetic etiology of NTDs.

Environmental influences also play an important role in the occurrence of NTDs (reviewed in [15]). Reduced maternal intake of folic acid has long been established as an NTD risk factor [81-84]. Supplementation with 4 mg/day of folic acid has been shown to reduce the frequency of NTDs by 70% [83]; however, lower doses of folic acid (0.4 mg/day) only reduce risk by 36% [85], demonstrating that there is a dose-related effect. These findings led to mandatory folic acid fortification of wheat flour in Canada beginning in 1998 in order to ensure all women of childbearing age were receiving a minimum of 400 μg/day [86]. This action led to a 46% decrease in the occurrence of NTDs [87]. Folic acid fortification was also implemented in the United States in 1998 [88], resulting in a 20% reduction in NTD occurrence [89]. Similar fortification campaigns have been implemented in over 75 countries, reducing the incidence of NTDs by an estimated 15-25% worldwide [90]. As most fortified foods are wheat products, the differences in percent reduction of NTDs between countries may be due to cultural differences in diet. For example, some of the highest rates of NTDs in the United States are amongst Hispanic groups who

preferentially use corn masa flour in lieu of wheat flour in their diets; therefore, there has been a recent push to fortify corn masa flour in the United States [91]. Other significant environmental risk factors for human NTDs are maternal diabetes [92], maternal obesity [93], and maternal hyperthermia [94,95]. As only a fraction of pregnancies affected by folate deficiency, diabetes, obesity, and hyperthermia result in NTDs, it is most likely that these environmental effects do not act alone, but rather interact with other environmental factors and/or pre-existing genetic variants in the embryo. Epidemiological studies have shown that folic acid fortification can alleviate the risk of NTDs in diabetic pregnancies [96,97], thereby exemplifying an environment-environment interaction. Candidate genes involved in biological processes affected by folic acid and diabetes, such as folate and glucose metabolism, have directed research to look for genetic associations between these candidate genes and NTDs. Variants within 9 genes involved in folate transport or metabolism, as well as variants within 4 genes involved in glucose transport and glycolysis, have been associated with NTDs in humans (Table 1.3.1). A primary example of a gene-environment interaction is that the risk associated with the folate metabolism gene *MTHFR C677T* variant has been shown to be alleviated with folic acid supplementation [98]. The authors looked at 98 mothers that were homozygous for the 'T' allele and in which serum folate levels were known, and found that the risk of anencephaly affected pregnancies reduced by 18% for each 1 ng/ml increment in serum folate [98]. Studies in mice have yet to show that folate deficiency is sufficient to cause NTDs in the absence of a genetic mutation; however, folate deficiency can exacerbate NTD frequency in some mice already genetically predisposed to NTDs (reviewed in [15]).

Environmental effects can also act through alterations of the epigenome, which can lead to global changes in gene expression that can affect neurulation. The epigenome refers to all modifications to the genome that can be inherited without altering the genetic code. Epigenetic mechanisms can be grouped into three major types: DNA methylation, histone modification, and nucleosome positioning. DNA methylation is an epigenetic modification that occurs at CpG dinucleotides through the action of DNA methyltransferases. Although DNA methylation of the vast majority of the genome is accomplished and stably maintained before embryonic implantation, which occurs prior to neurulation, regions enriched in CpG dinucleotides (CpG islands) remain unmethylated and are the major targets of transcriptional regulation via DNA methylation [99]. Folate metabolism is important for the biosynthesis of S-adenosylmethionine (SAM), the universal methyl donor that is used for DNA and protein methylation [100]. As folate

deficiency has been shown to alter methylation patterns leading to changes in gene expression in wild-type mice [101], it is suggested that NTD prevention by folic acid fortification may partially act through preventing erroneous DNA methylation [15]. DNA is packaged around nucleosomes, which are protein complexes consisting of 8 histone proteins (detailed in section 1.6). Histone modification involves the covalent attachment of various chemical groups, including but not limited to methyl groups, acetyl groups, and ubiquitin, to histone tails. It is theorized that these modifications affect nucleosome interactions, which alters DNA packaging and ultimately DNA accessibility to transcription factors and other DNA/chromatin binding proteins [102]. A study in mice revealed that there are alterations in histone methylation and acetylation in embryonic neural stem cells isolated from chemically induced diabetic pregnancies [103]. Another study utilizing chemically induced diabetic mouse models demonstrated that gene expression is more variable in NTD affected embryos compared to unaffected littermates [104]. These two studies provide evidence that environmentally induced alterations in histone modification and/or gene expression may play a role in NTDs. Nucleosome repositioning is another epigenetic mechanism that can regulate gene expression. The movement or ejection of nucleosomes is achieved by ATP-dependent chromatin remodelers (detailed in section 1.6), which recognize histone modifications as binding sites. Therefore, environmental insults that alter histone modifications can result in inappropriate nucleosome positioning, and ultimately dysregulated gene expression. Another type of epigenetic regulation involves microRNAs, which do not encode for proteins but rather act to regulate the translation of targeted mRNAs into their protein products. Genetic mutation that alters the expression or binding capabilities of microRNAs to their target mRNAs could have implications regarding translation of one or more mRNAs derived from NTD genes. Gu et al. (2012) demonstrated that there is an alteration in microRNA expression profiles between normal pregnancies and NTD affected pregnancies [105]. These findings suggest that microRNAs are another epigenetic mechanism that plays a role in NTD etiology.

In humans, cranial NTDs are twice as common in females compared to males, and this phenomenon was also seen in all mouse models of exencephaly where the penetrance of exencephaly was analyzed between sexes (reviewed in [106]). Although it is still unknown what the underlying cause of this sex difference is, it is unlikely due to differences in gonadal hormones. Neurulation completes ~1 day before the gonads begin to differentiate and release hormones in mice and ~16 days before gonadal differentiation in humans [106]. Studies in mice have also

indicated that this sex difference is not due to delayed development of female embryos compared to males [107], or early male death [28,35,108]. Furthermore, studies in mice mutant for the autosomal gene, *Trp3*, demonstrated no protective effect conferred by the Y chromosome; however, there was an association between exencephaly and the presence of two X chromosomes [109]. These findings led Juriloff and Harris (2012) to propose an elegant hypothesis regarding the underlying mechanism of this sex difference [106]. They postulated that the requirement of an entire X chromosome to be inactivated via epigenetic mechanisms after each cell division could titrate out important epigenetic factors required for neurulation. Therefore the presence of an X chromosome requiring inactivation is suggested to be a susceptibility factor that can exacerbate existing deficiencies in neurulation.

Human primary neurulation is very dynamic and complex, with several processes involving many molecular pathways and hundreds of genes leading to the development of NTDs when disrupted. As research progresses, we are learning that NTDs result from an elaborate interplay of genetic, environmental, and epigenetic insults. This leads to the theory that there are many underlying causes of NTDs, where the causes of some NTD cases may not share any similarities with the causes of other cases.

1.4: Mice as models of neurulation and neural tube defects

Mice are commonly used as genetic models of NTDs, with over 300 NTD mouse models established to date (http://ntdwiki.wikispaces.com [15]). Mice are unique from all other commonly used NTD genetic models in that embryonic development occurs *in utero*, and therefore mouse early development, environment, and maternal contributions are most similar to humans in comparison to other models. The process of neurulation and the manifestation of NTD phenotypes are very similar in mice and humans (Figure 1.4.1). Most of the NTD mouse models contain known gene mutations that confer a significant NTD risk, with over 300 NTD genes identified in mice. A comprehensive and growing list of these mouse NTD genes, along with mouse models with inheritable NTDs where the gene has not been identified, is publicly available in the online repository http://ntdwiki.wikispaces.com [15]. Genes involved in almost every important cellular function have been implicated in NTDs. Mutations have been identified in several genes required for cell signaling, such as SHH, BMP, and FGF signaling, which play a

role in neural plate induction as well as hinge point formation. Many additional NTD genes are involved in primary cilia formation and/or function, where the primary cilium is essential for correct SHH signaling and planar cell polarity. NTD causing mutations in several genes within the PCP pathway, integral to convergent extension, have been identified in mice, as well as a number of genes important for cell adhesion, which is required for neural fold tip fusion and neuroepithelial integrity. There are also genetic mutations affecting cell cycle, proliferation, survival, cellular transport, metabolism, and the cytoskeleton that result in NTDs in mice, as well as within genes that play more global roles within the cell such as transcription, translation, and epigenetic regulation. Several mouse models of NTDs demonstrate gene-gene interactions, with some models requiring more than one mutation in order for NTDs to manifest, and some models showing changes in penetrance in the presence of genetic modifiers or on various genetic backgrounds (discussed in detail in Chapter 3).

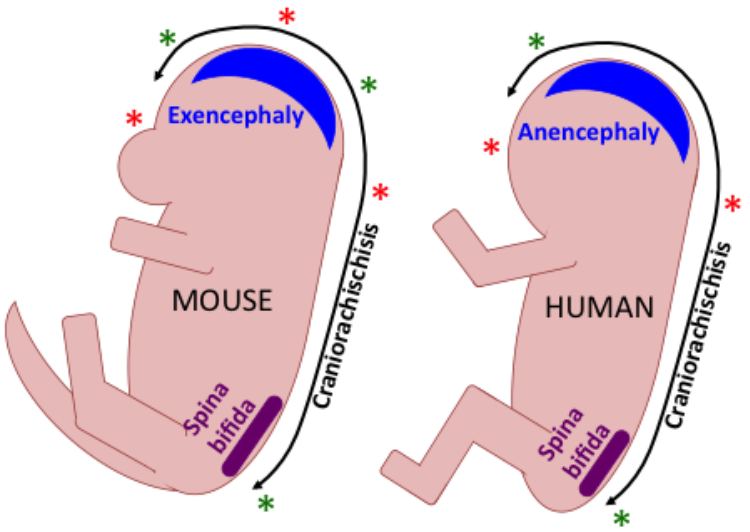


Figure 1.4.1: Mouse as a genetic model for neural tube defects. Of all commonly used models of neurulation, mouse most closely resembles human neurulation. The major difference lies in closure site 2 at the mid brain region, which is variable in mice and may or may not be present in humans. Red asterisks denote closure initiation sites, green asterisks denote neuropores.

NTDs in mice are also susceptible to environmental influences. Studies in several mouse genetic models of NTDs have shown that folic acid supplementation is capable of reducing risk for 10 mutant alleles tested, increasing risk for three mutant alleles tested, or having no effect in 11 mutant alleles tested (reviewed in [15]), exemplifying gene-environment interactions. It is

interesting to note that different mutant alleles for the same gene can respond differently to folic acid supplementation, where mice mutant for the gain-of-function *Lrp6*^{Cd} allele show lower rates of NTDs when supplemented with folic acid [110,111], whereas folic acid supplementation had a detrimental effect on mice mutant for the *Lrp6* null allele [112]. In another case, mice heterozygous for the *Grhl2*^{m1Nisw} allele had a detrimental response to folic acid supplementation but mice homozygous for this same allele did not show any response at all [113]. Folic acid supplementation in diabetic mice reduced the rate of NTDs from 28.4% to 6.0% [114], demonstrating an environment-environment interaction and corroborating the human epidemiological studies mentioned in section 1.3. Too much folic acid was also shown to be detrimental in mouse diabetic pregnancies, as high doses resulted in premature differentiation of neural progenitors [115]. Certain mouse NTD models are also affected by other environmental influences including maternal hyperthermia, inositol, fumonisin, and ethanol (reviewed in [15,116]). These findings support the multifactorial nature of NTDs in mice, which is similar to what is seen in humans, and also demonstrate the importance of taking environmental influences into consideration when performing genetic studies in mice.

1.5: Other models of neurulation and neural tube defects

Although *Drosophila melanogaster* (fruitflies) are invertebrates, and therefore do not possess a neural tube, many of the basic molecular pathways important for neurulation were originally characterized in *Drosophila*. The *Hedgehog (Hh)* gene was first identified in *Drosophila* by Nusslein-Volhard and Wieschaus (1980) [117], who later went on to win the 1995 Nobel Prize in Physiology or Medicine for their groundbreaking work in the genetics of *Drosophila* embryonic development. Since *Hh* was discovered, the majority of the Hh signaling pathway members were identified and characterized in *Drosophila* (reviewed in [118]). The genes and mechanisms involved in the PCP pathway were also discovered and characterized in *Drosophila* (reviewed in [119], detailed in Chapter 4). Although gross morphological processes in *Drosophila* are not easily extrapolated to humans, *Drosophila* has proven to be invaluable for elucidating the basic molecular function of genes and genetic pathways involved in embryonic development and important for vertebrate neurulation.

Neurulation is mechanistically different in non-mammalian vertebrates. In chicken, the sequence of closure events is different, with the initial contact site occurring at the midbrain region rather than at the hindbrain-cervical boundary as seen in mice and humans [120]. This is then proceeded by a second closure site characterized by multiple sites of contact between the neural folds, which occurs at the hindbrain-cervical boundary [120]. The neuropore formed between these two sites of closure quickly proceeds to zipper-up [120]. The remaining cranial neuropore then closes in a zipper-like fashion, and finally the caudal neuropore closes by forming a series of temporary button-like contact points between the apposing neural folds and then closing [120]. In *Xenopus*, closure of the neural tube occurs simultaneously along the entire rostral-caudal axis [121]. Zebrafish neurulation is a process that involves the formation of a neural keel that develops into a solid neural rod, which finally undergoes cavitation [122]. This process is not to be confused with mammalian secondary neurulation as the zebrafish neural keel is derived from a neural plate, which is more similar to mammalian primary neurulation, and not from mesenchymal tail-bud cells as is seen in mammalian secondary neurulation. Although less similar to human neurulation, these models are attractive for studying the dynamics of neurulation as embryonic development does not occur inside the mother and therefore allows for better visualization and manipulation. Also, many of the important molecular pathways and processes, such as PCP and convergent extension, are highly conserved.

As mentioned in section 1.3, many of the genetic variants identified in human NTD cases require functional studies in order to determine if these mutations affect protein function. Such studies can be performed *in vivo*, which is a common methodology performed in zebrafish [123]. Zebrafish *in vivo* studies usually involve the knockdown of the zebrafish homologue, analysis of the resulting phenotype, rescue with a wildtype copy of the human homologue, and comparison to rescue attempts with a human homologue containing the variant of interest. Functional studies can also be performed in cell culture, which typically involves the transfection of the variant containing human homologue into cell culture and comparison to cells transfected with the wildtype human homologue. Cell culture assays are usually designed to specifically suit the gene being tested, and can include comparing functional outputs such as enzymatic efficiencies, protein-binding efficiencies, or appropriate cellular localization capabilities.

1.6: ATP-dependent chromatin remodeling

The gene *Cecr2*, which functions in ATP-dependent chromatin remodeling, is central to this thesis. Therefore, the purpose of this section is to describe ATP-dependent chromatin remodeling and its importance in the process of neurulation.

DNA, the largest macromolecule in the cell, needs to overcome the major topological challenge of being packaged into the nucleus of the cell. This is accomplished by the formation of chromatin, which consists of DNA in complex with nucleosomes, where nucleosomes are protein complexes consisting of histone proteins or variant histone proteins. Canonical nucleosomes are octameric protein complexes composed of two copies of each of the H2A, H2B, H3, and H4 histone proteins. The histone octamer wrapped with 147 bp of DNA is referred to as a nucleosome. When the histone H1 or H5 linker protein is associated with the nucleosome, which most often occurs within more tightly packaged DNA, this structure is referred to as the chromatosome [124]. Histone variant proteins can replace the canonical histone proteins within the octamer resulting in more specialized nucleosomes that can distinguish some parts of the chromatin from others. For example, transcription start sites are flanked by nucleosomes that contain the histone variant H2A.Z in lieu of H2A [125]. Another example is H2AX, which is phosphorylated at serine 139 to form γH2AX following DNA damage [126] in order to recruit DNA repair machinery to the sites of DNA double-strand breaks [127].

The presence of chromatin can interfere with many processes involved in DNA metabolism, including DNA replication, transcription, accessibility to DNA binding proteins, repair, and recombination [125]. To overcome this, protein complexes called ATP-dependent chromatin remodelers are capable of relaxing DNA wrapped around histones to make the DNA more accessible, re-positioning of nucleosomes, ejection of nucleosomes, or exchange of histones within nucleosomes for histone variants [125]. The energy required to perform these tasks is obtained from the hydrolysis of ATP.

There are four families of ATP-dependent chromatin remodelers (reviewed in [125]). All four families share a common ATPase domain that is comprised of two parts, the Asp-Glu-xx box helicase (DExx) domain and the helicase superfamily c-terminal (HELICc) domain, separated by an insertion of a varying number of amino acids. Both of these domains provide the protein with helicase activity, which is the ability to unwind DNA or RNA. The SWI/SNF

(switching defective/sucrose non-fermenting) family also possesses a C-terminal helicase-SANT-associated (HSA) domain and an N-terminal bromodomain. The HSA domain is responsible for binding nuclear actin-related proteins to regulate ATP-dependent chromatin remodelers [128], and the bromodomain aids in targeting the chromatin remodeling complex to chromatin by recognizing acetylated lysine residues on histone tails [129]. The INO80 (inositol requiring 80) family is similar to SWI/SNF in that it contains a C-terminal HSA domain, but is unique in that it lacks a bromodomain and the insertion of amino acids between DExx and HELICc is much larger than in the other three families. The CHD (chromodomain, helicase, DNA binding) family is characterized by having a Tandem chromo domain in lieu of a bromodomain, and this is the only domain other than the ATPase domain. Lastly, the ISWI (imitation switch) family possesses only a SANT (Swi3, Ada2, N-cor, and TFIIB) domain and a SLIDE (SANT-like ISWI) domain in addition to the ATPase domain. The SANT domain is implicated in interacting with histones [130] and the SLIDE domain recognizes and binds to the DNA component of the nucleosome [131]. There are several chromatin remodeling complexes within each family, and many of these chromatin remodeling complexes are highly specialized and involved in unique biological processes.

Certain remodeling complexes play an important role in neurulation, demonstrated by the fact that genetic mutations in various subunits of ATP-dependent chromatin remodeling complexes have been shown to result in neural tube defects. The BAF complex, a member of the SWI/SNF family of remodelers, is one such remodeler involved in neurulation. Heterozygous loss-of-function of the ATPase domain containing subunit, *Brg1 (Smarca4)*, predisposes mice to the development of exencephaly, whereas homozygous mutation in *Brg1* results in lethality around the time of implantation [132]. Also, mice with a heterozygous null mutation affecting the BAF155 protein, another subunit of the same BAF complex, results in exencephaly with reduced penetrance [133]. Much like *Brg1*, a homozygous mutation affecting the BAF155 protein results in early embryonic lethality prior to neurulation [133], thereby demonstrating that the BAF complex plays a critical role in early embryonic development. Currently, there are two hypotheses as to how impairment in BAF function results in exencephaly. One hypothesis is that the maintenance and differentiation of neural progenitors are not properly regulated, resulting in decreased proliferation and therefore an overall reduction of neural progenitors [134,135]. The other hypothesis is based on recent evidence that the BAF complex occupies the promoter of *Vangl1* in embryonic stem cells [136], and therefore may be involved in regulating *Vangl1* transcription. *Vangl1* is a PCP

gene that has been associated with neural tube defects in humans [137-139], and has been shown to result in craniorachischisis in mice that contain a heterozygous mutation in *Vangl1* as well as in *Vangl2* [140]. Both of these hypotheses require further study before conclusions can be made regarding the mechanism by which mutations affecting the BAF complex result in neural tube defects.

The gene *Cecr2*, which is a focus within this thesis, encodes a subunit of the ISWI family CERF complex [141]. A homozygous mutation in *Cecr2* results in exencephaly in a proportion of mice [141]. Theories regarding the mechanism by which mutations in *Cecr2* result in exencephaly include misregulation of *Alx1* and *Dlx5*, genes known to be involved in neurulation [142], suggestive evidence that *Cecr2* may be involved in PCP [143], and the implication that *Cecr2* is involved in DNA double-strand-break (DSB) repair [144]. ATP-dependent chromatin remodelers function in DNA replication, DNA repair, and gene regulation, all of which have been shown to be important for the process of neurulation.

1.7: *Cecr2*

A homozygous loss-of-function mutation in *Cecr2* in our mouse model results in the NTD exencephaly [141,142] and inner ear defects [143], as well as kidney [145] and fertility [146] defects in mice that do not develop exencephaly. All *Cecr2* mutant mouse experiments described in this thesis utilized the mutant allele $Cecr2^{GT45bic}$, with the exception of one experiment involving a more severe mutant allele, $Cecr2^{tm1.1Hemc}$. The $Cecr2^{GT45bic}$ allele contains a genetrap insertion in intron 7, which results in a truncated CECR2 protein fused to β-galactosidase (Figure 1.7.1A-B). The β-galactosidase moiety allows for expression analyses through the utilization of X-gal staining, which demonstrated that *Cecr2* is expressed in the central nervous system, limbs, eyes, and nasal epithilium in E13.5 embryos [141], as well as the neural epithilium and head mesenchyme at the time of neurulation and the late embryonic cochlear duct portion of the inner ear (Figure 1.7.1C-H) [143]. There is an ~14-fold reduction of wildtype *Cecr2* transcript in homozygous mutant $Cecr2^{GT45bic}$ 11-14 somite neurulating embryos [142]. The presence of a small amount of full-length CECR2 protein in homozygous mutant $Cecr2^{GT45bic}$ mouse has been confirmed by western blot (Niri, unpublished data). The more severe $Cecr2^{tm1.1Hemc}$ mutant allele contains a deletion of *Cecr2* exon 1 and approximately 1 kb of upstream sequence, resulting in an ~260-fold reduction

in transcript compared to wildtype levels [142], with no visible CECR2 protein on western blot (Niri, unpublished data). The presence of some wildtype mRNA and a lower exencephaly penetrance in *Cecr2*^{GT45bic} homozygous mutant mice compared to *Cecr2*$^{tm1.1Hemc}$ homozygous mutant mice indicates that the *Cecr2*^{GT45bic} allele is hypomorphic, which is likely due to splicing around the genetrap cassette.

CECR2 has been shown to form at least two chromatin remodeling complexes, one of which was characterized in HEK293 cells and was shown to contain the catalytic ISWI protein, SNF2L [141]. The other CECR2 complex was characterized in testis and embryonic stem cells and was shown to include the catalytic ISWI protein, SNF2H [146], as well as probably SNF2L. It is not clear whether SNF2H and SNF2L are in two separate CECR2 complexes or the same complex. *Snf2l* and *Snf2h* are the only two mammalian members of the ISWI family of chromatin remodelers, both of which are expressed in the mouse embryo throughout development (80% identical at the nucleotide level in humans). However, *Snf2l* and *Snf2h* display differential expression patterns indicative of non-redundant roles in development [147]. Specifically, studies have indicated that *SNF2H* is predominantly expressed in undifferentiated progenitor cells, whereas *SNF2L* is expressed in differentiated cells [147,148]. Based on these findings and additional studies, Pépin et al. (2007) proposed that complexes containing SNF2H are likely responsible for assembling and spacing nucleosomes during replication in proliferating undifferentiated cells, and complexes containing SNF2L are likely involved in regulating gene transcription during terminal differentiation (reviewed in [149]). Since ISWI proteins lack a bromodomain, which recognizes and binds to chromatin, SNF2L and SNF2H form complexes that contain at least one member of the bromodomain adjacent to zinc finger (BAZ) family of proteins [150-152]. Along with targeting the ISWI complex to chromatin, BAZ proteins have also been shown to augment ISWI catalytic activity and enable chromatin assembly [153-155]. CECR2, along with RSF-1, BPTF, TIP5, ACF1, and WSTF, are the six members of the BAZ protein family in mammals. Members of the BAZ protein family share certain functional domains, which includes a DDT domain, at least one AT-hook, and a bromodomain and/or PHD type zinc finger domain. The DDT domain is a key motif in the BAZ family as it functions to bind to the catalytic ISWI component of the complex [156,157]. The AT-hook helps to stabilize the interaction between the ISWI complex and chromatin by binding to the minor groove of AT-rich DNA [158]. Bromodomains recognize and bind to acetylated lysines on histone tails of histones H3 and H4 [159-161], whereas PHD finger domains

recognize and bind to methylated histone tails [162,163]. CECR2 contains the characteristic DDT domain, a single AT-hook, and a bromodomain (Figure 1.7.1A).

The NTD phenotype in *Cecr2* mutant mice demonstrates a reduced penetrance in the FVB/N mouse strain compared to the BALB/cCrl mouse strain. This difference between strains suggests there are genetic modifiers, where one or more genetic variants that differ between these two strains renders BALB/cCrl susceptible and/or FVB/N resistant to the development of *Cecr2*-associated exencephaly. Previous work in the McDermid lab identified a region on chromosome 19 that contains at least two modifier genes [164,165]. Initial gene expression analyses between strains identified *Arhgap19* as a strong candidate modifier gene located within the chromosome 19 region [165].

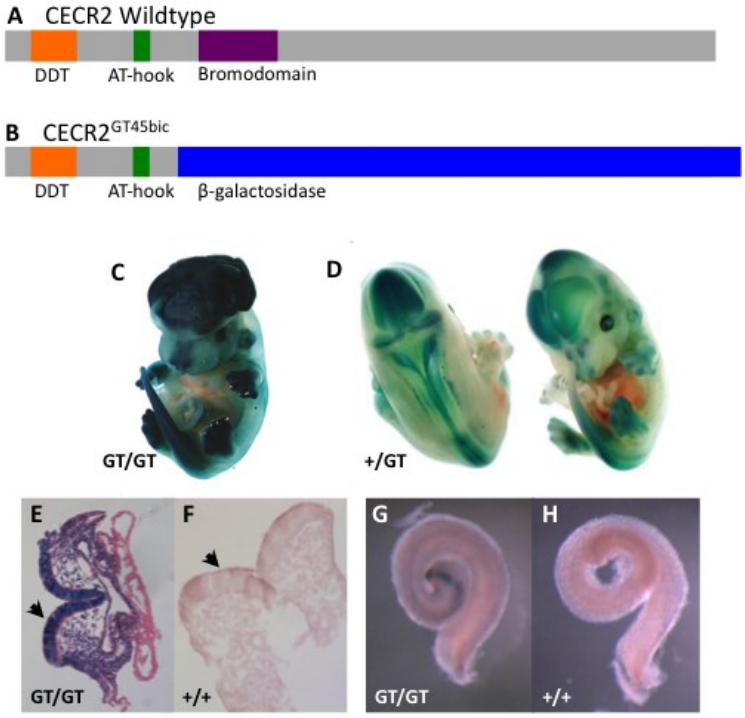


Figure 1.7.1: CECR2 and the CECR2^{GT45bic} protein. The CECR2 protein contains a DDT domain, an AT-hook, and a bromodomain (**A**). The *Cecr2*^{GT45bic} allele produces a truncated CECR2 protein lacking the bromodomain and fused to β-galactosidase (**B**). X-gal staining reveals *Cecr2* expression in the central nervous system, eye, limb-buds, and nasal epithelium in an E13.5 homozygous mutant embryo displaying exencephaly (**C**) and a phenotypically normal heterozygote (**D**). X-gal staining reveals *Cecr2* expression in the neural epithelium at the time of neurulation in a mutant embryo (**E**), with no staining observed in a wildtype age-matched control (**F**). *Cecr2* is also expressed in the late embryonic cochlear duct (**G**,**H**). GT represents the *Cecr2*^{GT45bic} allele, + represents the wildtype allele. Arrowheads point at the neural epithelium. **C**,**D** modified from Banting et al. (2005) [141], **E**-**H** modified from Dawe et al. (2011) [143].

Although a homozygous mutation in the gene *Cecr2* results in exencephaly in our mouse model for NTDs, a loss-of-function mutation in human *CECR2* has not yet been associated with a disease phenotype in humans. In humans, *CECR2* is one of the 14 genes found to be within the chromosome 22q11 region that is triplicated in cat eye syndrome (CES) [166], a condition characterized by defects in the eyes, ear auricles, heart, skeleton, kidneys, anus, and mental development. The CES phenotype is highly variable, but anencephaly or other NTDs have never been reported in patients with CES. This could indicate that overexpression of *CECR2* does not adversely affect neurulation on its own, or that CES has not yet been observed in any aborted or perinatal lethal cases of anencephaly. The specific effects of *CECR2* overexpression in cat eye syndrome are currently unknown, although a three-generation family with a triplicated region of only the 3 genes *CECR2*, *SLC25A18*, and *ATP6V1E1* show a subset of cat eye syndrome phenotypes, including renal, anorectal, and auricular malformations [167]. As the *Cecr2*$^{tm1.1Hemc}$ mutant allele results in renal abnormalities in FVB/N mice [145], it is possible that the triplication of *CECR2* contributed to the renal phenotype in this family. No mouse duplication or triplication including *CECR2* has been produced.

1.8: Hypothesis

The overarching hypothesis of this thesis is that there are identifiable genetic modifiers present within the mouse chromosome 19 modifier region that renders one mouse strain (BALB/cCrl) susceptible and another mouse strain (FVB/N) resistant to the development of *Cecr2*-associated exencephaly. The following objectives were employed in order to test this hypothesis:

- To further our understanding of how genetic modifiers affect exencephaly and small inner ear penetrance in *Cecr2* mutant mice, I characterized the penetrance of exencephaly and small inner ears in congenic BALB/cCrl (susceptible to exencephaly) and FVB/N (resistant to exencephaly) *Cecr2* mutant mouse strains, as well as sub-interval congenic *Cecr2* mutant mouse lines containing genomic modifier regions derived from FVB/N on an otherwise BALB/cCrl genetic background (Chapter 3).

- As a mutation in *Cecr2* is known to result in exencephaly in mice, we predict that variants within human *CECR2* may contribute to human NTD etiology. To test this, I identified and characterized mutations in human *CECR2* in a human cohort consisting of 156 probands (Chapter 3).

- *Arghap19*, which was the top candidate modifier gene, was tested as a modifier gene by performing genetic analyses in mouse (Chapter 3).

- Although *Arhgap19* was the top candidate modifier, more than one modifier gene is known to exist within the chomosome 19 modifier region. Therefore, I identified additional candidate modifier genes of *Cecr2*-associated exencephaly containing protein-coding mutations that differ between BALB/cCrl and FVB/N mouse strains, which will be added to the list of candidate modifier genes that are differentially expressed between BALB/cCrl and FVB/N (Chapter 3).

- The presence of DNA variants that alter protein function in human homologues of candidate modifier genes provide further support that these genes are involved in neurulation. Therefore, I identified and sought to determine if mutations in the human homologues of candidate modifier genes of *Cecr2* contribute to cranial NTD etiology in a human cohort consisting of 156 probands (Chapter 3).

- Understanding the basic molecular function of *Cecr2* would provide insight into how mutation in *Cecr2* disrupts neurulation. I performed phenotypic analyses of the *Drosophila melanogaster Cecr2* homologue, *dikar*, to determine if *Cecr2* is involved in planar cell polarity or DNA double-strand break repair (Chapter 4).

- A specific cohort of male mice containing homozygous mutation in *Arhgap19* developed abnormal circling behavior. Although this was not related to *Cecr2*-associated exencephaly, I explored the genetic and environmental contributions to this circling phenotype and uncovered a unique gene-environment interaction that required the presence of a running-wheel for the circling behavior to manifest (Chapter 5).

Chapter 2

Materials and methods

2.1: Mouse husbandry

Approval from the Animal Care and Use Committee of the University of Alberta was obtained for all experiments involving mice (University of Alberta AUP 00000094). Our mouse colony is maintained under the technical supervision provided by Sciences Animal Support Services (SASS). All mice were subjected to a 14 hour light/10 hour dark cycle and an ambient temperature of 22 ± 2 °C. Mice were housed in ventilated cages (cage dimensions 31.8 cm x 16.5 cm x 12.7 cm), with a maximum of 5 adult mice housed together in a single cage. Non-breeding mice were fed LabDiet Laboratory Rodent Diet 5001 and breeding mice were fed LabDiet Mouse Diet 9F 5020. For environmental enrichment, all mice were provided with cotton nestlet material, wood shavings and either a cardboard house (standard housing), or with a running-wheel attached to a dome for housing. Breeding mice were never housed with running-wheel domes. Individual mice were identifiable by ear notching, which was performed at 2 weeks of age. The ear-skin was clipped in a pattern that represented a particular number. Ear-skin biopsies were collected for genomic DNA extraction and genotyping.

Two mouse strains were used for this research. BALB/cCrl, a sub-strain of BALB/c, originated from Charles River Laboratories. A BALB/cCrl breeding colony has been maintained at the University of Alberta initially at the Health Sciences Laboratory Animal Services (HSLAS) unit for approximately 40 years and then at SASS from 2008 to present. BALB/cJ, another sub-strain of BALB/c, was purchased from Jackson Laboratories and bred within our colony. The second strain, FVB/N, was originally purchased from Jackson Laboratories and bred within our colony. The FVB/N strain used to generate the FVB/129P2 line discussed in Chapter 5 was purchased from Charles River Laboratories.

2.2: Embryo collection and exencephaly penetrance analysis

Pregnant mouse dams were euthanized by either CO_2 asphyxiation or cervical dislocation as per standard operating procedures, and embryos of various stages of development were dissected in 1X PBS or DEPC treated 1X PBS. Embryos of a desired developmental stage were obtained by setting up timed matings, which involved looking for the presence of a seminal plug to verify mating occurred. Tail biopsies or extra-embryonic membranes were taken for DNA

extraction. The presence of exencephaly was observed and recorded for embryos 10.5 days post-fertilization (E10.5) and older. Statistical analyses comparing the penetrance of exencephaly between two groups of mice were performed using the χ^2 test-of-independence. Some embryos were either stored at -80 °C for future RNA extraction or were prepped for histological examination.

2.3: Small inner ear penetrance analysis

Timed matings were set up to obtain E17.5 embryos. Pregnant mouse dams were then euthanized, and E17.5 embryos were dissected in 1X PBS or DEPC treated 1X PBS. Tail biopsies were taken for DNA extraction and genotyping. The presence of exencephaly was observed and recorded. Inner ears were removed from the embryo, fixed in 0.5 mL of 4% paraformaldehyde in 1X PBS and horizontally rotated on a Lab-Line Orbital Shaker rotator for 1.5 hours on ice. Inner ears then went through 3 washes in 1X PBS, 1 wash in 25% ethanol, 1 wash in 50% ethanol, 1 wash in 70% ethanol, and 2 washes in 100% ethanol. Each wash utilized 1 mL of solution and was 5 minutes long. Inner ears were stored in 100% ethanol at -20 °C until imaged. Imaging of inner ears was done on an Olympus SZ61 dissecting microscope with a Microscope Digital Camera MDC 320 and ScopePhoto software. Inner ears were measured from images using ImageJ software. Statistical analyses comparing the penetrance of small inner ears between groups of mice were performed using one-way ANOVA followed by Tukey's post-hoc.

2.4: DNA extraction

Genomic DNA was acquired from ear-skin biopsies (juvenile/adult mice), tail-skin biopsies (E10.5 to E18.5 mouse embryos), extra-embryonic membranes (E8.5 to E9.5 mouse embryos), or single adult *Drosophila*. Tissue was incubated overnight in a 0.5% SDS, 0.1 M NaCl, 50 mM Tris-Cl pH 8, 0.5 μM EDTA, and 0.0375% proteinase K lysis solution in a 60 °C water bath, precipitated with 8 M potassium acetate, chloroform extracted, precipitated with 100% ethanol, cleaned with 70% ethanol, and resuspended in TE buffer pH 8. Extracted genomic DNA was stored at -20 °C when not in use.

2.5: *Cecr2*^{GT45bic} genotyping

Generation of the *Cecr2*^{GT45bic} mutant mouse line was originally detailed in Banting et al. (2005) [141]. In short, a β-geo genetrap vector was introduced into *Cecr2* intron 7, resulting in a truncated CECR2 protein lacking exons 8 to 19 but encoding for a protein consisting of the first 7 CECR2 exons fused to β-galactosidase. A multiplex PCR reaction was performed on extracted genomic DNA samples in order to simultaneously amplify the *Cecr2* wildtype allele (376 bp amplicon), the *Cecr2*^{GT45bic} mutant allele (573 bp amplicon), and the male specific *Sry* gene for sexing (266 bp amplicon). The 20 µl PCR reaction contained 1X PCR buffer (25 mM Tris pH 9, 50 mM KCl, 1.5 mM $MgCl_2$, and 0.02 mg/mL bovine serum albumin (BSA) in milliQ water), 0.25 mM dNTPs, 1 µM of each primer (Cecr2 Intron7 F4, Cecr2 Intron7 R4, pGT1R4, SRY FOR, and SRY REV), 1 U of Taq polymerase (generated in house at MBSU by Dr. Michael Pickard), and ~75 ng of template DNA. Integrated DNA Technologies was the supplier for the primers (see Appendix A for primer sequences). PCR reactions were run on a Peltier Thermocycler-200 (MJ Research) with the following cycling conditions: (1) 94.0 °C for 1 minute and 30 seconds, (2) 94.0 °C for 15 seconds, (3) 60.0 °C for 20 seconds, (4) 68.0 °C for 40 seconds, (5) repeat steps 2-4 36 times, (6) 68.0 °C for 5 minutes, (7) 4.0 °C hold. Orange G loading buffer was added (1X final concentration) to the PCR products, which were then loaded onto a 2.0% agarose gel and underwent gel electrophoresis at 130 V for 1 hour. All genotyping PCR reactions were run alongside a no-template control PCR reaction.

2.6: *Cecr2*$^{tm1.1Hemc}$ genotyping

Generation of the FVB/N *Cecr2*$^{tm1.1Hemc}$ mutant mouse line was originally described in Fairbridge et al. (2010) [142]. Briefly, a germline deletion of exon 1 and 1 kb of upstream sequence was mediated by LoxP-Cre by crossing Chimeric mice heterozygous for a floxed *Cecr2* exon 1 to BALB/c-Tg(CMV-cre)1Cgn/J mice from Jackson laboratories. For genotyping, a multiplex PCR reaction was performed on extracted genomic DNA samples in order to simultaneously amplify the *Cecr2* wildtype allele (~200 bp amplicon) and the *Cecr2*$^{tm1.1Hemc}$ mutant allele (~450 bp amplicon). The 20 µl PCR reaction contained 1X PCR buffer (25 mM Tris pH 9, 50 mM KCl, 1.5 mM $MgCl_2$, and 0.02 mg/mL bovine serum albumin (BSA) in milliQ water), 0.25 mM dNTPs, 1

μM of each primer (Ingenious SDL2, LoxCECR2_DEL3R, and IngeniousLox 1), 1 U of Taq polymerase (generated in house at MBSU by Dr. Michael Pickard), and ~225 ng of template DNA. Integrated DNA Technologies was the supplier of the primers (see Appendix A for primer sequences). PCR reactions were run on a Peltier Thermocycler-200 (MJ Research) with the following cycling conditions: (1) 94.0 °C for 3 minutes, (2) 94.0 °C for 30 seconds, (3) 60.0 °C for 30 seconds, (4) 72.0 °C for 1 minute and 20 seconds, (5) repeat steps 2-4 35 times, (6) 72.0 °C for 7 minutes, (7) 4.0 °C hold. Orange G loading buffer was added (1X final concentration) to the PCR products, which were then loaded onto a 2.0% agarose gel and underwent gel electrophoresis at 130V for 1 hour. All genotyping PCR reactions were run alongside a no-template control PCR reaction.

2.7: Sanger sequencing

Unless otherwise noted, all Sanger sequencing reactions were performed as described here. A portion of the PCR products to be sequenced were first run on a 2.0% agarose gel to estimate DNA concentration by comparing to a sample of known concentration. PCR clean-up to remove leftover primers and dNTPs prior to the sequencing PCR was achieved by combining 5 μl of PCR product with ExoSAP mixture (10 U of Exonuclease I (Exo) and 5 U of Shrimp Alkaline Phosphatase (SAP) diluted in milliQ water to a final volume of 2 μl per reaction), and incubated at 37 °C for 30 minutes immediately followed by an 80 °C incubation for 15 minutes to deactivate the Exonuclease I and SAP enzymes. These incubations were performed on a Peltier Thermocycler-200 (MJ Research). Samples were then kept at 4 °C until further use. Based on the previous estimation of DNA concentration, ~150 ng of DNA template was used in the Sanger sequencing PCR reaction. The 10 μl PCR reaction also included 2 μl BigDye sequencing premix (version 1.1 for amplicons ≤ 200 bp, version 3.1 for all other amplicons), 3 μl of BigDye buffer, and 0.1 μM of the appropriate primer (either the forward or the reverse primer used in the initial PCR or a primer nested within the target sequence). The PCR reaction was run on a Peltier Thermocycler-200 (MJ Research) for 25 cycles of 95 °C for 30 seconds, 50 °C for 15 seconds, 60 °C for 1 minute and 30 seconds, and held at 4 °C. PCR product was transferred to a 1.5 mL centrifuge tube and DNA was precipitated by addition of 2 μl of NaOAc (3 M)/EDTA (125 mM) pH 8 and 80 μl of ice-cold 95% ethanol, and then incubated at -20 °C for 15 minutes. Samples

were then centrifuged at 14000 rpm for 15 minutes, supernatant was discarded, and pellets were washed with 500 µl of 70% ethanol. Samples were vortexed and centrifuged at 14000 rpm for 15 minutes, supernatant was discarded, and pellets were air-dried for 30 minutes. Samples were then delivered to the Molecular Biology Services Unit (MBSU) for resuspension in 10 µl of milliQ water, addition of 4 µl of resuspended template to 10 µl of highly deionized formamide, loading and running on an ABI 3730 automated sequencer (Applied Biosystems).

In the event that ≥48 samples underwent Sanger sequencing, initial PCR products were transferred to a 96 well plate, rather than a 1.5 mL centrifuge tube, and the rest of the procedure was performed with the following changes at the MBSU. The BigDye sequencing PCR was run on an eppendorf Mastercycler thermocycler under the same cycling conditions as described above. The samples were precipitated in a 96 well plate by the addition of 2 µl of NaOAc (3 M)/EDTA (125 mM), 50 µl of 95% ethanol, and incubation at room temperature rather than -20 °C. DNA was then spun down in a Sorvall® Legend RT centrifuge at 2500 X g for 30 minutes at 4 °C, and supernatant was removed by inverting the plate into a container lined with kimwipes and pulse-spun for 10-20 seconds at 185 X g. DNA pellets were washed with 70 µl of 70% ethanol and centrifuged at 1650 X g for 15 minutes at 4 °C. Supernatant was removed the same way as was done for the previous spin, and the pellets were dried in a Speedvac®Plus SC110A attached to a UVS400 Universal Vacuum System (Savant) for 5-10 minutes. Samples were then prepared and loaded on the ABI 3730 sequencer as described above.

Plasmid DNA underwent a simplified procedure, as the Economy Sequencing Service offered by the MBSU was utilized for these samples. Samples for sequencing were prepped by mixing 575 ng of plasmid (final concentration 57.5 ng/µl), 2.5 pmoles of sequencing primer (final concentration 0.25 µM), and nuclease free water (Integrated DNA Technologies) up to a total volume of 10 µl in a 1.5 mL centrifuge tube. Prepped samples were then delivered to MBSU for the Sanger sequencing service.

2.8: Whole exome sequencing of BALB/cCrl and FVB/N

Whole genomic DNA samples were obtained from a single adult female BALB/cCrl and a single adult female FVB/N mouse from our colony following the protocol given in section 2.4. These samples were then sent to Duke University where Deidre Krupp, a graduate student of Dr.

Allison Ashley-Koch and Dr. Simon Gregory, performed whole exome sequencing. Deidre utilized the Agilent SureSelect XT Mouse All Exon kit, which was based on the mm9 mouse genome assembly and was designed to capture 49.6 Mb of coding sequence that covers 221,784 exons within 24,306 genes, which is the complete mouse exome according to the UCSC mm9 genome build. The mouse genomic DNA was fragmented to ~250 bp, ligated with barcoded library prep adapters, then hybridized to biotinylated exon probes followed by streptavidin enrichment, and then sequenced on an Illumina HiSeq 2000 next generation sequencing platform. 100 bp paired end sequence reads were generated from the captured mouse DNA template and aligned to the mm9 (2007) mouse genome assembly using the alignment tool program Burrows-Wheeler Aligner (BWA) [168]. The alignment program produced SAM files, which were converted to BAM files using SAMtools [169]. The initial PCR step to amplify target DNA can introduce duplicate reads, which then can artificially inflate coverage at a site resulting in false positive variant sequence calls. These duplicate reads were removed from the alignments using the program, Picard (http://picard.sourceforge.net/). Multiple quality metrics were assessed, which included the fraction of reads that were mapped to the reference sequence, fraction of correctly paired reads, fraction of duplicate reads, percentage of usable bases on target (whole exome target ± 200 bp), depth of coverage metrics, mean base quality by cycle, insert size distribution, where the insert refers to the genomic DNA between the two adapters, and GC bias, using the programs BamValidator (http://genome.sph.umich.edu/wiki/BamUtil:_validate), Picard, and the Genome Analysis Toolkit (GATK) [170]. BAM files were then uploaded into GATK for further processing and variant calling following GATK's recommendations for "best practice variant detection", which involved local realignment around indels, as erroneous variant calls can result from misalignment around indels, base quality score recalibration, as recalibrated scores have been shown to be more informative and accurate compared to quality scores directly given by the sequencing machine, and variant calling by GATK's "Unified genotyper" within the targeted exome intervals (± 200 bp). Integrative Genomics Viewer (IGV) was then used to further evaluate the sequence alignment and variant calls [171]. Deidre then provided me with a summary of all SNV and indel calls, a summary of the quality metrics (Table 2.8.1), and BAM files.

Table 2.8.1: Summary of whole exome sequencing quality metrics.

Exome	% Mapped	% Duplicates	Coverage
BALB/cCrl	98.8	5.5	76.9x
FVB/N	98.7	6.0	72.1x

I then proceeded to select the best candidate variants and conduct further analyses, which was restricted to the chromosome 19 modifier region. Analyses were also restricted to protein coding DNA sequence, thereby complementing our existing microarray data, which already identified genes differentially expressed between the two mouse strains [165]. Analyses consisted of *in silico* predictions of deleteriousness by the web-based programs Sorting Tolerant from Intolerant (SIFT) [172], which uses evolutionary conservation to predict if an amino acid change is deleterious, Polyphen2 [173], which uses a combination of evolutionary conservation and biochemistry information to predict if an amino acid change is deleterious, and Genomic Evolutionary Rate Profiling (GERP) [174], which uses evolutionary conservation to predict if a nucleotide change is deleterious. The web-based software for SIFT can be found at http://sift.jcvi.org/. The web-based software for Polyphen2 can be found at http://genetics.bwh.harvard.edu/pph2/. GERP scores were assigned to variants by Deidre Krupp (Duke University), where she utilized the ANNOVAR [175] program, which can be downloaded from http://annovar.openbioinformatics.org/en/latest/.

2.9: Exonic sequencing of 25 candidate NTD genes in 156 human anencephaly samples

Next generation sequencing of the coding exons of the genes *CECR2, FOXD4, KANK1, GLIS3, TJP2, CSTF2T, RNLS, LIPJ, FAS, BTAF1, CPEB3, TCTN3, EXOSC1, MMS19, DNMBP, SCD, HIF1AN, HPS6, TMEM180, SFXN2, FAM160B1, TRUB1, PNLIPRP2, NANOS1,* and *FAM45A* was performed on 156 human cranial NTD probands in collaboration with Drs. Allison Ashley-Koch, Nicholas Katsanis, Simon Gregory, and Erica Davis at Duke University. Ethics approval was obtained from the University of Alberta Research Ethics Board, project name "Sequencing genes in a cohort of humans with a neural tube defect" (MS3_Pro00042452) in conjunction with Duke University Health System Institutional Review Board for Clinical Investigations, project name "Hereditary Basis of Neural Tube Defects" (CR5_Pro00016517). DNA samples from cranial NTD probands were acquired over a long period of time at the Center

for Human Genetics, Duke University. Specific cranial NTD phenotypes for all probands are provided in Appendix B. Proband genomic DNA samples were extracted from various fetal tissues from terminated pregnancies or from cord blood at the time of delivery, resuspended in nuclease free water, and allocated by the Duke DNA Bank as per their standard operating procedures. I assayed DNA concentration using Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen) and transferred 200 ng of DNA for each sample into individual wells of new plates. Research analyst Karen Soldano (Duke University) dried the DNA samples by speed vac and resuspended in 40 µl of nuclease free water to obtain a final concentration of 10 ng/µl for all samples. These samples were then supplied to Dr. Nicholas Katsanis' lab (Duke University), where research technician Natalie Mola performed the library preparation and sequencing. Primers to amplify coding exons were designed using QIAGEN® GeneRead DNAseq Custom Builder online tool. Coding sequence not captured in the primer design are displayed in Table 2.9.1. The online tool is programmed to design primers that are highly specific to the target region, meaning the exclusion of some coding regions was likely due to poor primer specificity to target sequence in these regions. Natalie Mola constructed the library by following the protocol NEBNext Fast DNA Library Prep Set for Ion Torrent (GeneRead DNAseq Gene Panel Handbook 11/2012). Natalie Mola also performed next generation sequencing using the Ion Torrent platform, aligned sequence data to UCSC hg19 (GRCh37), performed variant call analysis using both QIAGEN® GeneRead and Ion Torrent Suite, and shared these data with myself via DNAnexus (https://www.dnanexus.com/). I then downloaded and used the software ANNOVAR [175] to annotate the variant call files with RefSeq [176] gene names, dbSNP137 IDs [177], esp6500 European-American [178] minor allele frequencies (MAF), Genomic Evolutionary Rate Profiling (GERP) scores [174], Sorting Intolerant From Tolerant (SIFT) scores [172], and PolyPhen2 scores [173]. In order to include as many low frequency/rare coding variants as possible while keeping the number of variants manageable for analyses, a MAF cutoff of 0.03 was chosen for variants of interest. Although a MAF of 0.01 has arbitrarily been defined as "rare" in the literature, all variants contained herein with a MAF $\leq$ 0.03 are defined as "rare" for simplicity sake. I validated all variants with a MAF $\leq$ 0.03 or variants that occurred in the same proband as a variant with a MAF $\leq$ 0.03 by Sanger sequencing (one direction) in the proband and in available parental DNA samples. Primers for Sanger sequencing were designed using Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) to maximize primer specificity followed by

SNPCheck v3 (https://secure.ngrl.org.uk/SNPCheck/snpcheck.htm) to minimize the potential of known DNA sequence variants interfering with primer binding. I performed all PCR reactions for Sanger sequencing at Duke University using QIAGEN® HotStarTaq Plus Polymerase (2.5 U per reaction) and accompanying 10X PCR buffer (1X final concentration). Q-solution supplied along with the HotStarTaq Plus Polymerase was added to some PCR reactions (1 in 5 dilution) that performed better when Q-solution was present (see Appendix A master primer list for which PCR reactions included Q-solution). Primers (0.5 μM final concentration) were supplied by Integrated DNA Technologies and are listed in Appendix A. dNTP mix (200 μM final concentration) and template DNA (10-20 ng/μl final concentration) was also added. PCR reactions were run on a Mastercycler Pro 384 (Eppendorf) with the following touchdown cycling conditions: (1) 95.0 °C for 5 minutes, (2) 94.0 °C for 30 seconds, (3) 66.0 °C for 30 seconds, (4) 72.0 °C for 1 minute and 45 seconds, (5) repeat steps 2-4 13 times with -1.0 °C for step 3 each cycle, (6) 94.0 °C for 10 seconds, (7) 52.0 °C for 45 seconds, (8) 72.0 °C for 1 minute and 45 seconds, (9) repeat steps 6-8 25 times, (10) 72.0 °C for 8 minutes, (11) 4.0 °C hold. PCR product along with an aliquot of 5 μM primer in dH2O was then supplied to GENEWIZ, Inc, who then performed PCR clean-up and Sanger sequencing.

Table 2.9.1: Coding sequence that was not covered by QIAGEN® GeneRead DNAseq custom primers.

Gene	Chromosome	Genomic position (hg19)	Genomic size (bp)
FOXD4	9	117456-117469	14
KANK1	9	504751-504754	4
GLIS3	9	3824589-3824603	15
GLIS3	9	4118062-4118074	13
GLIS3	9	4299847-4300035	189
TJP2	9	71789140-71789193	54
TJP2	9	71836081-71836085	5
CSTF2T	10	53456951-53456964	14
CSTF2T	10	53457093-53457104	12
RNLS	10	90033663-90033676	14
RNLS	10	90044716-90044803	88
RNLS	10	90044925-90045101	177
RNLS	10	90341431-90341466	36
LIPJ	10	90353703-90353714	12
LIPJ	10	90356037-90356163	127
LIPJ	10	90356654-90356654	1
LIPJ	10	90362333-90362341	9

Gene	Chromosome	Genomic position (hg19)	Genomic size (bp)
FAS	10	90770341-90770357	17
FAS	10	90773100-90773109	10
FAS	10	90775264-90775542	279
BTAF1	10	93716982-93716998	17
BTAF1	10	93719542-93719556	15
BTAF1	10	93726394-93726412	19
BTAF1	10	93741400-93741402	3
BTAF1	10	93742344-93742352	9
BTAF1	10	93744139-93744161	23
BTAF1	10	93753451-93753457	7
BTAF1	10	93776206-93776213	8
BTAF1	10	93778677-93778692	16
BTAF1	10	93789112-93789215	104
BTAF1	10	93789341-93789358	18
CPEB3	10	93808501-93808701	201
CPEB3	10	93808818-93808834	17
CPEB3	10	93809179-93809250	72
CPEB3	10	93809523-93809571	49
CPEB3	10	93809674-93809782	109
CPEB3	10	93809882-93809905	24
CPEB3	10	93809990-93810076	87
CPEB3	10	93940720-93940758	39
CPEB3	10	93999495-93999632	138
EXOSC1	10	99198427-99198453	27
DNMBP	10	101769595-101769603	9
SCD	10	102124536-102124587	52
HIF1AN	10	102308777-102308881	105
TMEM180	10	104221170-104221186	17
TMEM180	10	104235960-104235971	12
TMEM180	10	104236058-104236268	211
SFXN2	10	104498007-104498320	314
FAM160B1	10	116581503-116581564	62
FAM160B1	10	116581875-116581882	8
FAM160B1	10	116595883-116595905	23
FAM160B1	10	116595962-116595966	5
FAM160B1	10	116602692-116602714	23
FAM160B1	10	116602978-116602985	8
FAM160B1	10	116606930-116606992	63
FAM160B1	10	116621674-116621682	9
FAM160B1	10	116621798-116621864	67
FAM160B1	10	116623026-116623043	18
FAM160B1	10	116623143-116623152	10
FAM160B1	10	116623339-116623344	6
FAM160B1	10	116659568-116659574	7
TRUB1	10	116735421-116735434	14

Gene	Chromosome	Genomic position (hg19)	Genomic size (bp)
TRUB1	10	116736581-116736600	20
TRUB1	10	116736924-116736951	28
TRUB1	10	116737020-116737439	420
NANOS1	10	120789228-120789324	97
NANOS1	10	120789553-120789945	393
NANOS1	10	120792424-120792699	276
NANOS1	10	120792872-120792883	12
FAM45A	10	120863611-120863681	71
FAM45A	10	120897096-120897099	4
FAM45A	10	120897189-120897197	9
CECR2	22	18033625-18033633	9
CECR2	22	18033737-18033741	5

2.10: X-gal staining

X-gal staining was performed on whole mount embryos or on embryonic brains to analyze the temporal and spatial expression of *Arhgap19* in mice carrying the *Arhgap19*$^{Gt(YHD020)Byg}$ allele. I harvested embryos/embryonic brains as described in section 2.2, and fixed these samples in 4% paraformaldehyde in 1X PBS for 1.5-3 hours. Embryo or brain samples were then washed 3 times (10 minutes each wash) in LacZ wash solution (2 mM $MgCl_2$, 0.01% deoxycholic acid, and 0.02% IGEPAL diluted in 1X PBS). Samples were then transferred to X-gal stain (5 mM potassium ferricyanide, 5 mM potassium ferrocyanide, and 5 mg/mL X-gal diluted in LacZ wash solution) and incubated at 37 °C for 1-2 days. Samples were washed 3 x 10 minutes in 1X PBS, transferred to 100% methanol by performing a series of 10 minute washes where each wash was incrementally increased by 25% of methanol diluted in 1X PBS, and were stored at -20 °C until imaging or paraffin embedding. Imaging of whole mount embryos and brains was performed at the Advanced Microscopy Facility (University of Alberta) using an Olympus SZ61 dissecting microscope with a Microscope Digital Camera MDC 320 and ScopePhoto software.

2.11: Paraffin embedding and histology of tissue samples

Fixed tissue samples were transferred from 1X PBS to either 70% ethanol or 100% methanol by performing a series of washes where each wash was incrementally increased by 25%

of ethanol/methanol diluted in 1X PBS. Washes were for 5 minutes for small samples, such as E8.5-E11.5 embryos, or 15 minutes for larger samples, such as E12.5-E16.5 embryos and brains. Samples were stored in 70% ethanol or 100% methanol at -20 °C. Originally, methanol was used only for X-gal stained samples (ethanol for all other samples) as it was believed that ethanol would wash out the X-gal stain. However, I compared X-gal stained samples stored in ethanol versus methanol and noted no obvious differences; therefore, all following samples were transferred to and stored in 70% ethanol regardless of whether they were X-gal stained or not. Samples were then brought to the Advanced Microscopy Facility (University of Alberta) for overnight processing into paraffin using a Fisher Histomatic Tissue Processor (Model 166). The next morning, I placed tissue samples in embedding moulds and allowed the paraffin to cool from 60 °C to room temperature. The tissue samples in paraffin blocks were cut into 5-10 μm sections using a Reichert-Jung 2040 microtome. Sections were transferred to a 42 °C water bath and mounted onto Fisherbrand® Superfrost® Plus Microscope Slides, dried overnight at 37 °C, and stored at room temperature until used for basic histological analysis or immunofluorescence. For basic histological analysis, tissue sections were de-paraffinized by 2 x 5 minute washes in toluene, and then were stained by 2 x 2 minute washes in 100% ethanol and 1 x 2 minute wash in 90% ethanol followed by 15-60 seconds in acidified Eosin, 2 x 2 minute washes in 100% ethanol, and 2 x 2 minute washes in toluene. Sections were then mounted with DPX mounting medium and a coverslip (Fisherbrand® Microscope Cover Glass 24x60-1) prior to imaging. Imaging of stained tissue sections was conducted on a Zeiss Scope A.1 AX10 microscope and Optronics camera with PictureFrame™ software.

2.12: Immunofluorescence of neurulating embryos

Wildtype FVB/N and homozygous mutant *Cecr2*^{GT45bic} FVB/N E 8.5 embryos were collected as described in section 2.2. Embryos were embedded in paraffin, sectioned, and put on slides as described above (section 2.11). Tissue sections were de-paraffinized by washing 3 x 5 minutes in toluene and rehydrated by a series of washes starting with 2 x 10 minutes in 100% ethanol, followed by 2 x 10 minutes in 95% ethanol, and finally 2 x 5 minutes in milliQ water. For antigen retrieval, slides were transferred into 1 mM EDTA pH 8, heated until boiling in a microwave, and allowed to sit for 10 minutes. This was done three times. Tissue sections were

blocked in 10% normal goat serum (Sigma) and 0.6% Triton®-X 100 (Sigma) diluted in 1X PBS for 1 hour at room temperature. Tissue sections were then transferred to primary antibody solution, which consisted of 1% bovine serum albumin (Sigma), 0.3% Triton®-X 100 (Sigma), and primary antibody diluted in 1X PBS, and incubated overnight at 4 °C in a humidifying chamber. Primary antibodies were affinity purified rabbit polyclonal anti-DNMBP (proteintech™) diluted to 1:1000, and affinity purified rabbit polyclonal anti-MMS19 (proteintech™) diluted to 1:300. Primary antibody was removed with 3 x 5 minute washes in 1X PBS and then incubated in secondary antibody solution, which consisted of 1% bovine serum albumin (Sigma), 0.3% Triton®-X 100 (Sigma), and secondary antibody (Invitrogen AlexaFluor® 488 goat anti-rabbit) diluted 1:200 in 1X PBS, for 2 hours at room temperature. This was followed by counterstaining for 5 minutes in a 0.1% DAPI in 1X PBS solution. Secondary antibody and DAPI were removed with 3 x 5 minute washes in 1X PBS. Flouromount G (SouthernBiotech) and a coverslip (Fisherbrand® Microscope Cover Glass 24x60-1) was then added to the slide and sealed with nail polish. Immunofluorescent sections were imaged under oil immersion at 40X magnification on a Nikon eclipse 80i confocal microscope with a CVI Melles Griot Ion Laser (green 488 nm and blue 561 nm), a Nikon C2-SHS camera, and NIS-Elements v4.0 imaging software.

2.13: Cloning of human *CECR2*

A previous graduate student in our lab, Graham Banting, inserted a 4395 bp clone of the human *CECR2* open reading frame (ORF) into a modified Insectselect™ protein expression vector pMIB/V5-His A (Invitrogen) and stored this clone in an *E. coli* glycerol stock at -80 °C [179]. I then revived the *E. coli* by streaking a small sample scraped off of the frozen glycerol stock on an LB-Amp plate (LB agar plate supplemented with 100 µg/mL Ampicillin) followed by overnight growth at 37 °C. An individual colony was selected and grown overnight in LB-Amp broth (LB broth (pH 7.0) supplemented with 100 µg/mL Ampicillin) for plasmid selection, plasmid was extracted using QIAquick® Spin Miniprep Kit according to the manufacturer's protocol, and plasmid DNA was quantified by Nanodrop 1000 (Thermoscientific). I employed Sanger sequencing to confirm the sequence of the *CECR2* ORF as described for plasmid DNA in section 2.7. Sequencing primers are listed in Appendix A.

Departmental technician, James MacLagan, cloned the *CECR2* ORF into the pENTR^{TM}11 Dual Selection Vector (Gateway® Life Technologies). Briefly, he amplified the ORF with the CECR2-KpnI-F forward primer that contained a KpnI restriction enzyme site directly upstream of sequence complementary to the beginning of the ORF, the CECR2-XhoI-R reverse primer that contained a XhoI restriction enzyme site directly downstream of sequence complementary to the end of the ORF (Appendix A) and Phusion® high-fidelity DNA polymerase. PCR template was then column-purified (QIAquick® PCR Purification Kit), digested with KpnI and XhoI restriction enzymes, and ligated to pENTR^{TM}11 vector linearized by KpnI Xho1 double digestion. Competent DH5α *E. coli* were transformed with the ligation reaction product and transformants were selected for by plating on LB-Kan plates (LB agar plates supplemented with 50 µg/mL Kanamycin) followed by overnight growth at 37 °C. James MacLagan selected four independent colonies, streaked them on LB-Kan agar plates, grew them overnight at 37 °C, and delivered them to me. I validated the ORF and Gateway® *att*L recombination sites with Sanger sequencing, which was achieved by using the same primers (Appendix A) and same procedure as was used for the original *CECR2* ORF described above with the exception that colonies were grown in LB-Kan broth (50 µg/mL) rather than LB-Amp broth. I prepared glycerol stocks of *E. coli* containing the *CECR2* pENTR^{TM}11 plasmid by adding 500 µl of overnight culture in LB-Amp broth to 500 µl of sterile 50% glycerol in a 2 mL sterile screw-cap tube for long-term storage at -80 °C. If needed, *E. coli* were revived from the glycerol stock as described above for the original *CECR2* clone with the remainder of the glycerol stock never undergoing a thaw cycle and always stored at -80 °C.

2.14: Site-directed mutagenesis of human *CECR2* plasmid constructs

All site-directed mutagenesis reactions were performed according to the protocol described in Niederriter et al. (2014) [123] on the human *CECR2* pENTR^{TM}11 plasmid to introduce DNA sequence alterations in the *CECR2* ORF or in upstream vector sequence. Briefly, I designed mutagenesis primers using the web-based software PrimerX (http://www.bioinformatics.org/primerx/) with default parameters, except that the minimum melting temperature was altered from 75 °C to 78 °C, and primers were supplied by Integrated DNA Technologies (Appendix A). Four mutagenesis reactions were set up for each DNA

sequence alteration with varying primer concentrations and presence/absence of DMSO (Table 2.14.1). Mutagenesis reactions were performed on a Peltier Thermocycler-200 (MJ Research) with the following cycling conditions: (1) 95 °C for 5 minutes (samples were added during this step to the PCR block once it reached 95 °C), (2) 95 °C for 50 seconds, (3) 55 °C for 1 minute, (4) 68 °C for 9 minutes, (5) repeat steps 2-4 18 times, (6) 68 °C for 7 minutes, (7) 4 °C hold. 4 µl of PCR product was electrophoresed on a 0.8% agarose gel, and the best mutagenesis reaction of the four, based on DNA band intensity, continued throughout the remainder of the procedure. The remaining three mutagenesis reactions were discarded. Original template DNA was degraded by addition of 0.5 µl DpnI restriction enzyme to the remainder of the mutagenesis reaction and incubated at 37 °C for 2 hours. Competent DH5α *E. coli* was then transformed with 2 µl of DpnI digested mutagenesis product, plated on LB-Kan plates, and grown overnight at 37 °C. Single colonies were picked for overnight growth in LB-Kan broth at 37 °C, plasmid extraction (QIAprep® Spin Miniprep Kit), quantification (Nanodrop 1000, Thermoscientific), and Sanger-sequence validation, which was performed as described for plasmid DNA above (section 2.7). Initially, only the targeted mutagenesis region were validated by Sanger sequencing in four clones, because duplication of the primer sequence often occurred (~1/2 of the clones) at the mutagenesis site. The rest of the ORF and *att*L recombination sites were validated by Sanger sequencing once a clone that was of the correct sequence at the mutagenesis site was identified. In the event that a second-site mutation was found elsewhere in the ORF or *att*L recombination sites, which was not very common, another clone would be Sanger-sequenced. ORF and flanking vector (including *att*L recombination sites) DNA sequence for all *CECR2* plasmid constructs generated by site-directed mutagenesis are provided in Appendix D. *E. coli* glycerol stocks were generated as described in section 2.13 for all *CECR2* pENTR^{TM}11 plasmid constructs and stored at -80 °C.

Table 2.14.1: Four site-directed mutagenesis reactions that vary in primer concentration and presence/absence of DMSO used for optimal generation of mutagenized *CECR2* plasmid constructs.

Reagent	Rxn 1	Rxn 2	Rxn 3	Rxn 4
Forward Primer (2.5 µM)	1 µl	1 µl	0.5 µl	0.5 µl
Reverse Primer (2.5 µM)	1 µl	1 µl	0.5 µl	0.5 µl
dNTP mix (40 mM)	0.5 µl	0.5 µl	0.5 µl	0.5 µl
5X High Fidelity Phusion® buffer*	5 µl	5 µl	5 µl	5 µl
CECR2 pENTR™11 DNA template (2 ng/µl)	1 µl	1 µl	1 µl	1 µl
milliQ H_2O	3.8 µl	3.2 µl	4.8 µl	4.2 µl
Phusion® High Fidelity DNA Polymerase*	0.2 µl	0.2 µl	0.2 µl	0.2 µl
Dimethyl Sulfoxide (DMSO)*	0 µl	0.6 µl	0 µl	0.6 µl
TOTAL	12.5 µl	12.5 µl	12.5 µl	12.5 µl

*Reagents were supplied together (New England Biolabs).
Starting concentrations of reagents are given in brackets.

2.15: Sub-cloning human *CECR2* plasmid constructs

Human *CECR2* DNA constructs were sub-cloned into the pEZY3 vector (Addgene plasmid #18672, gift from Dr. Fred Berry) [180] for future transfection and protein expression in HEK293 cells. Sub-cloning recombination reactions were performed using Gateway® LR Clonase II Enzyme Mix. I halved reagent volumes, with 0.75 µl of the *CECR2* pENTR™11 entry clone (100 ng/µl), 0.75 µl of pEZY3 destination vector (100 ng/µl), 2.5 µl of Tris-EDTA buffer pH 8 (Integrated DNA Technologies), and 1 µl of vortexed LR Clonase II Enzyme Mix combined in a 1.5 mL centrifuge tube and incubated at room temperature for 2 hours. The sub-cloning reaction was stopped by addition of 0.5 µl proteinase K (supplied with LR Clonase II Enzyme Mix) followed by a 10 minute incubation in a 37 °C waterbath. I then transformed competent DH5α *E. coli* with 1 µl of the sub-cloning reaction product, plated on LB-Amp plates, and grew *E. coli* overnight at 37 °C. An individual colony was selected for overnight growth in LB-Amp broth at 37 °C, plasmid extraction (QIAprep® Spin Miniprep Kit), quantification by Nanodrop 1000 (Thermoscientific), and confirmation of the presence of the *CECR2* ORF by Sanger sequencing of the start of the ORF as described for plasmid DNA (Section 2.7) using the T7-F primer (Appendix A). Plasmid size was also checked by gel electrophoresis. *E. coli* glycerol stocks were generated as described in section 2.13 for all *CECR2* pEZY3 plasmid constructs and stored at -80 °C.

Human *CECR2* DNA constructs were also sub-cloned into the pCSDest vector (Addgene plasmid #22423, gift from Drs. Nicholas Katsanis and Erica Davis) [181] following the same protocol above. These constructs were validated by Sanger sequencing (SP6-F primer, Appendix A) and sent to our collaborators, Drs. Erica Davis, Nicholas Katsanis, and Allison Ashley-Koch, for future *in vivo* analyses in zebrafish. *E. coli* glycerol stocks were generated as described in section 2.13 for all *CECR2* pCSDest plasmid constructs and stored at -80 °C.

2.16: HEK293 cell culture maintenance, harvest, and transfection

Dr. Roseline Godbout supplied us with a single cryogenic vial of HEK293 cells (1 mL), which was stored in liquid nitrogen until use. All cell culture work was performed in a biological safety cabinet to prevent contamination and anything placed in the biological safety cabinet was initially thoroughly decontaminated with 70% ethanol. I quickly thawed the cells by placing the vial in a 37 °C waterbath. The 1 mL of cells were then added to 5 mL of 37 °C cell culture media (89% Dulbecco's Modified Eagles Medium - High Glucose (Sigma), 10% Heat Inactivated Fetal Bovine Serum (Gibco), and 1% Penicillin/Streptomycin (Gibco) in a 15 mL falcon tube. Cells were spun down at 900 rpm for 5 minutes in a Jouan B3.11 centrifuge. Media was removed and replaced with 4 mL of fresh media, cells were resuspended, and then transferred to a T25 cell culture flask (Corning). The flask of cells was grown at 37 °C with 5% CO_2.

HEK293 cells were passaged by splitting 1/6 when cells reached ~90% confluency. As HEK293 cells are adherent, dissociation was accomplished by performing trypsinization, which consisted of the complete removal of media followed by addition of 2 mL of 0.25% Trypsin-EDTA phenol red (Gibco) and 2 minute incubation at 37°C. Once cells were completely dissociated from the flask, I transferred the 2 mL of Trypsin-EDTA cell mixture to a 15 mL falcon tube containing 6 mL of fresh media to deactivate the Trypsin, and spun down at 900 rpm for 5 minutes. The media-Trypsin solution was then aspirated off of the pellet of cells, and the cells were then resuspended in fresh media. I transferred cells in media were to 6 fresh T25 flasks (4 mL total of cell-media mixture per flask). Media and trypsin-EDTA were always pre-warmed to 37 °C in a waterbath prior to use.

TransIT-LT1 Transfection Reagent by Mirus was used and the manufacturer's protocol was followed for flasks of cells that underwent transfection with DNA plasmid. I performed all

transfections on cells grown to ≥70% confluency in T25 flasks. Briefly, 400 µl of Opti-MEM I Reduced-Serum Medium (Gibco), 12 µl if Trans-IT LT1 Reagent, and 6 µg of plasmid DNA (diluted in nuclease free water supplied by Integrated DNA Technologies) was warmed to room temperature, combined in a 1.5 mL sterile centrifuge tube, allowed to sit at room temperature for 20-30 minutes, and added drop-wise to media directly in the flask of cells. All plasmid DNA used for transfection was purified using the QIAGEN® Plasmid Maxi Kit according to the manufacturer's protocol to yield endotoxin-free plasmid preps. Plasmid DNA constructs consisted of the human *CECR2* ORF (wildtype or containing a single base-pair variation that results in a missense mutation) in the pEZY3 vector backbone (sections 2.14 and 2.15). The flask of cells was then grown at 37 °C and 5% CO_2 for ~36-48 hours and then harvested.

HEK293 cells were harvested at ≥80% confluency by first trypsinizing the cells and then adding the cells in trypsin to 6 mL of fresh media. Cells were spun down at 900 rpm for 5 minutes, media was removed, and pellet was washed with 3 mL of sterile 1X PBS. Cells were then centrifuged again at 900 rpm for 5 minutes, 1X PBS was removed, and the pellet was resuspended in 1 mL of 1X PBS. I then transferred this to a 1.5 mL microcentrifuge tube, spun down the cells at 4000 rpm for 5 min in an eppendorf 5417C centrifuge, and aspirated 1X PBS to leave only the pellet of cells behind. Cells were either flash frozen in liquid nitrogen and stored at -80 °C until further use, or immediately underwent protein extraction (section 2.17).

For long-term storage, HEK293 cells were kept in liquid nitrogen. To achieve this, a single T25 flask of cells at ~90% confluency was trypsinized, added to 6 mL of fresh media in a 15 mL falcon tube, and spun down at 900 rpm for 5 minutes. I then aspirated cell media, resuspended the cell pellet in cell media with 10% dimethyl sulfoxide (DMSO) and transferred this to a 2 mL screwcap cryogenic vial (Corning). Cryogenic vials containing cells were packaged by wrapping in absorbent material and placing in a styrofoam container prior to overnight incubation at -80 °C. This was done to achieve a slow-freeze to prevent damage to the cells. The next day, once cells were frozen, I transferred the cryogenic vials to liquid nitrogen for long-term storage. One vial of cells was revived to ensure cells were viable after freezing and storage in liquid nitrogen.

2.17: Protein extraction and quantification

At least 1 flask of cells (≥80% confluency) was used for all experiments requiring protein (1 flask will provide just enough protein for a single immunoprecipitation experiment, 1 flask will provide enough protein for multiple western blot experiments). I extracted protein from harvested HEK293 cells by resuspending cells in 300 µl of freshly prepared non-denaturing lysis buffer (20 mM Tris-HCl pH 8, 420 mM NaCl, 10% glycerol, 1% IGEPAL, 2 mM EDTA, 1X protease inhibitor) per T25 flask of cells. Cell lysate mixture was rocked at 4 °C for 30 minutes, and then centrifuged at 15000 rpm at 4 °C in a Beckman Microfuge E™. Protein extract (supernatant) was transferred to a clean 1.5 mL centrifuge tube and was either flash-frozen in liquid nitrogen and stored at -80 °C for future use or was used immediately for protein quantification followed by immunoprecipitation (section 2.18) and/or Western blot analysis (sections 2.19 and 2.20). Protein was also extracted from a single pooled sample of eleven E8.5 FVB/N mouse embryos that were flash-frozen in liquid nitrogen and stored at -80 °C. The embryos were thawed on ice and 250 µl of lysis buffer was added to the sample. I homogenized embryo tissue in a 1.5 mL centrifuge tube using a disposable homogenization pestle. The remainder of the procedure was conducted as described above for HEK293 cells.

To determine protein concentration, 10 µl of protein extract was diluted 1:10 in lysis buffer (100 µl final volume) and protein was quantified with the DC™ Protein Assay (Bio-Rad) according to the manufacturer's protocol. Absorbance (wavelength 750 nm) was measured on a Genesys 10uv (Thermo Electron Corporation) spectrophotometer. Protein extracts were then diluted to equal concentrations for future applications.

2.18: CECR2 immunoprecipitation

CECR2 containing complexes were immunoprecipitated from freshly extracted HEK293 protein. Initially, affinity purified rabbit polyclonal anti-CECR2 antibody (generated in the lab by Farshad Niri) was bound to Dynabeads® (Novex) by removal of supernatant from 10 µl of Dynabeads® in a 1.5 mL centrifuge tube with the aid of a magnet to sequester the Dynabeads®, resuspension of the Dynabeads® in 200 µl of 1X PBS containing 0.02% Tween-20, and addition of anti-CECR2 antibody to a final dilution of 1:2000. The Dynabeads® were incubated in

primary antibody solution for 10 minutes at room temperature with rocking, then the supernatant was discarded and washed with 200 µl of fresh 1X PBS containing 0.02% Tween-20, which was then removed from the Dynabeads®. Antigen solution, which contained IP buffer (20 mM Tris-HCl pH 8, 150 mM NaCl, 10% glycerol, 1% IGEPAL, 2 mM EDTA, 1X protease inhibitor) and protein extract (~1500 µg of protein), was then added to the Dynabeads®. The minimum total volume of antigen solution used per reaction was 500 µl to allow for sufficient motion of Dynabeads® during rocking, which was done at 4 °C for 2 hours to allow CECR2 containing complexes to bind anti-CECR2 antibody on the Dynabeads®. To remove excess protein extract, the Dynabeads® underwent 10 washes, 5 minutes each, in 500 µl of high salt lysis buffer containing Tween-20 (20 mM Tris-HCl pH 8, 420 mM NaCl, 10% glycerol, 1% IGEPAL, 2 mM EDTA, 1X protease inhibitor, 0.1% Tween-20) at 4 °C with rocking. CECR2 containing complexes were then eluted from the Dynabeads® and denatured by addition of 10 µl of 4X NuPAGE™ loading buffer (Invitrogen) and 10 µl of 1M β-mercaptoethanol followed by a 5 minute incubation in boiling water. The eluted sample was separated from the Dynabeads® by removal of the sample with the aid of a magnet. The sample was transferred to a fresh 1.5 mL centrifuge tube and either used immediately for protein electrophoresis or stored at -20 °C.

2.19: Protein electrophoresis

Protein extracts that did not undergo immunoprecipitation (described in previous section) were prepared for electrophoresis by combining extracts with 1X final concentration of NuPAGE™ loading buffer (Invitrogen) and 100 mM final concentration of β-mercaptoethanol. Protein was then denatured by incubating samples in boiling water for 5 minutes. Prepped samples were either used immediately for protein electrophoresis or stored at -20 °C.

Protein samples were electrophoresed according to the previously described Tris-glycine SDS-PAGE method [182]. Protein samples (5-10 µg) were loaded onto and electrophoresed through a polyacrylamide gel that consisted of a 4% acrylamide stacking gel (12.5% 37.5:1 acrylamide/bisacrylamide (BioRad), 50 mM Tris-HCl pH 6.8, 0.1% SDS, 0.1% ammonium persulfate, and 0.1% TEMED (Sigma-Aldrich)) and a 7.5% resolving gel (18.75% 37.5:1 acrylamide/bisacrylamide (BioRad), 50 mM Tris-HCl pH 8.8, 0.1% SDS, 0.1% ammonium persulfate, and 0.1% TEMED (Sigma-Aldrich)). An aliquot of PageRuler Prestained Protein

Ladder (Thermo Scientific) was loaded into a minimum of 1 well for each electrophoresis run. Running buffer consisted of 25 mM Tris, 0.19 M glycine, and 1% SDS. Voltage was set to 120 V for electrophoresis through the stacking gel and 175 V for electrophoresis through the resolving gel. I used a Mini Protean III™ electrophoresis system (Bio-Rad) apparatus by following the manufacturer's instructions.

2.20: Western blot analysis

Protein samples separated by electrophoresis were transferred to 0.45 µm pore size polyvinylidene fluoride (PVDF) membrane (Immobilon®, Millipore) for 30 minutes at 350 mA in transfer buffer (25 mM Tris, 190 mM glycine, 10% methanol). The PVDF membrane was first activated by rinsing the membrane in 100% methanol followed by a 5 minute incubation in transfer buffer. The transfer was set up in a Mini Protean III™ submerged tank wet transfer unit (Bio-Rad) following the manufacturer's protocol and was run for 30 minutes at 350 mA. Successful transfer was visualized by the presence of PageRuler Prestained Protein Ladder (Thermo Scientific) markers on the PVDF membrane. The membrane was washed in TBST (Tris-buffered saline, 0.05% Tween-20) for 5 minutes, and then blocked for 45 minutes in 5% skim milk (Carnation) in TBST. After blocking, the membrane was rinsed in TBST, and then incubated in primary antibody diluted in 5% skim milk in TBST overnight at 4 °C. Primary antibodies used for western blot analysis were (1) affinity purified rabbit polyclonal anti-CECR2 antibody diluted to 1:50000, (2) sheep polyclonal anti-SNF2L antibody (a gift from Dr. David J. Picketts) diluted to 42 µg/mL, (3) affinity purified rabbit polyclonal anti-DNMBP (proteintech™) diluted to 1:5000, (4) affinity purified rabbit polyclonal anti-MMS19 (Proteintech™) diluted to 1:500, and (5) mouse monoclonal anti-α-Tubulin (Sigma) diluted to 1:10000 as a loading control. Antibodies were used separately. In the event that a single PVDF membrane was exposed to multiple primary antibodies, the membrane was cut and the appropriate sections were placed in the corresponding primary antibody solutions. The next morning, I washed the membrane 3 X 5 minutes in TBST, and then incubated it in secondary antibody diluted in 5% skim milk in TBST for 1 hour at room temperature. All secondary antibodies used for western blot analysis were supplied already conjugated to horseradish peroxidase (HRP) for detection purposes. Secondary antibodies used were (1) goat anti-rabbit IgG-HRP conjugate (Bio-Rad) diluted to 1:5000, (2)

rabbit anti-sheep IgG-HRP conjugate (Santa Cruz Biotechnology) diluted to 1:5000, and (3) goat anti-mouse IgG-HRP conjugate (Sigma) diluted to 1:10000. All antibodies were used separately. The membrane was rinsed in TBST, and washed 3 X 10 minutes in TBST. HRP was detected by incubating the membrane in a 1:1 ratio of SuperSignal® West Pico Stable Peroxide Solution and SuperSignal® West Pico Luminol/Enhancer Solution (Thermo Scientific) for 5 minutes. The membrane was then wrapped in Saran wrap, exposed to X-ray film (FUJIFILM) and developed in a Kodak X-OMAT 2000 automated developer.

2.21: MOD 5/31, MOD 5/+, and MOD 31/+ crosses

Megan Kooistra, a former graduate student, created the MOD 5 and MOD 31 sub-interval congenic mouse lines as previously described [165]. Each sub-interval congenic line contained a segment of the chromosome 19 modifier region that is homozygous FVB/N on an otherwise BALB/cCrl genetic background (~98.4% BALB/cCrl) with the MOD 5 and MOD 31 segments partially overlapping. I performed the MOD 5/31 cross by crossing MOD 5 and MOD 31 mice heterozygous for the *Cecr2*^{GT45bic} allele to each other to produce embryos that were all at least heterozygous FVB/N for the entire chromosome 19 modifier region (homozygous FVB/N for the overlapping region only) and were also either homozygous wildtype, heterozygous, or homozygous mutant for the *Cecr2*^{GT45bic} allele. For the MOD 5/+ cross, I crossed MOD 5 mice heterozygous for the *Cecr2*^{GT45bic} allele to congenic *Cecr2*^{GT45bic} BALB/cCrl mice heterozygous for the *Cecr2*^{GT45bic} allele to produce embryos that were all heterozygous for the FVB/N MOD 5 segment and were either homozygous wildtype, heterozygous, or homozygous mutant for the *Cecr2*^{GT45bic} allele. I conducted the MOD 31/+ cross in the same way as the MOD 5/+ cross. Embryos were collected and exencephaly penetrance was determined as described in section 2.2 and embryos were genotyped for the *Cecr2*^{GT45bic} allele as described in section 2.5. The crossing schemes for producing MOD 5/31, MOD 5/+ and MOD 31/+ progeny is shown in figure 2.21.1, and genetic backgrounds of the chromosome 19 modifier region for MOD 5, MOD 31, and the embryos resulting from the above crosses are displayed in Figure 2.21.2.

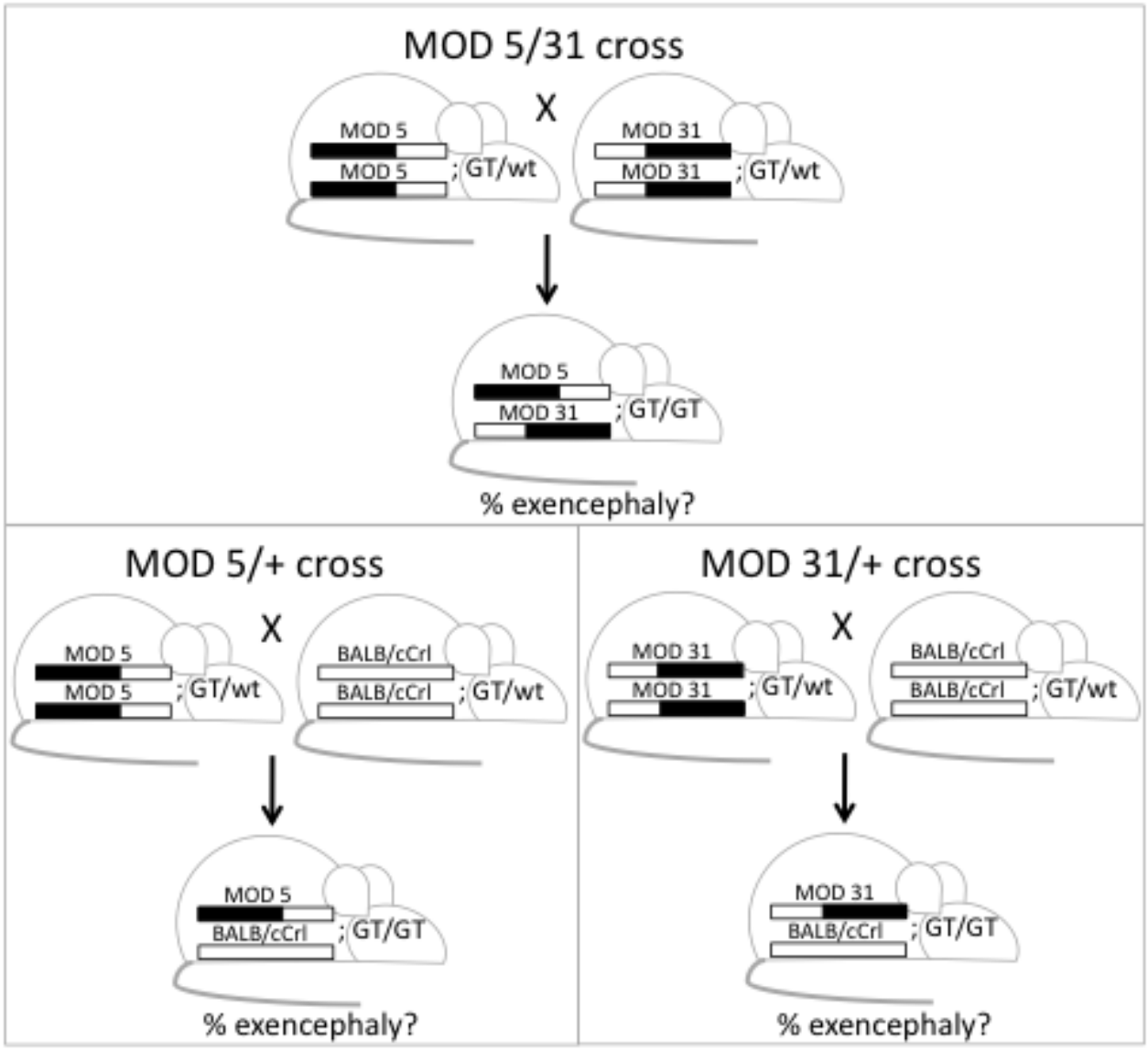


Figure 2.21.1: MOD 5/31, MOD 5/+, and MOD 31/+ crossing scheme. Black and/or white bars represent the chromosome 19 modifier region. Black color represents FVB/N genetic background and white color represents BALB/cCrl genetic background. In the MOD 5 and MOD 31 lines, the remainder of chromosome 19 and the rest of the genome is BALB/cCrl. The BALB/cCrl mice (all-white bars) used for the MOD 5/+ and MOD 31/+ crosses were congenic *Cecr2*^{GT45bic} BALB/cCrl. Only progeny homozygous for the *Cecr2*^{GT45bic} allele are shown, as exencephaly was not seen in any of the heterozygous or homozygous wildtype progeny. The *Cecr2*^{GT45bic} allele is represented by "GT". The wildtype *Cecr2* allele is represented by "wt".

Marker	Location (Mb)	MOD 5	MOD 31	MOD 5/31	MOD 5/+	MOD 31/+
rs13483563	21.647					
rs6259521	24.021					
rs31272775	26.796					
rs6381572	31.478					
D19Mit88	37.331					
rs3677115	40.658					
rs30714351	40.748					
rs30418423	41.104					
rs8250750	43.873					
rs13483644	45.853					
rs30659530	47.627					
rs31199765	48.286					
rs13483654	48.457					
rs13483669	52.277					
rs3709671	55.691					
rs3712604	60.010					

Figure 2.21.2: MOD 5, MOD 31, MOD 5/31, MOD 5/+, MOD 31/+ chromosome 19 modifier region genetic background. Black represents FVB/N and white represents BALB/cCrl. Half black-half white boxes represent regions heterozygous for FVB/N and BALB/cCrl. The remainder of chromosome 19 and the rest of the genome were of BALB/cCrl origin.

2.22: Congenic *Cecr2*^{GT45bic} BALB/cCrl x BALB/cJ crossing scheme

Previous work in our lab has shown *Arhgap19* to contain a single base pair insertion in exon 6 resulting in a homozygous nonsense mutation (herein referred to as *Arhgap19*^{Ex6non}) in BALB/cCrl but not in the closely related sub-strain BALB/cJ obtained from Jackson laboratories [165]. To test if this nonsense mutation is a modifier of *Cecr2*-associated exencephaly, a dihybrid cross was performed with the aid of Leanne Donnelly, a former undergraduate student in the lab.

We crossed congenic *Cecr2*^{GT45bic} BALB/cCrl heterozygous for the *Cecr2*^{GT45bic} allele and homozygous for the *Arhgap19* nonsense mutation to BALB/cJ, which are homozygous wildtype for both *Cecr2* and *Arhgap19*. We then selected progeny that were heterozygous at both alleles and crossed them to each other (Figure 2.22.1). The resulting embryos were collected, scored for exencephaly, and genotyped for the *Cecr2*^{GT45bic} allele. Only embryos that were homozygous mutant for the *Cecr2*^{GT45bic} allele were genotyped for the *Arhgap19* nonsense mutation. This was done by PCR amplification of exon 6 with primers Ar19Ex6-F and Ar19Ex6-R (Appendix A), which generated a 206 bp amplicon, followed by Sanger sequencing in the forward direction (section 2.7). Exencephaly penetrance in *Cecr2*^{GT45bic} homozyogous mutant embryos that contained at least one wildtype copy of *Arhgap19* was compared to *Cecr2*^{GT45bic} mutant embryos that were also homozygous mutant for the *Arghap19* nonsense allele. Leanne Donnelly provided much assistance with setting up matings, dissections, phenotyping, genotyping, and preliminary analyses.

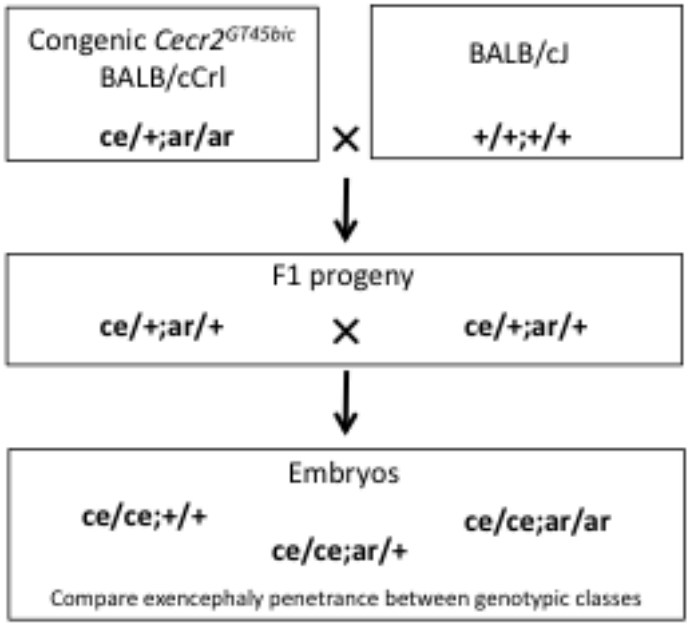


Figure 2.22.1: Crossing scheme to test if the *Arhgap19* nonsense mutation contributes to exencephaly susceptibility. Only relevant F1 progeny and embryo genotypes are shown. The *Cecr2*^{GT45bic} allele is represented by "ce", the *Arhgap19*^{Ex6non} allele is represented by "ar", wildtype alleles are represented by "+".

2.23: Total RNA extraction and cDNA amplification for quantitative real-time PCR (qRT-PCR)

A previous graduate student, Megan Kooistra, collected embryos at the time of neurulation (~12-14 somites), individually placed embryos in 2 mL centrifuge tubes, flash-froze embryos in liquid nitrogen, and stored embryos at -80 °C. I then extracted total RNA from frozen embryos using an RNeasy Lipid Tissue MiniKit (QIAGEN®) according to the manufacturer's protocol. RNA was quantified using a Nanodrop 1000 (Thermoscientific) and quality was assessed with a BioAnalyzer (Agilent Technologies, Inc).

I synthesized cDNA from RNA extracts by using a Superscript® VILO™ cDNA Synthesis Kit (Invitrogen) according to the manufacturer's protocol. 1 µg of RNA from a single embryo was used for each 20 µl cDNA synthesis reaction. Reactions were conducted on a Peltier Thermocycler-200 (MJ Research) and cycling conditions were 25 °C for 10 minutes, 42 °C for 1 hour, 85 °C for 5 minutes, and 4 °C hold. cDNA was stored at -20 °C while not in use.

2.24: Quantitative real-time PCR (qRT-PCR) of candidate modifier genes

Assays for qRT-PCR were designed using the Universal ProbeLibrary Assay Design Center web-based software by Roche Life Science using default settings (https://lifescience.roche.com/webapp/wcs/stores/servlet/CategoryDisplay?catalogId=10001&tab=Assay+Design+Center&identifier=Universal+Probe+Library&langId=-1). Dr. Kirst King-Jones (University of Alberta) provided qRT-PCR probes, which were originally obtained from the Universal Probe Library from Roche Life Science, and Integrated DNA Technologies supplied the primers (see Appendix A for primer sequences and accompanying probes). As all qRT-PCR experiments were conducted to compare gene expression in neurulating embryos between BALB/cCrl and FVB/N mouse strains, I Sanger-sequenced cDNA regions targeted by the primer/probe assay in both strains to ensure no DNA sequence variants were present that could interfere with primer/probe binding (sequencing primers are listed in Appendix A). This sequence validation was not performed for three assays (see Appendix A); however, no known variants exist in mouse under the primer/probe sets for these assays (www.ensembl.org).

All qRT-PCR experiments were run using a 96-well plate format on a StepOnePlus Real-Time PCR System (Applied Biosystems). Reactions were all a total volume of 10 µl and

contained 1X TaqMan Universal PCR Mastermix (Applied Biosystems), 400 nM of each primer (forward and reverse), 100 nM of probe, and 2.5 µl of diluted cDNA sample (generated as described in section 2.23). Reactions were prepared in a MicroAmp Fast Optical 96-Well Reaction Plate (Applied Biosystems), loaded onto the StepOnePlus Real-Time PCR System, and cycling conditions were (1) 95 °C for 10 minutes, (2) 95 °C for 15 seconds, (3) 60 °C for 1 minute, (4) repeat steps 2-3 40 times, and (5) 4 °C hold. I first tested qRT-PCR assays for each target gene for similar amplification efficiency as the endogenous control gene *Elmo2*. This was accomplished by generating a standard dilution curve, which consisted of a single cDNA sample undergoing four dilutions (1:4, 1:16, 1:64 and 1:256) and qRT-PCR reactions for each dilution (three technical replicates for each qRT-PCR reaction) being run for *Elmo2* and each target gene tested. Successfully validated qRT-PCR assays were then used for qRT-PCR experiments comparing BALB/cCrl and FVB/N gene expression differences in neurulating embryos. This was accomplished by setting up qRT-PCR reactions for the target gene along-side the *Elmo2* endogenous control and running the qRT-PCR experiment as described above for assay validation with the exception that 3 biological replicates for each mouse strain (BALB/cCrl and FVB/N) were tested and all cDNA templates were diluted 1:10. There were three technical replicates for each qRT-PCR reaction. Results for qRT-PCR assay validations and experiments were analyzed using StepOne™ Software v2.0.

I performed a qRT-PCR experiment for the target gene *Foxd4* on four mouse strains, BALB/cCrl, FVB/N, 129S2, and C57BL/6. Megan Kooistra designed primers for *Foxd4* as described above, prepared cDNA from three biological replicates for each mouse strain at the time of neurulation (age-matched by somite number), and stored cDNA at -20 °C. Using these samples, I set up three technical replicates for each qRT-PCR reaction. For this assay, I used SYBR® Green Dye instead of TaqMan®, therefore a probe was not required. Each 10 µl reaction consisted of 1X SYBR® Green mastermix (Kapa Biosystems), 400 nM of each primer, and 2.5 µl of cDNA template (1:10 dilution). Results were analyzed with StepOne™ Software v2.0.

To determine gene expression differences, cycle threshold (CT) data obtained from the StepOne™ Software v2.0 analysis for the three technical replicates was averaged for each biological replicate. The difference in CT (ΔCT) between the target gene and the endogenous control, *Elmo2*, was calculated by $CT_{\text{target gene}} - CT_{\text{Elmo2}}$. Each qRT-PCR experiment required a reference sample, which was one of the three BALB/cCrl samples for all qRT-PCR experiments.

The ΔCT for each biological replicate was then compared to the ΔCT of the reference sample in order to obtain the ΔΔCT (ΔCT_{sample} - $\Delta CT_{reference}$) and ΔΔCT for the three biological replicates were averaged. This average was used to obtain fold-change (FC) between strains, where FC = $2^{\wedge\Delta\Delta CT}$. Statistical analysis of gene expression was calculated using type 2, 2-tailed t-Test and 95% confidence intervals were plotted as error bars.

2.25: Generation of FVB/129P2 (*Arhgap19*$^{Gt(YHD020)Byg}$) circling mice

The FVB/129P2 line was homozygous for the genetrap mutation *Arhgap19*$^{Gt(YHD020)Byg}$ in intron 2. The founder male was generated by BayGenomics as previously described [165]. A pedigree showing the generation of the line is shown in Figure 5.2.1 (Chapter 5). I crossed the chimeric founder male mouse to FVB/N females (Charles River Laboratories) to generate progeny heterozygous for the *Arhgap19*$^{Gt(YHD020)Byg}$ allele. These progeny were then backcrossed to FVB/N mice or interbred to produce homozygotes. The FVB/129P2 line resulted from a single breeder pair homozygous for the *Arhgap19*$^{Gt(YHD020)Byg}$ allele. The genetic background of this line consists of 81.25% FVB/N and 18.75% 129P2 (from the original 129P2 ES cell line containing the *Arhgap19* mutation).

2.26: *Arhgap19*$^{Gt(YHD020)Byg}$ genotyping

In order to determine the genotype of mice carrying the *Arhgap19*$^{Gt(YHD020)Byg}$ allele, I performed two PCR reactions on extracted genomic DNA. Attempts to generate a 3 primer multiplexed PCR were unsuccessful; therefore, two individual PCR reactions were performed on each sample. One PCR reaction utilized primers Ar19In2WT-F and Ar19In2WT-R within intron 2 to amplify the wildtype allele (457 bp amplicon). The other PCR reaction utilized primers Engeo-F and Engeo-R located within the genetrap cassette to amplify the mutant allele (960 bp amplicon). Primers were supplied by Integrated DNA Technologies and primer sequences can be found in Appendix A. PCR cycling programs are displayed in Table 2.26.1. PCR products were pooled and electrophoresed on a 2% agarose gel.

Table 2.26.1: PCR programs for *Arhgap19*$^{Gt(YHD020)Byg}$ genotyping. Programs are named after their respective primers.

Ar19In2WT			**Engeo**		
Temperature	Time (min:sec)	Repeat	Temperature	Time (min:sec)	Repeat
95°C	3:00	--	95°C	3:00	--
94°C	0:20		94°C	0:20	
67°C; -1°C per cycle	0:20	11X	64°C; -1°C per cycle	0:20	10X
72°C	1:20		72°C	1:00	
94°C	0:20		94°C	0:20	
57°C	0:20	31X	55°C	0:20	30X
72°C	1:20		72°C	1:00	
72°C	10:00	--	72°C	10:00	--
4°C	HOLD	--	4°C	HOLD	--

2.27: Phenotyping of circling behavior in FVB/129P2 mice

Male mice in our colony are most often housed in social groups of 2-5 mice per cage. In June 2011 we introduced igloo-style running-wheels for environmental enrichment to all male and female non-breeder cages in our mouse colony, regardless of age after weaning. In November 2011 we noted the first circling FVB/129P2 male, and eventually determined that a percentage of the FVB/129P2 male mice housed with running-wheels developed circling behavior. Once we suspected a link to the running-wheels, we removed the wheel from most of the colony cages, except for the experiments discussed in Chapter 5 with FVB/129P2 mice, which were either given wheels (n = 278) or cardboard houses (n = 43) upon weaning, or cardboard houses upon weaning followed by wheels at 6 to 12 weeks of age (n = 40). I observed male FVB/129P2 mice for circling behavior at least three times a week starting at age of weaning and up until 3 months of age. All observations were made during the light cycle. Although mice are less active during the light cycle, they do often wake up and become active when researchers and technicians are in the room. A mouse with circling behavior almost always displayed circling behavior while awake, which was easily observed without any special efforts made by technicians or researchers in the room. However, when specifically looking for circling behavior, home cages were taken into a biosafety cabinet, opened, and mice were observed for approximately 5-10 minutes. A mouse was scored to have developed circling behavior if it demonstrated frequent bouts of unidirectional circling when active and preferentially or always

turned in one direction. Once established in the mouse, the phenotype was obvious and permanent. Ear-notch numbers were used to identify individual circling mice and ear-notch skin biopsies were used for DNA extraction and genotyping (sections 2.4 and 2.5). It is possible that more mice would have circled had they been observed longer than 3 months – this was not tested. The male FVB/129P2 cages were also tagged so that technicians and other lab members could leave a note if any abnormal behavior was apparent. Significant differences in penetrance of the circling phenotype between different groups of mice were calculated using the χ^2 test-of-independence.

Increased fighting within home cages also occurred in FVB/129P2 cages of males housed with wheels. This observation was originally noted as wounding, particularly to the hind trunk and/or tail. Animals in these cages were separated and singly housed with wheels. To avoid injury to subsequent animals housed with wheels, technicians observed all socially housed males daily and particularly watched males after cage changing. If mice were fighting, all the males were separated and housed singly with wheels before wounding occurred. There was no attempt to determine which of the cage-mates were the aggressors, victims or spectators; therefore, aggression was not scored for individual mice. However, it was recorded if a mouse was in a cage that required separation due to fighting.

2.28: Phenotyping pipeline to determine inability to sense gravity due to inner ear defects

Three severely-circling male mice and four male control mice underwent a series of behavioral tests to screen for vestibular dysfunction using methods described in an established phenotyping pipeline [183]. One of the three circling mice was the male that parented the FVB/129P2 line and two of his brothers served as controls. The other two circling mice and two control mice were F1 progeny of the breeder pair that founded the FVB/129P2 line. In brief, the mice underwent (1) the trunk curl test, where mice with vestibular defects will preferentially curl their trunks rather than reach for a presented surface when held by their tail; (2) the contact righting test, where mice with vestibular defects will not right themselves when turned onto their backs while in a plastic tube; and (3) the swim test, where mice with vestibular defects demonstrate underwater tumbling. With the aid of Kacie Norton, a graduate student in the lab, all behavioral tests were video recorded. I reviewed videos to calculate time spent engaging in each

type of behavior scored. For the trunk curl test, I held mice by the tail and the first 18 seconds were scored for the amount of time spent curling their trunk forward, curling their trunk sideways, or hanging straight down/reaching towards presented surface. To be considered curling, the mouse trunk had to be bent by 90 degrees or fore and hind paws clasped. For the contact-righting test, mice were placed inside a clear plastic tube open at both ends with a diameter of 2.8 cm. Once the mouse was inside, I turned the tube so that the mouse was on its back. If the mouse righted itself, then the tube was turned again so that the mouse was once again positioned in its back. Time spent with the head upside down, right side up, or sideways was scored over a total of 20 seconds starting from the time the mouse was turned onto its back. For the swim test, mice were placed in a tank of water (tank dimensions 50.8 cm x 40.6 cm x 20.3 cm, water temp. 25 °C) and allowed to swim for 2 minutes. Time spent swimming normally, swimming in circles, or immobile floating in the first minute was scored. Amount of time spent underwater tumbling was not scored, as the mouse would be immediately rescued if in this situation. Only underwater tumbling was looked for in the second minute, which did not occur for any of the mice tested.

2.29: High performance liquid chromatography (HPLC) analysis of biogenic amines and amino acids in left and right brain hemispheres

FVB/129P2 mouse whole brain samples were taken from six male mice that circled (housed with wheels), five male mice that did not circle (housed with wheels), and six male mice that did not circle and were housed with cardboard houses. All mice were 90 days old at the time of sample collection. Mice were socially housed (2-5/cage), but were separated and housed singly if fighting was seen in the cage, as described in section 2.27. All 6 circling mice housed with wheels were separated due to aggression, where 3 of the 6 circling mice spent their last 21 days singly housed, and the other 3 mice spent their last 9 days singly housed. All non-circling mice except one that were housed with wheels required separation due to fighting within the cage, where 2 of them spent their last 45 days singly housed, one spent his last 21 days singly housed, and one spent his last 9 days singly housed. The one non-circling mouse housed with a wheel remained socially housed with one circling sibling (which was not included in this experiment) for the entire 90 days. Fighting was not observed in all 6 mice with cardboard housing and

therefore they were socially housed for the entire 90 days. Mice were euthanized by cervical dislocation. I then removed the whole brain and cut it into left and right hemispheres. The brain samples were then flash frozen in isopentane cooled on dry ice and stored at -80 °C until further use. I brought brain samples to Dr. Glen Baker's lab where technician Gail Rauw conducted sample processing for HPLC. Left and right brain hemispheres were independently homogenized in ice-cold water and then divided into two portions. In order to study biogenic amines, one-tenth the volume of 1N perchloric acid was added to one portion and precipitated protein was removed with centrifugation. Concentrations of biogenic amines and their metabolites (dopamine, 3,4-dihydroxyphenylacetic acid, noradrenaline, 3-methoxytyramine, homovanillic acid, serotonin, 5-hydroxyindoleacetic acid) and the amino acids tyrosine and tryptophan were measured by HPLC with electrochemical detection as previously described [184]. To study additional amino acids, ethanol was added to the other portion, and following removal of precipitated protein by centrifugation, concentrations of the amino acids (aspartate, glutamate, L-serine, D-serine, glutamine, glycine, arginine, taurine, alanine, gamma-aminobutyric acid) were measured using HPLC with fluorimetric detection following derivatization, as previously described [185]. Gail Rauw provided the results to me, and I determined statistically significant differences in biogenic amines, amino acids, and metabolites between groups of mice by averaging left and right hemisphere concentrations for each mouse and then performing a series of one-way ANOVAs followed by Bonferroni correction.

2.30: *Drosophila* stocks and husbandry

All *Drosophila* used in this study were maintained in foam capped vials (95 mm height, 25 mm radius) containing ~5 mL food. The fly food recipe contained the following ingredients: 50 L of water, 550 g of agar, 215.2 g of sodium phosphate monobasic (FW = 137 g/mol), 135.1 g of sodium phosphate dibasic (FW = 141.96 g/mol), 787.2 g of yeast, 3321 g of cornmeal, 2097 g of malt, 455 g of soy flour, 3850 mL of molasses, 145 mL of propionic acid, 145 g of methyl paraben, and 500 mL of ethanol, which was combined, cooked for 40 minutes, and distributed into vials. I maintained fly stocks by tipping flies into new vials on a weekly basis.

A summary of all fly stocks is provided in Table 2.30.1. All GAL4 driver stocks were a gift from Dr. Shelagh Campbell (University of Alberta), who originally ordered them from the

Bloomington Stock Center (Indiana University). All RNAi stocks were a gift from Dr. John Locke (University of Alberta), who originally obtained them from the Vienna RNAi stock center. The remaining stocks were a gift from Dr. John Locke, and/or were generated in house from existing stocks.

Experiments contained herein were directed towards the characterization of the *Drosophila Cecr2* homologue, *dikar*, which contains 10 exons and is located on the left arm of chromosome 3. The fly stocks that carried either the $dikar^{3}$ mutant allele, which is a 1423 bp deletion of exons 1A and 1B that encodes the 5' untranslated region, plus upstream sequence that may contain promoter elements, or the $dikar^{5}$ allele, which is a 1581 bp deletion that encompasses all of exons 2, 3, 4 and 13 bp of exon 5, were both generated in this lab as previously described [186]. The $dikar^{3}$ and $dikar^{5}$ alleles are homozygous viable; however, to produce heterozygotes, I maintained secondary stocks with the third chromosome balancer *TM3* present in a proportion of the flies ($dikar^{3}$/*TM3* and $dikar^{5}$/*TM3*).

The parental stock containing the original *P{SUPor-P}dikar*^{KG00884} P element used to generate the $dikar^{3}$ and $dikar^{5}$ stocks was lost in the lab, but re-ordered from Bloomington. However, the P element bearing chromosome was not homozygous viable in this re-ordered strain, suggesting that the chromosome had acquired a homozygous lethal second-site mutation. This "parental" stock was not representative or useful in this state; the homozygous lethal needed to be removed. Therefore, I regenerated a homozygous viable *P{SUPor-P}dikar*^{KG00884} parental line from this stock by performing a series of backcrosses to $dikar^{3}$ and selecting for recombination events retaining the *P{SUPor-P}dikar*^{KG00884} allele (had a *w*+ marker) but without the second-site mutation (Figure 2.30.1). Backcrossing to $dikar^{3}$ for recombination also improved the likelihood of a similar genetic background on chromosome 3 and throughout the rest of the genome, which was desirable for future comparisons between all *dikar* stocks. I generated two new lines from 2 independent male founders that were homozygous viable for the *P{SUPor-P}dikar*^{KG00884} allele. Two secondary *P{SUPor-P}dikar*^{KG00884} stocks, one from each of the founding male stocks, were maintained with the third chromosome balancer *TM3* present in a proportion of the flies (*P{SUPor-P}dikar*^{KG00884}/*TM3*). These two *P{SUPor-P}dikar*^{KG00884} stocks, and not the original homozygous lethal stock, were used for future crosses and experiments.

The original chromosome without the *P{SUPor-P}dikar*^{KG00884} insert was not available, so I precisely excised the transgene insert from the *P{SUPor-P}dikar*^{KG00884} stock. Two precise excision stocks homozygous for a wildtype allele of *dikar* (*dikar*^{WT}) that were on a similar genetic background as *dikar*^{3} and *dikar*^{5} for comparison purposes were generated (Figure 2.30.2). The precise excision event was validated with Sanger sequencing (primers used are listed in Appendix A). Two new *dikar*^{WT} stocks were also maintained with the third chromosome balancer *TM3* present in a proportion of the flies (*dikar*^{WT}/*TM3*).

The *Df(3L)ED212/TM3* stock supplied by Bloomington was supposed to contain a deficiency that encompassed *dikar* (along with several other genes). PCR analysis using primers where at least one forward or reverse primer was within the *dikar*^{3} or *dikar*^{5} deletion indicated that the *Df(3L)ED212* deficiency does contain a wildtype allele of *dikar* (Appendix A). Therefore, a second deficiency stock, *Df(3L)ZN47*/TM3, was used for experiments requiring a deficiency.

Table 2.30.1: *Drosophila* stocks. All GAL4 stocks were a gift from Dr. Shelagh Campbell (University of Alberta). The rest of the stocks were a gift from Dr. John Locke (University of Alberta).

Name	Source	Notes
RNAi CG32394 24738/GD	Vienna	Targets *dikar*
RNAi CG32394 100383/KK	Vienna	Targets *dikar*
RNAi CG32393 107312/KK	Vienna	Targets *dikar*
RNAi CG17697 43077/GD	Vienna	Targets *frizzled*
RNAi CG17697 105493/KK	Vienna	Targets *frizzled*
RNAi CG8075 100819/KK	Vienna	Targets *Vang*
RNAi CG8075 7376/GD	Vienna	Targets *Vang*
HS-2-GAL4	Bloomington	Heat-shock promoter on 2nd chr.
HS-3-GAL4	Bloomington	Heat-shock promoter on 3rd chr.
en-GAL4	Bloomington	engrailed promoter
T279-GAL4	Bloomington	Maternal tubulin promoter
eyeless-GAL4	Bloomington	eyeless promoter
actin-2-GAL4	Bloomington	actin promoter on 2nd chr.

Name	Source	Notes
actin-3-GAL4	Bloomington	actin promoter on 3rd chr.
prd-GAL4	Bloomington	prd promoter on 3rd chr.
Df(3L)ZN47/TM3	Bloomington	Chromosome 3 deletion that includes *dikar*
Df(3L)ED212/TM3	Bloomington	Chromosome 3 deletion that was shown not to include *dikar*
Sbe Δ2-3/TM6	Bloomington	Transposase source to excise P element
ash^1 *e*/TM3	In house	Source of TM3 balancer moved into *dikar* stocks
Xa/TM6	In house	Used for P element precise excision cross
$dikar^3$	In house	Homozygous viable
$dikar^3$/TM3	In house	Used for IR
$dikar^5$	In house	Homozygous viable
$dikar^5$/TM3	In house	Used for IR
P{SUPor-P}dikar^{KG00884}	In house	Two stocks from 2 single homozygous viable founding males
P{SUPor-P}dikar^{KG00884}/TM3	In house	Two stocks created from *P{SUPor-P}dikar*^{KG00884}, used for IR
$dikar^{WT}$	In house	Two stocks created from *P{SUPor-P}dikar*^{KG00884}
$dikar^{WT}$/TM3	In house	Two stocks created from $dikar^{WT}$, used for IR

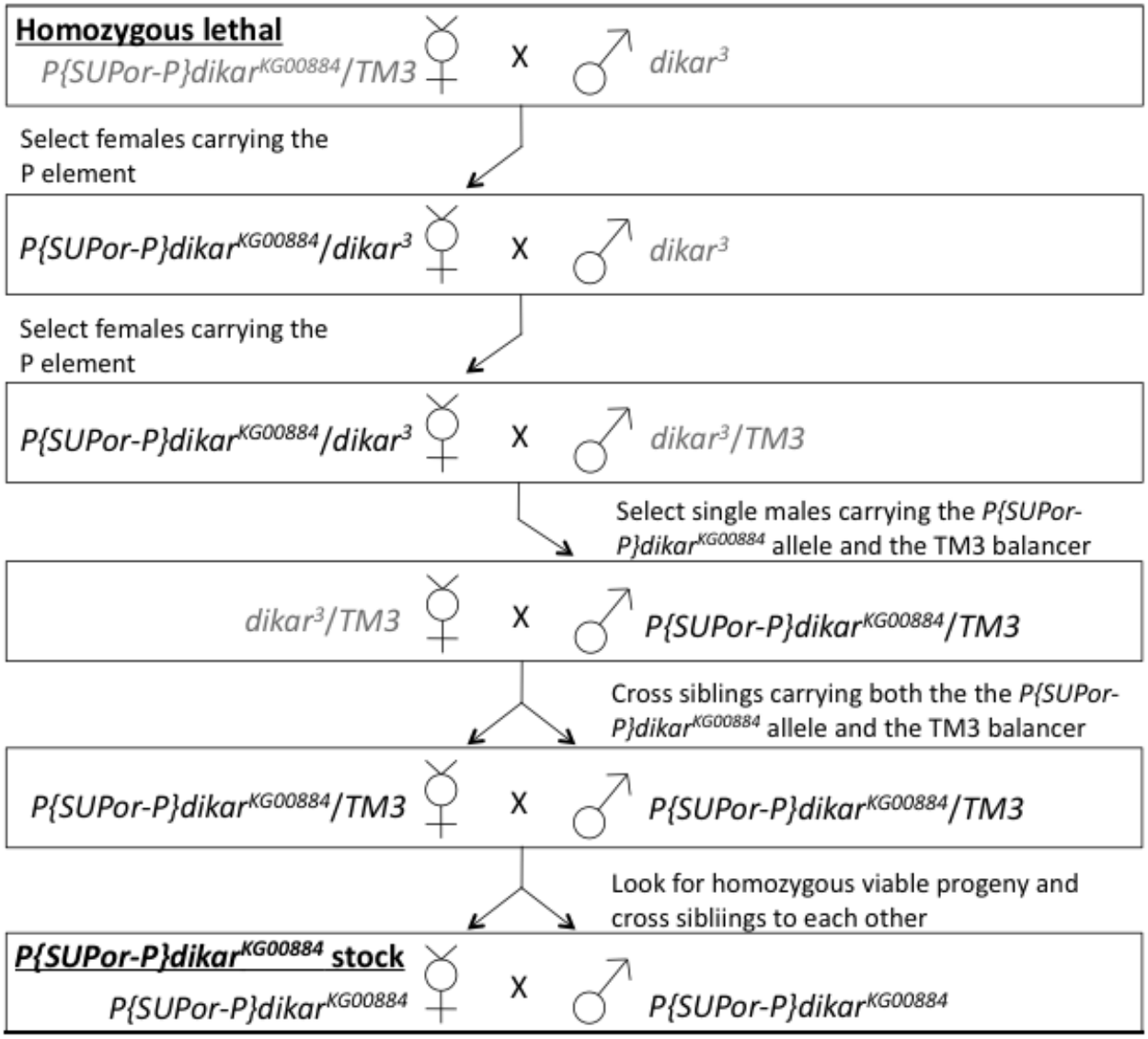


Figure 2.30.1: Crossing scheme to generate the homozygous viable *P{SUPor-P}dikar^{KG00884}* stock to act as a control. The *P{SUPor-P}dikar^{KG00884}* stock is the parental stock from which the *dikar^{3}* and *dikar^{5}* deletion mutant stocks were originally generated. The *P{SUPor-P}dikar^{KG00884}* allele contains a mini-*white* gene (red eyes) and a mini-*yellow* gene (tan body) as selectable markers. Arrows point to progeny selected from the previous cross. Genotypes in black text denote progeny selected from the previous cross, genotypes in grey text represent flies from an outside stock used to cross to the selected progeny.

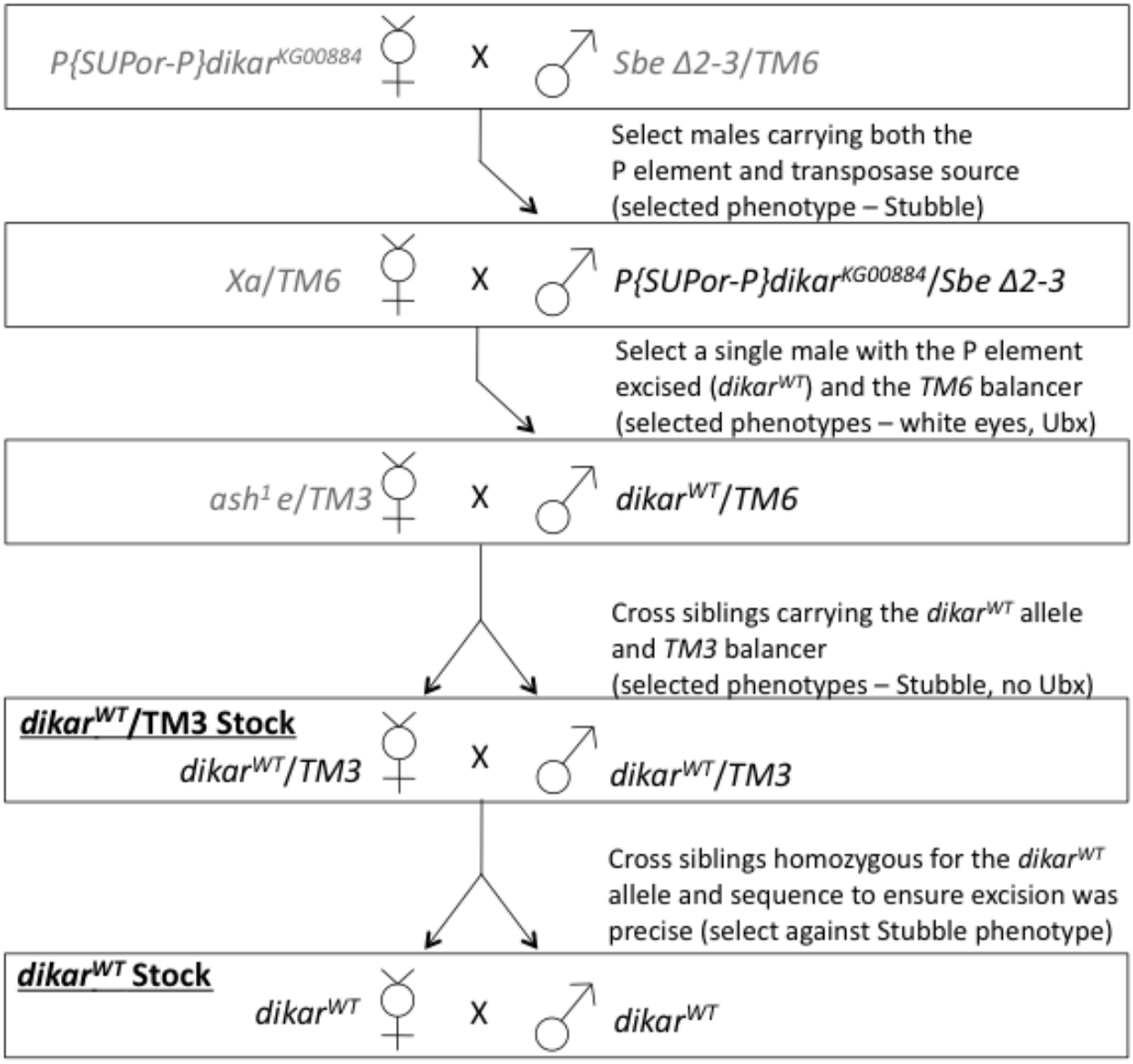


Figure 2.30.2: Crossing scheme to precisely excise the *P{SUPor-P}dikar^{KG00884}* P element in order to generate the *dikar^{WT}* stock. Selectable markers are indicated by "selected phenotypes" in brackets, with the Stubble phenotype consisting of short bristles, the white eye phenotype is white eyes instead of red, and the Ubx phenotype is an enlarged haltere. Arrows point to progeny selected from the previous cross. Genotypes in black text denote progeny selected from the previous cross, genotypes in grey text represent flies from an outside stock used to cross to the selected progeny.

2.31: RNAi in *Drosophila* to knock-down *dikar* mRNA and characterize phenotype

RNAi-mediated targeted gene silencing was achieved via utilization of the GAL4/UAS system in *Drosophila*. Briefly, I set up reciprocal crosses consisting of ~3 males and ~5 virgins for each RNAi experiment. One parent was homozygous or heterozygous for the GAL4 "driver" gene in which a tissue specific, conditional, or ubiquitously active promoter drove expression. The other parent was homozygous for a "responder" element containing a 300-400 bp inverted repeat DNA sequence that is complementary to the mRNA targeted for degradation. Recognition and binding of the driver GAL4 transcriptional activator to the upstream activation sequence (UAS) of the responder results in expression of the inverted repeat. The transcribed inverted repeat in the progeny heterozygous for both the GAL4 gene and the inverted repeat sequence results in the production of hairpin RNAs, which are then processed and bound to the targeted gene transcript by the RNAi machinery, resulting in gene-specific mRNA degradation. *Drosophila* stocks used for these crosses are summarized in Table 2.30.1 above, and a general crossing scheme and experimental concept is summarized in Figure 2.31.1.

Two of the GAL4 driver lines required heat-shock in order to express GAL4. All crosses requiring heat-shock were set up on the same day. Parents were removed from vials 9 days later and progeny were heat-shocked by submerging the bottom half of the vials in a 38 °C water bath and incubating for 30 minutes. Progeny were scored after pupal eclosure for planar cell polarity mutant phenotypes, which consisted of whorls in the wing bristles, misorientation of small thorax bristles, and misorientation of abdomen bristles. Other obvious phenotypes, such as pupal lethality and morphological defects, were also noted if observed in any progeny.

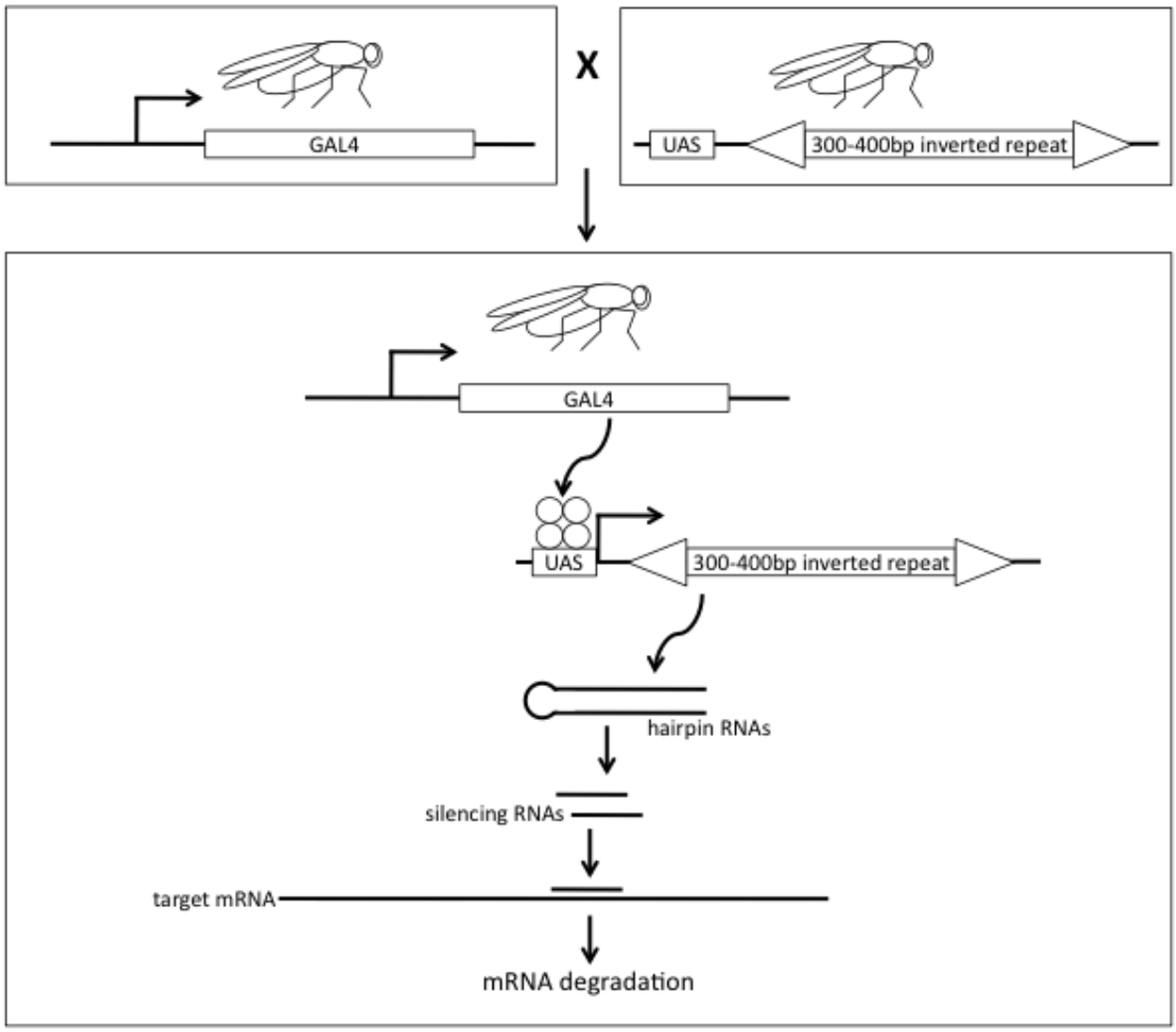


Figure 2.31.1: Crossing scheme and summary of RNAi-mediated gene silencing in *Drosophila.* All crosses were set up in reciprocal.

2.32: Gamma irradiation (IR) of *Drosophila* embryos/larvae to test *dikar* function in DNA double-strand break repair

All IR experiments were performed at the Edmonton Cross Cancer institute, and utilized the irradation apparatus model CC2-20, type B, which contained Cobalt-60 as the IR source. Embryonic and/or larval stages were irradiated in their vials with food.

Kathryn Wilfong-Pritchard, an undergraduate student in the lab, aided in determining sensitivity to IR at various ages and IR dosages. She set up heterozygous crosses for both the

dikar^{3} and *dikar*^{5} alleles with 5-8 males and 5-8 virgins per cross. For the crosses, both male and virgin parents were of the genotype *dikar*^{3}/*TM3* to test the IR sensitivity of the *dikar*^{3} allele, and the same was done for the *dikar*^{5} allele. The expected ratio of the resulting viable progeny was 2 *dikar*$^{(3\ or\ 5)}$/*TM3* for every 1 *dikar*$^{(3\ or\ 5)}$/*dikar*$^{(3\ or\ 5)}$. Parents were tipped into new vials every 2 days, resulting in a series of 3 vials from the same cross that contained progeny ranging from 0-2 days, 2-4 days, and 4-6 days old, all of which underwent a single dose of irradiation (either 15 gray or 30 gray). Control crosses were set up, tipped, and transported to and from the Cross Cancer Institute alongside the experimental crosses, but did not receive any irradiation (0 gray).

I then performed large-scale IR experiments on 2-4 day old larvae, which were dosed with 22.5 gray of irradation. The crosses were *dikar*^{3}/TM3 x *dikar*^{3}/TM3, *dikar*^{5}/TM3 x *dikar*^{5}/TM3, *dikar*^{3}/*dikar*^{3} x Df(3L)ZN47/*TM3*, *dikar*^{5}/*dikar*^{5} x Df(3L)ZN47/*TM3*, *P{SUPor-P}dikar*^{KG00884}/*TM3* x *P{SUPor-P}dikar*^{KG00884}/*TM3* (for both independent lines generated), and *dikar*^{WT}/*TM3* x *dikar*^{WT}/*TM3* (for both independent lines generated) (Table 2.32.1). Ten vials were set up for each cross, 8 of which were experimental (22.5 gray), and 2 of which were control (0 gray). In order to improve number of progeny per vial due to high levels of generalized death in vials that underwent IR, each cross was set up with 10-13 male and 20-30 virgin female parents. Each of the 10 vials was tipped 4 times (for *dikar*^{3}/*TM3* and *dikar*^{5}/*TM3* crosses) or 5 times (for *dikar*^{3}/*dikar*^{3} x Df(3L)ZN47/*TM3*, *dikar*^{5}/*dikar*^{5} x Df(3L)ZN47/*TM3*, *P{SUPor-P}dikar*^{KG00884}/*TM3*, and *dikar*^{WT}/*TM3* crosses). This resulted in a total of 48 irradiated and 8 control vials for the *dikar*^{3}/*TM3* and *dikar*^{5}/*TM3* crosses, and 50 irradiated and 10 control vials for the *dikar*^{3}/*dikar*^{3} x Df(3L)ZN47/*TM3*, *dikar*^{5}/*dikar*^{5} x Df(3L)ZN47/*TM3*, *P{SUPor-P}dikar*^{KG00884}/*TM3*, and *dikar*^{WT}/*TM3* crosses. Statistical analyses comparing the ratio of progeny between irradiated flies and non-irradiated flies were performed using the χ^2 test-of-independence.

Table 2.32.1: Fly crosses set up for IR experiments. Progeny homozygous for the balancer chromosome *TM3* were not viable. Crosses 5 and 6 were set up in duplicate as two independent stocks were generated *P{SUPor-P}*^{dikarKG00884}*/TM3* for and *dikar*^{WT}*/TM3*.

Cross	Viable genotypes	Expected ratio	Purpose of experiment
dikar^{3}*/TM3* x *dikar*^{3}*/TM3*	*dikar*^{3}*/dikar*^{3}, *dikar*^{3}*/TM3*	1:2	Determine if *dikar*^{3} homozygotes are IR sensitive
dikar^{5}*/TM3* x *dikar*^{5}*/TM3*	*dikar*^{5}*/dikar*^{5}, *dikar*^{5}*/TM3*	1:2	Determine if *dikar*^{5} homozygotes are IR sensitive
dikar^{3}*/dikar*^{3} x *Df(3L)ZN47/TM3*	*dikar*^{3}*/Df(3L)ZN47*, *dikar*^{3}*/TM3*	1:1	Determine if *dikar*^{3} over a deficiency encompassing *dikar* is IR sensitive (positive control)
dikar^{5}*/dikar*^{5} x *Df(3L)ZN47/TM3*	*dikar*^{5}*/Df(3L)ZN47*, *dikar*^{5}*/TM3*	1:1	Determine if *dikar*^{5} over a deficiency encompassing *dikar* is IR sensitive (positive control)
P{SUPor-P}^{dikarKG00884}*/TM3* x *P{SUPor-P}*^{dikarKG00884}*/TM3*	*P{SUPor-P}*^{dikarKG00884}*/P{SUPor-P}dikar*^{KG00884}, *P{SUPor-P}*^{dikarKG00884}*/TM3*	1:2	Determine if parental P element homozygotes are IR sensitive
dikar^{WT}*/TM3* x *dikar*^{WT}*/TM3*	*dikar*^{WT}*/dikar*^{WT}, *dikar*^{WT}*/TM3*	1:2	Determine if *dikar*^{WT} homozygotes are IR sensitive (negative control)

2.33: DNA sequence analysis of *Drosophila* genes *dikar* and *Gen* to detect possible second site mutations

The exons and splice sites of the genes *dikar* and *Gen* were Sanger-sequenced to look for differences between the *TM3* balancer chromosome and the third chromosome that contained the *dikar* deletion alleles. Genomic DNA was extracted as described in section 2.4 from a single *dikar*^{WT}*/TM3* heterozygous fly, which was used for the template for PCR and sequencing. Initial PCR reactions were performed on a Peltier Thermocycler-200 (MJ Research) with cycling conditions (1) 95.0 °C for 3 minutes, (2) 94.0 °C for 20 seconds, (3) 65.0 °C for 20 seconds, (4) 72.0 °C for 1 minute, (5) repeat steps 2-4 11 times with -1.0 °C for step 3 each cycle, (6) 94.0 °C for 20 seconds, (7) 55.0 °C for 20 seconds, (8) 72.0 °C for 1 minute, (9) repeat steps (6)-(8) 30 times, (10) 72.0 °C for 10 minutes, (11) 4.0 °C hold. DNA concentration of the PCR product was approximated and cleaned up with ExoSAP as described in section 2.7. I prepped samples for

sequencing by combining ~225 ng of PCR template DNA, 2.5 pmoles of primer, and nuclease free water (Integrated DNA Technologies) to 10 µl. Samples were submitted to MBSU where Sanger sequencing was carried out via their Economy Sequencing Service.

I analyzed chromatograms for the presence of double (heterozygous) peaks, which indicated a difference between the two alleles in heterozygous flies. Identified DNA sequence variants were compared to the reference sequence (www.ensembl.org) to determine if any variants resulted in missense mutations at the protein level. DNA sequence variants of interest were then Sanger-sequenced in a homozygous $dikar^{WT}/dikar^{WT}$ fly to determine which variant belonged to which allele. Missense variants were analyzed with SIFT [172] to predict if there were any functional differences between the protein products of the two alleles. Protein modeling comparing the protein products of the two alleles was done using the web-based software I-TASSER [187-189].

Chapter 3

Cecr2 and candidate modifier genes of *Cecr2*-associated exencephaly in mice and humans

Preface

Some of the work in this chapter was published:

Kooistra, M.K., R.Y.M. Leduc, C.E. Dawe, N.A. Fairbridge, J. Rasmussen, J.H. Man, M. Bujold, D. Juriloff, K. King-Jones and H.E. McDermid. (2012). Strain-specific modifier genes of Cecr2-associated exencephaly in mice: genetic analysis and identification of differentially expressed candidate genes. Physiol. Genomics 44:35-46.

Some of the work in this chapter was done as part of various collaborations.

Section 3.5: I contributed MOD 5/31 and MOD 31/+ exencephaly penetrance data for publication. **Section 3.6:** Megan Kooistra performed quantitative real-time PCR (qRT-PCR) validation for 6 of the candidate modifier genes, whereas I performed qRT-PCR validation for an additional 10 candidate modifier genes. I contributed Sanger sequencing of qRT-PCR primer-probe binding sites for publication. **Section 3.7:** Whole exome sequencing of two mouse strains was performed in collaboration. I prepared DNA samples for the mouse whole exome sequencing experiment, and Deidre Krupp performed the library preparation, sequencing, initial filtering/analysis, and annotation at Duke University. I followed up with final analysis and validation of variants. **Section 3.8:** An undergraduate student, Leanne Donnelly, aided with the BALB/cCrl x BALB/cJ dihybrid cross under my supervision. Leanne set up crosses, dissected mice, assessed embryonic phenotypes and analyzed data for the first two thirds of embryo collection. I finished by setting up crosses, dissecting, and assessing phenotypes for the remaining embryo collection and I performed final analyses. I contributed the *Arhgap19*$^{GT(YHD020)Byg}$ mouse line generation, genotyping, and E9.5 embryo X-gal staining for publication. **Sections 3.9 and 3.10:** The work that utilized human DNA samples in this chapter was done in collaboration. I prepared DNA samples for the human sequencing experiment, and Natalie Mola performed the library preparation, sequencing, and initial filtering/analysis at Duke University. I followed up with annotation, final analysis, and Sanger-sequence validation of variants. Drs. Allison Ashley-Koch and Simon Gregory provided human DNA samples, lab space, and expertise regarding experimental design and data analysis. Drs. Nicholas Katsanis and Erica Davis shared expertise and provided all resources required for next generation sequencing.

3.1: Introduction

Neurulation is a highly dynamic process that involves the complex interplay of hundreds of genes in many genetic pathways. *Cecr2* is only one of over 300 NTD genes identified in mice. However, as a chromatin remodeling protein, CECR2 likely plays critical roles in global regulation of gene transcription as well as in important DNA metabolic processes including DNA replication and repair. The disruption of global gene expression, DNA replication/cell proliferation, and/or DNA repair at the time of neurulation could have a major impact on several of the important pathways and processes integral to neurulation. Understanding how *Cecr2* functions during neurulation will further enlighten the molecular underpinnings of the essential developmental process of neurulation.

In humans, NTDs are considered to be multifactorial, with many genetic variants in combination with environmental factors contributing to the phenotypic output in a single patient. It is currently hypothesized that human NTDs fall under the multifactorial threshold model [190,191]. The multifactorial threshold model theorizes that a combination of several genetic and environmental factors act in an additive fashion to predispose an individual to the development of a disease, with only those who have accumulated enough of these predisposing factors breaching the "threshold" and developing the disease (Figure 3.1.1). Multifactorial threshold traits are characterized by non-Mendelian inheritance. Multifactorial traits demonstrate a higher risk of disease development in 1st degree relatives that diminishes as the degree of relation becomes further apart. For example, the risk for a first-degree relative of someone with an NTD is ~2.5% and the risk for a second-degree relative is ~1% [192].

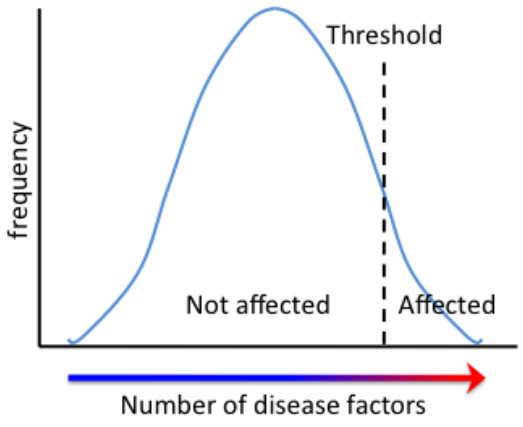


Figure 3.1.1: The threshold model of multifactorial disease inheritance. Threshold traits are believed to be qualitative traits (presence/absence of the disease) that have an underlying quantitative etiology. Only individuals that possess a number of disease factors that exceeds the "threshold" go on to develop the disease.

As most mouse models of NTDs demonstrate Mendelian inheritance with complete or reduced penetrance, original schools of thought were that neurulation in mice is a simpler process than in humans and therefore is not the best approximation of human neurulation. However, more studies are focusing on gene-gene, gene-environment, and environment-environment interactions in mice. These studies are demonstrating that NTDs in mice are more complex than previously thought and that mice are currently, and will continue to be, powerful models of the multifactorial inheritance of NTDs (reviewed in [116]). Examples of mouse models demonstrating gene-environment and environment-environment interactions were discussed in Chapter 1.4. Here, the focus is on mouse models of gene-gene interactions, which appear to be an important component of mouse NTD etiology. In this case, a gene-gene interaction refers to a situation where a single genetic lesion does not cause NTDs on its own, but can result in NTDs in combination with one or more additional genetic lesions.

Gene-gene interactions in mice have been shown to occur in the same biological pathway or process, such as the interaction of *Vangl1*^{gt} and *Vangl2*^{Lp} affecting convergent extension [140]. The *Vangl2*^{Lp} allele results in craniorachischisis with 100% penetrance in homozygous mutant mice and although heterozygous mice do not develop craniorachischisis, they do display urogenital defects, a looped tail, and occasionally spina bifida [193,194]. All mice heterozygous or homozygous for the *Vangl1*^{gt} allele were shown to develop normally, albeit with minor alterations in stereociliary bundle organization in homozygous mice [140]. To test if these alleles genetically interact, Torban et al. (2008) crossed *Vangl2*$^{Lp/+}$ heterozygous mice to *Vangl1*$^{gt/gt}$ homozygous mice. A genetic interaction between the two alleles was revealed, with craniorachischisis developing in >60% (20/32) of embryos that were heterozygous for both the *Vangl2*^{Lp} allele and the *Vangl1*^{gt} allele [140].

Gene-gene interactions can also occur in biological processes that are discrete from one another, such as the interaction of *Vangl2*^{Lp} (convergent extension in the neural epithelium) and *Grhl3*^{ct} (hindgut endoderm proliferation) [194]. *Vangl2*$^{Lp/+}$mice only occasionally develop spina bifida [193] and *Grhl3*$^{ct/ct}$ mice develop spina bifida and/or tail flexion defects ~40% of the time [195]. Stiefel et al. (2003) were the first to publish that embryos with a *Vangl2*$^{Lp/+}$*;Grhl3*$^{ct/ct}$ genotype develop spina bifida close to 100% [194]. However, the authors did not attempt to explain the relationship between PCP governed convergent extension in the neural epithelium and hindgut endoderm proliferation. Gustavsson et al. (2007) postulated that *Grhl3*$^{ct/ct}$ embryos are

predisposed to spina bifida because posterior neuropore closure is mechanically difficult due to the excessive ventral curvature in the caudal region, which is a result of reduced proliferation in the hindgut endoderm [195]. Therefore, it can be hypothesized that *Vangl2*$^{Lp/+}$ heterozygous mice possess a slightly shorter and wider neural tube that almost always successfully closes, and *Grhl3*$^{ct/ct}$ homozygous mice are capable of overcoming the mechanical opposition induced by reduction in gut endoderm ~60% of the time; however, embryos that have both a slightly shorter/wider neural plate as well as a reduction in gut endoderm are unable to successfully close their neural tube up to 100% of the time as they possess both predisposing factors.

The above examples of gene-gene interactions represent just two of more than 30 mouse NTD models represented on the NTD Wiki database (http://ntdwiki.wikispaces.com) that require a mutation in more than one gene (digenic or trigenic) in order for phenotype to manifest or increase in penetrance [15]. Furthermore, most monogenic NTD genes have been studied in congenic mouse strains, where individual variation within a strain is almost eliminated. The penetrance of NTDs resulting from a single gene mutation can vary from mouse strain to mouse strain, indicating the presence of other contributing genetic factors and perhaps modeling the individual variation in the human population. For example, the *Pax3 splotch* mutation showed reduced penetrance when moved from the original C57BL/6J mouse strain to a partial FVB/N background, with the penetrance of exencephaly dropping from 25.7% (9/35) to 11.5% (3/26) and the penetrance of spina bifida dropping from 100% (35/35) to 73.1% (19/26) [164]. The same experiment was done for the *shroom*^{GtROSA53Sor} genetrap mutation, where although the penetrance of exencephaly remained at 100% for both genetic backgrounds, the penetrance of spina bifida dropped from 87.5% (7/8) to 8.7% (2/23) [164]. Although the C57BL/6 genetic background is more susceptible compared to FVB/N for the above two examples, the C57BL/6 genetic background demonstrates resistance relative to the 129/SvEv genetic background when mice are homozygous mutant for *Alx1*, with exencephaly penetrance dropping from 100% on a 129/SvEv background to 65% on a C57BL/6:129/SvEv hybrid background [196]. An NTD modifier gene within a susceptible genetic background has been successfully identified in the past. De Castro et al. (2012) demonstrated that the presence of a lamin B1 variant found within the curly tail mouse (mutant for *Grhl3*) genetic background increased the susceptibility of these mice to the development of NTDs [197]. The introduction of a wildtype lamin B1 allele into the curly tail mouse genetic background reduced the NTD rate by ~3 fold. The authors went on to show that the lamin B1

variant had a functional impact at the molecular level, which ultimately affected cell cycle progression. Genetic background influences on NTD penetrance are also present in our *Cecr2* mouse model, which is a focus of this chapter.

Homozygous mutation in the mouse gene *Cecr2* results in exencephaly with reduced penetrance that varies between the mouse strains BALB/cCrl and FVB/N. Original exencephaly penetrance analyses performed by Banting et al. (2005) demonstrated that homozygous mutant $Cecr2^{GT45bic}$ BALB/cCrl embryos (generation N6) developed exencephaly 74.5% (35/47) of the time, but that when the $Cecr2^{GT45bic}$ allele was crossed into the FVB/N background for five generations, exencephaly penetrance dropped to 0% (0/45) in homozygous mutant embryos [141]. A deletion of the first exon of *Cecr2* and upstream sequence, resulting in the more severe mutant allele $Cecr2^{tm1.1Hemc}$, also displayed variable penetrance in the two mouse strains [142]. When the $Cecr2^{tm1.1Hemc}$ allele was crossed into the BALB/cCrl genetic background for three to four generations, homozygous mutant embryos displayed an exencephaly penetrance of 96.0% (48/50); however, when crossed into the FVB/N genetic background for the same number of generations, penetrance was only 31.4% (11/35) in $Cecr2^{tm1.1Hemc}$ mutant embryos [142]. This extreme difference in exencephaly penetrance suggests the presence of genetic modifiers, where one or more genetic variants that differ between the two mouse strains renders BALB/cCrl susceptible and/or FVB/N resistant to the development of exencephaly.

When BALB/cCrl and FVB/N mice carrying the $Cecr2^{GT45bic}$ allele were crossed to generate BALB/cCrl:FVB/N hybrid embryos, homozygous mutant embryos with exencephaly were reduced to 2.9% (1/35 mutants) [164]. The significant reduction in exencephaly penetrance in these homozygous mutant embryos demonstrated that at least most of the FVB/N resistance alleles are dominant. Therefore, whole genome linkage analysis was performed in the McDermid lab to identify the FVB/N genomic regions containing the resistance loci [164]. The non-penetrant homozygous mutant BALB/cCrl:FVB/N hybrid embryos were used in the experiment because, not only were they viable and fertile, they were also more likely carrying the FVB/N resistance alleles the authors were interested in. For linkage analysis, these mice were backcrossed to heterozygous $Cecr2^{GT45bic}$ BALB/cCrl mice to generate exencephalic embryos along with non-exencephalic homozygous mutant embryos (penetrance of 28.1%, 101/360 mutant embryos) [164]. Results demonstrated that there was an ~40.1 Mb region on mouse chromosome 19 with an LOD score of 4.35 that contains one or more major modifiers. Suggestive linkage peaks were also

found on chromosome 2 and on the X chromosome, demonstrating that there are multiple modifier loci that differ between the two strains. It is also possible that other loci outside of these peaks contribute small effects that the linkage analysis did not detect.

Further insight into the chromosome 19 modifier region revealed the presence of more than one modifier locus [165]. This was established by the creation of three sub-interval congenic mouse lines (herein referred to as MOD 4, MOD 5, and MOD 31), which broke up the chromosome 19 modifier region into 3 segments, resulting in 4 genetically distinct sub-regions (Figure 3.1.2, Figure 3.1.3). Each sub-interval congenic line was homozygous FVB/N for a different segment of the chromosome 19 modifier region on an otherwise BALB/cCrl genetic background, which also carried the *Cecr2*^{GT45bic} allele. The homozygous FVB/N chromosome 19 sub-regions in MOD 4 and MOD 31 lines do not overlap; however, both lines demonstrated a significant drop in exencephaly penetrance when compared to the originally established penetrance of 74.5% in incipient congenic *Cecr2*^{GT45bic} BALB/cCrl. Exencephaly penetrance in the MOD 5 line shows a similar drop in penetrance as MOD 4 and MOD 31. The MOD 5 homozygous FVB/N segment completely overlaps with the MOD 4 segment and partially overlaps with the MOD 31 segment.

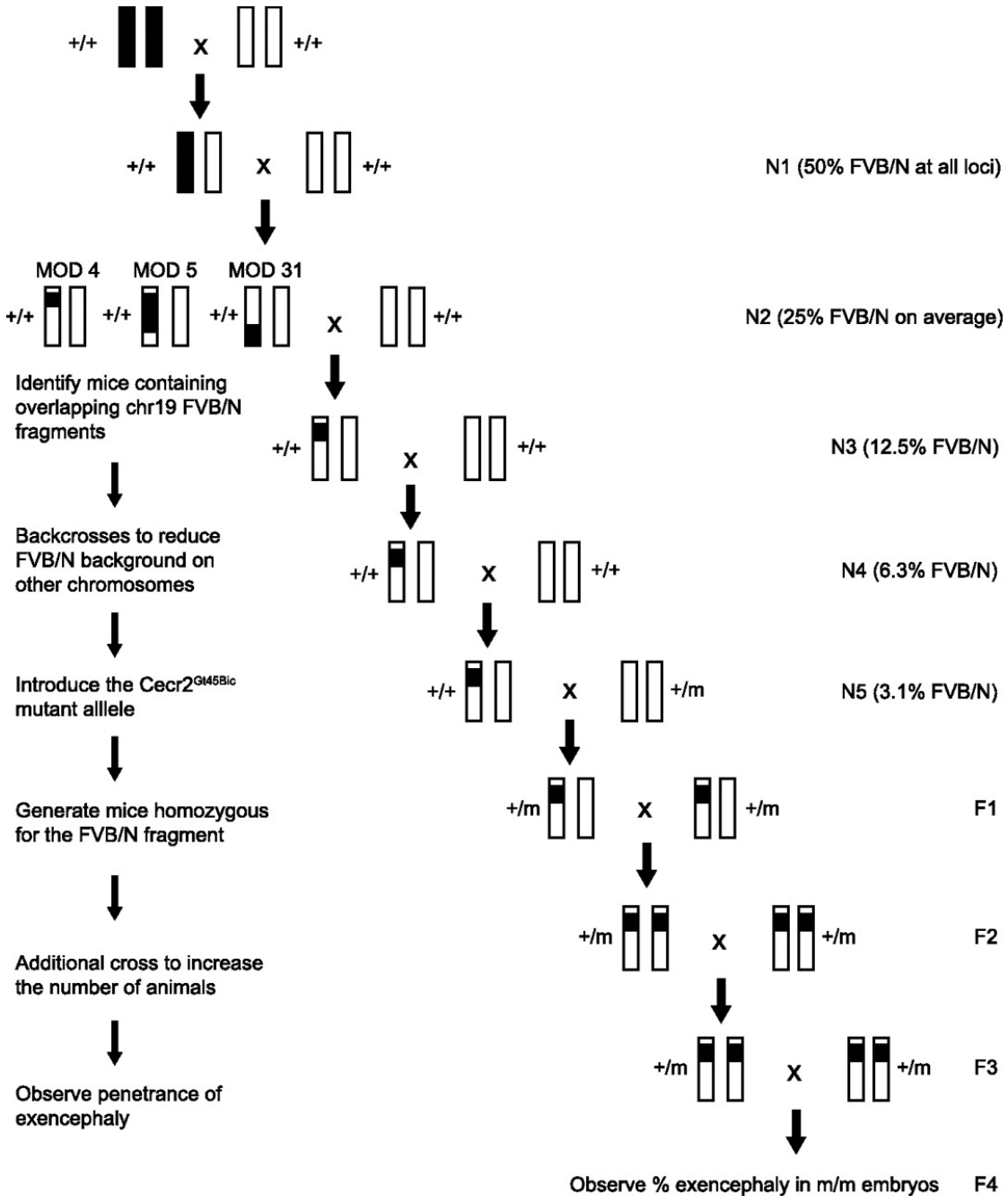
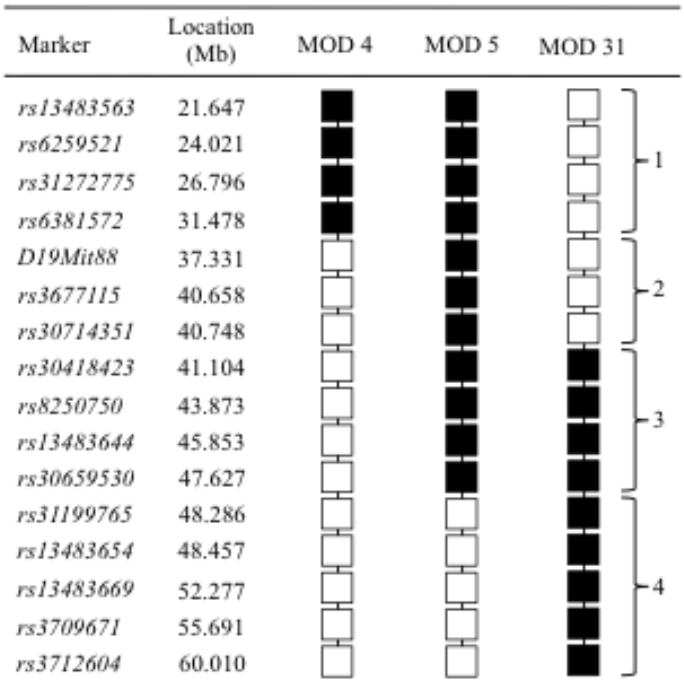


Figure 3.1.2: Crossing scheme to generate sub-interval congenic mouse lines MOD 4, MOD 5, and MOD 31. Black shading represents FVB/N genetic background, white shading represents BALB/cCrl genetic background. Mouse crosses were performed by Megan Kooistra. '+' indicates wildtype *Cecr2* allele, 'm' indicates *Cecr2*^{GT45bic} allele. Image acquired from Kooistra et al. (2012) [165].



Line	MOD 4	MOD 5	MOD 31
Exencephaly	25	12	36
Normal	44	12	42
Total mutants, *n*	69	24	78
Penetrance, %	36.2	50.0	46.1

Figure 3.1.3: Exencephaly penetrance analyses of sub-interval congenic mouse lines demonstrate more than one modifier locus within the chromosome 19 modifier region. Black boxes signify homozygous FVB/N genetic background, white boxes signify homozygous BALB/cCrl genetic background. The remainder of the genome is of BALB/cCrl origin. These data were collected by Megan Kooistra. Image modified from Kooistra et al. (2012) [165].

There are over 500 genes within the chromosome 19 modifier region. To identify candidate modifier genes, a whole genome microarray analysis (Affymetrics Mouse 430 2.0) was performed on BALB/cCrl and FVB/N embryos at 11-14 somites (~E8.5) in order to find genes within the chromosome 19 modifier region that are differentially expressed between the two strains during neurulation [165]. The microarray data was analyzed using two different statistical methods, the GC-RMA/Student's t-test and the RMA/ANOVA, to identify genes where expression differed by at least 1.5 fold. A total of 22 genes were identified by GC-RMA/Student's t-test analysis (Table 3.1.1), and 11 genes were identified by RMA/ANOVA analysis with 9 of these genes also represented in the GC-RMA/Student's t-test analysis (Table 3.1.2). A caveat to the microarray was that not all genes within the genome were represented on the array. One such gene, *Pax2*, is within the chromosome 19 modifier region and was a good candidate modifier as it results in exencephaly when mutated in mice [198]. However, quantitative real-time PCR (qRT-PCR) demonstrated that this gene is not differentially expressed between the two strains. Another caveat of the microarray was that it only detected candidate modifier genes

based on differential expression, but neglected candidate genes that contained DNA sequence variants that may have altered protein function without affecting expression.

Table 3.1.1: Genes within the chromosome 19 modifier region that differ in expression by at least 1.5 fold using GC-RMA/Student's t-test analysis. Data obtained from whole genome microarray using 11-14 somite embryos. Table acquired from Kooistra et al. (2012) [165].

FVB/N vs BALB/cCRL wildtype					FVB/N vs BALB/cCRL *Cecr2*$^{Gt45Bic/Gt45Bic}$				
Affy probe set	Gene symbol	MOD region	FC	*P* value	Affy probe set	Gene symbol	MOD region	FC	*P* value
1434911_s_at	*Arhgap19*	3	21.43	3.27E-08	1434911_s_at	*Arhgap19*	3	20.15	1.26E-05
1444218_at	*D19Ertd737e*	4	4.97	2.31E-04	1431982_at	*Arhgap19*	3	3.37	5.02E-04
1422318_at	*Foxd4*	1	2.65	3.08E-02	1444218_at	*D19Ertd737e*	4	2.84	6.71E-05
1423875_at	*Fam160b1*	4	2.27	4.89E-04	1418473_at	*Cutc*	3	1.94	1.16E-04
1423876_at	*Fam160b1*	4	2.24	3.09E-04	1439561_at	*2010012O05Rik*	3	1.85	7.08E-06
1426867_at	*Dmrt2*	1	2.11	4.07E-02	1423876_at	*Fam160b1*	4	1.80	7.09E-03
1439561_at	*2010012O05Rik*	3	1.87	8.38E-03	1431337_a_at	*Fam45a*	4	1.78	2.40E-04
1418473_at	*Cutc*	3	1.73	6.06E-04	1452012_a_at	*Exosc1*	3	1.77	1.29E-03
1431982_at	*Arhgap19*	3	1.77	1.50E-02	1422318_at	*Foxd4*	1	1.74	4.65E-02
1438454_at	*Pten*	1	1.64	3.45E-02	1418873_at	*Sfxn4*	4	1.66	3.82E-03
1448186_at	*Pnliprp2*	4	1.64	3.02E-02	1440707_at	*Dmrt3*	1	-1.50	3.48E-02
1431337_at	*Fam45a*	4	1.61	1.27E-04	1438561_x_at	*Tmem180*	3	-1.64	2.02E-04
1424983_at	*2700078E11Rik*	4	1.51	1.24E-04	1423023_at	*Sfrp5*	3	-1.75	1.95E-03
1452845_at	*Hif1an*	3	1.50	2.42E-04	1436075_at	*Sfrp5*	3	-1.76	1.31E-02
1435896_at	*Sfxn2*	3	-1.51	6.04E-04	1446838_at	*Atad1*	1	-1.80	1.12E-02
1436075_at	*Sfrp5*	3	-1.53	5.86E-02	1449853_at	*Sfxn2*	3	-1.82	9.09E-03
1415964_at	*Scd1*	3	-1.54	1.82E-04	1417900_a_at	*Vldlr*	1	-1.84	1.03E-03
1453180_at	*Rnls*	1	-1.57	2.07E-02	1434465_x_at	*Vldlr*	1	-1.95	5.87E-04
1438561_x_at	*Tmem180*	3	-1.73	3.08E-05	1423141_at	*Lipa*	1/2	-2.27	8.37E-03
1415965_at	*Scd1*	3	-1.87	8.66E-05	1429910_at	*March5*	2	-2.43	2.58E-04
1428139_at	*Tmem180*	3	-2.04	2.80E-02	1438258_at	*Vldlr*	1	-2.66	4.25E-06
1423141_at	*Lipa*	1/2	-2.41	7.30E-04	1415824_at	*Scd2*	3	-2.83	5.65E-04
1415824_at	*Scd2*	3	-2.67	4.03E-04	1432052_at	*Exosc1*	3	-2.84	1.33E-03
1429910_at	*March5*	2	-2.78	2.24E-05	1438207_at	*Gbf1*	3	-3.13	8.84E-03
1432052_at	*Exosc1*	3	-2.99	3.47E-07	1433914_at	*Lipo1*	1	-5.70	8.35E-05
1433914_at	*Lipo1*	1	-6.26	4.81E-05					

FC – fold change

Table 3.1.2: Genes within the chromosome 19 modifier region that differ in expression by at least 1.5 fold using RMA/ANOVA analysis. Data obtained from whole genome microarray using 11-14 somite embryos. All genes within this table except *Vldlr* and *Xpnpep1* are also included in Table 3.1.2. Table acquired from Kooistra et al. (2012) [165].

			FVB/N vs BALB/cCRL WT		FVB/N *vs* BALB/cCRL *Cecr2* $^{Gt45Bic/Gt45Bic}$		
Affy probe set	Gene symbol	MOD region	Fold change	*P* value	Fold change	*P* value	ANOVA
1434911_s_at	*Arhgap19*	3	6.12	2.11E-08	5.88	2.14E-07	8.89E-09
1429910_at	*March5*	2	-2.04	5.58E-06	-1.69	9.81E-05	5.89E-04
1424983_a_at	*2700078E11Rik*	4	1.45	2.15E-05	1.39	3.24E-05	7.96E-04
1428139_at	*Tmem180*	3	-1.29	1.52E-04	-1.38	7.40E-06	1.51E-03
1433914_at	*Lipo1*	1	-3.74	1.03E-04	-3.52	1.75E-05	2.83E-03
1438258_at	*Vldlr*	1	-1.55	1.02E-03	-2.31	9.78E-06	4.03E-03
1422443_at	*Xpnpep1*	4	-1.29	1.61E-04	-1.27	3.79E-05	1.01E-02
1444218_at	*D19Ertd737e*	4	3.25	4.72E-05	2.27	2.49E-04	1.03E-02
1432052_at	*Exosc1*	3	-2.33	2.25E-05	-2.12	3.09E-04	1.10E-02
1418473_at	*Cutc*	3	1.34	7.76E-04	1.55	1.14E-04	2.61E-02
1423875_at	*Fam160b1*	4	1.66	7.60E-06	1.21	1.53E-02	3.82E-02

FC – fold change

Arhgap19 was a candidate gene of particular interest as it had 21.43-fold lower expression in BALB/cCrl relative to FVB/N according to microarray GC-RMA/Student's t-test analysis (6.12-fold lower according to RMA/ANOVA analysis). It was also discovered that the wildtype BALB/cCrl mouse strain is homozygous for a 'T' insertion within exon 6 that results in the introduction of a nonsense mutation (herein referred to as *Arhgap19*^{Ex6non}) [165,199]. *Arhgap19* belongs to the GTPase activating protein (GAP) family, which act as negative regulators of Rho proteins by stimulating the intrinsic GTP hydrolysis activity of Rho [200]. Members of the RhoGAP family are involved in regulating many cellular processes, including cell adhesion, cell migration, gene expression, cytoskeletal dynamics, cell proliferation, cell differentiation, and planar cell polarity [201,202]. All of these processes have been shown to result in NTDs if disrupted [203]. The *Arhgap19*^{Ex6non} allele produces a truncated protein product that still contains the RhoGAP domain; however, we hypothesize that the reduced levels of transcript in BALB/cCrl is due to nonsense-mediated decay.

In this thesis, I reanalyzed and expanded on the penetrance analyses of the wildtype and MOD lines, solidifying the conclusion that more than one modifier exists on chromosome 19, and discovering that these modifiers do not act in an additive manner. As part of further characterizing the congenic *Cecr2*^{GT45bic} BALB/cCrl line, penetrance of the small inner ear

phenotype was also analyzed, with results indicating that small inner ears are secondary to the development of exencephaly. Small inner ear penetrance analyses were also performed for both MOD 5 and MOD 31, with results corroborating the observations in BALB/cCrl. In the pursuit of finding one or more modifier genes of *Cecr2*-associated exencephaly, I identified additional candidate modifier genes by whole exome sequencing (WES) of the BALB/cCrl and FVB/N strains. This resulted in an updated list of candidate genes, which was validated and refined based on additional information, including what is previously known about gene expression and function. I excluded the top candidate gene, *Arhgap19*, based on genetic analysis in mouse.

Although studies in our mouse model have established *Cecr2* as an NTD gene, variants within human *CECR2* have not yet been associated with NTDs in humans. Here, for the first time, human *CECR2* was sequenced in a cranial NTD cohort consisting of 156 probands, where 9 protein-coding variants of interest were identified. In an effort to identify the top candidate modifier genes in mouse and the possible contribution of these genes to human NTDs, the human homologues of the remaining candidate modifier genes (excluding *Arhgap19*) were sequenced in the human cranial NTD cohort consisting of 156 probands. *DNMBP*, *MMS19*, and *TJP2* were identified as the three top candidate modifier genes.

RESULTS

3.2: Exencephaly penetrance in congenic *Cecr2*^{GT45bic} BALB/cCrl mice is lower than the penetrance in the earlier N6 generation

An increase in homozygous viable mutant mice was perceived in our colony over a period of 8 years; therefore, it was important that we re-established what the penetrance of exencephaly is in the *Cecr2* mutant congenic lines. Over time, modifying mutations can accumulate within an inbred mouse line, which can partially protect these mice from developing exencephaly. Changes in the environment could also influence the penetrance of exencephaly.

The *Cecr2*^{GT45bic} allele was originally generated in a 129P2/Ola mouse genetic background, and was then backcrossed into either the BALB/cCrl or FVB/N genetic backgrounds [141]. The most obvious phenotypic consequence of the *Cecr2*^{GT45bic} homozygous mutation when backcrossed into the BALB/cCrl mouse strain was exencephaly. This was analyzed mainly in the incipient congenic (between N5 and N9 generations) BALB/cCrl line at generation N6, with

some inclusion of generation N5 progeny. A penetrance of 74.5% (35/47) exencephaly was detected [141]. The *Cecr2*^{GT45bic} allele was then further backcrossed into the BALB/cCrl genetic background until a combined total of 10 generational backcrosses was achieved (N10), which resulted in a congenic mouse line. A perceived increase in homozygous viable *Cecr2*^{GT45bic} mutant mice led me to re-analyze exencephaly penetrance in this congenic line. A minimum of 8 years passed between these two analyses, with the original analysis being published in 2005 and the recent analysis on congenic mice being performed in 2013. Results from this re-analysis yielded a drop in exencephaly penetrance from 74.5% (35/47) observed in the original analysis in the incipient congenic BALB/cCrl *Cecr2*^{GT45bic} line to 54.1% (73/135) in the more recently produced congenic BALB/cCrl *Cecr2*^{GT45bic} line (Table 3.2.1). A χ^2 test-of-independence demonstrated that the exencephaly penetrance in congenic BALB/cCrl was significantly lower than the exencephaly penetrance obtained from the previous analysis on generation N6 incipient congenic BALB/cCrl ($P = 0.014$). Also, the *Cecr2*^{GT45bic} allele did not demonstrate Mendelian inheritance according to a χ^2 goodness-of-fit test (homozygous mutants = 135, heterozygotes = 310, homozygous wildtypes = 183, $P = 0.0242$). Banting et al. (2005) also reported fewer homozygous mutants than expected in the previous analysis, with 47 homozygous mutants, 110 heterozygotes, and 70 homozygous wildtypes; however, this was not statistically significant ($P = 0.0873$). In the recent analysis, the observed number of 310 heterozygote embryos was close to the expected number of 314 embryos. If the expected number of mutant and wildtype embryos were derived from the observed number of heterozygotes (310/2), we would expect 155 mutant and 155 wildtype embryos. Therefore, results indicate both the loss of mutant embryos as well as an excess of wildtype embryos. Exencephaly was not seen in any heterozygous or homozygous wildtype embryos in both the current and previous analyses.

An increased predisposition to the development of exencephaly in females has been established in many mouse genetic NTD models, as well as in humans (reviewed in [106]). While Banting et al. (2005) did not record the sex of the embryos for the original penetrance analysis; Davidson et al. (2007) collected a series of exencephalic *Cecr2*^{GT45bic} BALB/cCrl homozygous mutant embryos for a different study. There were 61 females and 33 males among the mutant exencephalic embryos, with 64.9% being female and 35.1% being male, which implied that female *Cecr2*^{GT45bic} mutant mice also had a higher predisposition to developing exencephaly [164]. The recent analysis of exencephaly penetrance in congenic *Cecr2*^{GT45bic} BALB/cCrl confirms this

female predisposition to the development of exencephaly in our mouse model, with 67.6% (50/74) of *Cecr2*^{GT45bic} BALB/cCrl mutant females developing exencephaly, and only 37.7% (23/61) of *Cecr2*^{GT45bic} BALB/cCrl mutant males developing exencephaly (Table 3.2.1). A χ^2 test-of-independence revealed that this difference between sexes is significant ($P = 0.00053$).

Table 3.2.1: Congenic *Cecr2*^{GT45bic} BALB/cCrl homozygous mutant exencephaly penetrance analysis.

	Female	Male	Combined sexes	N6 Balb/cCrl (old analysis) [141]
Exencephaly	50	23	73	35
Normal	24	38	62	12
Penetrance (%)	67.6	37.7	54.1	74.5

3.3: Small inner ears are secondary to the exencephaly phenotype

Phenotypic characterization of *Cecr2*^{GT45bic} BALB/cCrl homozygous mutant mice revealed that inner ears were also affected, with smaller cochlea and disorganization of stereociliary bundles observed in mutant mice with exencephaly [143]. Here, the small inner ear phenotype was followed up on, but organization of stereociliary bundles was not assessed for these embryos. Dawe et al. (2011) addressed the small inner ear phenotype; however, the penetrance of this phenotype was not quantified and I was interested in determining if small inner ear penetrance differed between BALB/cCrl and the MOD lines (MOD inner ears are addressed in section 3.5). The sizes of both inner ears were measured and averaged per embryo for congenic *Cecr2*^{GT45bic} BALB/cCrl homozygous wildtype E17.5 embryos ($n = 14$) and compared to *Cecr2*^{GT45bic} BALB/cCrl homozygous mutant E17.5 embryos with exencephaly ($n = 5$) and without exencephaly ($n = 9$). Overall, inner ear pairs from the same embryo appeared similar to one another. The smaller inner ear phenotype was only seen in exencephalic *Cecr2*^{GT45bic} BALB/cCrl homozygous mutant embryos, suggesting that smaller inner ears are secondary to the development of exencephaly (Figure 3.3.1). One-way ANOVA followed by Tukey's post-hoc revealed no significant difference when comparing inner ear measurements of wildtype embryos to homozygous mutant embryos without exencephaly ($P > 0.05$, Table 3.2.1). However, there was a significant inner ear size difference between wildtype embryos and homozygous mutant

exencephalic embryos ($P < 0.001$), as well as homozygous mutant embryos without exencephaly compared to homozygous mutant embryos with exencephaly ($P < 0.001$, Table 3.3.1).

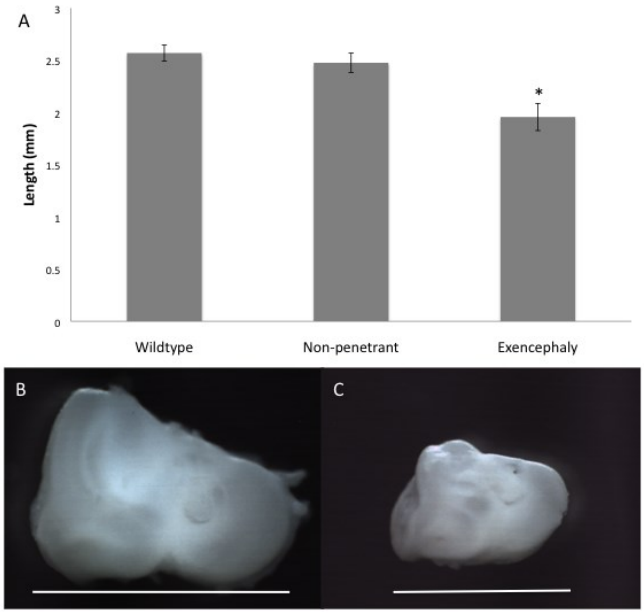


Figure 3.3.1: Small inner ears are secondary to the exencephaly phenotype in congenic *Cecr2*^{GT45bic} BALB/cCrl mutant embryos. Both inner ears were removed from E17.5 day embryos, fixed in paraformaldehyde, and imaged. Length was measured using ImageJ and the average length of the two inner ears per mouse was used for statistical analysis. (**A**) Homozygous mutant embryos with exencephaly have significantly smaller inner ears (denoted with asterisk) than homozygous wildtype embryos and homozygous mutant embryos without exencephaly. Error bars represent the 95% confidence interval. (**B**) Image of an inner ear from a wildtype embryo and (**C**) a homozygous mutant embryo with exencephaly. White lines indicate where measurement was taken.

Table 3.3.1: Statistical analysis of inner ear size comparing E17.5 congenic *Cecr2*^{GT45bic} BALB/cCrl embryos wildtype for *Cecr2* (WT), homozygous mutant for the *Cecr2*^{GT45bic} allele but not penetrant for exencephaly (NP), and homozygous mutant for the *Cecr2*^{GT45bic} allele with exencephaly (E). Inner ear length was averaged for each mouse and then analyzed with one-way ANOVA followed by a Tukey's post-hoc test.

Comparison	*P*-value
WT : NP	0.683
WT: E	0.000
NP : E	0.000

3.4: Exencephaly penetrance of congenic *Cecr2*$^{tm1.1Hemc}$ FVB/N mice is lower than the penetrance in the earlier N3 generation

The *Cecr2*$^{tm1.1Hemc}$ mutant allele resulted in a more severe exencephaly phenotype than the *Cecr2*^{GT45bic} allele in the BALB/cCrl background, with 96.0% of 3rd – 4th generation BALB/cCrl *Cecr2*$^{tm1.1Hemc}$ homozygous mutant mice developing exencephaly [142]. This difference in penetrance between the two alleles holds true for FVB/N, as incipient congenic *Cecr2*^{GT45bic} FVB/N homozygous mutant mice do not develop exencephaly [141], but an exencephaly penetrance analysis of *Cecr2*$^{tm1.1Hemc}$ homozygous mutant mice that were generation N3 (~87.5% FVB/N) yielded a penetrance of 31.4% [145]. As we saw a drop in exencephaly penetrance in congenic *Cecr2*^{GT45bic} BALB/cCrl compared to analyses performed on earlier generations, we were interested in determining if this same phenomenon occurred in congenic *Cecr2*$^{tm1.1Hemc}$ FVB/N.

I collected and analyzed a total of 73 homozygous mutant embryos, with results indicating an exencephaly penetrance of 12.3% (9/73) when sexes were combined (Table 3.4.1). The *Cecr2*$^{tm1.1Hemc}$ allele was inherited in a perfectly Mendelian fashion (1:2:1 ratio), with a total of 292 embryos collected (73 homozygous mutants, 146 heterozygotes, and 73 homozygous wildtype embryos), which demonstrated that there was no significant loss of mutant embryos prior to implantation or early in development relative to heterozygous or homozygous wildtype siblings. Exencephaly was not seen in any of the heterozygous or homozygous wildtype embryos. To determine if there was a sex difference in exencephaly penetrance, as was seen in congenic *Cecr2*^{GT45bic} BALB/cCrl homozygous mutant mice, I analyzed sexes separately. Five out of 29 females (14.7%) and 4 out of 35 males (10.3%) developed exencephaly (Table 3.4.1). There was no significant difference between sexes according to a χ^2 test-of-independence. However, it is possible that this difference may become statistically significant if more embryos were looked at, as there were only a total of 9 embryos that developed exencephaly. Exencephaly penetrance dropped from the 31.4% (11/35) seen in generation N3 FVB/N mice to the 12.3% seen here in congenic FVB/N. A χ^2 test-of-independence demonstrated that the exencephaly penetrance in congenic FVB/N was significantly lower than the exencephaly penetrance obtained from the previous analysis on generation N3 FVB/N ($P = 0.017$). These results confirm that the FVB/N genetic background harbors a resistance towards the development of exencephaly relative to

BALB/cCrl. However, the FVB/N genetic background could not fully protect embryos that were homozygous mutant for the more severe allele, *Cecr2*$^{tm1.1Hemc}$, from developing exencephaly.

Table 3.4.1: Congenic *Cecr2*$^{tm1.1Hemc}$ FVB/N homozygous mutant exencephaly penetrance analysis.

	Female	Male	Combined sexes	N3 FVB/N (old analysis) [142]
Exencephaly	5	4	9	11
Normal	29	35	64	24
Penetrance (%)	14.7	10.3	12.3	31.4

3.5: Modifier genes within the chromosome 19 modifier region are not additive

Initial exencephaly penetrance analyses demonstrated significantly lower exencephaly penetrance in MOD 4, MOD 5, and MOD 31 sub-interval congenic mouse lines compared to the original exencephaly penetrance in incipient congenic *Cecr2*^{GT45bic} BALB/cCrl of 74.5% [165], confirming that the FVB/N chromosome 19 region contains resistance loci. However, in light of my more recent exencephaly penetrance analysis yielding a penetrance of 54.1% in congenic *Cecr2*^{GT45bic} BALB/cCrl, I performed another exencephaly penetrance analysis on the MOD 5 and MOD 31 sub-interval congenic mouse lines. I was unable to perform a penetrance analysis on the MOD 4 line as it was no longer available. Results indicated a lower penetrance than previously reported, with MOD 5 penetrance dropping from the previously reported 50.0% to 33.3% (24/72) and MOD 31 penetrance dropping from the previously reported 46.1% to 31.4% (11/35) (Figure 3.5.1, Table 3.5.2). A χ^2 test-of-independence demonstrated that exencephaly penetrance was significantly lower in MOD 5 and MOD 31 compared to the 54.1% penetrance seen in congenic *Cecr2*^{GT45bic} BALB/cCrl ($P < 0.05$) (Table 3.5.1).

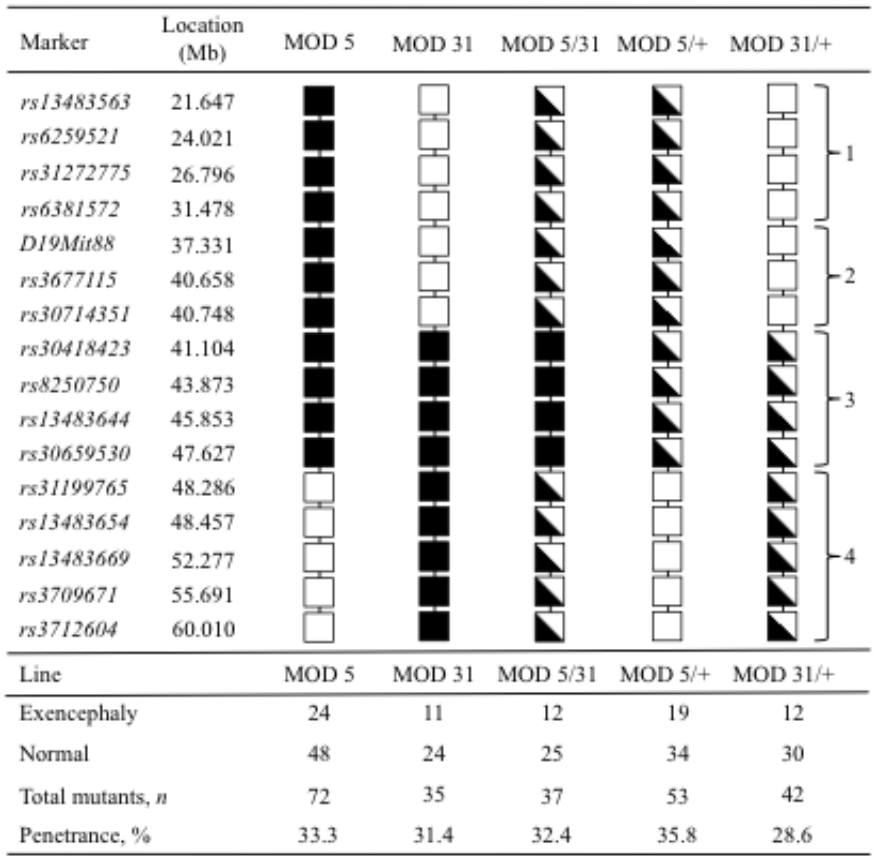
Previous data from Kooistra et al. (2012) showed that there is more than one modifier locus within the chromosome 19 region, as there was no genomic region of FVB/N origin common to all three MOD lines tested, which all demonstrated a similar drop in exencephaly penetrance (Figure 3.1.3). To confirm this, I performed a control cross. Since there was no line available in which the entire FVB chromsome 19 modifier region was present on a BALB/cCrl background, MOD 5 and MOD 31 were crossed to each other to generate progeny that were heterozygous for regions 1, 2 and 4, and homozygous for region 3 (herein referred to as MOD

5/31). FVB/N resistance loci have already been shown to be dominant. Therefore, the entire region is at least heterozygous for the dominant FVB/N resistance loci, and so should mimic the entire region being homozygous for FVB/N. Exencephaly penetrance in these embryos was determined to be 32.4% (12/37) which is similar to the exencephaly penetrance in MOD 4, MOD 5, and MOD 31 (Figure 3.5.1, Table 3.5.2) [165].

Thus far, all MOD lines demonstrated a similar penetrance in exencephaly. If there are two modifiers, one in region 1 and one in region 4, it is possible that they are acting in a co-dominant, additive manner. This means that it is possible that a single FVB/N modifier allele confers a certain amount of resistance, and having two alleles, either both in region 1 (MOD 4 and MOD 5), both in region 4 (MOD 31), or one in region 1 and the other in region 4 (MOD 5/31) would all result in similar frequencies of exencephaly. This hypothesis excludes the possibility of a modifier locus existing in region 3, as this would result in three FVB/N modifier alleles in MOD 5/31 embryos or four modifier alleles in MOD 5 embryos. If there was a modifier in region 3, then the proposed hypothesis would predict a reduced exencephaly penetrance in MOD 5 and MOD 5/31 compared to the MOD 31 line – but this was not the case. Under this hypothesis of only 2 modifier loci, one in each of region 1 and 4, we would predict that having only one allele in a heterozygous state would not provide the same amount of resistance as two alleles. This would therefore result in an increased frequency of exencephaly compared to the MOD lines. To test this hypothesis, MOD 5 (homozygous FVB/N for region 1, heterozygous for the *Cecr2*^{GT45bic} allele) was crossed to congenic *Cecr2*^{GT45bic} BALB/cCrl heterozygotes to produce *Cecr2*^{GT45bic} mutants that only contain one FVB/N modifier allele in region 1 (herein referred to as MOD 5/+). Under the co-dominant, additive model, we would expect an exencephaly penetrance intermediate between congenic BALB/cCrl (54.1%) and MOD 5 (33.3%). The results did not support this model, as MOD 5/+ embryos displayed an exencephaly penetrance of 35.5% (n = 53) (Figure 3.5.1, Table 3.5.2). This analysis was also performed for MOD 31 (region 4), where the exencephaly penetrance of MOD 31/+ embryos was 28.6% (n = 42) (Figure 3.5.1, Table 3.5.2) [165]. There was no significant difference in exencephaly penetrance between any of the MOD lines or crosses (Table 3.5.1). There was also no significant difference when all combinations of any of these penetrance analyses were compared to the original penetrance analyses performed on MOD 4, MOD 5, and MOD 31 (Figure 3.1.3). This may be due to small sample sizes, as some statistical comparisons did come close to significance. Also,

the original penetrance in MOD 4 of 36.2% was very similar to the penetrance seen in this recent analysis. The *Cecr2*^{GT45bic} allele was inherited in a Mendelian fashion for all MOD crosses according to χ^2 goodness-of-fit tests ($P > 0.05$). Exencephaly was not seen in any heterozygous or homozygous wildtype embryos.

Based on these results, we concluded that one of the modifiers is located in region 1. Since the modifiers do not appear to be additive, the location of the second modifier, if there are indeed only two, could be in either region 3 (meaning that MOD 5 would contain both) or 4. This is important, since *Arhgap19* and other candidate genes are in region 3.



Line	MOD 5	MOD 31	MOD 5/31	MOD 5/+	MOD 31/+
Exencephaly	24	11	12	19	12
Normal	48	24	25	34	30
Total mutants, *n*	72	35	37	53	42
Penetrance, %	33.3	31.4	32.4	35.8	28.6

Figure 3.5.1: Exencephaly penetrance analyses in MOD 5, MOD 31, MOD 5/31, MOD 5/+, and MOD 31/+. The FVB/N modifier alleles are dominant and not additive. There is at least one modifier gene within region 1 (based on MOD 4 data in Figure 3.1.3) and at least one modifier gene within the combined regions 3 and 4. Black boxes indicate homozygous FVB/N genetic background, white boxes indicate homozygous BALB/cCrl genetic background, and half-black half-white boxes indicate genomic regions heterozygous for the FVB/N and BALB/cCrl genetic backgrounds.

Table 3.5.1: χ^2 test-of-independence comparing exencephaly penetrance between congenic *Cecr2*^{GT45bic} BALB/cCrl, MOD 5, MOD 31, MOD 5/31, MOD 5/+, and MOD 31/+ mouse embryos homozygous mutant for the *Cecr2*^{GT45bic} allele. All mouse lines containing at least one copy of the FVB/N modifier region (MOD) show significantly lower exencephaly penetrance than congenic *Cecr2*^{GT45bic} BALB/cCrl. Exencephaly penetrances for all MOD mouse lines were not significantly different from each other.

Comparison	*P*-value
BALB/cCrl : MOD 5	4.39E-03
BALB/cCrl : MOD 31	1.69E-02
BALB/cCrl : MOD 5/31	1.97E-02
BALB/cCrl : MOD 5/+	2.45E-02
BALB/cCrl : MOD 31/+	3.86E-03
MOD 5 : MOD 31	0.844
MOD 5 : MOD 5/31	0.925
MOD 5 : MOD 5/+	0.770
MOD 5 : MOD 31/+	0.598
MOD 31 : MOD 5/31	0.927
MOD 31 : MOD 5/+	0.669
MOD 31 : MOD 31/+	0.785
MOD 5/31 : MOD 5/+	0.737
MOD 5/31 : MOD 31/+	0.710
MOD 5/+ : MOD 31/+	0.452

As sex differences were seen in congenic *Cecr2*^{GT45bic} BALB/cCrl, exencephaly penetrance between males and females for each MOD penetrance analysis was also compared using the χ^2 test-of-independence. There did not appear to be any sex differences for MOD 5/31 (male 33.3%, 6/18; female 31.6%, 6/19) or for MOD 31/+ (male 28.0%, 7/ 25; female 29.4%, 5/17) ($P > 0.05$). Sex differences were observed for MOD 5 (male 29.0%, 9/31; female 36.6%, 15/41), and MOD 31 (male 18.8%, 3/16; female 42.1%, 8/19) but were not significant ($P > 0.05$). A significant difference was seen in MOD 5/+ (male 18.2%, 4/22; female 48.4%, 15/31) ($P < 0.05$). A summary of all exencephaly penetrance analyses, which include *P*-values for sex comparisons, is displayed in Table 3.5.2.

Table 3.5.2 Summary of exencephaly penetrance of males, females, and combined sexes for congenic *Cecr2*^{GT45bic} BALB/cCrl, MOD 5, MOD 31, MOD 5/31, MOD 5/+, and MOD 31/+. Differences in exencephaly penetrance between sexes were analyzed using a χ^2 test-of-independence.

Mouse line	Female Penetrance	Percent	Male Penetrance	Percent	Sex differences (*P*-value)	Combined sex penetrance	Percent
BALB/cCrl	50/74	67.6%	23/61	37.7%	5.30E-04	73/135	54.1%
MOD 5	15/41	36.6%	9/31	29.0%	0.501	24/72	33.3%
MOD 31	8/19	42.1%	3/16	18.8%	0.138	11/35	31.4%
MOD 5/31	6/19	31.6%	6/18	33.3%	0.909	12/37	32.4%
MOD 5/+	15/31	48.4%	4/22	18.2%	2.39E-02	19/53	35.8%
MOD 31/+	5/17	29.4%	7/25	28.0%	0.921	12/42	28.6%

Small inner ears are an additional phenotype seen in homozygous mutant congenic *Cecr2*^{GT45bic} BALB/cCrl embryos. Prior to establishing that the small inner ear phenotype was secondary to the exencephaly phenotype, as presented in section 3.3, I was interested in determining if the FVB/N resistance modifier loci also reduced the penetrance of the small inner ear phenotype. Therefore, I also measured inner ears in *Cecr2*^{GT45bic} homozygous mutant MOD 5 and MOD 31 embryos. As was shown for congenic *Cecr2*^{GT45bic} BALB/cCrl, the small inner ear phenotype in MOD 5 and MOD 31 embryos was only apparent in homozygous mutant embryos that developed exencephaly (Figure 3.5.2, Table 3.5.3). *Cecr2*^{GT45bic} homozygous mutant E17.5 MOD 5 embryos with exencephaly ($n = 6$) had significantly smaller inner ears than *Cecr2*^{GT45bic} homozygous mutant E17.5 MOD 5 embryos without exencephaly ($n = 7$, $P < 0.001$) and E17.5 MOD 5 embryos wildtype for *Cecr2* ($n = 13$, $P < 0.001$). The same was true for MOD 31, where *Cecr2*^{GT45bic} homozygous mutant E17.5 embryos with exencephaly ($n = 3$) had significantly smaller inner ears than *Cecr2*^{GT45bic} homozygous mutant E17.5 embryos without exencephaly ($n = 9$, $P < 0.001$) and E17.5 embryos wildtype for *Cecr2* ($n = 6$, $P < 0.001$). There was no significant difference in inner ear size between the congenic *Cecr2*^{GT45bic} BALB/cCrl, MOD 5, and MOD 31 mouse lines when comparing E17.5 embryos wildtype for *Cecr2* to each other, *Cecr2*^{GT45bic} homozygous mutant but not penetrant for exencephaly to each other, and *Cecr2*^{GT45bic} homozygous mutant with exencephaly to each other ($P > 0.05$).

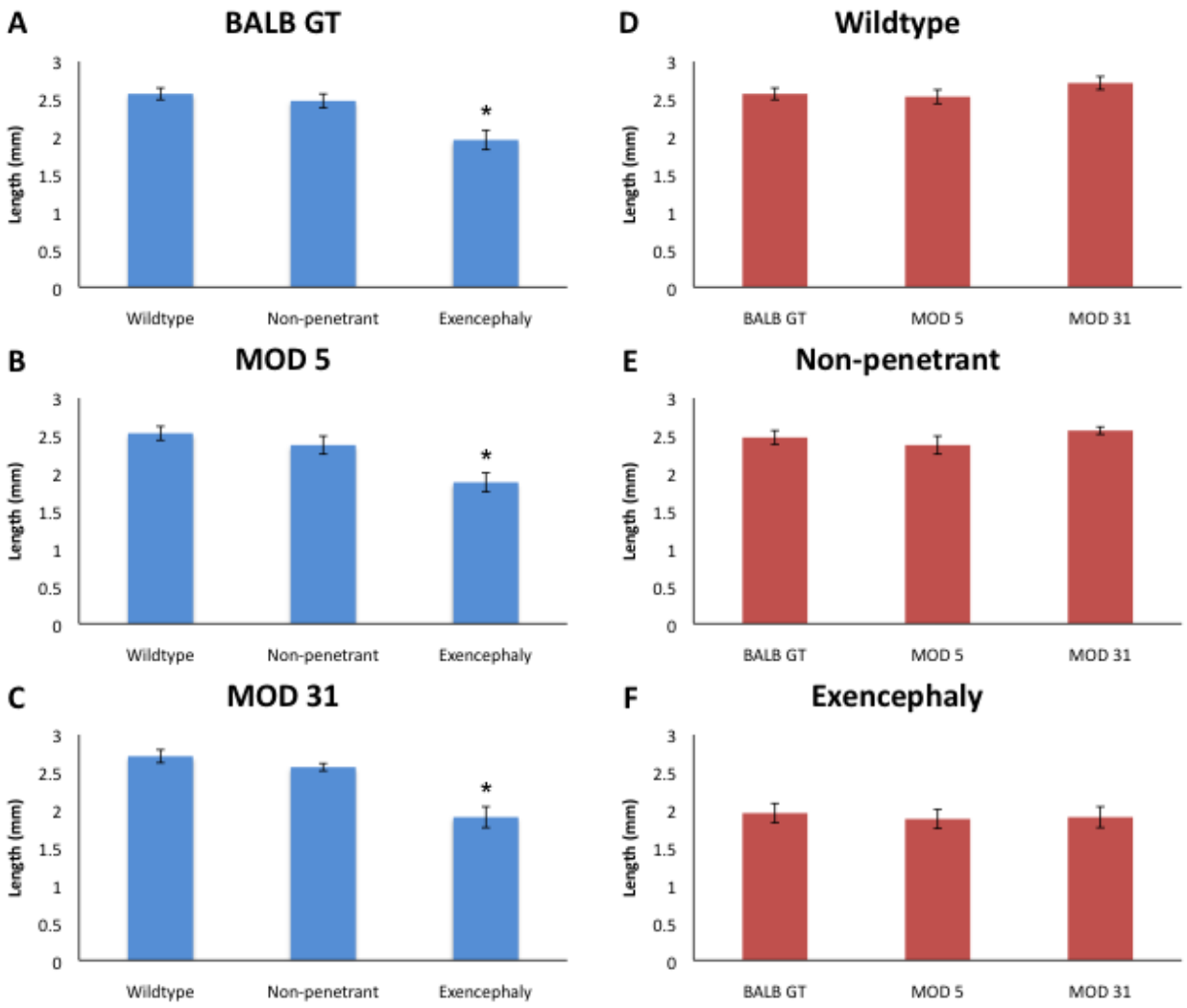


Figure 3.5.2: Inner ears are significantly smaller in embryos with exencephaly. Both inner ears were removed from E17.5 day embryos, fixed in paraformaldehyde, and imaged. Length was measured using ImageJ and the average length of the two inner ears per mouse was used for statistical analysis. One-way ANOVA followed by Tukey's post-hoc revealed that embryos with exencephaly have significantly smaller inner ears than embryos without exencephaly (**A, B, C**). There was no difference in inner ear size between the three lines analyzed when comparing wildtype (**D**), non-penetrant mutant (**E**), or exencephalic embryos (**F**). Error bars represent the 95% confidence interval. Asterisk represents a *P*-value < 0.05.

Table 3.5.3: Statistical analysis of inner ear size comparing E17.5 embryos wildtype for *Cecr2* (WT), homozygous mutant for the *Cecr2*^{GT45bic} allele but not penetrant for exencephaly (NP), and homozygous mutant for the *Cecr2*^{GT45bic} allele with exencephaly (E), between congenic *Cecr2*^{GT45bic} BALB/cCrl, MOD 5, and MOD 31 mouse lines. Inner ear length was averaged for each mouse and then analyzed with one-way ANOVA followed by a Tukey's post-hoc test.

Comparison	*P*-value
MOD 5 (WT) : MOD 5 (NP)	0.150
MOD 5 (WT) : MOD 5 (E)	0.000
MOD 5 (NP) : MOD 5 (E)	0.000
MOD 31 (WT) : MOD 31 (NP)	0.379
MOD 31 (WT) : MOD 31 (E)	0.000
MOD 31 (NP) : MOD 31 (E)	0.000
BALB/cCrl (WT) : MOD 5 (WT)	0.996
BALB/cCrl (WT) : MOD 31 (WT)	0.294
MOD 5 (WT) : MOD 31 (WT)	0.082
BALB/cCrl (NP) : MOD 5 (NP)	0.771
BALB/cCrl (NP) : MOD 31 (NP)	0.802
MOD 5 (NP) : MOD 31 (NP)	0.059
BALB/cCrl (E) : MOD 5 (E)	0.979
BALB/cCrl (E) : MOD 31 (E)	1.000
MOD 5 (E) : MOD 31 (E)	1.000

3.6: qRT-PCR validation of microarray data yields a list of 9 candidate modifier genes differentially expressed between BALB/cCrl and FVB/N

Whole genome microarray GC-RMA/Student's t-test analysis identified 22 genes differentially expressed by at least 1.5 fold between BALB/cCrl and FVB/N wildtype strains (Table 3.1.1). In collaboration with Megan Kooistra, gene expression was validated by qRT-PCR for 16 of these genes, of which 9 showed a significant difference in expression between mouse strains ($P < 0.05$) (Table 3.6.1). I did not validate gene expression for the additional 6 genes because further analysis of the binding location of the microarray probe revealed that the probe did not fall on a transcribed region or the gene was not a good candidate based on what was known about expression and/or function. Seven of the 9 genes validated by qRT-PCR, *Lipo1*, *Arhgap19*, *Foxd4*, *Tmem180*, *Sfxn2*, *Rnls*, and *Scd1*, all showed a fold-change (FC) greater than 1.5. Although statistically significant, *Scd2* was only 1.47 fold lower in FVB/N relative to

BALB/cCrl, and *Pnliprp2* was only 1.38 fold higher in FVB/N relative to BALB/cCrl. On the microarray, *Lipa* demonstrated one of the largest gene expression differences between strains; however, qRT-PCR validation demonstrated that there was no expression difference. I found that there existed several SNVs that differed between BALB/cCrl and FVB/N within the region that the microarray probe bound to, which led to inefficient probe binding in one strain compared to the other resulting in erroneous microarray data. Therefore, to ensure that the remaining qRT-PCR results would not present with a similar artifact, I checked the regions where qRT-PCR primer-probe sets bound to for the presence of SNVs by Sanger sequencing or by referencing the Ensembl online database.

Table 3.6.1: qRT-PCR validation of candidate modifier genes identified by whole genome microarray. Presented microarray fold-change (FC) was determined by GC-RMA/Student's t-test analysis, except for *Vldlr*, which was determined by RMA/ANOVA analysis. Asterisk denotes validation performed by Megan Kooistra [199]. The line in the middle of the table separates significant qRT-PCR *P*-values from those that were not significant. A subset of these data has previously been published [165].

Gene Symbol	MOD region	FC microarray	*P*-value	FC qRT-PCR	95% CI	*P*-value
Lipo1	1	-6.26	4.18E-05	-7.15	0.36	5.40E-04
*Arhgap19**	3	21.43	3.27E-08	5.41	0.15	1.01E-03
*Foxd4**	1	2.65	3.08E-02	2.68	0.31	2.25E-04
*Tmem180**	3	-1.73	3.08E-05	-1.96	0.07	4.07E-06
Sfxn2	3	-1.51	6.04E-04	-1.78	0.06	4.29E-03
Rnls	1	-1.57	2.07E-02	-1.70	0.08	1.42E-02
Scd1	3	-1.54	1.82E-04	-1.57	0.07	1.00E-02
*Scd2**	3	-2.67	4.03E-04	-1.47	0.13	8.33E-03
Pnliprp2	4	1.64	3.02E-02	1.38	0.14	3.57E-02
*Dmrt2**	1	2.11	4.07E-02	-1.03	0.19	>0.05
*Exosc1**	3	-2.99	3.47E-07	-1.02	0.19	>0.05
Lipa	1	-2.41	7.30E-04	1.08	0.06	>0.05
Vldlr	1	-1.55	1.02E-03	1.03	0.11	>0.05
Fam160b1	4	2.27	4.89E-04	-1.26	0.08	>0.05
Fam45a	4	1.61	1.27E-04	1.12	0.07	>0.05
Hif1an	3	1.50	2.42E-04	-1.11	0.05	>0.05

Foxd4 was of particular interest because of its possible role in regulating transcription of *Zic2* [204], a known NTD gene in mouse [205]. Therefore, I compared *Foxd4* gene expression between four strains, BALB/cCrl, FVB/N, 129S2, and C57BL/6 to determine which of the two strains,

BALB/cCrl or FVB/N, is unique in *Foxd4* expression (Figure 3.6.1). FVB/N and C57BL/6 showed no significant difference in *Foxd4* expression. However, *Foxd4* expression was significantly lower by 2.28 fold in BALB/cCrl relative to FVB/N ($P = 0.024$), and by 1.38 fold in 129S2 relative to FVB/N ($P = 0.040$).

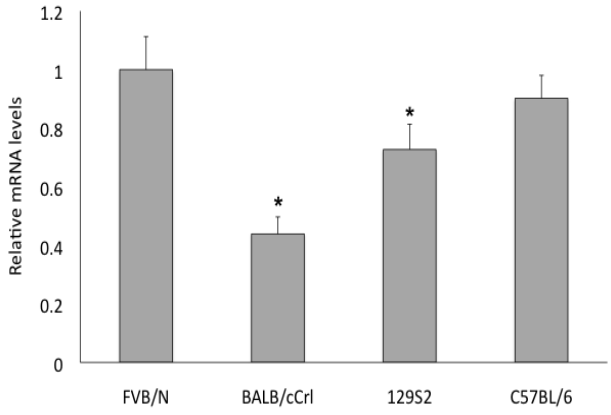


Figure 3.6.1: *Foxd4* expression is significantly down-regulated in BALB/cCrl and 129S2, but not C57BL/6, relative to FVB/N. Gene expression of *Foxd4* was analyzed in embryos at the time of neurulation by qRT-PCR. Asterisks denote significant *P*-values relative to FVB/N. Error bars represent the standard error.

3.7: Whole exome sequencing of BALB/cCrl and FVB/N identifies coding variants of interest in 19 genes within the chromosome 19 candidate modifier region

A modifier gene may be expressed differentially between strains, or may have similar expression levels but contain a functional mutation, or both. To complement microarray data by looking for sequence variants that may alter function without affecting expression, we performed whole exome sequencing (WES) on a single BALB/cCrl female and a single FVB/N female to identify any single nucleotide variants (SNVs) or small insertions or deletions (INDELs) that exist between these two strains. The WES experiment was done via collaboration with Drs. Allison Ashley-Koch and Simon Gregory at Duke University.

The analyses done at Duke were able to identify both SNVs and INDELs. WES identified a total of 2456 SNVs affecting 371 genes and 434 INDELs affecting 147 genes (Table 3.7.1). After filtering out intronic, synonymous, untranslated region (3'UTR or 5'UTR) variants, as well as candidates within modifier region 2, and accounting for what is currently known about gene expression/function, we were left with a total of 19 candidate genes within the modifier region, 15 of which contained SNVs (Table 3.7.2), and 4 of which contained INDELs (Table 3.7.3). One of the four INDELs identified by WES was the previously identified *Arhgap19*^{Ex6non} novel

nonsense mutation that is unique to BALB/cCrl, thereby serving as a good positive control for the INDEL analysis performed at Duke University. I analyzed SNVs *in silico* with three predictors of deleteriousness. GERP scores reflect the evolutionary conservation of a particular nucleotide by comparing its rate of substitution to the neutral rate of substitution in the mouse genome. It is important to look at conservation at the nucleotide level because although conservative amino acid changes or silent mutations may have little to no effect on protein function, a particular nucleotide may be highly conserved as mutations may result in erroneous splicing events [206]. SNVs resulting in amino acid changes were analyzed with the protein-sequence-based predictors of deleteriousness Polyphen2, which takes into account evolutionary conservation along with biochemical data such as amino acid properties and structural information, and SIFT, which takes into account evolutionary conservation (Table 3.7.2).

Table 3.7.1: Summary of SNVs and INDELs identified by WES in the chromosome 19 modifier region. The variants are within the genomic coordinates 19:21630117-61203228 (mm9 genome assembly).

Mutation class	**SNVs**	**Number of genes affected**
Nonsynonymous	90	44
Synonymous	268	81
Stop lost	1	1
Splice	5	5
UTR	115	58
Intronic	1831	143
Non-coding RNA	6	4
Intergenic	143	38
TOTAL	**2459**	**374**
Mutation class	**INDELs**	**Number of genes affected**
Frameshift	1	1
Non-frameshift	5	4
Splice	4	4
UTR	29	20
Intronic	394	117
Non-coding RNA	1	1
Intergenic	0	0
TOTAL	**434**	**147**
GRAND TOTAL	**2893**	**521**

Table 3.7.2: Candidate modifier genes containing single nucleotide variants (SNVs) that alter the amino acid sequence that differ between BALB/cCrl and FVB/N wildtype mice. All variants are homozygous in BALB/cCrl and FVB/N.

Gene	dbSNP ID	REF	Balb/cCrl	FVB/N	AA change	GERP	SIFT	PP2 (HDIV)
Btaf1	rs30319309	C	G	C	p.Pro769Arg	-1.99	0.07	*0.897*
Cpeb3	rs242552627	A	A	G	p.Ile48Thr	0	0.91	0
Dnmbp	rs37627946	G	A	G	p.Pro948Leu	-0.345	**0.01**	**0.989**
	rs30860292	C	T	C	p.Val700Met	-4.25	0.33	*0.497*
Fam160b1	rs31128381	T	T	A	p.Ser548Thr	0.18	0.31	*0.782*
Fas	rs47662284	A	A	G	p.Asp281Gly	-1.41	1	0.008
Hif1an	rs51053298	T	T	G	p.Arg177Leu	0.67	0.33	0.006
Hps6	rs30714718	T	T	G	p.Phe384Leu	1.57	0.08	**0.98**
Mms19	rs39183008	T	T	C	p.Ile102Val	1.37	1	0.026
Nanos1	rs30854475	C	A	C	p.Pro111Thr	**2.27**	0.29	**0.997**
Tctn3	rs13483625	A	G	A	p.Met186Thr	0.461	1	0
Trub1	rs13467218	C	C	T	p.Ser89Phe	0.461	**0.01***	0.005
Foxd4	rs3023492	C	C	T	p.Ala371Thr	0.205	0.05	**0.962**
	rs37460493	C	T	C	p.Cys51Tyr	0	1	*0.731*
	rs50746028	T	G	T	p.Glu41Asp	**3.15**	0.23	0.001
Glis3	rs48544080	T	T	G	p.Lys95Thr	0.964	**0***	*0.816*
	rs30347506	C	C	T	p.Gly379Glu	0	0.17	*0.775*
Kank1	rs51029258	A	A	G	p.Gln818Arg	0.841	0.12	*0.841*
Tjp2	rs37013001	G	T	G	p.Thr731Asn	1.3	0.64	0

Abbreviations – REF: Reference sequence (C57BL/6). AA: amino acid. GERP: genomic evolutionary rate profiling. SIFT: sorting intolerant from tolerant. PP2: Polyphen2.
Asterisk – Low confidence prediction due to insufficient sequence diversity.
Bold numbers – Predicted to be damaging (SIFT <0.05, PP2 (HDIV) >0.956) or evolutionary conserved for GERP (>2).
Italicized numbers – Predicted to be possibly damaging (PP2 (HDIV) between 0.453 and 0.956).

Table 3.7.3: Candidate modifier genes containing an insertion or deletion (INDEL) that alter the amino acid sequence that differ between BALB/cCrl and FVB/N wildtype mice. All variants are homozygous in BALB/cCrl and FVB/N.

Gene	dbSNP ID	REF	Balb/cCrl	FVB/N	Mutation	AA change
Arhgap19	NA	--	T	--	Frameshift	p.Lys309_Ala310insX
Cstf2t	rs237707176	--	--	AGAGGGATGGAAGCA	Nonframeshift	p.Thr454_Arg455insArgGlyMetGluAla
Exosc1	rs230635656	--	AAACAAAC	--	Splicing	NA
Fam45a	rs258299840	--	--	TCT	Nonframeshift	p.Asp21_Ser22insSer

Abbreviations – REF: Reference sequence (C57BL/6). AA: amino acid.

Genes differentially expressed (microarray) and/or that contained protein-coding variants (WES) were combined into a prioritized candidate gene list. In order to be included in the prioritized candidate gene list, a gene must (a) fall within the chromosome 19 MOD region 1, 3 or 4, (b) show a significant difference in expression (at least 1.5 fold) based on microarray data and be verified by qRT-PCR and/or (c) contain a functional mutation based on whole exome data that was validated by existing in at least one database (ex. db SNP) and/or Sanger sequencing. Candidate genes identified by WES must also meet two of the three following criteria: (a) *in silico* prediction of a deleterious mutation, (b) predicted or known gene function could be involved in neurulation, and (c) the gene is predicted to be expressed around the time of neurulation based on the literature and/or database. The combination of WES and microarray analyses yielded a total of 26 prioritized candidate modifiers in mouse (Table 3.7.3).

Table 3.7.3: Refined list of candidate modifier genes that differ in expression and/or amino acid sequence between BALB/cCrl and FVB/N.

Mouse gene	Chr. 19 modifier region	Differs in expression	Differs in amino acid sequence	Human homologue
Arhgap19	3	Yes	Yes	*ARHGAP19*
Foxd4	1	Yes	Yes	*FOXD4*
Lipo1	1	Yes	×	*LIPJ*
Tmem180	3	Yes	×	*TMEM180*
Sfxn2	3	Yes	×	*SFXN2*
Rnls	1	Yes	×	*RNLS*
Scd1	3	Yes	×	*SCD*
Scd2	3	Yes	×	
Pnliprp2	4	Yes	×	*PNLIPRP2*
Btaf1	1	×	Yes	*BTAF1*
Cpeb3	1	×	Yes	*CPEB3*
Dnmbp	3	×	Yes	*DNMBP*
Fam160b1	4	×	Yes	*FAM160B1*
Fas	1	×	Yes	*FAS*
Hif1an	3	×	Yes	*HIF1AN*
Hps6	3	×	Yes	*HPS6*
Mms19	3	×	Yes	*MMS19*
Nanos1	4	×	Yes	*NANOS1*
Tctn3	3	×	Yes	*TCTN3*
Trub1	4	×	Yes	*TRUB1*
Glis3	1	×	Yes	*GLIS3*
Kank1	1	×	Yes	*KANK1*
Tjp2	1	×	Yes	*TJP2*
Cstf2t	1	×	Yes	*CSTF2T*
Exosc1	3	×	Yes	*EXOSC1*

Mouse gene	Chr. 19 modifier region	Differs in expression	Differs in amino acid sequence	Human homologue
Fam45a	4	×	Yes	*FAM45A*

3.8: *Arhgap19* is not a modifier of *Cecr2*-associated exencephaly

Arhgap19 was the top candidate modifier gene as it is located within the chromosome 19 modifier region 3 and *Arhgap19* expression was significantly lower in BALB/cCrl compared to FVB/N, C57BL/6, and 129S2. Also, BALB/cCrl was shown to be homozygous for a nonsense mutation within *Arhgap19* exon 6 ($Arhgap19^{Ex6non}$) that resulted in a truncated protein product just after the RhoGAP domain. Therefore, I tested *Arhgap19* as a modifier of *Cecr2*-associated exencephaly by performing two genetic analyses in mice. Although much of the $Arhgap19^{Ex6non}$ transcript in wildtype BALB/cCrl was likely degraded through nonsense-mediated decay, some of the transcript may have still produced some functional protein product. Therefore, we acquired a mouse with a different, potentially more severe, mutant allele of *Arhgap19.* This allele consisted of a pGT01xf vector genetrap cassette inserted within intron 2 that produced a truncated protein product devoid of the functional RhoGAP domain and fused to β-galactosidase, herein referred to as $Arhgap19^{GT(YHD020)Byg}$ (Figure 3.8.1A). Based on qRT-PCR analysis, there was an approximately 15-fold reduction of wildtype *Arhgap19* transcript in E10.5 homozygous mutant embryos, indicative of a hypomorphic allele with some wildtype transcript present due to splicing around the genetrap cassette (Figure 3.8.1B). Therefore, both *Arhgap19* mutant alleles may have residual activity; however, the $Arhgap19^{GT(YHD020)Byg}$ allele produces a β-galactosidase fusion protein product that allows for X-gal staining, which can be used to assess *Arhgap19* expression. X-gal staining of embryos throughout development revealed what appeared to be ubiquitous *Arhgap19* expression in as early as E9.5 whole-mount embryos, which is around the time of neurulation. There was strong specific expression in the telencephalon and midbrain of E12.5 and E13.5 embryos; however, there was no longer detectable expression in the brain of E14.5 embryos (Figures 3.8.2 and 3.8.3). Embryos and adult mice homozygous mutant for the $Arhgap19^{GT(YHD020)Byg}$ allele all appeared to be morphologically normal.

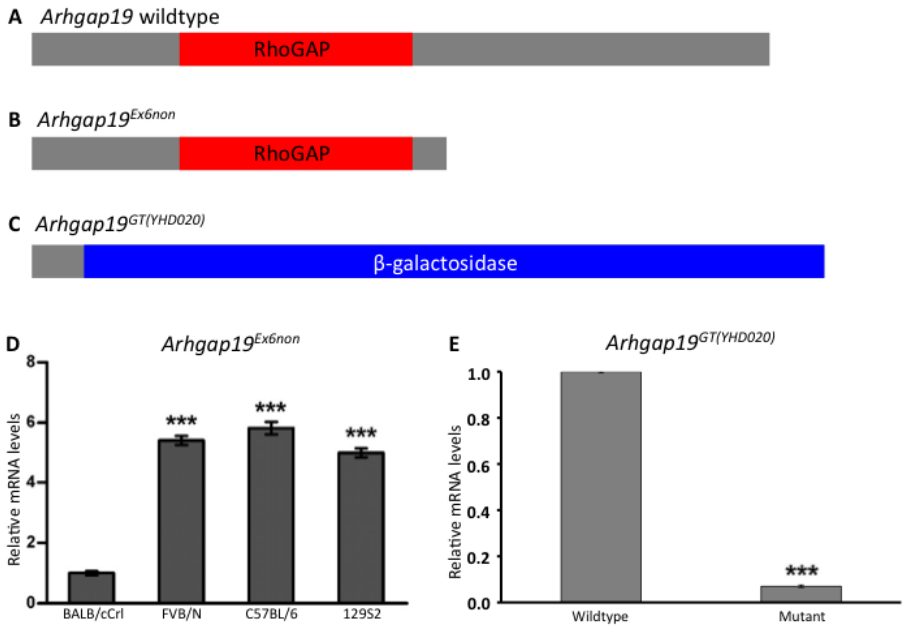


Figure 3.8.1: Wildtype and mutant alleles of *Arhgap19*. Wildtype *Arhgap19* contains a RhoGAP domain (**A**). The *Arhgap19*^{Ex6non} allele is specific to BALB/cCrl and contains a 'T' insertion in exon 6 that results in truncation of the protein just after the RhoGAP domain (**B**). Expression of the *Arhgap19*^{Ex6non} allele is approximately 5-fold lower in BALB/cCrl neurulating embryos relative to FVB/N, C57BL/6, and 129S2 neurulating embryos (Megan Kooistra generated this expression data, this section is an image modified from Kooistra et al. (2012) [165]), error bars represent the standard deviation from the mean (**D**). The *Arhgap19*$^{GT(YHD020)Byg}$ allele contains a genetrap cassette in intron 2, which produces a truncated protein product devoid of the RhoGAP domain and fused to β-galactosidase (**C**). Expression of the *Arhgap19*$^{GT(YHD020)Byg}$ allele as determined by qRT-PCR is approximately 15-fold lower in E10.5 homozygous mutant embryos compared to wildtype littermates, error bars represent the 95% confidence interval (**E**). Embryos were a mixed FVB/N:129P2 genetic background (~87.5% FVB/N).

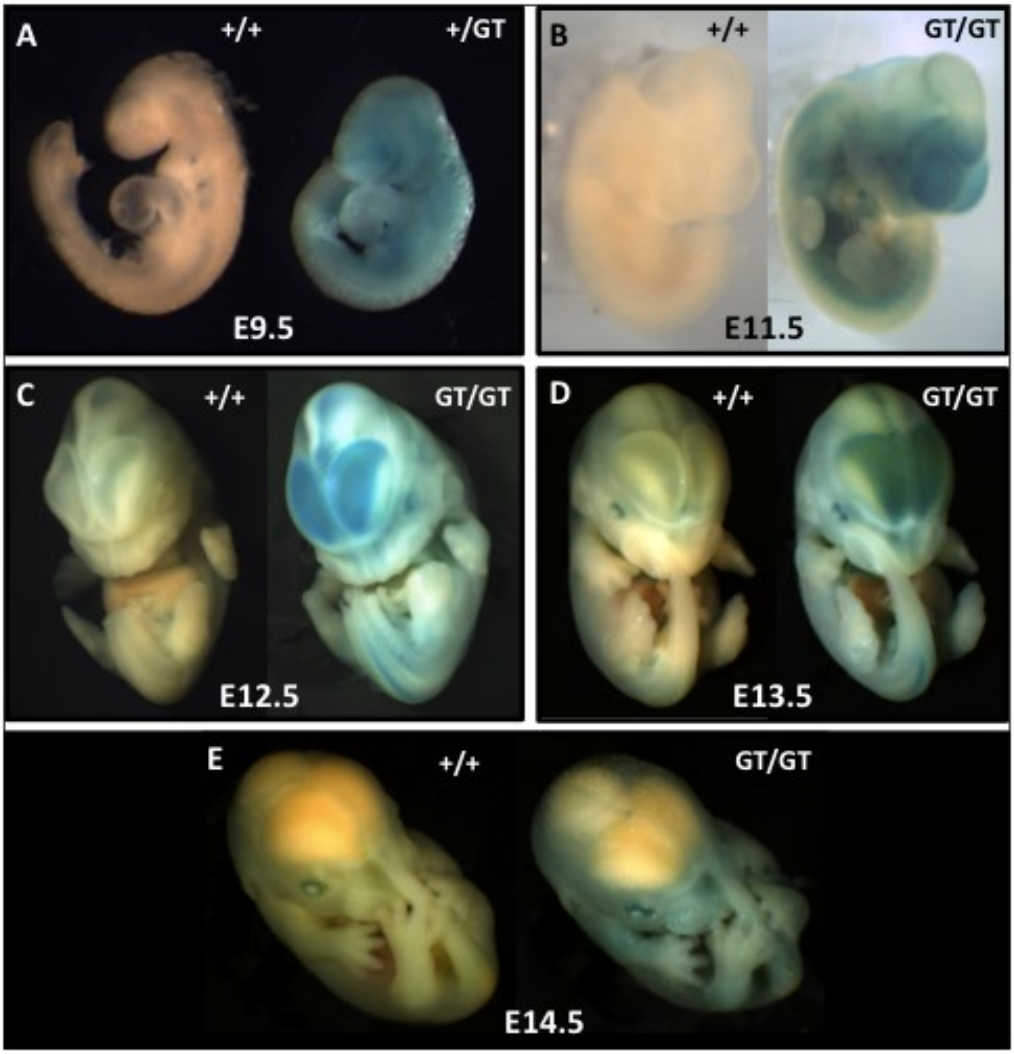


Figure 3.8.2: *Arhgap19* expression in whole embryos. X-gal staining of an E9.5 embryo heterozygous for the *Arhgap19*$^{GT(YHD020)Byg}$ allele shows non-specific expression throughout the embryo compared to a wildtype embryo control (**A**). X-gal staining of an E11.5 *Arhgap19*$^{GT(YHD020)Byg}$ homozygous mutant embryo shows expression in the telencephalon and the caudal neural tube compared to a wildtype embryo control (**B**). X-gal staining of an E12.5 homozygous mutant embryo shows expression in the telencephalon, midbrain, eye, and caudal neural tube compared to a wildtype embryo control (**C**). X-gal staining of an E13.5 homozygous mutant embryo displays a similar expression pattern as the E12.5 embryo, and is compared to a wildtype control (**D**). X-gal staining of an E14.5 homozygous mutant embryo no longer shows staining in the brain, and is compared to a wildtype control (**E**). '+' denotes wildtype, 'GT' denotes the *Arhgap19*$^{GT(YHD020)Byg}$ allele.

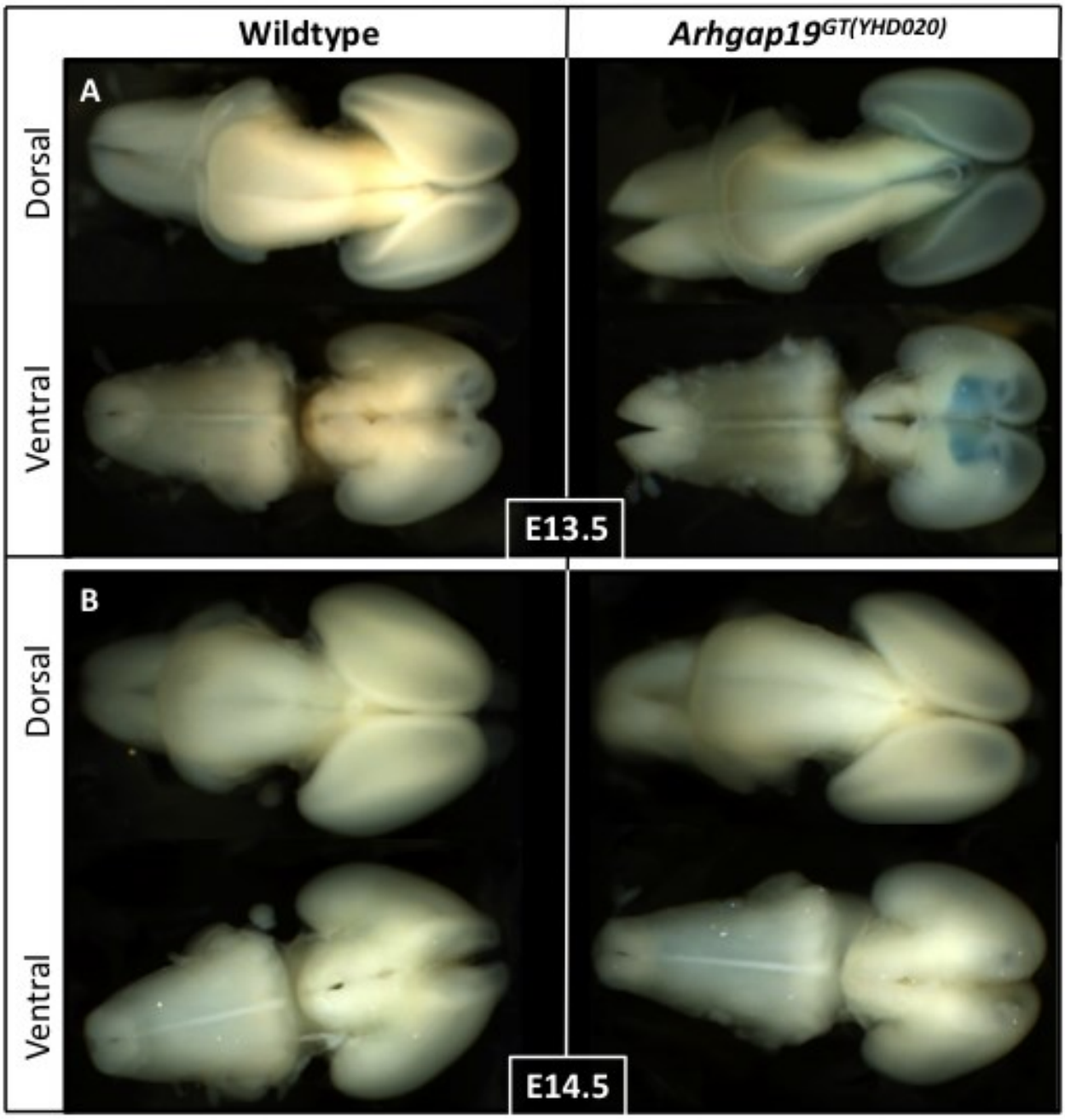


Figure 3.8.3: *Arhgap19* expression in E13.5 and E14.5 embryo brains. Dorsal view of a X-gal stained E13.5 *Arhgap19*$^{GT(YHD020)Byg}$ homozygous mutant embryo brain shows expression in the telencephalon and midbrain, and when viewed ventrally, more specific expression is seen medially and rostrally in the cortex. No expression is seen in the wildtype control (**A**). X-gal stained E14.5 homozygous mutant brain no longer shows expression within these regions, expression is similar in the E14.5 mutant and wildtype control brain (**B**).

The availability of two substrains of BALB/c, one of which naturally contained a mutation in *Arhgap19,* allowed me to genetically test whether *Arhgap19* is a modifier of *Cecr2*-associated exencephaly. One substrain originally comes from Charles River Laboratories (BALB/cCrl) and one from Jackson Laboratories (BALB/cJ). These substrains originated from a common strain but have been bred separately for approximately 70 years. The *Arhgap19*^{Ex6non} allele is unique to BALB/cCrl, as it was shown by our lab to not be present in BALB/cJ, FVB/N, C57BL/6, or 129S2 [165]. To test if the *Arhgap19*^{Ex6non} allele confers susceptibility to the development of exencephaly, congenic *Cecr2*^{GT45bic} BALB/cCrl mice were crossed to BALB/cJ mice in order to introduce the wildtype *Arhgap19* allele. If *Arhgap19* is a modifier, the presence of the wildtype copy should decrease exencephaly penetrance. If this experiment were performed using two different strains, then we would not be testing the variant in *Arhgap19* specifically, but rather all variants in all genes present in the donor strain linkage disequilibrium region around *Arhgap19*. Even after reaching congenic status, a linkage disequilbrium region can be a considerable size and contain many neighbouring genes from the donor line. My experiment relies on the assumption that the donor BALB/cJ strain will differ from the recipient BALB/cCrl strain by only the *Arhgap19* mutation. Although not genetically identical, the BALB/cCrl and BALB/cJ substrains should be very similar as they are both of BALB/c origin. The BALB/cJ genome has been published; however, the BALB/cCrl genome has not. Therefore, as part of the sequence alignment performed at Duke University, Deidre Krupp aligned the BALB/cCrl exome we generated to the BALB/cJ genome to compare SNVs. There were 78 SNVs that differed between the substrains within the chromosome 19 modifier region (Table 3.8.1). Only 4 of these SNVs were exonic, and only one of these was nonsynonymous. An additional 2 SNVs were within untranslated sequence (3'UTR), and the remaining SNVs were intronic or intergenic. This number of SNVs is much lower than the 2459 SNVs that differed between BALB/cCrl and FVB/N, of which 91 affected the amino acid sequence (Table 3.7.1). Interestingly, 55 of the 78 SNVs identified were heterozygous in one of the strains. This means that the two substrains still share at least one allele at each of these loci. As these substrains are interbred, all loci should theoretically be homozygous. Therefore, it is also possible that these variants were sequencing artifacts; however, Sanger validation would be required in order to determine this. Unfortunately, this analysis was not performed for INDELs. It is unclear as to whether or not any of these variants could contribute to NTD susceptibility differences between strains.

Table 3.8.1: SNVs that differ between BALB/cCrl and BALB/cJ within the chromosome 19 modifier region. BALB/cCrl variants were identified by WES (section 3.7) and compared to the published BALB/cJ genome (http://www.sanger.ac.uk/resources/mouse/genomes/). Red text denotes heterozygous genotypes, blue text highlights genotypes that are homozygous in both BALB/c substrains.

Position (mm9)	dbSNP ID	BALBc/Crl	BALB/cJ	Gene	CLASS_REF
19:40076322	rs46059991	G/A	A/A	*Cyp2c37*	exonic - nonsynonymous
19:41904672	.	A/A	G/G	*Frat1*	exonic - synonymous
19:41904702	.	A/A	G/G	*Frat1*	exonic - synonymous
19:40076360	rs48095172	A/T	T/T	*Cyp2c37*	exonic-synonymous
19:40086670	rs259668402	G/A	A/A	*Cyp2c37*	UTR3
19:40086688	rs223887888	A/G	G/G	*Cyp2c37*	UTR3
19:24011198	.	T/T	C/C	*Apba1*	intronic
19:24059422	rs50905186	C/C	C/T	*Fam189a2*	intronic
19:24060867	rs107808350	T/T	G/T	*Fam189a2*	intronic
19:24188398	rs47697646	T/C	C/C	*Tjp2*	intronic
19:24188404	rs108221931	G/T	T/T	*Tjp2*	intronic
19:24209502	.	C/C	A/A	*Tjp2*	intronic
19:24420597	rs38445782	G/G	A/A	*Pip5k1b*	intronic
19:24420598	.	G/G	T/T	*Pip5k1b*	intronic
19:25125806	.	A/A	G/G	*Dock8*	intronic
19:25125810	.	T/T	G/T	*Dock8*	intronic
19:25125816	.	C/C	G/G	*Dock8*	intronic
19:25125820	.	A/A	C/C	*Dock8*	intronic
19:25160653	rs50219959	C/C	T/T	*Dock8*	intronic
19:25258148	rs30349254	T/C	C/C	*Dock8*	intronic
19:25421037	rs30854984	G/G	A/A	*Kank1*	intronic
19:26721335	.	T/T	A/T	*Smarca2*	intronic
19:32127618	rs52549068	T/T	A/T	*Asah2*	intronic
19:34310671	.	C/C	A/C	*Stambpl1*	intronic
19:34389577	rs256659143	T/C	C/C	*Fas*	intronic
19:35003690	rs49732642	T/T	A/A	*Kif20b*	intronic
19:37070946	.	C/C	A/C	*Btaf1*	intronic
19:37070947	.	C/C	A/C	*Btaf1*	intronic
19:37070948	.	C/C	A/C	*Btaf1*	intronic
19:37076780	.	A/G	G/G	*Btaf1*	intronic
19:38215815	rs30818033	C/C	T/C	*Pde6c*	intronic
19:38288803	rs231278137	A/A	C/A	*Fra10ac1*	intronic
19:39399349	rs265346774	A/A	A/T	*Cyp2c29*	intronic
19:39861236	rs31102250	C/A	A/A	*Cyp2c40*	intronic
19:39955923	rs45706161	A/A	C/A	*Cyp2c69*	intronic
19:40076294	rs46104641	C/T	T/T	*Cyp2c37*	intronic
19:40086170	rs47225207	G/C	C/C	*Cyp2c37*	intronic

Position (mm9)	dbSNP ID	BALBc/Crl	BALB/cJ	Gene	CLASS_REF
19:40437771	rs221101595	A/G	G/G	*Sorbs1*	intronic
19:41133301	.	C/C	T/T	*Dntt*	intronic
19:41848810	.	A/A	G/G	*Arhgap19*	intronic
19:41859088	rs47375704	A/A	G/G	*Arhgap19*	intronic
19:41859193	.	A/A	G/G	*Arhgap19*	intronic
19:42676491	rs50621525	T/A	A/A	*Loxl4*	intronic
19:42676498	.	A/T	T/T	*Loxl4*	intronic
19:4328812	rs47168238	C/C	A/C	*Kdm2a*	intronic
19:43688656	rs221539566	G/A	G/G	*Nkx2-3*	intronic
19:43688657	rs249720970	A/T	A/A	*Nkx2-3*	intronic
19:43688658	rs264283337	T/G	T/T	*Nkx2-3*	intronic
19:44374345	.	A/A	C/A	*Scd2*	intronic
19:44375529	.	T/T	A/A	*Scd2*	intronic
19:44583056	.	A/A	T/T	*Wnt8b*	intronic
19:44583057	.	G/G	A/A	*Wnt8b*	intronic
19:44583060	.	C/C	T/T	*Wnt8b*	intronic
19:46238220	.	A/A	A/G	*Gbf1*	intronic
19:46462162	rs107814177	A/G	A/A	*Actr1a*	intronic
19:46626195	.	A/A	A/G	*Arl3*	intronic
19:47603285	rs36819904	T/T	C/T	*Obfc1*	intronic
19:47734594	.	T/T	T/A	*Col17a1*	intronic
19:47806478	rs223221951	A/A	G/A	*Sfr1*	intronic
19:48024047	rs239879173	C/C	T/C	*Ccdc147*	intronic
19:4802521	.	A/A	A/G	*Rbm14*	intronic
19:53693387	rs46734564	G/G	A/G	*Smc3*	intronic
19:5382557	rs217161253	G/A	A/A	*Sart1*	intronic
19:53842177	.	T/T	T/C	*Rbm20*	intronic
19:53842178	.	C/C	C/T	*Rbm20*	intronic
19:53907081	rs30452877	T/T	G/T	*Rbm20*	intronic
19:55698187	rs52121707	A/A	G/A	*Vti1a*	intronic
19:5686754	.	C/C	G/C	*Pcnxl3*	intronic
19:5687048	rs50337696	G/G	C/G	*Pcnxl3*	intronic
19:57113837	rs225747905	T/T	T/C	*Ablim1*	intronic
19:57154200	.	G/G	A/G	*Ablim1*	intronic
19:58865175	.	A/A	A/G	*1700019N19Rik*	intronic
19:59347548	rs30801249	A/G	A/A	*Slc18a2*	intronic
19:60887738	rs241280490	A/A	G/A	*Fam45a*	intronic
19:24140454	.	T/T	A/A	*Fam189a2,Tjp2*	intergenic
19:40287843	.	G/G	A/G	*Cyp2c70,Pdlim1*	intergenic

Using a dihybrid crossing scheme, we crossed BALB/cJ mice, which are homozygous wildtype for *Cecr2* and *Arhgap19*, to congenic BALB/cCrl mice heterozygous for the *Cecr2*^{GT45bic} allele and homozygous for the *Arhgap19*^{Ex6non} allele. The resulting progeny were

50% BALB/cJ and 50% BALB/cCrl. We selected for progeny that were heterozygous for both the *Cecr2*^{GT45bic} and *Arhgap19*^{Ex6non} alleles and crossed them to each other in order to generate embryos homozygous mutant for the *Cecr2*^{GT45bic} allele that were either homozygous wildtype, heterozygous, or homozygous mutant for the *Arhgap19*^{Ex6non} allele (see Chapter 2.22 for crossing scheme). The penetrance of exencephaly was compared between the embryos containing these three genotypes. If homozygosity for the *Arhgap19*^{Ex6non} allele predisposes homozygous mutant *Cecr2*^{GT45bic} mice to developing exencephaly, as we predict, then we would expect to see an exencephaly penetrance that is reduced in *Arhgap19*$^{+/+}$*;Cecr2*$^{GT45bic/GT45bic}$ embryos, as well as in *Arhgap19*$^{Ex6non/+}$*;Cecr2*$^{GT45bic/GT45bic}$ embryos. We would also predict that exencephaly penetrance in *Arhgap19*$^{Ex6non/Ex6non}$*;Cecr2*$^{GT45bic/GT45bic}$ embryos would be similar to the 54.1% penetrance seen in congenic *Cecr2*^{GT45bic} BALB/cCrl embryos, as congenic *Cecr2*^{GT45bic} BALB/cCrl embryos are also homozygous for the *Arhgap19*^{Ex6non} allele. If homozygosity for the *Arhgap19*^{Ex6non} allele does not predispose to the development of exencephaly, then we would expect to see a similar exencephaly penetrance in all three genotypic classes. Leanne Donnelly, an undergraduate Biology 499 student, provided assistance with setting up matings, dissections, phenotyping, genotyping, and preliminary analyses of the dihybrid cross. The experiment involved setting up 137 dihybrid crosses and generating 845 offspring, of which only 19.3 % of offspring were homozygous for *Cecr2*^{GT45bic} and therefore used in analysis. Results indicated an exencephaly penetrance not significantly different from the 54.1% penetrance seen in congenic *Cecr2*^{GT45bic} BALB/cCrl, with embryos heterozygous or homozygous wildtype for *Arhgap19* displaying an exencephaly penetrance of 50.0% (40/80) and 53.3% (24/45) respectively (Table 3.8.2). Exencephaly penetrance did not significantly differ between any of the three genotypic classes. As expected, the exencephaly penetrance of 42.1% (16/38) seen in embryos homozygous mutant for both *Cecr2*^{GT45bic} and *Arhgap19*^{Ex6non} was not significantly different from congenic *Cecr2*^{GT45bic} BALB/cCrl homozygous mutant embryos (Table 3.8.2), which was also homozygous mutant for the same *Cecr2* and *Arhgap19* alleles. These results indicated that *Arhgap19* is not a modifier of *Cecr2*-associated exencephaly, since the presence of a wildtype *Arhgap19* allele was unable to reduce the penetrance of exencephaly on a mixed BALB/cCrl:BALB/cJ genetic background.

Table 3.8.2: Summary of exencephaly penetrance in BALB/c (50% Charles River substrain, 50% Jackson substrain) *Cecr2*^{GT45bic} homozygous mutant embryos that are homozygous mutant, heterozygous, or homozygous wildtype for the *Arhgap19*^{Ex6non} nonsense mutation and compared to exencephaly penetrance of congenic *Cecr2*^{GT45bic} BALB/cCrl mutant embryos. Statistical comparisons were performed using a χ^2 test-of-independence. + wildtype *Arhgap19* allele, - *Arhgap19*^{Ex6non} allele.

Genotype	Penetrance	Percent	Comparison	*P*-value
Arhgap19 -/-	16/38	42.1%	BALB/cCrl: *Arhgap19* -/-	0.192
Arhgap19 +/-	40/80	50.0%	BALB/cCrl: *Arhgap19* +/-	0.563
Arhgap19 +/+	24/45	53.3%	BALB/cCrl: *Arghap19* +/+	0.931
Arhgap19 total	80/163	49.1%	BALB/cCrl: *Arhgap19* total	0.391
BALB/cCrl	73/135	54.1%	*Arhgap19* -/-: *Arhgap19* +/-	0.422
			Arhgap19 -/-: *Arhgap19* +/+	0.308
			Arhgap19 +/-: *Arhgap19* +/+	0.720

3.9: Exonic sequencing of *CECR2* in 156 human cranial NTD samples identifies protein-coding variants of interest

As shown above and elsewhere, *Cecr2* is a known NTD gene in mouse. However, *CECR2* has not yet been included in human NTD association or sequencing studies. Therefore, we sequenced the coding exons and flanking sequence of *CECR2* in 156 unrelated human cranial NTD probands in order to identify DNA sequence variants that alter the amino acid sequence (missense, nonsense/stop-gain, stop-loss, or frameshift), and therefore potentially affect protein function. This human NTD cohort is unique in that all probands have some form of cranial NTD, the majority of which had anencephaly (Appendix B), and is relevant to studying *CECR2*, which results in exencephaly (the murine equivalent of anencephaly) when mutant in mouse. Human variant data for *CECR2* from normal populations were acquired from the esp6500 whole exome sequencing project and the 1KG 1000 genomes project for comparison purposes.

A total of 9 variants of interest were identified in *CECR2* and are listed in Table 3.9.1. A summary of the 17 probands that these 9 variants were identified in, along with the proband and parental genotypes, are provided in Table 3.9.2. All variants of interest were also identified in at least one parent. The presence of DNA variants in unaffected parents does not necessarily mean the variant is functionally benign. The proband may have additional susceptibility factors not present in the parent, which in combination resulted in the development of a cranial NTD in the

proband. An enrichment of any of the identified variants within the cranial NTD cohort compared to the minor allele frequencies (MAFs) in normal populations (available in esp6500 or 1KG human variation databases) would provide evidence for the variant contributing to the development of NTDs.

As rare variants are more likely to be deleterious, I focused analyses on variants with a minor allele frequency (MAF) < 0.03. Seven such rare variants were identified and confirmed with Sanger sequencing in 8 probands and available parental DNA samples (Table 3.9.2, Figure 3.9.1). Six of the 7 rare variants were shown to be evolutionarily conserved according to GERP scores, and 5 of the 7 rare alleles were predicted to be deleterious by either SIFT or Polyphen2. In the event that more than one variant was identified in the same proband, regardless of MAF, I confirmed both variants with Sanger sequencing in the proband and available parental DNA samples, leading to the inclusion of an additional 2 variants of interest (Table 3.9.2, Figure 3.9.1). Validation of more than one variant in a proband and in both parents was done in order to determine if the proband was a compound heterozygote, meaning that both *CECR2* alleles in the proband contain an alteration to the amino acid sequence. If both variants are deleterious, then it is more likely that a phenotype would manifest due to the complete absence of wildtype protein product. This analysis identified four confirmed and four possible compound heterozygotes. All four confirmed compound heterozygotes and two of the four possible compound heterozygotes contained the same two variants, one of which altered an arginine to a histidine (rs5747211, MAF = 0.11) and the other that altered a proline to a leucine (rs1296794, MAF = 0.19). Although both variants are common, they are both predicted to be deleterious by SIFT or Polyphen2. Another proband that is a possible compound heterozygote contained a variant that was novel, meaning that this variant has not yet been identified in another human sample. This novel variant resulted in a glutamate to lysine alteration (E32K) within the ISWI binding DDT functional domain (Figure 3.9.1). The altered nucleotide is highly conserved, with a GERP score of 6.08; however, SIFT and Polyphen2 predictions suggest that this alteration is only possibly deleterious. The proband containing this novel variant also contained a second variant that resulted in a proline to leucine alteration (P632L, rs1296794, MAF = 0.19), which is highly conserved (GERP = 4.64) and is predicted to be deleterious by Polyphen2. It is possible that this proband was a compound heterozygote; however, parental genotypes acquired from Sanger sequencing were not informative. Therefore, it is also possible that the identified variants were within the same

CECR2 allele in this proband. There were also an additional three probands that were homozygous for the proline to leucine variant, meaning that all CECR2 protein in these probands contained this alteration. However, the number of probands that were homozygous for this variant was below what would be expected based on the established MAF (observed frequency = 3/156 = 0.019, expected frequency = 0.19^2 = 0.0361). According to the 1KG human variant database, a frequency of ~0.03 of homozygotes for this variant is seen in normal European populations (www.ensembl.org). The only identified variants that demonstrated an enrichment in the cranial NTD cohort compared to the expected MAFs were the five rare variants rs201912432, rs199780601, rs199565531, rs142851999, and rs181553013, as well as the single novel variant, all of which had a frequency within the cranial NTD cohort of 0.00321. However, as only one proband contained each variant, it is not possible to determine if this enrichment has any significance; a much larger sample size would be required. As homozygous mutation in mouse *Cecr2* is known to cause exencephaly, the 9 protein-coding variants of interest identified in *CECR2* are strong candidate susceptibility factors for the development of anencephaly in these 17 probands.

Table 3.9.1: List of 9 variants of interest identified in *CECR2* in the human cranial NTD cohort.

Gene	Position hg19	dbSNP137 ID	Ref	Alt	AA change	MAF esp6500	SIFT	PP2 (HDIV)	GERP
CECR2	22:17956663	.	G	A	E32K	.	0.14	*0.792*	**6.08**
CECR2	22:17983978	rs201912432	G	A	R245Q	0.00012	0.63	**1**	**5.35**
CECR2	22:18028672	rs199780601	C	G	P1210R	0.000483	**0.02**	*0.868*	0.659
CECR2	22:18022111	rs199565531	A	G	Y738C	0.001941	0.07	**1**	**4.37**
CECR2	22:18029251	rs142851999	C	T	S1261L	0.001948	0.11	0.335	**4.73**
CECR2	22:18031793	rs181553013	C	T	P1430L	0.002269	**0.01**	**0.983**	**3.79**
CECR2	22:18021936	rs62623401	A	G	M680V	0.023169	0.57	0.005	0.44
CECR2	22:17990852	rs5747211	G	A	R271H	0.113241	**0.04**	**0.999**	1.5
CECR2	22:18021604	rs1296794	C	T	P632L	0.191566	0.08	**1**	**4.64**

Abbreviations – AA: amino acid. MAF: minor allele frequency. SIFT: sorting intolerant from tolerant. PP2: Polyphen2. GERP: genomic evolutionary rate profiling.

Bold text – Predicted to be damaging (SIFT <0.05, PP2 (HDIV) >0.956) or evolutionary conserved for GERP (>2).

Italicized text – Predicted to be possibly damaging (PP2 (HDIV) between 0.453 and 0.956).

Table 3.9.2: *CECR2* protein-coding sequence variants of interest. 9 variants were identified in 17 cranial NTD probands by next generation sequencing, and were Sanger-sequence validated in probands and available parental DNA samples. Probands containing the lowest frequency variants (MAF esp6500) are listed first. Proband phenotype information is provided in Appendix B.

		Annotation				Genotype (reference/alternate)			Predictions of deleteriousness			
	Proband ID	Position hg19	dbSNP137 ID	AA change	MAF esp6500	Proband	Maternal	Paternal	SIFT	PP2 (HDIV)	GERP	Compound heterozygote
1	20-0602035	17956663 18021604	Novel rs1296794	E32K P632L	NA 0.191566*	G/A C/T	G/G C/T	G/A T/T	0.14 0.08	*0.792* **1**	**6.08** **4.64**	Possibly
2	Z0156471	17983978	rs201912432	R245Q	0.00012	G/A	G/G	G/A	0.63	**1**	**5.35**	-
3	A1021263	18028672	rs199780601	P1210R	0.000483	C/G	C/C	C/G	**0.02**	*0.868*	0.659	-
4	A1024167	18022111	rs199565531	Y738C	0.001941	A/G	A/A	A/G	0.07	**1**	**4.37**	-
5	20-0504243	18029251	rs142851999	S1261L	0.001948	C/T	C/C	C/T	0.11	0.335	**4.73**	-
6	A1021271	18031793	rs181553013	P1430L	0.002269	C/T	C/C	C/T	**0.01**	**0.983**	**3.79**	-
7	20-0714522	18021936 18021604	rs62623401 rs1296794	M680V P632L	0.023169 0.191566*	A/G C/T	A/A C/C	A/G C/T	0.57 0.08	0.005 **1**	0.44 **4.64**	No
8	20-0800216	18021936 18021604	rs62623401 rs1296794	M680V P632L	0.023169 0.191566*	A/G C/T	A/G C/T	A/A C/T	0.57 0.08	0.005 **1**	0.44 **4.64**	Possibly
9	20-0212267	17990852 18021604	rs5747211 rs1296794	R271H P632L	0.113241* 0.191566*	G/A C/T	A/A C/T	G/A C/C	**0.04** 0.08	**0.999** **1**	1.5 **4.64**	Possibly
10	A1015381	17990852 18021604	rs5747211 rs1296794	R271H P632L	0.113241* 0.191566*	G/A C/T	G/G C/T	G/A C/C	**0.04** 0.08	**0.999** **1**	1.5 **4.64**	Yes
11	20-0421416	17990852 18021604	rs5747211 rs1296794	R271H P632L	0.113241* 0.191566*	G/A C/T	G/G C/T	G/A C/C	**0.04** 0.08	**0.999** **1**	1.5 **4.64**	Yes
12	20-0601809	17990852 18021604	rs5747211 rs1296794	R271H P632L	0.113241* 0.191566*	G/A C/T	G/A C/T	G/G C/T	**0.04** 0.08	**0.999** **1**	1.5 **4.64**	Possibly
13	20-0608812	17990852 18021604	rs5747211 rs1296794	R271H P632L	0.113241* 0.191566*	G/A C/T	G/A C/C	G/G C/T	**0.04** 0.08	**0.999** **1**	1.5 **4.64**	Yes
14	20-0711091	17990852 18021604	rs5747211 rs1296794	R271H P632L	0.113241* 0.191566*	G/A C/T	G/G T/T	G/A C/C	**0.04** 0.08	**0.999** **1**	1.5 **4.64**	Yes
15	20-0515262	18021604	rs1296794	P632L	0.191566*	T/T	C/T	C/T	0.08	**1**	**4.64**	-
16	20-0601092	18021604	rs1296794	P632L	0.191566*	T/T	T/T	T/T	0.08	**1**	**4.64**	-
17	A1024158	18021604	rs1296794	P632L	0.191566*	T/T	C/T	C/T	0.08	**1**	**4.64**	-

Abbreviations – AA: amino acid. MAF: minor allele frequency. SIFT: sorting intolerant from tolerant. PP2: Polyphen2. GERP: genomic evolutionary rate profiling.

Position hg19 is on chromosome 22.

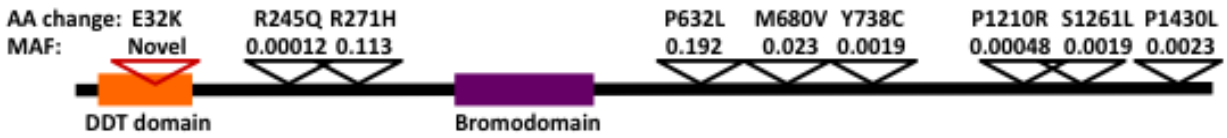
Asterisk – variants with MAF > 0.03 were only Sanger-sequence validated in probands that were homozygous for the alternate allele or were possible compound heterozygotes.

Green – homozygous reference, Yellow – heterozygous, Red – homozygous alternate.

Bold text – Predicted to be damaging (SIFT <0.05, PP2 (HDIV) >0.956) or evolutionary conserved for GERP (>2).

Italicized text – Predicted to be possibly damaging (PP2 (HDIV) between 0.453 and 0.956).

Figure 3.9.1: Location of all variants of interest within the CECR2 protein. Novel mutations (not in esp6500 or 1KG human variation databases) are marked with a red triangle. Note that the novel mutation is located within the functional DDT domain. AA change – amino acid change, MAF – minor allele frequency (esp6500).



3.10: Exonic sequencing of 24 *CECR2* candidate modifier genes in 156 human cranial NTD samples identifies protein-coding variants of interest, particularly in *DNMBP*, *MMS19*, and *TJP2*

We sequenced the coding exons of 24 candidate modifier genes identified in mouse (listed in Table 3.7.3, excluding *Arhgap19*) in 156 unrelated human cranial NTD probands in an effort to identify DNA sequence variants that alter the amino acid sequence (missense, nonsense/stop-gain, stop-loss, or frameshift), and therefore potentially affect protein function. The purpose of this experiment was to narrow down the candidate gene list by identifying variants of interest within a human cranial NTD cohort. Any of the 24 candidate genes that also possess variants affecting protein function in humans with cranial NTDs would be stronger candidate NTD genes (in both mice and humans) compared to sequenced genes that do not contain functional variants. I identified a total of 43 variants of interest (MAF < 0.03) in 17 of the 24 candidate modifier genes, all of which I verified by Sanger sequencing (Table 3.10.1). There were 10 mutations not listed in esp6500 or 1KG, and were therefore annotated as novel. None of the novel variants were *de novo*, as all novel variants were also present in one parent. This is a common occurrence when identifying variants that contribute to multifactorial disease. I included an additional 10 variants of interest with a MAF > 0.03 due to the possibility that they contribute to compound heterozygosity or homozygosity for the alternative allele. There were no obvious enrichments of common or rare variants in the human anencephaly cohort. However, with over 300 NTD genes identified in mouse and due to the multifactorial nature of neurulation defects in humans, the lack

of enrichment in this one subset of genes may be due to a small sample size of 156 unrelated probands, many of which likely have a unique etiology. Table 3.10.1 lists all of the 53 variants of interest identified in the 17 candidate modifiers. Table 3.10.2 is an additional summary displaying which variants were identified in each proband; therefore, a variant that was seen in more than one proband is displayed more than once in this summary. This summary also includes the 9 variants of interest identified in *CECR2*.

Three of the candidate modifier genes, *DNMBP*, *MMS19*, and *TJP2,* stood out as strong candidates based on the number of identified variants, the predicted deleteriousness of these variants, and what is currently known about gene function. Each of these three genes is addressed in greater detail below. Variants identified in these three candidates, along with *CECR2*, are closest to the top of Table 3.10.2 as these variants are of greatest interest. A summarized list of the 24 sequenced candidate modifier genes that include the number of variants identified in the human cranial NTD cohort in conjunction with what is known about gene expression and protein function is provided in Table 3.10.3.

Table 3.10.1: List of 53 variants of interest identified in candidate NTD modifier genes in the human cranial NTD cohort. The coding exons of 24 human homologues of candidate modifier genes of *Cecr2*-associated exencephaly in mice were sequenced in 156 human cranial NTD probands. Variants of interest were identified in 17 of the 24 candidate modifier genes.

Gene	Position hg19	dbSNP137 ID	Ref	Alt	AA change	MAF esp6500	SIFT	PP2 (HDIV)	GERP
DNMBP	10:101646214	Novel	C	T	R1154Q	.	0.59	**1**	**5.82**
DNMBP	10:101654735	Novel	G	A	R1042X	.	.	.	**3.85**
DNMBP	10:101639989	rs372003127	C	T	R1376H	0.000116	0.07	**1**	**5.42**
DNMBP	10:101667847	rs114927649	G	A	P820L	0.001744	0.62	0.015	**3.96**
DNMBP	10:101639731	rs147752816	G	A	T1462M	0.002209	0.09	**0.986**	**3.49**
DNMBP	10:101667814	rs17854134	A	G	M831T	0.03186	0.45	*0.673*	**5.99**
DNMBP	10:101639877	rs11190305	A	C	C1413W	0.380349	0.19	*0.6*	0.604
DNMBP	10:101716112	rs35924554	G	C	N373K	0.04	0.89	0.049	-1.33
MMS19	10:99237155	Novel	G	A	R187X	.	.	.	**5.48**
MMS19	10:99218991	rs200490757	G	A	P985S	0.000581	0.06	**1**	**5.9**
MMS19	10:99218456	rs36023427	C	T	G1029D	0.008721	0.16	**1**	**5.86**
MMS19	10:99225846	rs17112809	C	T	V526I	0.010465	0.31	0.155	1.33
MMS19	10:99225645	rs12360068	G	A	A558V	0.043023	0.37	**1**	**4.64**
MMS19	10:99220707	rs3740526	C	T	G790D	0.41389	0.62	0.103	**3.64**
TJP2	9:71850989	Novel	T	C	I609T	.	**0**	**1**	**5.43**

Gene	Position hg19	dbSNP137 ID	Ref	Alt	AA change	MAF esp6500	SIFT	PP2 (HDIV)	GERP
TJP2	9:71866149	rs199892018	G	T	V1064L	0.000349	0.94	0	-0.183
TJP2	9:71851040	rs149911553	C	G	T626S	0.001628	0.16	0.008	**3.55**
TJP2	9:71835842	rs41305539	C	A	Q128K	0.019316	1	0.001	**2.89**
TJP2	9:71851877	rs34774441	G	A	M668I	0.064884	**0.02**	*0.928*	**5.77**
TJP2	9:71863235	rs77236826	A	G	H992R	0.07511	0.12	0	-9.87
TJP2	9:71865988	rs41277907	C	T	S1010F	0.078256	**0.02**	0.001	**5.58**
CPEB3	10:93811989	rs140779166	C	T	V679I	0.000581	1	0.19	**5.65**
CSTF2T	10:53457701	rs143644186	T	C	S537G	0.00024*	0.74	.	0.131
CSTF2T	10:53457703	rs142002882	A	C	V536G	.	0.55	*0.824*	.
CSTF2T	10:53458828	Novel	A	C	M161R	.	**0**	**0.997**	**4.99**
EXOSC1	10:99200958	rs141001349	G	A	P94L	0.000116	0.67	*0.944*	**6.08**
FAM45A	10:120867671	rs149569390	A	C	K83Q	0.000465	0.24	0.098	**2.14**
FAS	10:90771767	rs56006128	G	A	E194K	0.003372	0.3	0.011	.
GLIS3	9:4118634	rs143051164	G	C	P282A	0.002442	1	*0.649*	**3.76**
GLIS3	9:4118585	rs148572278	G	T	S298Y	0.003256	**0.02**	**0.999**	**5.59**
GLIS3	9:4117933	rs72687988	C	G	E515D	0.005	0.31	0.01	**5.51**
GLIS3	9:4118540	rs35154632	C	G	G313A	0.009186	0.16	**1**	**5.59**
HPS6	10:103825929	rs36078476	T	G	L233R	0.009767	0.81	0	**2.39**
PNLIPRP2	10:118385551	rs200056143	C	T	A100V	0.000117	0.11	*0.525*	**4.97**
PNLIPRP2	10:118383557	rs62623669	G	C	D51H	0.017154	0.14	0.002	1.39
PNLIPRP2	10:118397894	rs4751996	A	G	I360V	0.479667	1	0.013	**2.61**
PNLIPRP2	10:118397884	rs4751995	A	G	X357W	0.481849	.	.	**5.48**
RNLS	10:90332762	Novel	A	C	W142G	.	**0.02**	**1**	**5.78**
RNLS	10:90342837	rs2296545	C	G	E37D	0.434302	0.07	0.425	-0.832
SCD	10:102116427	Novel	G	T	W262C	.	**0**	**1**	**5.43**
SCD	10:102107936	rs150416868	C	G	P48R	0.001047	0.16	0.002	**4.06**
SFXN2	10:104486433	Novel	C	T	R14C	.	**0**	**0.999**	**5.49**
SFXN2	10:104486843	Novel	T	A	N87K	.	0.3	**1**	-10.9
SFXN2	10:104486493	rs201068739	C	T	R34C	0.000116	**0.02**	0.075	**3.68**
SFXN2	10:104493338	rs151088981	C	T	Q267X	0.003023	.	.	**5.37**
TCTN3	10:97445357	rs55859130	C	A	A327S	0.022313	0.84	0.003	0.118
TCTN3	10:97446261	rs147928670	C	G	Q311H	0.000116	0.11	**0.993**	**4.46**
TMEM180	10:104233399	rs371861479	C	T	A333V	0.000116	0.17	*0.588*	**5.71**
TMEM180	10:104231098	rs149406506	G	A	R258Q	0.000814	0.25	0.435	**3.33**
KANK1	9:710972	rs61737969	C	T	P69L	0	0.67	0.01	**5.16**
KANK1	9:732509	Novel	T	C	M1046T	.	**0.01**	0.001	**3.55**
KANK1	9:744588	rs143775530	C	T	P251L	0.000581	**0**	**1**	**6.02**

Abbreviations – AA: amino acid. MAF: minor allele frequency. SIFT: sorting intolerant from tolerant. PP2: Polyphen2. GERP: genomic evolutionary rate profiling.

Asterisk – MAF from 1000 genomes European descent database.

Bold text – Predicted to be damaging (SIFT <0.05, PP2 (HDIV) >0.956) or evolutionary conserved for GERP (>2).

Italicized text – Predicted to be possibly damaging (PP2 (HDIV) between 0.453 and 0.956).

Table 3.10.2: Protein-coding sequence variants of interest in cranial NTD probands.
Variants identified by next generation sequencing in cranial NTD probands were Sanger-sequence validated in probands and available parental DNA samples. Specific phenotype information for each proband is provided in Appendix B.

		Annotation				Genotype (reference/alternate)			Predictions of deleteriousness		
Proband ID	Gene	Position hg19	dbSNP137 ID	AA change	MAF esp6500	Proband	Maternal	Paternal	SIFT	PP2 (HDIV)	GERP
A1024167	*CECR2*	22:18022111	rs199565531	Y738C	0.001941	A/G	A/A	A/G	0.07	**1**	**4.37**
	DNMBP	10:101639731	rs147752816	T1462M	0.002209	G/A	G/A	G/G	0.09	**0.986**	**3.49**
20-0212267	*CECR2*	22:17990852	rs5747211	R271H	0.113241*	G/A	A/A	G/A	**0.04**	**0.999**	1.5
		22:18021604	rs1296794	P632L	0.191566*	C/T	C/T	C/C	0.08	**1**	**4.64**
	DNMBP	10:101716112	rs35924554	N373K	0.04*	G/C	G/G	G/C	0.89	0.049	-1.33
		10:101639877	rs11190305	C1413W	0.380349*	C/C	FS	C/C	0.19	*0.6*	0.604
20-0504243	*CECR2*	22:18029251	rs142851999	S1261L	0.001948	C/T	C/C	C/T	0.11	0.335	**4.73**
	DNMBP	10:101716112	rs35924554	N373K	0.04*	G/C	C/C	G/G	0.89	0.049	-1.33
		10:101639877	rs11190305	C1413W	0.380349*	A/C	A/C	C/C	0.19	*0.6*	0.604
A1015381	*CECR2*	22:17990852	rs5747211	R271H	0.113241*	G/A	G/G	G/A	**0.04**	**0.999**	1.5
		22:18021604	rs1296794	P632L	0.191566*	C/T	C/T	C/C	0.08	**1**	**4.64**
	MMS19	10:99225846	rs17112809	V526I	0.010465	C/T	C/C	C/T	0.31	0.155	1.33
20-0421416	*CECR2*	22:17990852	rs5747211	R271H	0.113241*	G/A	G/G	G/A	**0.04**	**0.999**	1.5
		22:18021604	rs1296794	P632L	0.191566*	C/T	C/T	C/C	0.08	**1**	**4.64**
20-0601809	*CECR2*	22:17990852	rs5747211	R271H	0.113241*	G/A	G/A	G/G	**0.04**	**0.999**	1.5
		22:18021604	rs1296794	P632L	0.191566*	C/T	C/T	C/T	0.08	**1**	**4.64**
20-0608812	*CECR2*	22:17990852	rs5747211	R271H	0.113241*	G/A	G/A	G/G	**0.04**	**0.999**	1.5
		22:18021604	rs1296794	P632L	0.191566*	C/T	C/C	C/T	0.08	**1**	**4.64**
	GLIS3	9:4118540	rs35154632	G313A	0.009186	C/G	C/C	C/G	0.16	**1**	**5.59**
	SFXN2	10:104486843	Novel	N87K	-	T/A	T/A	T/T	0.3	**1**	-10.9
20-0711091	*CECR2*	22:17990852	rs5747211	R271H	0.113241*	G/A	G/G	G/A	**0.04**	**0.999**	1.5
		22:18021604	rs1296794	P632L	0.191566*	C/T	T/T	C/C	0.08	**1**	**4.64**
Z0156471	*CECR2*	22:17983978	rs201912432	R245Q	0.00012	G/A	G/G	G/A	0.63	**1**	**5.35**
A1021271	*CECR2*	22:18031793	rs181553013	P1430L	0.002269	C/T	C/C	C/T	**0.01**	**0.983**	**3.79**
A1021263	*CECR2*	22:18028672	rs199780601	P1210R	0.000483	C/G	C/C	C/G	**0.02**	*0.868*	0.659
20-0714522	*CECR2*	22:18021936	rs62623401	M680V	0.023169	A/G	A/A	A/G	0.57	0.005	0.44
		22:18021604	rs1296794	P632L	0.191566*	C/T	C/C	C/T	0.08	**1**	**4.64**
20-0800216	*CECR2*	22:18021936	rs62623401	M680V	0.023169	A/G	A/G	A/A	0.57	0.005	0.44
		22:18021604	rs1296794	P632L	0.191566*	C/T	C/T	C/T	0.08	**1**	**4.64**
20-0515262	*CECR2*	22:18021604	rs1296794	P632L	0.191566*	T/T	C/T	C/T	0.08	**1**	**4.64**
20-0601092	*CECR2*	22:18021604	rs1296794	P632L	0.191566*	T/T	T/T	T/T	0.08	**1**	**4.64**
	PNLIPRP2	10:118383557	rs62623669	D51H	0.017154	G/C	G/C	G/G	0.14	0.002	1.39
		10:118397894	rs4751996	I360V	0.479667*	G/G	A/G	G/G	1	0.013	**2.61**
		10:118397884	rs4751995	X357W	0.481849*	G/G	A/G	G/G	-	-	**5.48**
A1024158	*CECR2*	22:18021604	rs1296794	P632L	0.191566*	T/T	C/T	C/T	0.08	**1**	**4.64**
20-0602035	*CECR2*	22:17956663	Novel	E32K	-	G/A	G/G	G/A	0.14	*0.792*	**6.08**
		22:18021604	rs1296794	P632L	0.191566*	C/T	C/T	T/T	0.08	**1**	**4.64**
20-0620476	*DNMBP*	10:101716112	rs35924554	N373K	0.04*	G/C	G/G	G/C	0.89	0.049	-1.33
		10:101639877	rs11190305	C1413W	0.380349*	A/C	A/A	C/C	0.19	0.6	0.604
	MMS19	10:99225846	rs17112809	V526I	0.010465	C/T	C/T	C/C	0.31	0.155	1.33
	TJP2	9:71865988	rs41277907	S1010F	0.078256*	T/T	C/T	C/T	**0.02**	0.001	**5.58**
	FAS	10:90771767	rs56006128	E194K	0.003372	G/A	G/G	G/A	0.3	0.011	-
20-0521275	*DNMBP*	10:101667814	rs17854134	M831T	0.03186*	A/G	A/A	A/G	0.45	*0.673*	**5.99**
	MMS19	10:99237155	Novel	R187X	-	G/A	G/G	G/A	-	-	**5.48**
		10:99225645	rs12360068	A558V	0.043023*	G/A	G/A	G/G	0.37	**1**	**4.64**
	TJP2	9:71835842	rs41305539	Q128K	0.019316	C/A	C/C	C/A	1	0.001	**2.89**
		9:71851877	rs34774441	M668I	0.064884*	G/A	G/G	G/A	**0.02**	*0.928*	**5.77**
20-0421455	*DNMBP*	10:101667814	rs17854134	M831T	0.03186*	A/G	A/G	A/A	0.45	*0.673*	**5.99**
	TCTN3	10:97445357	rs55859130	A327S	0.022313	C/A	C/A	C/C	0.84	0.003	0.118
20-0504690	*DNMBP*	10:101667814	rs17854134	M831T	0.03186*	A/G	A/G	A/A	0.45	*0.673*	**5.99**
20-0801162	*DNMBP*	10:101667814	rs17854134	M831T	0.03186*	A/G	A/G	A/A	0.45	*0.673*	**5.99**
A1021253	*DNMBP*	10:101667814	rs17854134	M831T	0.03186*	A/G	A/G	A/A	0.45	*0.673*	**5.99**
	SFXN2	10:104486433	Novel	R14C	-	C/T	C/C	C/T	0	**0.999**	**5.49**
A1024197	*DNMBP*	10:101667814	rs17854134	M831T	0.03186*	A/G	A/A	A/G	0.45	*0.673*	**5.99**
A1024256	*DNMBP*	10:101667847	rs114927649	P820L	0.001744	G/A	G/G	G/A	0.62	0.015	**3.96**
20-0426714	*DNMBP*	10:101667814	rs17854134	M831T	0.03186*	A/G	A/G	G/G	0.45	*0.673*	**5.99**
		10:101716112	rs35924554	N373K	0.04*	G/C	G/G	C/C	0.89	0.049	-1.33
		10:101639877	rs11190305	C1413W	0.380349*	A/A	A/C	C/C	0.19	*0.6*	0.604

		Annotation				Genotype (reference/alternate)			Predictions of deleteriousness		
Proband ID	Gene	Position hg19	dbSNP137 ID	AA change	MAF esp6500	Proband	Maternal	Paternal	SIFT	PP2 (HDIV)	GERP
20-0613589	*DNMBP*	10:101667814	rs17854134	M831T	0.03186*	A/G	A/G	A/A	0.45	*0.673*	**5.99**
		10:101716112	rs35924554	N373K	0.04*	G/C	G/C	G/C	0.89	0.049	-1.33
		10:101639877	rs11190305	C1413W	0.380349*	A/C	A/C	A/A	0.19	*0.6*	0.604
	GLIS3	9:4117933	rs72687988	E515D	0.005	C/G	C/C	C/G	0.31	0.01	**5.51**
20-0410539	*DNMBP*	10:101716112	rs35924554	N373K	0.04*	G/C	G/C	G/C	0.89	0.049	-1.33
		10:101639877	rs11190305	C1413W	0.380349*	A/C	A/C	A/A	0.19	*0.6*	0.604
20-0512275	*DNMBP*	10:101716112	rs35924554	N373K	0.04*	G/C	G/C	NA	0.89	0.049	-1.33
		10:101639877	rs11190305	C1413W	0.380349*	A/C	A/C	NA	0.19	*0.6*	0.604
	EXOSC1	10:99200958	rs141001349	P94L	0.000116	G/A	G/G	NA	0.67	*0.944*	**6.08**
20-0514451	*DNMBP*	10:101716112	rs35924554	N373K	0.04*	G/C	G/G	G/C	0.89	0.049	-1.33
		10:101639877	rs11190305	C1413W	0.380349*	A/C	A/A	C/C	0.19	*0.6*	0.604
20-0601184	*DNMBP*	10:101639989	rs372003127	R1376H	0.000116	C/T	C/C	FS	0.07	**1**	**5.42**
	SCD	10:102116427	Novel	W262C	-	G/T	G/G	T/T	**0**	**1**	**5.43**
20-0806627	*DNMBP*	10:101646214	Novel	R1154Q	-	C/T	C/C	C/T	0.59	**1**	**5.82**
20-027140	*DNMBP*	10:101654735	Novel	R1042X	-	G/A	G/A	G/G	-	-	**3.85**
20-0302223	*MMS19*	10:99225846	rs17112809	V526I	0.010465	C/T	C/C	C/T	0.31	0.155	1.33
	TJP2	9:71835842	rs41305539	Q128K	0.019316	C/A	C/C	C/A	1	0.001	**2.89**
A1024176	*MMS19*	10:99225846	rs17112809	V526I	0.010465	C/T	C/T	C/C	0.31	0.155	1.33
	TJP2	9:71865988	rs41277907	S1010F	0.078256*	T/T	C/T	C/T	**0.02**	0.001	**5.58**
	GLIS3	9:4117933	rs72687988	E515D	0.005	C/G	C/C	C/G	0.31	0.01	**5.51**
20-0700015	*MMS19*	10:99218991	rs200490757	P985S	0.000581	G/A	G/G	G/A	0.06	**1**	**5.9**
		10:99220707	rs3740526	G790D	0.41389*	C/T	C/T	T/T	0.62	0.103	**3.64**
	GLIS3	9:4118585	rs148572278	S298Y	0.003256	G/T	G/T	G/G	**0.02**	**0.999**	**5.59**
	PNLIPRP2	10:118383557	rs62623669	D51H	0.017154	C/C	C/C	G/C	0.14	0.002	1.39
		10:118397894	rs4751996	I360V	0.479667*	G/G	G/G	A/G	1	0.013	**2.61**
		10:118397884	rs4751995	X357W	0.481849*	G/G	G/G	A/G	-	-	**5.48**
20-0300929	*MMS19*	10:99225846	rs17112809	V526I	0.010465	C/T	C/T	C/C	0.31	0.155	1.33
	SFXN2	10:104493338	rs151088981	Q267X	0.003023	C/T	C/T	C/C	-	-	**5.37**
	TMEM180	10:104231098	rs149406506	R258Q	0.000814	G/A	G/A	G/G	0.25	0.435	**3.33**
20-0604224	*MMS19*	10:99225846	rs17112809	V526I	0.010465	C/T	C/T	C/C	0.31	0.155	1.33
	GLIS3	9:4118634	rs143051164	P282A	0.002442	G/C	G/G	G/C	1	*0.649*	**3.76**
A1030655	*MMS19*	10:99225846	rs17112809	V526I	0.010465	C/T	C/C	NA	0.31	0.155	1.33
Z0167301	*MMS19*	10:99225846	rs17112809	V526I	0.010465	C/T	C/T	C/T	0.31	0.155	1.33
20-0427674	*MMS19*	10:99218456	rs36023427	G1029D	0.008721	C/T	FS	C/C	0.16	**1**	**5.86**
		10:99220707	rs3740526	G790D	0.41389*	T/T	FS	T/T	0.62	0.103	**3.64**
20-0708160	*MMS19*	10:99218456	rs36023427	G1029D	0.008721	C/T	C/C	C/T	0.16	**1**	**5.86**
		10:99220707	rs3740526	G790D	0.41389*	T/T	T/T	C/T	0.62	0.103	**3.64**
20-0705228	*TJP2*	9:71851040	rs149911553	T626S	0.001628	C/G	C/C	C/G	0.16	0.008	**3.55**
	PNLIPRP2	10:118385551	rs200056143	A100V	0.000117	C/T	C/T	C/C	0.11	*0.525*	**4.97**
Z0082501	*TJP2*	9:71850989	Novel	I609T	-	T/C	T/C	T/T	**0**	**1**	**5.43**
A1024159	*TJP2*	9:71851877	rs34774441	M668I	0.064884*	G/A	NA	NA	**0.02**	*0.928*	**5.77**
		9:71863235	rs77236826	H992R	0.07511*	A/G	NA	NA	0.12	0	-9.87
A1021228	*TJP2*	9:71851877	rs34774441	M668I	0.064884*	G/A	NA	NA	**0.02**	*0.928*	**5.77**
		9:71865988	rs41277907	S1010F	0.078256*	C/T	NA	NA	**0.02**	0.001	**5.58**
A1024188	*TJP2*	9:71835842	rs41305539	Q128K	0.019316	C/A	C/C	NA	1	0.001	**2.89**
		9:71865988	rs41277907	S1010F	0.078256*	C/T	C/T	NA	**0.02**	0.001	**5.58**
20-0518590	*TJP2*	9:71835842	rs41305539	Q128K	0.019316	C/A	C/C	C/A	1	0.001	**2.89**
20-0609211	*TJP2*	9:71835842	rs41305539	Q128K	0.019316	C/A	C/C	C/A	1	0.001	**2.89**
20-0613565	*TJP2*	9:71835842	rs41305539	Q128K	0.019316	C/A	C/C	C/A	1	0.001	**2.89**
A1034284	*TJP2*	9:71835842	rs41305539	Q128K	0.019316	C/A	C/A	C/C	1	0.001	**2.89**
20-0426206	*TJP2*	9:71863235	rs77236826	H992R	0.07511*	A/G	A/G	NA	0.12	0	-9.87
		9:71865988	rs41277907	S1010F	0.078256*	C/T	C/C	NA	**0.02**	0.001	**5.58**
	HPS6	10:103825929	rs36078476	L233R	0.009767	T/G	T/G	NA	0.81	0	**2.39**
20-0517865	*TJP2*	9:71863235	rs77236826	H992R	0.07511*	G/G	A/G	G/G	0.12	0	-9.87
A1030628	*TJP2*	9:71866149	rs199892018	V1064L	0.000349	G/T	G/T	G/G	0.94	0	-0.183
		9:71851877	rs34774441	M668I	0.064884*	G/A	G/A	G/G	**0.02**	*0.928*	**5.77**
	CPEB3	10:93811989	rs140779166	V679I	0.000581	C/T	C/T	C/C	1	0.19	**5.65**

		Annotation				Genotype (reference/alternate)			Predictions of deleteriousness		
Proband ID	Gene	Position hg19	dbSNP137 ID	AA change	MAF esp6500	Proband	Maternal	Paternal	SIFT	PP2 (HDIV)	GERP
20-0422758	*CSTF2T*	10:53458828	Novel	M161R	-	A/C	A/C	A/A	**0**	**0.997**	**4.99**
	PNLIPRP2	10:118383557	rs62623669	D51H	0.017154	G/C	G/G	G/C	0.14	0.002	1.39
		10:118397894	rs4751996	I360V	0.479667*	G/G	A/G	A/G	1	0.013	**2.61**
		10:118397884	rs4751995	X357W	0.481849*	G/G	A/G	A/G	-	-	**5.48**
A1021274	*CSTF2T*	10:53457701	rs143644186	S537G	0.00024	T/C	NA	NA	0.74	-	0.131
		10:53457703	rs142002882	V536G	-	A/C	NA	NA	0.55	*0.824*	-
	KANK1	9:744588	rs143775530	P251L	0.000581	C/T	NA	NA	**0**	**1**	**6.02**
A1034295	*FAM45A*	10:120867671	rs149569390	K83Q	0.000465	A/C	NA	NA	0.24	0.098	**2.14**
	TCTN3	10:97445357	rs55859130	A327S	0.022313	C/A	NA	NA	0.84	0.003	0.118
Z0082809	*GLIS3*	9:4118585	rs148572278	S298Y	0.003256	G/T	G/G	G/T	**0.02**	**0.999**	**5.59**
20-0608051	*GLIS3*	9:4118540	rs35154632	G313A	0.009186	C/G	C/C	C/G	0.16	**1**	**5.59**
20-0424645	*GLIS3*	9:4118540	rs35154632	G313A	0.009186	C/G	C/C	NA	0.16	**1**	**5.59**
20-0802143	*HPS6*	10:103825929	rs36078476	L233R	0.009767	T/G	T/T	T/G	0.81	0	**2.39**
20-0602679	*PNLIPRP2*	10:118383557	rs62623669	D51H	0.017154	G/C	G/G	G/C	0.14	0.002	1.39
		10:118397894	rs4751996	I360V	0.479667*	A/G	A/G	A/G	1	0.013	**2.61**
		10:118397884	rs4751995	X357W	0.481849*	A/G	A/G	A/G	-	-	**5.48**
A1030639	*PNLIPRP2*	10:118383557	rs62623669	D51H	0.017154	G/C	NA	NA	0.14	0.002	1.39
		10:118397894	rs4751996	I360V	0.479667*	G/G	NA	NA	1	0.013	**2.61**
		10:118397884	rs4751995	X357W	0.481849*	G/G	NA	NA	-	-	**5.48**
Z0174261	*PNLIPRP2*	10:118383557	rs62623669	D51H	0.017154	G/C	NA	NA	0.14	0.002	1.39
		10:118397894	rs4751996	I360V	0.479667*	G/G	NA	NA	1	0.013	**2.61**
		10:118397884	rs4751995	X357W	0.481849*	G/G	NA	NA	-	-	**5.48**
20-0404771	*PNLIPRP2*	10:118383557	rs62623669	D51H	0.017154	G/C	G/C	G/G	0.14	0.002	1.39
		10:118397894	rs4751996	I360V	0.479667*	A/G	A/G	A/G	1	0.013	**2.61**
		10:118397884	rs4751995	X357W	0.481849*	A/G	A/G	A/G	-	-	**5.48**
A1030634	*RNLS*	10:90332762	Novel	W142G	-	A/C	NA	NA	**0.02**	1	**5.78**
		10:90342837	rs2296545	E37D	0.434302*	C/G	NA	NA	0.07	0.425	-0.832
20-0800185	*SCD*	10:102107936	rs150416868	P48R	0.001047	C/G	C/G	C/C	0.16	0.002	**4.06**
A1021240	*SFXN2*	10:104486493	rs201068739	R34C	0.000116	C/T	NA	NA	**0.02**	0.075	**3.68**
Z0082701	*TCTN3*	10:97446261	rs147928670	Q311H	0.000116	C/G	NA	NA	0.11	**0.993**	**4.46**
20-0300594	*TCTN3*	10:97445357	rs55859130	A327S	0.022313	C/A	C/C	C/C	0.84	0.003	0.118
20-014072	*TCTN3*	10:97445357	rs55859130	A327S	0.022313	C/A	C/A	C/C	0.84	0.003	0.118
A1015387	*TMEM180*	10:104233399	rs371861479	A333V	0.000116	C/T	NA	NA	0.17	*0.588*	**5.71**
20-028661	*KANK1*	9:732509	Novel	M1046T	-	T/C	NA	NA	**0.01**	0.001	**3.55**
Z0156466	*KANK1*	9:710972	rs61737969	P69L	0	C/T	NA	NA	0.67	0.01	**5.16**

Abbreviations – AA: amino acid. MAF: minor allele frequency. SIFT: sorting intolerant from tolerant. PP2: Polyphen2. GERP: genomic evolutionary rate profiling. NA: sample not available. FS: sample failed to sequence.

Asterisk – variants with MAF > 0.03 were only Sanger-sequence validated in probands that were homozygous for the alternate allele or were possible compound heterozygotes.

Blue number – MAF from 1000 genomes European descent database.

Red text – Stop gain (nonsense) mutation. Green text – Stop loss mutation.

Green highlight – homozygous reference, Yellow highlight – heterozygous, Red highlight – homozygous alternate, Grey highlight – information not available.

Bold numbers – Predicted to be damaging (SIFT <0.05, PP2 (HDIV) >0.956) or evolutionary conserved for GERP (>2).

Italicized numbers – Predicted to be possibly damaging (PP2 (HDIV) between 0.453 and 0.956).

Table 3.10.3: List of candidate modifier genes based on variants identified in a human cranial NTD cohort, protein function, and gene expression. Asterisk represents data generated in this thesis. All other information for function and expression was obtained from online databases. "Unknown" means this information was not found in online databases. Online databases include PubMed (http://www.ncbi.nlm.nih.gov/pubmed), GeneCards Human Gene Database (http://www.genecards.org/), and MGI Gene eXpression Database (http://www.informatics.jax.org/expression.shtml).

Candidate gene	Variants of interest (#)	Function	Expression in mouse neural tube at neurulation
DNMBP	8	CDC42 activation [207], Epithelial junction integrity [208], Ciliogenesis [209]	*Yes (Figure 3.10.1)
MMS19	6	Iron-sulfur cluster assembly [210], Genome integrity [210], Nucleotide excision repair [211], RNAPII transcription [211]	*Yes (Figure 3.10.3)
TJP2	7	Epithelial junction integrity [212]	Yes [213]
TCTN3	2	Shh signaling [214], Ciliogenesis [215]	Unknown
SCD	2	Fatty acid biosynthesis [216]	Mouse *Scd1* - Unknown, Mouse *Scd2* - Yes [217]
KANK1	3	Cytoskeletal dynamics [218]	Unknown
GLIS3	4	Transcription regulator involved in pancreatic beta cell, thyroid, and kidney development [219-221]	Yes [222]
FAS	1	Apoptosis [223], Immune function [224]	Unknown
SFXN2	4	Mitochondrial tricarboxylate carrier [225]	Unknown
CSTF2T	3	mRNA 3' processing [226], Male fertility [227]	Unknown
CPEB3	1	Long-term memory [228], mRNA binding [229]	Little to none [230]
EXOSC1	1	Component of exosome involved in RNA processing and degradation [231]	Unknown
HPS6	1	Lysosome biogenesis [232,233]	Unknown
RNLS	2	Oxidoreductase activity [234], Cardiovascular maintenance [235]	Unknown
PNLIPRP2	4	Fat digestion [236]	Unknown
TMEM180	2	Unknown	Unknown
FAM45A	1	Unknown	Unknown
FOXD4	0	Neural transcription factor [237]	Yes [238,239]
BTAF1	0	Regulation of transcription [240]	Yes (E8.5 whole embryo) [241]
HIF1AN	0	Oxygen sensor for hypoxic response pathway [242]	Unknown
NANOS1	0	Translational repressor [243]	Unknown
TRUB1	0	Pseudouridine synthase [244]	Unknown
LIPJ	0	Unknown (lipase)	Unknown
FAM160B1	0	Unknown	Unknown

DNMBP contains an SH3 domain that binds dynamin, a protein that plays a key role in fission of endocytic buds, a BAR domain that binds highly curved membranes typical of sites undergoing endocytosis, and a DH domain that specifically activates the rhoGTPase CDC42 [245]. WES analysis in mice identified two missense mutations in *DNMBP*, one of which altered a

proline (FVB/N) to leucine (BALB/cCrl) within the conserved DH domain that was predicted to be damaging by both SIFT and Polyphen2. The other variant in mouse altered a valine (FVB/N) to a methionine (BALB/cCrl) that was predicted to be possibly damaging by Polyphen2 in a coiled-coil, which is a motif involved in protein-protein interactions (Table 3.7.1). Western blot and immunofluorescence staining indicated that DNMBP protein was present in mouse embryos at the time of neurulation, including within the neural epithilium (Figure 3.10.1). DNMBP protein appeared to be cytoplasmic, although nuclear localization cannot be excluded according to these images (Figure 3.10.1B,C). Staining without 1° antibody, but with 2° antibody, demonstrated that the α-DNMBP signal (green) was not due to non-specific staining from the 2° antibody (Figure 3.10.1E). Mouse DNMBP is predicted to be 177 kDa in size, and western blot confirmed that the DNMBP protein expressed in E8.5 FVB/N embryos was ~177 kDa (Figure 3.10.1G). Interestingly, another band was seen in E8.5 FVB/N embryos that was ~94 kDa (Figure 3.10.1G). Although only one DNMBP isoform has been annotated in mouse, two isoforms have been annotated in humans. The larger human isoform is predicted to be 177 kDa and the smaller isoform is predicted to be 94 kDa. These results suggest that the smaller DNMBP isoform is also expressed in mouse. A third band intermediate in size of these two isoforms was also present on the western blot. It is inconclusive as to whether this band was specific to a novel isoform of DNMBP, was a breakdown product of the larger DNMBP isoform, or was a cross-reaction of the α-DNMBP antibody with another protein.

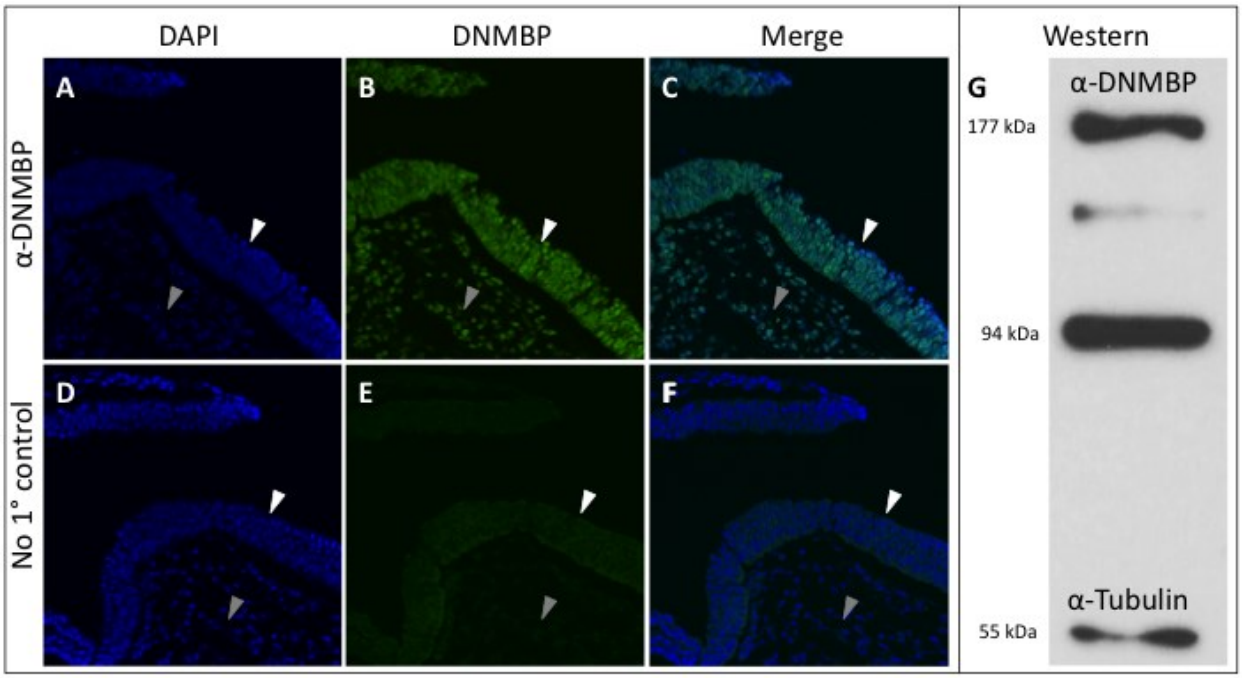


Figure 3.10.1: DNMBP is present in the neural epithelium at the time of neurulation. Immunofluorescent staining of wildtype E8.5 FVB/N embryo head sections showed DNMBP presence in the neuroepithelium and head mesenchyme (**B**). DAPI was used to counterstain the nuclei (**A**). The merged image of DAPI and DNMBP suggests that DNMBP is mainly cytoplasmic (**C**). The no 1° control demonstrated that staining was not due to non-specific binding of the 2° antibody (**D-F**). White arrowheads within panels **A-F** point at the neural epithelium. Grey arrowheads within panels **A-F** point at head mesenchyme. Immunofluorescent images were taken at 400X magnification. Western blot of a pooled sample of eleven E8.5 FVB/N embryos showed the presence of two DNMBP protein isoforms, the larger isoform at the molecular weight of ~177 kDa and the smaller isoform at ~94 kDa (**G**), α-tubulin served as the loading control.

Sequencing of *DNMBP* in the cranial NTD human cohort identified 8 variants of interest within 13 probands (Table 3.10.4, Figure 3.10.2). Five variants with a MAF < 0.03 were found in 5 different probands (Table 3.10.4). Two of these variants were novel, both of which were located within the BAR domain, one of which introduced a premature stop codon and therefore can be considered a loss-of-function mutation. None of these five probands contained additional coding variants within *DNMBP* and therefore were not compound heterozygotes. However, two of these probands also contained mutations in another candidate gene. Having variants in more

than one candidate gene in a single proband is worthy of note because NTDs are multifactorial, with the potential of many genetic factors contributing to the phenotype. Therefore, it is possible that these variants are contributing to the development of NTDs in an additive manner. Proband 20-0601184 contained a variant within *DNMBP* that altered an arginine to a histidine (MAF = 0.000116), which was predicted to be conserved by GERP (5.82) and was predicted to be damaging by polyphen2 (Table 3.10.4). This same proband, 20-0601184, also contained a novel variant within *SCD* that altered a tryptophan to a cysteine, which was predicted to be conserved by GERP (5.43) and was predicted to be damaging by both SIFT and Polyphen2 (Table 3.10.4). The other proband, A1024167, contained a variant in *DNMBP* that altered a threonine to a methionine (MAF = 0.002209), which was predicted to be conserved by GERP (3.49) and was predicted to be damaging by polyphen2 (Table 3.10.4). This same proband, A1024167, also contained a mutation in *CECR2* that altered a tyrosine to a cysteine (MAF = 0.001941), which was predicted to be conserved (GERP = 4.37) and was predicted to be damaging by Polyphen2 (Table 3.10.4). The remaining 8 probands contained more than one of the remaining 3 variants, all of which had a MAF > 0.03 (Table 3.10.4). Three probands were confirmed to be compound heterozyotes. Two probands inherited both variants from the same parent and therefore were not compound heterozyotes. The parental genotypes were uninformative for the remaining 2 probands, meaning it was inconclusive as to whether they were compound heterozygotes or not. *DNMBP* was originally a strong candidate modifier gene in mouse, and sequencing in humans found 8 variants, two of which were novel, that further strengthened *DNMBP* as a top candidate NTD gene in both mice and humans. The potential role of *DNMBP* in neurulation is further supported by current knowledge of gene function, as well as evidence for expression during neurulation (Figure 3.10.1).

Table 3.10.4: *DNMBP* protein-coding sequence variants of interest. Variants identified by next generation sequencing in cranial NTD probands were Sanger-sequence validated in probands and available parental DNA samples. Probands containing the lowest frequency variants (MAF esp6500) are listed first. Proband phenotype information is provided in Appendix B.

	Annotation				Genotype (reference/alternate)			Predictions of deleteriousness			
Proband ID	Position hg19	dbSNP137 ID	AA change	MAF esp6500	Proband	Maternal	Paternal	SIFT	PP2 (HDIV)	GERP	Compound heterozyg.
20-027140	101654735	Novel	R1042X	-	G/A	G/A	G/G	-	-	**3.85**	-
20-0806627	101646214	Novel	R1154Q	-	C/T	C/C	C/T	0.59	**1**	**5.82**	-
20-0601184	101639989	rs372003127	R1376H	0.000116	C/T	C/C	FS	0.07	**1**	**5.42**	-
A1024256	101667847	rs114927649	P820L	0.001744	G/A	G/G	G/A	0.62	0.015	**3.96**	-
A1024167	101639731	rs147752816	T1462M	0.002209	G/A	G/A	G/G	0.09	**0.986**	**3.49**	-
20-0426714	101667814	rs17854134	M831T	0.03186*	A/G	A/G	A/A	0.45	*0.673*	**5.99**	Yes
	101716112	rs35924554	N373K	0.04*	G/C	G/G	G/C	0.89	0.049	-1.33	
	101639877	rs11190305	C1413W	0.380349*	C/C	C/C	C/C	0.19	0.6	0.604	
20-0613589	101667814	rs17854134	M831T	0.03186*	A/G	A/G	A/A	0.45	*0.673*	**5.99**	Possibly
	101716112	rs35924554	N373K	0.04*	G/C	G/C	G/C	0.89	0.049	-1.33	
	101639877	rs11190305	C1413W	0.380349*	A/C	A/C	A/A	0.19	*0.6*	0.604	
20-0512275	101716112	rs35924554	N373K	0.04*	G/C	G/C	NA	0.89	0.049	-1.33	Possibly
	101639877	rs11190305	C1413W	0.380349*	A/C	A/C	NA	0.19	*0.6*	0.604	
20-0410539	101716112	rs35924554	N373K	0.04*	G/C	G/C	G/C	0.89	0.049	-1.33	Possibly
	101639877	rs11190305	C1413W	0.380349*	A/C	A/C	A/A	0.19	*0.6*	0.604	
20-0514451	101716112	rs35924554	N373K	0.04*	G/C	G/G	G/C	0.89	0.049	-1.33	No
	101639877	rs11190305	C1413W	0.380349*	A/C	A/A	C/C	0.19	*0.6*	0.604	
20-0620476	101716112	rs35924554	N373K	0.04*	G/C	G/G	G/C	0.89	0.049	-1.33	No
	101639877	rs11190305	C1413W	0.380349*	A/C	A/A	C/C	0.19	*0.6*	0.604	
20-0504243	101716112	rs35924554	N373K	0.04*	G/C	C/C	G/G	0.89	0.049	-1.33	Yes
	101639877	rs11190305	C1413W	0.380349*	C/C	A/C	C/C	0.19	*0.6*	0.604	
20-0212267	101716112	rs35924554	N373K	0.04*	G/C	G/G	G/C	0.89	0.049	-1.33	Yes
	101639877	rs11190305	C1413W	0.380349*	C/C	FS	C/C	0.19	*0.6*	0.604	

Abbreviations – AA: amino acid. MAF: minor allele frequency. SIFT: sorting intolerant from tolerant. PP2: Polyphen2. GERP: genomic evolutionary rate profiling. NA: sample not available. FS: sample failed to sequence. Heterozyg.: heterozygosity

Position hg19 is on chromosome 10.

Asterisk – variants with MAF > 0.03 were only Sanger-sequence validated in probands that were homozygous for the alternate allele or were possible compound heterozygotes.

Red text – Stop gain (nonsense) mutation.

Green highlight – homozygous reference, Yellow highlight – heterozygous, Red highlight – homozygous alternate, Grey highlight – information not available.

Bold numbers – Predicted to be damaging (SIFT <0.05, PP2 (HDIV) >0.956) or evolutionary conserved for GERP (>2).

Italicized numbers – Predicted to be possibly damaging (PP2 (HDIV) between 0.453 and 0.956).

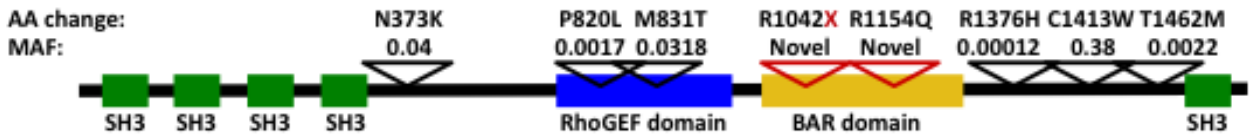


Figure 3.10.2: Variants identified in the candidate modifier gene *DNMBP* in the human cranial NTD cohort. Novel mutations (not in esp6500 or 1KG) are marked with a red triangle. Nonsense mutations are marked with a red 'X'. Note that the two novel mutations, one of which is a nonsense mutation, are located within the functional BAR domain. AA change – amino acid change, MAF – minor allele frequency (esp6500).

The MMS19 homologue in fission yeast has been shown to form a large complex that includes the proteins Dos2, Rik1, and cdc20, which functions in binding to and regulating RNA polymerase II transcription activity [246]. MMS19 contains an N-terminal Dos2-interacting domain and a C-terminal RNAPII transcription regulator domain, which are involved in aiding in complex formation and RNAPII regulation respectively [246]. It is important to note that there does not appear to be a known human homologue of Dos2; therefore, it is currently unknown what the function of the Dos-2 interacting conserved domain is in humans. The C-terminus of human and mouse MMS19 also contains a series of HEAT (huntingtin, elongation factor 3, protein phosphatase 2A, kinase TOR1) repeats [247], which are motifs that have been shown to function as scaffolds for the assembly of large multi-subunit protein complexes [248-250]. *MMS19* was selected for sequencing in the human cohort as *Mms19* was found to contain one variant that differed between BALB/cCrl and FVB/N that altered a proline to a valine (Table 3.7.1). This variant was not predicted to be evolutionarily conserved (GERP = 1.37) and neither was it predicted to be damaging by Polyphen2. SIFT was unable to make a prediction. Although the predictors of deleteriousness were helpful in filtering for good candidates, it is still possible that this variant has a functional effect. Therefore, *Mms19* remained a candidate gene despite these predictions, as *Mms19* has been implicated in important biological functions such as maintaining genomic integrity [210]. Western blot and immunofluorescence staining indicated that MMS19 was expressed in neurulating mouse embryos, including within the neural epithelium and head mesenchyme (Figure 3.10.3). The majority of MMS19 protein appeared to be cytoplasmic; however, a small proportion of cells within the neural epithelium and mesenchyme displayed heightened levels of MMS19 relative to other cells and some of the MMS19 protein appeared to

co-localize with DAPI in the nuclei (Figure 3.10.3B,C). Staining without 1° antibody, but with 2° antibody, demonstrated that the α-MMS19 signal (green) was not due to non-specific staining from the 2° antibody (Figure 3.10.3E). Mouse MMS19 is predicted to be 113 kDa in size [247], and western blot confirmed that the MMS19 protein expressed in E8.5 FVB/N embryos was ~113 kDa (Figure 3.10.3G).

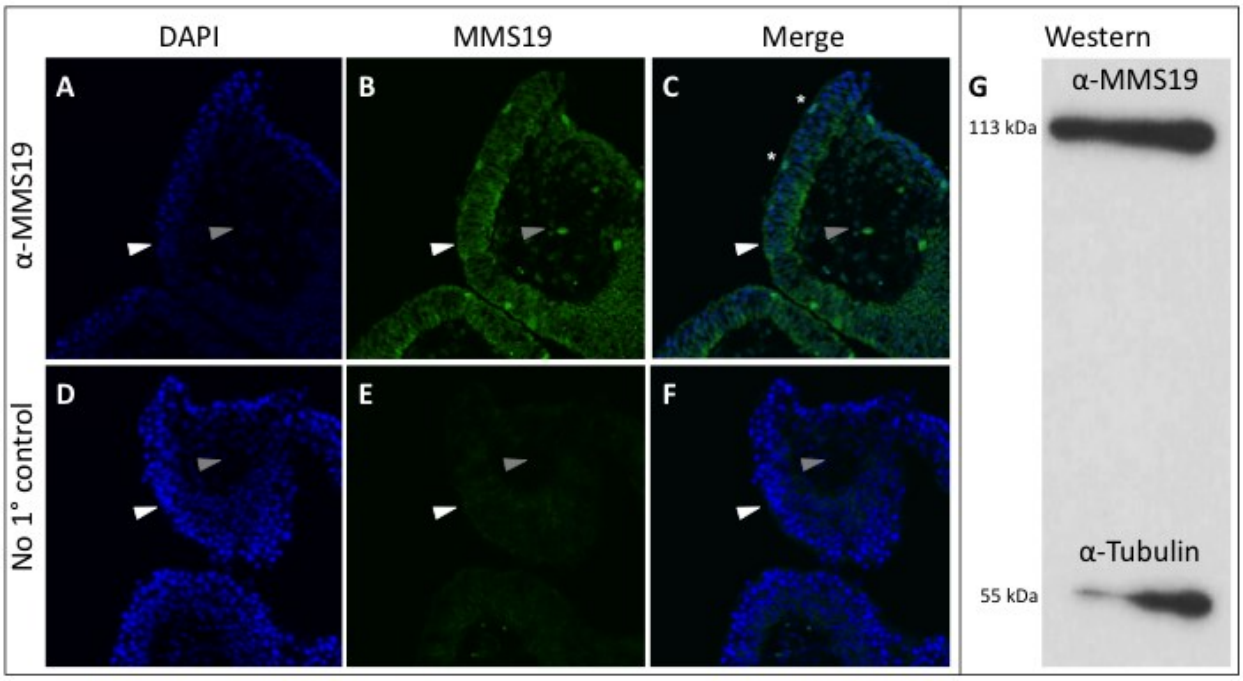


Figure 3.10.3: MMS19 is present in the neural epithelium at the time of neurulation. Immunofluorescent staining of wildtype E8.5 FVB/N embryo head sections showed MMS19 presence in the neuroepithelium and head mesenchyme (**B**). DAPI was used to counterstain the nuclei (**A**). The merged image of DAPI and MMS19 revealed that the majority of MMS19 is mainly cytoplasmic, with a small number of cells within the neural epithelium (denoted with asterisk) showing higher levels of MMS19 expression as well as possible nuclear localization (**C**). The no 1° control demonstrated that staining was not due to non-specific binding of the 2° antibody (**D-F**). White arrowheads within panels **A-F** point at the neural epithelium. Grey arrowheads within panels **A-F** point at head mesenchyme. Immunofluorescent images were taken at 400X magnification. Western blot of a pooled sample of eleven E8.5 FVB/N embryos showed the presence of MMS19 protein at the molecular weight of ~113 kDa (**G**), α-tubulin served as the loading control.

Sequencing of *MMS19* in the cranial NTD human cohort identified 6 variants of interest (1 novel, 3 rare, 2 common) within 12 probands (Table 3.10.5, Figure 3.10.4). All 12 probands contained at least one novel or rare variant (MAF < 0.03). The novel variant resulted in the introduction of a stop codon early within the protein in the Dos2-interacting domain, and was found in proband 20-0521275 (Table 3.10.5, Figure 3.10.4). This same proband, 20-0521275, also contained another variant, rs12360068, that altered an alanine to a valine, with a MAF = 0.043023, high conservation (GERP = 4.64), and a prediction to be damaging by polyphen2 (Table 3.10.5). This proband, 20-0521275, was confirmed to be a compound heterozygote. Interestingly, this same proband also contained a variant in *DNMBP* (rs17854134, MAF = 0.03186) and two variants in *TJP2* (rs41305539, MAF = 0.019316; rs34774441, MAF = 0.064884) (Table 3.10.5), all of which are evolutionary conserved according to GERP. Another proband, 20-0700015, contained a rare variant (MAF = 0.000581) in *MMS19* that altered a proline to a serine, was predicted to be conserved (GERP = 5.9), and was predicted to be damaging by Polyphen2. This same proband, 20-0700015, was possibly a compound heterozygote for the common variant rs3740526 (MAF = 0.41389), which was also shown to be conserved (GERP = 3.64) but was not predicted to be damaging; however, parental genotypes were uninformative. This proband, 20-0700015, also contained mutations in other candidate genes, which included a variant in *GLIS3* (rs148572278, MAF = 0.003256, evolutionarily conserved and predicted to be damaging) and homozygosity for three variants in *PNLIPRP2* (rs62623669, MAF = 0.017154; rs4751996, MAF = 0.479667; rs4751995, MAF = 0.481849), two of which are shown to be evolutionarily conserved (Table 3.10.2). Two probands, 20-0427674 and 20-0708160, were discovered to contain the same rare variant, rs36023427, that has a MAF = 0.008721, with high conservation (GERP = 5.86), and was predicted to be damaging by Polyphen2. Both probands containing this variant, rs36023427, were also confirmed compound heterozygotes for the common variant rs3740526 discussed above for proband 20-0700015. The remaining rare variant, rs17112809, was found in 8 probands. rs17112809 results in the alteration of a valine to an isoleucine and is located just upstream of the RNAPII transcription regulator domain (Figure 3.10.4). Interestingly, rs17112809 has a MAF of 0.010465 in the esp6500 database, but it was enriched by ~2.5 fold in the human cranial NTD cohort with a frequency of 8/312 (0.0256). However, rs17112809 was not predicted to be conserved (GERP = 1.33) and neither was it predicted to be damaging by SIFT or Polyphen2. Six of the 8 probands containing

this variant also contained variants in other candidate genes (Table 3.10.2). One proband, A1015381, was also a compound heterozygote for two variants in *CECR2* (rs5747211, MAF = 0.113241; rs1296794, MAF = 0.191566), both of which are predicted to be deleterious. Two other probands, 20-0620476 and 20-0521275, contained additional variants in both *DNMBP* and *TJP2*. Yet another two probands, 20-0302223 and A1024167, contained variants in *TJP2*, the latter proband of which was homozygous for the alternate allele in *TJP2*. Of the 12 probands with variants in *MMS19*, four did not contain variants of interest in any other candidate genes. These four probands were A1030655, Z0167301, 20-0427674, and 20-0708160. These results, along with gene expression and function make *MMS19* a strong candidate NTD gene in humans.

Table 3.10.5: *MMS19* protein-coding sequence variants of interest. Variants identified by next generation sequencing in cranial NTD probands were Sanger-sequence validated in probands and available parental DNA samples. Probands containing the lowest frequency variants (MAF esp6500) are listed first. Proband phenotype information is provided in Appendix B.

	Annotation				Genotype (reference/alternate)			Predictions of deleteriousness			
Proband ID	Position hg19	dbSNP137 ID	AA change	MAF esp6500	Proband	Maternal	Paternal	SIFT	PP2 (HDIV)	GERP	Compound heterozygosity
20-0521275	99237155	Novel	R187X	-	G/A	G/G	G/A	-	-	**5.48**	Yes
	99225645	rs12360068	A558V	0.043023*	G/A	G/A	G/G	0.37	**1**	**4.64**	
20-0700015	99218991	rs200490757	P985S	0.000581	G/A	G/G	G/A	0.06	**1**	**5.9**	Possibly
	99220707	rs3740526	G790D	0.41389*	C/T	C/T	T/T	0.62	0.103	**3.64**	
20-0427674	99218456	rs36023427	G1029D	0.008721	C/T	FS	C/C	0.16	**1**	**5.86**	Yes
	99220707	rs3740526	G790D	0.41389*	T/T	FS	T/T	0.62	0.103	**3.64**	
20-0708160	99218456	rs36023427	G1029D	0.008721	C/T	C/C	C/T	0.16	**1**	**5.86**	Yes
	99220707	rs3740526	G790D	0.41389*	T/T	T/T	C/T	0.62	0.103	**3.64**	
Z0167301	99225846	rs17112809	V526I	0.010465	C/T	C/T	C/T	0.31	0.155	1.33	-
20-0604224	99225846	rs17112809	V526I	0.010465	C/T	C/T	C/C	0.31	0.155	1.33	-
20-0302223	99225846	rs17112809	V526I	0.010465	C/T	C/C	C/T	0.31	0.155	1.33	-
20-0300929	99225846	rs17112809	V526I	0.010465	C/T	C/T	C/C	0.31	0.155	1.33	-
A1015381	99225846	rs17112809	V526I	0.010465	C/T	C/C	C/T	0.31	0.155	1.33	-
20-0620476	99225846	rs17112809	V526I	0.010465	C/T	C/T	C/C	0.31	0.155	1.33	-
A1024176	99225846	rs17112809	V526I	0.010465	C/T	C/T	C/C	0.31	0.155	1.33	-
A1030655	99225846	rs17112809	V526I	0.010465	C/T	C/C	NA	0.31	0.155	1.33	-

Abbreviations – AA: amino acid. MAF: minor allele frequency. SIFT: sorting intolerant from tolerant. PP2: Polyphen2. GERP: genomic evolutionary rate profiling. NA: sample not available. FS: sample failed to sequence.

Position hg19 is on chromosome 10.

Asterisk – variants with MAF > 0.03 were only Sanger-sequence validated in probands that were homozygous for the alternate allele or were possible compound heterozygotes.

Red text – Stop gain (nonsense) mutation.

Green highlight – homozygous reference, Yellow highlight – heterozygous, Red highlight – homozygous alternate, Grey highlight – information not available.

Bold numbers – Predicted to be damaging (SIFT <0.05, PP2 (HDIV) >0.956) or evolutionary conserved for GERP (>2).

Figure 3.10.4: Variants identified in the candidate modifier gene *MMS19* in the human cranial NTD cohort. Novel mutations (not in esp6500 or 1KG) are marked with a red triangle. Nonsense mutations are marked with a red 'X'. Note that the novel nonsense mutation is located within the functional Dos2-interacting domain. AA change – amino acid change, MAF – minor allele frequency (esp6500).

TJP2 contains 3 PDZ (post synaptic density protein, *Drosophila* disc large tumor suppressor, zonula occludens-1) domains, a single SH3 domain, and a guanylate kinase domain. PDZ domains function in forming protein-protein interactions (reviewed in [251]). The guanylate kinase domain is catalytically inactive in TJP2, but instead acts in conjunction with the neighboring SH3 domain [252]. This SH3-GK tandem domain has been shown to function in interacting with various proteins [253-257]. *TJP2* was selected for sequencing in the human cohort as mouse *Tjp2* was found to contain one variant that differed between BALB/cCrl and FVB/N that altered a threonine to an asparagine (Table 3.7.1). Much like the variant identified in mouse *Mms19*, this variant was not predicted to be evolutionarily conserved (GERP = 1.3) and neither was it predicted to be damaging by Sift or Polyphen2. However, *Tjp2* remained a good candidate as previous work has used *Tjp2* as a marker of epithelial junctions in the neural epithelium at the time of neurulation in mouse [213], which demonstrated that *Tjp2* is expressed and localized to neural epithelial junctions at the time of neurulation.

Sequencing of *TJP2* in the cranial NTD human cohort identified 6 variants of interest within 16 probands (Table 3.10.6, Figure 3.10.5). One novel variant that altered an isoleucine to a threonine was identified in proband Z0082501. This novel variant showed high conservation (GERP = 5.43), was predicted to be damaging by both SIFT and Polyphen2, and was located within an SH3 protein domain (Figure 3.10.5). The proband containing this novel variant did not have additional variants of interest in *TJP2* or any other sequenced candidate genes. Three probands containing rare variants (MAF < 0.03) in *TJP2* also contained one additional variant in *TJP2* (MAF ≈ 0.07); however, two of these probands were confirmed to not be compound heterozygotes, and compound heterozygosity for the third proband could not be determined due

to lack of a paternal DNA sample. Three additional probands also contained more than one variant in *TJP2*; however, both variants were common (MAF ≈ 0.07) and compound heterozygosity could not be determined for all three due to lack of one or more parental DNA samples. Proband 20-0517865 was homozygous for rs77236826 (MAF = 0.07511); however, this variant did not show conservation (GERP = -9.87) and was not predicted to be damaging by either SIFT or Polyphen2. Two probands, 20-0620476 and A1024176, were homozygous for rs41277907 (MAF = 0.078256) that altered a serine to a phenylalanine, demonstrated high conservation (GERP = 5.58), and was predicted to be damaging by SIFT. It is interesting to note that based on the MAF of rs41277907, we would predict a frequency of homozygotes = 0.0061; however, there were 2 homozygotes in the human cranial NTD cohort of 156 probands (frequency = 0.013), demonstrating an ~ 2-fold enrichment. Unfortunately, a small sample size prevents any definitive conclusions regarding this enrichment. In addition to the four probands discussed above for *MMS19* (20-0620476, 20-0302223, 20-0521275, A1024176), there were three more probands containing variants in *TJP2* that also contained variants of interest in additional candidate genes (Table 3.10.2). One of these probands, 20-0426206, which was possibly a compound heterozygote for *TJP2*, also contained a rare variant in *HSP6* (rs36078476, MAF = 0.009767) that was predicted to be conserved (GERP = 2.39) but was not predicted to be damaging. The second proband, 20-0705228, contained a rare variant in *TJP2* (rs149911553, MAF = 0.001628) as well as in *PNLIPRP2* (rs200056143, MAF = 0.000117) that was shown to be conserved (GERP = 4.97) and predicted to be possibly damaging by Polyphen2. The third proband, A1030628 contained two intra-allelic variants in *TJP2* (rs199892018, MAF = 0.000349; rs34774441, MAF = 0.064884), as well as a rare variant in *CPEB3* (rs140779166, MAF = 0.000581) that altered a valine to an isoleucine, was shown to be conserved (GERP = 5.65), but was not predicted to be damaging. *TJP2* is known to function in epithelial junction integrity and is expressed within the neural epithelium at the time of neurulation. This, in conjunction with the identified variants, makes *TJP2* a strong NTD candidate in mice and humans.

Table 3.10.6: *TJP2* protein-coding sequence variants of interest. Variants identified by next generation sequencing in cranial NTD probands were Sanger-sequence validated in probands and available parental DNA samples. Probands containing the lowest frequency variants (MAF esp6500) are listed first. Proband phenotype information is provided in Appendix B.

	Annotation				Genotype (reference/alternate)			Predictions of deleteriousness			
Proband ID	Position hg19	dbSNP137 ID	AA change	MAF esp6500	Proband	Maternal	Paternal	SIFT	PP2 (HDIV)	GERP	Compound heterozygosity
Z0082501	71850989	Novel	I609T	-	T/C	T/C	T/T	**0**	**1**	**5.43**	-
A1030628	71866149	rs199892018	V1064L	0.000349	G/T	G/T	G/G	0.94	0	-0.183	No
	71851877	rs34774441	M668I	0.064884*	G/A	G/A	G/G	**0.02**	*0.928*	**5.77**	
20-0705228	71851040	rs149911553	T626S	0.001628	C/G	C/C	C/G	0.16	0.008	**3.55**	-
20-0521275	71835842	rs41305539	Q128K	0.019316	C/A	C/C	C/A	1	0.001	**2.89**	No
	71851877	rs34774441	M668I	0.064884*	G/A	G/G	G/A	**0.02**	*0.928*	**5.77**	
A1024188	71835842	rs41305539	Q128K	0.019316	C/A	C/C	NA	1	0.001	**2.89**	Possibly
	71865988	rs41277907	S1010F	0.078256*	C/T	C/T	NA	**0.02**	0.001	**5.58**	
A1034284	71835842	rs41305539	Q128K	0.019316	C/A	C/A	C/C	1	0.001	**2.89**	-
20-0518590	71835842	rs41305539	Q128K	0.019316	C/A	C/C	C/A	1	0.001	**2.89**	-
20-0302223	71835842	rs41305539	Q128K	0.019316	C/A	C/C	C/A	1	0.001	**2.89**	-
20-0609211	71835842	rs41305539	Q128K	0.019316	C/A	C/C	C/A	1	0.001	**2.89**	-
20-0613565	71835842	rs41305539	Q128K	0.019316	C/A	C/C	C/A	1	0.001	**2.89**	-
A1024159	71851877	rs34774441	M668I	0.064884*	G/A	NA	NA	**0.02**	*0.928*	**5.77**	Possibly
	71863235	rs77236826	H992R	0.07511*	A/G	NA	NA	0.12	0	-9.87	
A1021228	71851877	rs34774441	M668I	0.064884*	G/A	NA	NA	**0.02**	*0.928*	**5.77**	Possibly
	71865988	rs41277907	S1010F	0.078256*	C/T	NA	NA	**0.02**	0.001	**5.58**	
20-0426206	71863235	rs77236826	H992R	0.07511*	A/G	A/G	NA	0.12	0	-9.87	Possibly
	71865988	rs41277907	S1010F	0.078256*	C/T	C/C	NA	**0.02**	0.001	**5.58**	
20-0517865	71863235	rs77236826	H992R	0.07511*	G/G	A/G	G/G	0.12	0	-9.87	-
20-0620476	71865988	rs41277907	S1010F	0.078256*	T/T	C/T	C/T	**0.02**	0.001	**5.58**	-
A1024176	71865988	rs41277907	S1010F	0.078256*	T/T	C/T	C/T	**0.02**	0.001	**5.58**	-

Abbreviations – AA: amino acid. MAF: minor allele frequency. SIFT: sorting intolerant from tolerant. PP2: Polyphen2. GERP: genomic evolutionary rate profiling. NA: sample not available.

Position hg19 is on chromosome 9.

Asterisk – variants with MAF > 0.03 were only Sanger-sequence validated in probands that were homozygous for the alternate allele or were possible compound heterozygotes.

Green highlight – homozygous reference, Yellow highlight – heterozygous, Red highlight – homozygous alternate, Grey highlight – information not available.

Bold numbers – Predicted to be damaging (SIFT <0.05, PP2 (HDIV) >0.956) or evolutionary conserved for GERP (>2).

Italicized numbers – Predicted to be possibly damaging (PP2 (HDIV) between 0.453 and 0.956).

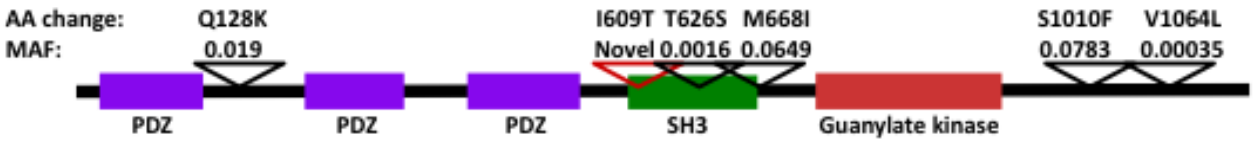


Figure 3.10.5: Variants identified in the candidate modifier gene *TJP2* in the human cranial NTD cohort. Novel mutations (not in esp6500 or 1KG) are marked with a red triangle. Note that the novel mutation is located within a functional SH3 domain. AA change – amino acid change, MAF – minor allele frequency (esp6500).

DNMBP, *MMS19*, and *TJP2* are strong candidate modifier genes based on the number and location of the variants of interest identified, expression within the neural tube during neurulation, and what is known about protein function (Table 3.10.3). However, the other 14 candidate modifiers with variants of interest should still be kept in consideration (Figure 3.l0.6). I identified 6 novel variants in the five genes *CSTF2T*, *RNLS*, *SCD*, *SFXN2* (2 novel variants), and *KANK1*. The 5 novel variants within the first 4 aforementioned genes were located within protein functional domains (functional domain annotations were acquired from www.ensembl.org). All except for one novel variant (SFXN2-N87K) were predicted to be evolutionarily conserved by GERP, and all 6 novel variants were predicted to be deleterious by SIFT and/or Polyphen2. As these genes potentially modify the phenotypic impact of *Cecr2* mutation in mice, variants within these genes identified in probands that also contain variants of interest in *CECR2* deserve some attention. Probands containing variants in *CECR2* along with variants in *DNMBP*, *MMS19*, and/or *TJP2* are mentioned above. There were two probands that contained a variant of interest in *CECR2* as well as at least one variant of interest in one of the remaining 14 candidate modifier genes (Table 3.10.2). The proband 20-0608812 is a compound heterozygote for the *CECR2* alleles rs5747211 (MAF = 0.113241) and rs1296794 (MAF = 0.191566), both of which are predicted to be deleterious. This proband also contains a rare variant in *GLIS3* (rs35154632, MAF = 0.009186) that is highly conserved and predicted to be deleterious, as well as the novel mutation *SFXN2*-N32K, which is predicted to be deleterious. Another proband, 20-0601092, was homozygous for the *CECR2* variant rs1296794 (MAF = 0.191566), and also contained three variants in *PNLIPRP2* (rs62623669, MAF = 0.017154; rs4751996, MAF = 0.479667; rs4751995, MAF = 0.481849). Of the three variants in *PNLIPRP2*, two are common variants. However, both of these variants are predicted to be highly conserved by GERP, and the proband was homozygous for the alternative allele for both of these variants. Variants of interest were not identified in 7 of the 24 candidate modifier genes sequenced, making them lower priority candidate NTD genes (Table 3.10.3). These 7 genes were *LIPJ*, *FOXD4*, *BTAF1*, *HIF1AN*, *FAM160B1*, *TRUB1*, and *NANOS1*.

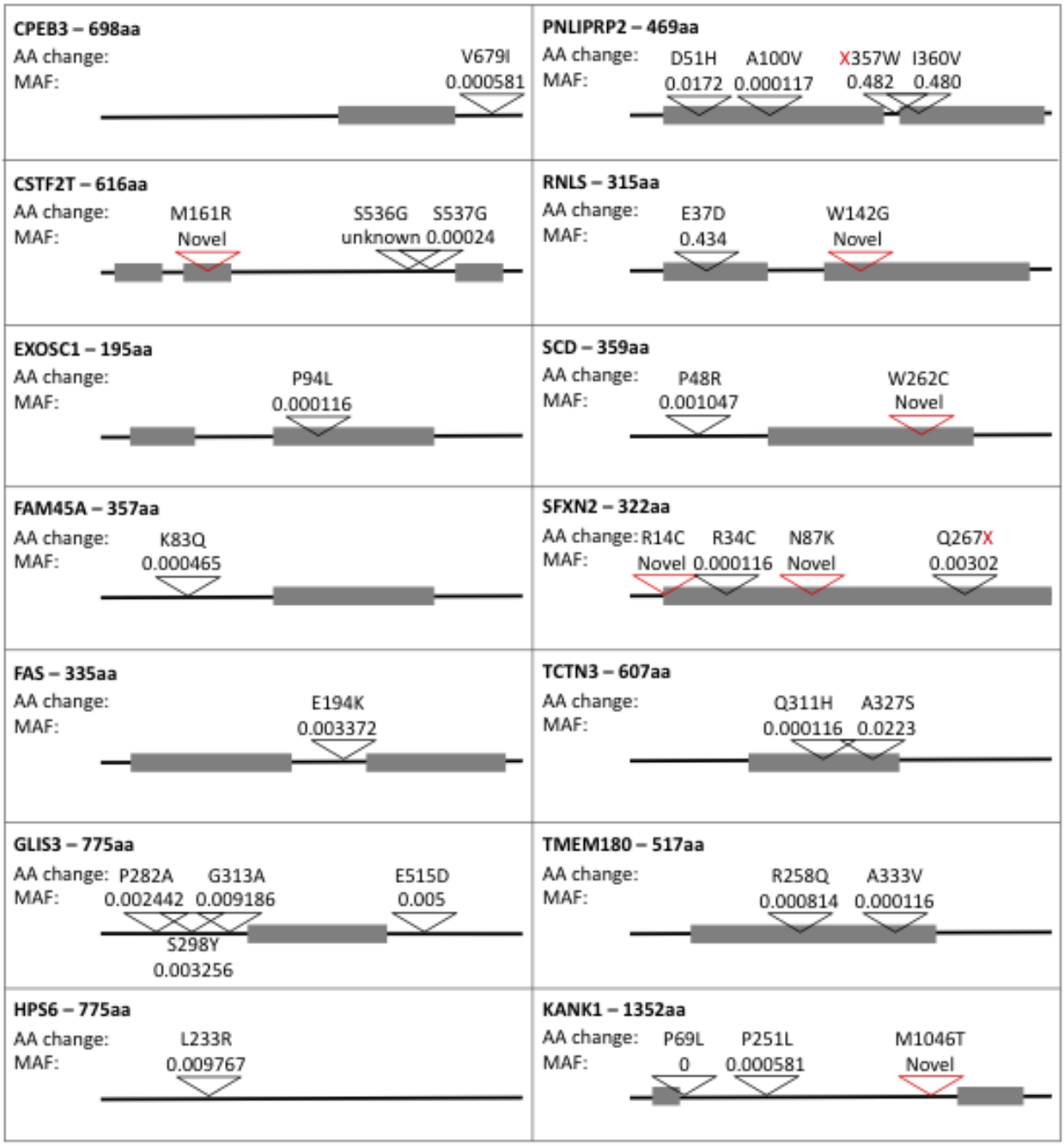


Figure 3.10.6: Variants identified in the remaining 14 candidate modifier genes in the human cranial NTD cohort. Novel mutations (not in esp6500 or 1KG) are marked with a red triangle. Stop codons are marked with a red 'X'. Grey bars denote annotated functional domains (www.ensembl.org). Note that all except one of the novel nonsense mutations are located within a protein functional domain. AA change – amino acid change, MAF – minor allele frequency (esp6500 for all variants except CSTF2T-S537G, which is from 1KG).

3.11: Discussion

In *Cecr2* mutant mice, the manifestation of exencephaly is reliant on the presence of other susceptibility factors. This is exemplified by *Cecr2*-associated exencephaly demonstrating reduced penetrance for two different mutant alleles in multiple mouse lines. Exencephaly penetrance also changed over time, meaning that the mutant mice were exposed to additional genetic and/or environmental resistance factors that were not present at the time of the original exencephaly penetrance analyses. The difference in exencephaly penetrance between mouse strains indicates the presence of genetic background effects, thereby also supporting the multifactorial nature of *Cecr2*-associated exencephaly in mice. Efforts to identify the genes contributing to these genetic background effects led to the discovery of a modifier region on chromosome 19 that contains more than one modifier gene. Further analyses of this region resulted in the identification of 25 candidate modifier genes in mouse. Sequencing of human *CECR2* and 24 candidate modifier genes of *CECR2* in a cranial NTD cohort identified protein-coding sequence variants that may contribute to cranial NTD etiology in humans.

Exencephaly penetrance re-analyses demonstrated that penetrance dropped over time

Updated exencephaly penetrance analyses were completed for four different mouse lines (congenic *Cecr2*^{GT45bic} BALB/cCrl, congenic *Cecr2*$^{tm1.1Hemc}$ FVB/N, *Cecr2*^{GT45bic} MOD 5, and *Cecr2*^{GT45bic} MOD 31), all of which demonstrated an exencephaly penetrance that dropped by ~15-20% (Table 3.11.1). Exencephaly penetrance analyses performed on the two mouse strains, BALB/cCrl and FVB/N, congenic for two different *Cecr2* mutant alleles, *Cecr2*^{GT45bic} and *Cecr2*$^{tm1.1Hemc}$ respectively, demonstrated a significant drop in exencephaly penetrance relative to previous analyses on earlier generations. It was not surprising that the penetrance of exencephaly dropped in congenic *Cecr2*$^{tm1.1Hemc}$ FVB/N relative to N3 *Cecr2*$^{tm1.1Hemc}$ FVB/N because it has already been established that FVB/N is resistant to the development of *Cecr2*-associated exencephaly and that the resistance (modifier) loci are dominant [141,164]. Also, various analyses of exencephaly penetrance suggested that the FVB/N genetic background harbors a general resistance to the development of NTDs [141,145,164]. Specifically, the penetrance of NTDs associated with mutations not only in *Cecr2*, but also in *Pax3* and *Shroom*, dropped when moved into an FVB/N genetic background [164]. Therefore, the drop in exencephaly penetrance in congenic

FVB/N was most likely due to additional FVB/N resistance loci being acquired over the remaining 7 backcrosses into the FVB/N genetic background.

As BALB/cCrl are not a resistant strain, the drop in exencephaly penetrance in congenic *Cecr2*^{GT45bic} BALB/cCrl was likely influenced by something different than what was seen in congenic FVB/N. Exencephaly penetrance caused by the *Cecr2*^{GT45bic} allele was also characterized in embryos that were 75% C57BL/6 and 25% BALB/cCrl, which yielded a penetrance of 69.7% (23/33) [199]. This was not significantly different from the original 74.5% seen in N6 BALB/cCrl ($P = 0.638$) or from the 54.1% seen in congenic BALB/cCrl ($P = 0.10$). Another previous penetrance analysis performed on embryos that were 75% 129/S2 and 25% BALB/cCrl demonstrated a penetrance of 48.5% (16/33) [199]. This penetrance was also not statistically significant from the congenic BALB/cCrl penetrance of 54.1% ($P = 0.56$) but was significantly lower than the original N6 BALB/cCrl penetrance of 74.5% ($P = 0.017$). These additional penetrance analyses further support that BALB/cCrl has a similar or higher susceptibility to *Cecr2*-associated exencephaly when compared to other strains. Possible reasons for the drop in exencephaly penetrance in congenic BALB/cCrl relative to the N6 generation are that several years passed in between each analysis; therefore, the mice may have acquired genetic changes that provided some resistance to the development of exencephaly, or environmental changes within the colony may have altered the susceptibility to exencephaly.

Although not significant, we also saw a drop in exencephaly penetrance in the MOD 5 and MOD 31 mouse lines. We did not out-cross the MOD 5 and MOD 31 mouse lines to other mouse lines in the colony within the time-frame that both old and new penetrance analyses were performed, meaning genetic background from other mouse lines were not introduced into MOD 5 or MOD 31. Therefore, the drop in exencephaly penetrance can only be explained by genetic changes that happened independently within each line or by environmental changes within the mouse colony. The latter explanation is most parsimonious as it can also explain the drop in penetrance seen in BALB/cCrl and FVB/N. It is more likely that the drop in penetrance was due to the same variable for all four mouse lines, rather than four independent variables for each mouse line. One major environmental change introduced to our mouse colony was a move from Health Sciences Laboratory Animal Services (HSLAS) to Sciences Animal Support Services (SASS) in the summer of 2008. This is more likely to be a factor for the old N6 BALB/cCrl analysis compared to the new congenic BALB/cCrl analysis, as the old analysis was performed in

HSLAS. However, the original penetrance analyses performed on the MOD lines were done between the fall of 2008 and summer of 2009, which was shortly after the mouse colony was moved from HSLAS to SASS. Therefore, less time has passed between the original and current analyses, meaning there was less time for a gradual change in penetrance to take place. Also, if the move of the colony introduced something that altered NTD susceptibility, it is possible that the NTD penetrance in mice used for the original MOD analysis was in the process of changing to the lower penetrance seen in the current analyses. In particular, animals used early on in the fall of 2008 may have spent a portion of their lives at HSLAS. Another environmental change within the colony was changing from autoclaved food to irradiated food. These sterilization practices may alter the nutrients within the food in different ways, which may result in differing environmental influences on neurulating embryos. Two independent pilot studies previously performed in the lab indicated that *Cecr2*-associated exencephaly penetrance is not sensitive to changes in dietary folic acid (Banting, unpublished; Norton, unpublished). Therefore, if the differences between autoclaved and irradiated food are contributing to changes in exencephaly penetrance, it is not likely due to changes in diatery folic acid. It cannot be ruled out that the penetrance in the two MOD lines simply did not drop over time, since there was not a significant difference between the old and the new penetrance analyses. To confirm this, a larger sample size of both the original and current analyses would be needed.

Table 3.11.1: Exencephaly penetrance associated with homozygous mutation in *Cecr2* dropped over time in four different mouse lines. The old penetrance analyses performed on FVB/N was at generation N3 [142] and BALB/cCrl was at generation N6 [141]. The new analyses performed on FVB/N and BALB/cCrl were at least generation N10. The genetic background for MOD 5 and MOD 31 should theoretically be the same for both old [165] and new analyses, as both lines were maintained by interbreeding within each line. The drop in exencephaly penetrance was statistically significant for FVB/N and BALB/cCrl, but not for MOD 5 and MOD 31.

	FVB/N ($Cecr2^{tm1.1Hemc}$)	BALB/cCrl ($Cecr2^{GT45bic}$)	MOD 5 ($Cecr2^{GT45bic}$)	MOD 31 ($Cecr2^{GT45bic}$)
Old penetrance (%)	31.4	74.5	50.0	46.1
New penetrance (%)	12.3	54.1	33.3	31.4
Drop in penetrance (%)	19.1	20.4	16.7	14.7

It has been well established that exencephaly has a higher incidence in females compared to males in both mice and humans, and *Cecr2*-associated exencephaly is not an exception to this trend. When homozygous mutant for *Cecr2*^{GT45bic}, ~1.8 female mice for every 1 male mouse develops exencephaly, a difference that has been shown to be statistically significant. A similar trend was seen in mice homozygous mutant for the *Cecr2*$^{tm1.1Hemc}$ allele, with ~1.4 female mice for every 1 male mouse developing exencephaly. However, this trend was not statistically significant, most likely due to the low exencephaly penetrance in congenic *Cecr2*$^{tm1.1Hemc}$ FVB/N producing a small number of exencephalic embryos. As this phenomenon has been witnessed in several NTD mouse models as well as in humans, it is likely that the female predisposition is a general susceptibility that is in addition to other susceptibility factors for most or all types of NTDs. This fits with the hypothesis devised by Juriloff and Harris (2012), which postulated that the presence of an additional X chromosome in females titrates out important epigenetic factors required for neurulation [106]. One possible theory is that because CECR2 itself is a chromatin binding protein, a significant reduction in CECR2 (~14-fold in *Cecr2*^{GT45bic} mutant mice) compounded with an additional X chromosome to titrate out the little wildtype CECR2 that is present could be a mechanism by which mutation in *Cecr2* results in a higher incidence of exencephaly in female mice.

Small inner ears in Cecr2^{GT45bic} *mutant embryos are secondary to the exencephaly phenotype*

Dawe et al. (2011) previously characterized inner ear defects in *Cecr2* homozygous mutant embryos [143]. The focus of the previous analysis was on stereociliary bundle misorientation, with results concluding that a greater proportion of stereociliary bundle cells were misoriented in homozygous mutant embryos, with an intermediate misoriented stereociliary bundle phenotype in heterozygous embryos [143]. The authors also observed a few non-penetrant embryos to have a milder phenotype than exencephalic embryos but more severe than the heterozygous embryos; however, this was not quantified due to a small sample size of non-penetrant embryos in this analysis [143]. Smaller inner ears was another phenotype addressed in Dawe et al. (2011), with the authors concluding that inner ears were smaller in exencephalic embryos but similar in size in non-penetrant mutant embryos, heterozygous embryos, and wildtype embryos; however, these data were not quantified [143]. Here, inner ear size was measured and compared in wildtype embryos, mutant embryos not penetrant for exencephaly, and mutant

embryos with exencephaly, with results indicating that the small inner ear phenotype is specific to embryos with exencephaly. This suggests that the small inner ear phenotype is secondary to the exencephaly phenotype. Inner ear size may be restricted in exencephalic embryos because the cranium fails to form beginning directly above the inner ears due to gross eversion of brain tissue. Also, the small inner ear phenotype is likely independent from the stereociliary bundle misorientation phenotype, as Dawe et al. (2011) already established that misoriented stereociliary bundles are seen in non-penetrant mutant embryos as well as heterozygous embryos. Therefore, misorientation of stereociliary bundles can and do occur in normally sized inner ears.

In an effort to determine if the small inner ear size previously observed in BALB/cCrl was also partially rescued by FVB/N modifiers within the chromosome 19 region, I also measured inner ear size in embryos generated from the MOD 5 and MOD 31 exencephaly penetrance, which was performed concurrently with the congenic *Cecr2*^{GT45bic} BALB/cCrl exencephaly and small inner ear penetrance analyses. Small inner ears were only seen in embryos with exencephaly, corroborating the conclusion that the small inner ear phenotype is secondary to exencephaly and not an independent mutant phenotype produced by the *Cecr2*^{GT45bic} allele. Stereociliary bundle misorientation, a phenotype that also occurs in non-penetrant *Cecr2*^{GT45bic} mutant BALB/cCrl embryos as well as to a lesser degree in heterozygous embryos, has not been analyzed in MOD 5 or MOD 31 embryos. Therefore, it is currently unknown whether or not the FVB/N modifiers are capable of rescuing the stereociliary bundle misorientation phenotype to a similar degree as the exencephaly phenotype.

The FVB/N chromosome 19 modifier loci are dominant and not additive

NTDs demonstrate a complex, multifactorial mode of inheritance in both mice and humans. Our mouse model for NTDs exemplifies this by the drastic difference in exencephaly penetrance between congenic *Cecr2*^{GT45bic} BALB/cCrl (54.1%) and *Cecr2*^{GT45bic} FVB/N (~0%). Previous work identified a region on chromosome 19 that contains more than one modifier gene that renders BALB/cCrl susceptible and FVB/N resistant to the development of exencephaly [165]. Previous work demonstrated and my work confirmed that there is at least one modifier gene within region 1, and another modifier gene within the combined regions 3 and 4 (Figure 3.1.3, Figure 3.5.1) [165]. Although it is possible that more than two modifier genes exist within the chromosome 19 region, the simplest explanation is that there are two modifiers, which is the

hypothesis underlying the following discussion. The genetic analyses performed in the MOD mouse lines also determined that the FVB/N alleles of these two modifiers are dominant and not additive. The FVB/N allele being dominant means that only one FVB/N allele is required to reduce exencephaly penetrance. The exencephaly penetrance in mice heterozygous for the FVB/N resistance alleles (MOD 5/+, MOD 31/+, and MOD 5/31) was similar to the penetrance seen in mice homozygous for FVB/N resistance alleles (MOD 4, MOD 5 and MOD 31), which demonstrated that the FVB/N alleles are dominant. Whether or not the two modifier alleles are additive was also tested. Under this hypothesis, any combination of two FVB/N resistance alleles should result in a similar exencephaly penetrance. Therefore, one modifier would have to be located in region 1 and the other in region 4 in order to explain the similar penetrance in MOD 4, MOD 5, MOD 31, and MOD 5/31. This also means that if this were the case, the MOD 5/+ and MOD 31/+ embryos would only contain one FVB/N resistance allele and should display an exencephaly penetrance intermediate between MOD 5 or MOD 31 and congenic BALB/cCrl. However, this was not the case, with MOD 5/+ and MOD 31/+ displaying a similar exencephaly penetrance as all other MOD lines (MOD 4, MOD 5, MOD 31, and MOD 5/31). The fact that the modifiers were shown to not be additive means that there is a similar drop in exencephaly penetrance regardless of how many FVB/N modifier alleles are present. Therefore, the modifier genes appear to be epistatic, which suggests that the modifier genes are in the same pathway or process.

A refined candidate modifier gene list based on microarray and whole exome sequence data

A list of candidate modifier genes within the chromosome 19 modifier region was originally generated by the employment of whole genome microarray to identify genes differentially expressed between BALB/cCrl and FVB/N [165]. Although microarray analysis identified genes that are differentially expressed between strains, it neglected to identify genes containing DNA sequence variants that do not affect gene expression but may alter gene function. I used a new and complementary method to refine this list. I used whole exome sequencing to identify genes that contain DNA sequence variants that alter the amino acid sequence and therefore may affect function with or without affecting expression levels. It is possible that whole exome sequencing can capture variants that affect gene expression; however, only variants close to or within coding exons would be captured by this method and the ability of these variants to

affect gene expression would require additional testing. By combining these two complementary methods, I produced a list that contains candidate modifier genes that are differentially expressed between BALB/cCrl and FVB/N and/or contain protein-coding variants that differ between the two strains.

Genetic analyses in mice demonstrated that Arhgap19 is not a modifier gene

The top candidate modifier gene, *Arhgap19*, showed significantly lower expression by ~5-fold in BALB/cCrl compared to FVB/N, most likely due to nonsense-mediated decay of the mRNA as a result of the protein-coding variant, $Arhgap19^{Ex6non}$, that introduced a stop-codon in *Arhgap19* exon 6 in BALB/cCrl [165]. *Arhgap19* expression analysis via utilization of the genetrap $Arhgap19^{GT(YHD020)Byg}$ allele demonstrated that *Arhgap19* is expressed at the time of neurulation, which further supported the potential role of *Arhgap19* in neurulation, as well as in specific brain regions after neurulation up until embryonic day E14.5. The discovery that the $Arhgap19^{Ex6non}$ allele was unique to the BALB/cCrl substrain provided us with the opportunity to introduce a wildtype *Arhgap19* allele from the BALB/cJ substrain into the BALB/cCrl substrain, while keeping the genetic background "BALB/c". A dihybrid cross that resulted in the production of $Cecr2^{GT45bic}$ homozygous mutant embryos that were either homozygous wildtype, heterozygous, or homozygous mutant for the $Arhgap19^{Ex6non}$ allele demonstrated no difference in exencephaly penetrance between any of the three genotypic classes (Table 3.8.2). This led to the conclusion that *Arhgap19* is not a modifier of *Cecr2*-associated exencephaly. A caveat to this experiment was that although BALB/cJ was genetically similar to BALB/cCrl, we could not know whether or not any of the genetic differences between the two substrains can alter the penetrance of exencephaly. Nevertheless, no differences in exencephaly penetrance were observed, which suggests that the two substrains do have a similar exencephaly penetrance. All of the other DNA sequence variants identified by whole exome sequencing that differ between BALB/cCrl and FVB/N do not differ between BALB/cCrl and BALB/cJ, meaning that a similar genetic analysis could not be performed for the other candidate modifier genes. Attempting a similar experiment with an unrelated strain would not work because the allele being tested would remain linked to variants within the chromosome 19 region that originate from the unrelated strain, even if backcrossed into the BALB/cCrl genetic background to generate a congenic line. Therefore, the results would be inconclusive as alterations in exencephaly penetrance may be due to variants

that are linked to the allele being tested. Another option to test candidate modifier genes in mice would involve generating a *Cecr2*^{GT45bic} BALB/cCrl mouse containing a BAC transgene of the FVB/N modifier allele in order to determine if the FVB/N allele can partially rescue the exencephaly phenotype in BALB/cCrl.

Sequencing 25 genes in a human cranial NTD cohort identifies variants of interest

After eliminating *Arhgap19* as a candidate modifier gene, we sequenced human *CECR2* and the remaining 24 candidate genes identified by microarray and whole exome sequencing in a human cranial NTD cohort consisting of 156 probands in an effort to narrow down the candidate gene list as well as to identify novel candidate NTD genes in humans. As mutation in mouse *Cecr2* results in the cranial NTD exencephaly, the Duke human cohort is particularly relevant as all probands contain some sort of cranial NTD (129 anencephaly, 13 acrania, 10 encephalocele, 3 craniorachischisis, 1 cranial myelomeningocele) (Appendix B). The Duke cohort is also unique in that the majority of probands had anencephaly. Although anencephaly and spina bifida both occur with a similar frequency of 1-10/1000 births in humans [40-42], very few NTD studies have large numbers of individuals affected by anencephaly as part of their cohort. Another large cranial NTD cohort based in China contains a total of 105 probands with some form of cranial NTD; however, only 16 of these probands had anencephaly [68,71,258]. Therefore, work in this thesis included the first NTD sequencing study that focused on the anencephaly NTD sub-type.

In an effort to conserve resources, we did not sequence a control cohort as part of this study. Instead, I obtained the MAF of human variants in normal populations of European-American ancestry from the esp6500 database. A benefit to the use of this database is the large number of individuals that are included (~4300 European-American individuals), which is much larger than defined control cohorts in other studies. A caveat is the lack of access to information regarding these individuals. For example, an important consideration when selecting control individuals for NTD studies is to ensure that there are no known first-degree relatives with NTDs. Seven of the rare variants identified in this study in the cranial NTD cohort have only been identified in a heterozygous state in one normal individual in the esp6500 database (MAF $\leq$ 0.00012). We do not know whether any of these 'normal' individuals are related to someone with a NTD. As 'novel' variants are defined as not being present in the control cohort, using a large control cohort, such as the esp6500 database, reduces the likelihood of finding novel variants.

Our use of the esp6500 database as a control cohort is a valid strategy because it is much larger than any control cohort we could independently collect and sequence. Also, sequence information from an additional control cohort would unlikely aid in the elimination of variants of interest, particularly since our study design is inclusive of variants with low frequencies in normal populations.

To assess the functional impact of the identified variants, *in silico* predictions of evolutionary conservation (GERP) and functional deleteriousness (SIFT, Polyphen2) were used. An analysis by Flanagan et al. (2010) was performed on the ability of SIFT and PolyPhen to accurately predict deleteriousness of 131 pathogenic and 8 benign missense variants [259]. Their findings determined that SIFT and PolyPhen had a reasonably high sensitivity (68% and 69% respectively) but a low specificity (13% and 16%). They also found that assessing a missense variant by evolutionary conservation was the most reliable method to predict deleteriousness. Although *in silico* predictions are commonly used to prioritize variants identified in sequencing studies, a study by Jordan et al. (2015) argued that *in silico* predictions based on evolutionary conservation are constrained due to a lack of ability to put genomic context into consideration [260]. Jordan et al. (2015) focused on a phenomenon referred to as compensated pathogenic deviation, whereby a pathogenic variant is compensated for by another variant in the genome, potentially in several species, and could therefore be predicted as benign. However, loss of the compensatory variant in a species would uncover the pathogenicity of the other variant, which would therefore only be pathogenic in that species. Ultimately, although *in silico* predictions of deleteriousness are useful in prioritizing identified variants, functional assays are required in order to draw conclusions regarding the functional impact of these variants. Functional studies of the variants identified in this study are currently underway.

Previous sequencing studies of PCP genes in NTD cases placed emphasis on novel protein-coding variants [68,70,72-74,76-78]. Novel variants are particularly noteworthy because their absence in normal populations is suggestive of high conservation and therefore functional importance. In addition to novel variants, I still considered rare variants, as well as a sub-set of common variants, that are present in normal populations to be relevant in this study for two major reasons. Firstly, the vast majority of novel variants identified in previous NTD sequencing studies where parental DNA was available demonstrated that the variant is also present in a normal parent. This is indicative of reduced penetrance, meaning that even these novel variants,

which are predicted to be more likely deleterious due to their absence in controls, do not always result in a disease phenotype. Based on this logic, it is likely that variants present in control populations can also contribute to NTDs in susceptible individuals. It can be theorized that a NTD susceptibility variant is easily maintained in a normal population with low frequency if it is selected against in those individuals that carry other susceptibility factors and develop a NTD, but allowed to persist in and be passed on by non-penetrant individuals. This theory is in agreement with Juriloff and Harris (2012), who addressed the fact that the majority of novel variants identified in NTD sequencing studies were inherited from a normal parent, meaning that these variants are present in the normal population [261]. They also stated that a focus on very low frequency (<0.5%) or novel variants, which are predicted to have the lowest genetic fitness, would increase the likelihood of identifying highly penetrant variants; however, this would exclude other, more common, variants that contribute to NTDs. The second reason for including known variants in this analysis was because of their potential contribution to compound heterozygosity or homozygosity within a single proband. Such variants would be allowed to persist in a heterozygous state in the normal population, as they may only exert deleterious effects when present in conjunction with another variant-containing allele.

In this study, I identified a combined total of 62 variants of interest in the human cranial NTD cohort, 37 of which were rare (MAF < 0.03) and 11 of which were novel. Of these 62 variants, 48 were shown to be evolutionarily conserved (GERP > 2), and 26 variants were predicted *in silico* to be deleterious by SIFT and/or Polyphen2. None of these variants were *de novo*, meaning that the variant was present in one of the parents of the proband. However, whether or not the two novel variants identified in *RNLS* and *KANK1* were *de novo* could not be determined due to lack of parental DNA samples. This analysis focused strictly on variants that altered the amino acid sequence. Therefore, it is possible that synonymous or UTR variants that affect gene function exist in the genes sequenced in this cohort, meaning that possibly damaging variants could have been overlooked. This is true of other sequencing studies in the literature as well. Fourteen variants with a MAF > 0.03 were also of interest due to the possible contribution of compound heterozygosity or homozygosity in some probands. Compound heterozygous or homozygous cases are particularly remarkable because both alleles of the gene contain a protein-coding variant. In the event that these variants are in fact functionally deleterious, these probands would not have produced any wildtype protein product. Reports of compound heterozygosity in

previous NTD sequencing studies are very low. This may be due to the fact that analyses focused on novel/rare variants, therefore neglecting to account for common nonsynonymous variants in the same proband, and/or due to the lack of parental DNA samples. However, two studies focusing on the PCP gene *CELSR1* did find cases of compound heterozygosity. Allache et al. (2012) identified two compound heterozygous individuals [74] and Robinson et al. (2012) identified three possibly compound heterozygous individuals [69]. In our study, a total of 13 confirmed and 16 possible compound heterozygotes were identified. The vast majority of probands were heterozygous for the variants of interest, particularly for all novel and rare variants. This is similar to what has been found in other sequencing studies performed on NTD genes. However, one proband was homozygous for a rare allele (MAF = 0.017154), and the inclusion of common variants in this analysis identified 13 probands that were homozygous for the alternate allele, 5 of which were homozygous for more than one alternate allele (Table 3.10.2). As NTDs are considered to be multifactorial, it is more likely that compound heterozygosity within a single gene or an accumulation of heterozygous mutations within multiple genes, rather than the extremely rare event of a rare/novel homozygous mutation in a single gene, is what is driving the genetic susceptibility to NTDs. This theory is further corroborated by other studies involving the sequencing of additional PCP genes in the same NTD cohort, which are beginning to identify more patients with mutations in more than one PCP gene [74,78]. Although 62 variants of interest were identified in 18 of the 25 genes sequenced in 156 cranial NTD probands, further evidence for a biological significance of these variants will only come to light once functional analyses have been performed. Expanded sequencing efforts would also aid in determining the relevance of these genes in human NTD etiology.

Sequencing CECR2 in a human cranial NTD cohort identified nine variants of interest

To date, *CECR2* SNVs have not been associated with human NTDs and *CECR2* mutations have not been identified in human NTD cases. This is mainly because *CECR2* was simply not one of the candidate genes that were sequenced in the majority of these studies. The only exceptions to this are two studies that employed a genome-wide approach, which were the whole exome sequencing study performed by Lemay et al. (2015) [66] and the comparative genome hybridization CNV study performed by Chen et al. (2013) [80]. However, even in these studies, anything identified in *CECR2* was not mentioned in the publications. In this thesis, the first

sequencing effort of the coding regions of human *CECR2* in 156 cranial NTD probands identified 9 variants of interest. Only one novel variant was identified, which altered a glutamate to a lysine at position 32 within the functional DDT domain. As the DDT domain is responsible for mediating complex formation with ISWI binding partners, this mutation may affect the interactions between CECR2 and SNF2L or SNF2H, which could be tested in transfected cell lines. Another variant was identified in *CECR2* in the same proband with the novel E32K variant. This other variant altered a proline to a leucine at position 632, and although this is a more common variant with a MAF of ~0.19, it may contribute to compound heterozygosity in this proband. If both the E32K variant and the P632L variant were indeed functionally deleterious and this proband was a compound heterozygote, then this proband would not have had any normally functioning CECR2 protein. The P632L variant is interesting in that despite its high frequency in the population, it was shown to be conserved and was predicted to be damaging. In addition to possibly contributing to compound heterozygosity in the proband containing the novel E32K variant, the P632L variant contributed to compound heterozygosity in four probands and possibly contributed to compound heterozygosity in three other probands. The P632L variant also contributed to homozygosity in an additional three probands, which were homozygous for the alternate leucine amino acid.

Functional assays are required in order to determine if these two variants or any of the additional seven variants of interest affect protein function. Since the E32K variant is within the DDT domain, it would be interesting to determine if this variant affects CECR2 complex formation with SNF2L or SNF2H. One way to test protein-protein interactions is by conducting co-immunoprecipitation assays, which involves isolating CECR2-containing complexes with an α-CECR2 antibody and using western blot analysis to determine if binding partners such as SNF2L or SNF2H are present in these complexes. Robinson et al. (2012) used a similar assay to test whether identified variants in human SCRIB affected protein-protein interactions with binding partners VANGL2 and LGL2, with results demonstrating that these interactions remained intact [69]. Lei et al. (2015) also performed a co-immunoprecipitation assay to test the impact variants identified in human *LRP6* have on the ability of LRP6 to bind MESD. The authors found impaired complex formation for one of the variants identified in *LRP6* [79].

HEK293 cells were chosen as a human cell line to perform CECR2 co-immunoprecipitation analyses as Banting et al. (2005) characterized the CERF complex (CECR2

and SNF2L) in HEK293 cells [141]. I have made constructs for wildtype and variant CECR2 proteins to test this, as well as an EGFP control. In a pilot study described in Appendix C, expression of wildtype *CECR2* alongside expression of *CECR2*-E32K and *CECR2*-P632L in transfected HEK293 cells demonstrated that SNF2L could be co-immunoprecipitated along with CECR2. Conclusions could not be made regarding differences in complex formation because the immunoprecipitation was unable to pull down detectable levels of CECR2 except from HEK293 cells transfected with wildtype *CECR2* (Appendix C). It is unclear as to why CECR2 could not be detected in these immunoprecipitations; optimization of this experiment is required to determine if CECR2 can be detected in α-CECR2 immunoprecipited HEK293 protein. It is important that there are detectable levels of CECR2, as this is needed for comparison between the ability of wildtype and variant CECR2-containing complexes to bring down binding partners such as SNF2L. Results also showed that there was enough endogenous CECR2 in HEK293 cells to pull down detectable amounts of SNF2L in non-transfected or EGFP-transfected cells (Appendix C). The use of a tagged *CECR2* construct in HEK293 cells may overcome this problem; however, native protein is more ideal as the tag may interfere with protein function. Another possibility would be to perform the experiment in a cell line known to not express CECR2, but this cell line would have to express CECR2 binding partners SNF2L or SNF2H, as well as be transfectable. A caveat to this assay is that it may not be sensitive enough to pick up subtle differences in complex formation. A different method to look at CECR2 protein-protein interactions would be the yeast two-hybrid assay, which has been successful for assessing *Vangl1* and *Vangl2* loss-of-function variants identified in human NTD samples [67,68].

Ideal functional assays would strive to analyze a functional output, such as enzymatic activity or phenotype rescue, rather than look at a specific feature of the protein, such as protein complex formation or cellular localization. An assay measuring functional output would be able to tell you whether or not the protein has lost function regardless of what feature of the protein is affected, thereby allowing for conclusions to be made on negative results. For example, if the enzymatic function of a protein variant is similar to wildtype, it is easier to conclude that the variant does not affect protein function. However, if the functional assay addresses only one specific feature of the protein, then conclusions can only be made regarding that specific feature and therefore cannot conclude that the variant is completely benign based on a negative result. This is because although the specific feature being analyzed may not be affected (ie. complex

formation), other features not being assayed may be impaired (ie. DNA binding, cellular localization, etc). This was exemplified in a study by Robinson et al. (2012), which demonstrated that two SCRIB variants functioned normally in binding the SCRIB partners VANGL2 and LGL2 based on a co-immunoprecipitation assay, but were unable to appropriately localize to the plasma membrane [69]. An *in vitro* functional assay for *CECR2* to measure functional output would be an ATP-dependent chromatin remodeling assay. This can be achieved by purifying CECR2-containing protein complexes from a stably transfected cell line and performing the assay as previously described in Banting et al. (2005) [141]. Briefly, this assay involves the incubation of the purified CECR2-containing complex with a DNA nucleosomal array and ATP followed by restriction digest to determine if nucleosome rearrangement occurred. An increase in digested DNA is indicative of nucleosome rearrangement. A caveat to this assay is that it is performed *in vitro*, which is a much simpler system than the complexity of the cellular environment. For example, additional factors within the cell that may alter the efficiency of the CECR2 complex would not be present, and if the CECR2 variant being analyzed affects CECR2 function by altering interactions with these additional factors then this *in vitro* assay would not be able to identify such an impairment. Another caveat is that the chromatin remodeling assay may not be sensitive enough to pick up subtle differences in chromatin remodeling activity between a wildtype CECR2 containing complex and a variant CECR2 containing complex; therefore, one may only be able to identify severe loss-of-function mutations with this assay. Another type of assay to measure functional output would be an *in vivo* assay that would measure the ability of a variant *CECR2* to rescue a mutant phenotype compared to a wildtype *CECR2* rescue. Such assays are regularly performed in zebrafish, and for our purposes would involve knockdown of the zebrafish *CECR2* homologue, analysis of the resulting phenotype, rescue of the phenotype with a wildtype human *CECR2* mRNA construct, and comparison to rescue attempts with a human *CECR2* construct containing the variant of interest. This assay not only relies on zebrafish containing a homologue of the protein of interest, but also that knockdown of the zebrafish homologue produces a score-able phenotype that can be rescued by expression of the human protein. Our collaborators at Duke University, Drs. Nicholas Katsanis and Erica Davis, are currently performing these *in vivo* assays in zebrafish for all 9 variants of interest identified in *CECR2*.

Sequencing candidate modifiers of CECR2 in a human cranial NTD cohort identified variants of interest: DNMNP, MMS19, and TJP2 are strong candidate NTD genes

In the literature, more and more known mouse NTD genes are being implicated in human NTD etiology. This is especially the case for many PCP genes, including but not limited to *VANGL1*, *VANGL2*, *SCRIB*, *PRICKLE1*, *FZD6*, and *CELSR1*. These findings exemplify that the molecular underpinnings of mouse and human neurulation are likely conserved. Therefore, the identification of DNA sequence variants in a human cranial NTD cohort that alter gene function of any of the mouse *Cecr2* candidate modifier genes will further support a role of these genes in both mouse and human neurulation. After the elimination of *Arghap19* as a modifier gene, we sequenced the human homologues of the remaining 24 candidate modifier genes in the human cranial NTD cohort. Several candidate modifier genes of *Cecr2* were of particular interest based on expression differences or protein-coding variants in mice, what was known about gene function, and/or protein-coding variants identified in humans. Mouse *Foxd4* was initially of particular interest because it demonstrated a significant ~2.7-fold reduction in mRNA levels in BALB/cCrl compared to FVB/N (Table 3.6.1), as well as contained 3 protein-coding variants that differed between the two strains (Table 3.7.2). As exencephaly penetrance is much lower in FVB/N compared to BALB/cCrl, as well as C57BL/6 and 129S2, we were interested in finding out if the higher *Foxd4* expression seen in FVB/N was also specific to the FVB/N strain (Figure 3.6.1). However, *Foxd4* expression was not significantly lower in C57BL/6. *Foxd4* expression was significantly lower in 129S2 relative to FVB/N, but expression levels in 129S2 were intermediate between FVB/N and BALB/cCrl. In the event that *Foxd4* was actually a modifier, the similar *Foxd4* expression levels between FVB/N and C57BL/6 would suggest that different modifier genes are contributing to the difference in exencephaly penetrance between these two strains. It would also be difficult to conclude if *Foxd4* modifies exencephaly penetrance in 129S2 based on these results. *Foxd4* encodes for a neural transcription factor, which functions early in the process of neurulation where it is responsible for promoting proliferation within the immature neural ectoderm in order to expand the neural plate [262]. The *Xenopus* homologue of *Foxd4* (*foxd4/5*) has been shown to regulate *zic2* expression [204], a gene that encodes for another neural transcription factor, which results in exencephaly and spina bifida in mice when knocked down [205]. Unpublished data by Lee et al. (2014) found that mouse *Foxd4* also demonstrated evolutionary conservation with *Xenopus foxd4* as ectopic expression of mouse *Foxd4* had the

same effects as ectopic expression of *Xenopus Foxd4* in the developing epidermal lineage of *Xenopus* embryos (reviewed in [237]). As *foxd4* has been shown to regulate *zic2* expression in *Xenopus*, it would be expected that *Zic2* would also show significantly lower expression levels in BALB/cCrl relative to FVB/N; however, according to microarray data, *Zic2* demonstrated a 1.22-fold reduction in BALB/cCrl relative to FVB/N that was not significant (*P*-value = 0.214). This would need to be confirmed by qRT-PCR. The microarray was performed on 11-14 somite mouse embryos, where the neural tube has begun the process of folding and zippering up. As *Foxd4* and *Zic2* play a role in neural plate expansion, which occurs prior to 11-14 somites, it would be interesting to analyze *Foxd4* and *Zic2* expression at an earlier stage of neurulation (~7-10 somites). Although *Foxd4* is an interesting candidate modifier gene in mouse, sequencing in the human cranial NTD cohort did not identify any coding variants of interest in *FOXD4*. Although it is possible that there were variants that alter *FOXD4* expression levels, the human sequencing study design did not allow for robust capture and identification of expression altering variants. Also, where there is only one *Foxd4* in mouse, a primate specific expansion of the *FOXD4* gene resulted in six *FOXD4* paralogues in humans (*FOXD4*, *FOXD4L1*, *FOXD4L3*, *FOXD4L4*, *FOXD4L5*, and *FOXD4L6*) [263], all of which demonstrate high sequence similarity at the nucleotide level (*FOXD4L1* shares 74% identity with *FOXD4*, the other four paralogues share 85% identity with *FOXD4*). In humans, this means that there may be functional redundancy between these paralogues, which would decrease the likelihood of a loss-of-function mutation in *FOXD4* having a negative impact on neurulation. Although *Foxd4* function along with expression differences and protein-coding variants identified in our mice make *Foxd4* a good candidate modifier gene in our mouse model, the lack of variants of interest in the human cranial NTD cohort in conjunction with several closely related human paralogs make *FOXD4* a lower priority candidate NTD gene in humans. Results from the human sequencing data identified variants of interest in 17 of the 24 genes sequenced (Table 3.10.1), and resulted in a shift in focus to three other strong candidate genes, which are discussed below.

After *Arhgap19* and *Foxd4*, the top candidate modifier gene identified in mouse was *Dnmbp*, which was shown to contain two potentially deleterious functional mutations in BALB/cCrl, one of which altered a proline to a leucine within the functional DH domain. DNMBP functions to regulate actin cytoskeleton and vesicle movement in cells. DNMBP contains an N-terminal SH3 domain that binds dynamin, a protein that plays a key role in fission

of endocytic buds, a C-terminal SH3 domain that binds the cytoskeleton associated protein MENA, a BAR domain that binds highly curved membranes typical of sites undergoing endocytosis, and a DH domain that specifically activates the rhoGTPase CDC42 [245]. DNMBP is present in the early developing mouse brain [264], suggesting a possible role in brain formation. Here, I have also shown that DNMBP is present in the neural epithelium at the time of neurulation in mouse (Figure 3.10.1). DNMBP has been well characterized in culture, where knockdown to ~5% of wildtype levels in cell culture resulted in an ~80% decrease in CDC42 activation as well as defective centrosomes and aberrant mitotic spindles, which eventually lead to cell death [207]. CDC42 is a functional component of the PAR polarity complex, which is responsible for reorganizing the cytoskeleton and reorienting the centrosome to allow for cell planar polarizition and migration [265]. Therefore, DNMBP may function within the neural plate by regulating the PAR polarity complex via activation of CDC42, which then contributes to planar cell polarization. DNMBP has also been shown to colocalize with the apical junction protein ZO-1 and to be involved in proper apical junction formation in epithelial cells [208]. Here, I did not see apical localization of DNMBP in the E9.5 mouse neural epithelium. This may be due to several reasons. Two possible reasons could be that DNMBP does not localize to the apical region in the mouse E9.5 neural epithelium, or DNMBP does localize to the apical region but the immunofluorescence assay was not sensitive enough or unable to detect it (ie. apically localized DNMBP may be bound in protein complexes that protect the DNMBP antigen from antibody detection). There are several possible ways that DNMBP may function in the process of neurulation. For example, cranial neural plate folding requires the formation of hinge points through apical constriction of cells. In the bending neural tube, apical junctions are partially disassembled via endocytosis to allow for the mobility required to undergo apical constriction, which has been shown by the movement of ZO-1 and the PAR polarity complex component PAR-3 from apical junctions into the cytosol in the neural hinge points [17]. DNMBP interacts with dynamin [245], a protein with a key role in endocytic fission, and activates CDC42 [207], a core component of the PAR polarity complex that is located at apical junctions. Therefore, DNMBP may also play a role in hinge point formation during neurulation. Also, the DNMBP C-terminal SH3 domain physically interacts with the actin regulatory protein MENA [245], a protein that is involved in neurulation [266,267]. Studies have also shown that knockdown of DNMBP in MDCK cell culture results in impaired ciliogenesis [209], a process that is integral to successful neural tube

closure. A complete knockout of DNMBP in mice has not yet been done, and based on DNMBP function we predict that a complete knockout would likely be an early lethal involving multiple systems.

We identified several protein-coding variants of interest in DNMBP in the human cranial NTD cohort (Table 3.10.4, Figure 3.10.2). One of these variants was a novel mutation that introduced a premature stop-codon within the functional BAR domain. It is probable that the majority of DNMBP transcript from this allele is degraded by nonsense-mediated decay. Transcript that evades degradation and is translated into protein likely produces an aberrant protein product that is non-functional. Therefore, this proband likely only had 50% of normal levels of DNMBP present at the time of neurulation, which, along with other susceptibility factors present in the proband, may have contributed in the development of a cranial NTD. Another novel variant was identified in a different proband that altered an arginine to a glutamine. This variant is highly conserved and is also located within the functional BAR domain. An additional three rare variants were identified in three different probands that are also highly conserved. The remaining three variants were more common in the normal population but contributed to homozygosity and compound heterozygosity in some probands.

Functional assays are required in order to confirm if any of the DNMBP variants identified in mice and humans are in fact deleterious to protein function. *In vitro* CDC42 activation assays comparing wildtype cell culture to DNMBP RNAi knockdown cell culture has successfully been performed in the past [207]; however, such assays involve measuring the amount of activated CDC42 in whole-cell lysates. For our purposes, wildtype and variant DNMBP would be transfected into cell culture, resulting in an over-expression rather than a knockdown of DNMBP. This means that endogenous DNMBP present in cell lysate may interfere with the ability to recognize differences in CDC42 activation efficiencies. One strategy to overcome this would be to find a cell type that has low levels of endogenous DNMBP expression. This cell type would ideally also have moderate to high levels of CDC42 expression, unsaturated CDC42 activation, and should be easily transfectable. Another strategy would be to knockdown endogenous DNMBP mRNA via RNAi targeted to untranslated regions not present in DNMBP mRNA transcribed from the transfected plasmid DNA. A third strategy would be to stably integrate the transfected DNMBP constructs into the genome and then knockout the endogenous DNMBP gene via CRISPR technology. The DNMBP gene constructs would need to be stably

introduced prior to endogenous DNMBP knockout because it has been previously reported that DNMBP knockout in cell culture leads to cell death [207]. In addition to CDC42 activation assays, other protein features can be assessed, such as protein-protein interactions with MENA or sub-cellular localization in Caco-2 epithelial cells. We also intend to perform future *in vivo* functional assays in zebrafish using the same principle as was discussed above for *CECR2* via collaboration with Drs. Nicholas Katsanis and Erica Davis at Duke University.

MMS19 is also a good candidate gene based on protein function and the variants identified within the human cranial NTD cohort. MMS19 is a component of the cytosolic iron-sulfur protein assembly machinery, which functions to facilitate insertion of iron-sulfur clusters into apoproteins involved in methionine biosynthesis, DNA replication and repair, and maintenance of telomeres [210]. Iron-sulfur cluster assembly is a process that occurs in the cytoplasm, and the vast majority of MMS19 has previously been shown to be cytoplasmic [268]. However, MMS19 has also been shown to localize to the nucleus, specifically heterochromatin, during the S-phase [246]. Here, I have shown that MMS19 is expressed in mouse within the neural epithelium during neurulation and observed that the majority of MMS19 is cytoplasmic with a subset of cells demonstrating nuclear localization (Figure 3.10.3), which agrees with what has been found in the literature. Previous studies have also demonstrated that knockout of MMS19 in cell culture resulted in nucleotide excision repair and RNA polymerase II transcription deficiencies [211], and knockout in mouse resulted in embryonic lethality prior to implantation [268].

The single protein-coding variant identified in *Mms19* that differed between BALB/cCrl and FVB/N was not predicted to be highly conserved or damaging, meaning that *Mms19* was included as a candidate gene for sequencing in the human cranial NTD cohort mainly due to protein function. We identified six protein-coding variants of interest in *MMS19* in the cranial NTD cohort (Table 3.10.5, Figure 3.10.4), one of which was a novel mutation that resulted in the introduction of a premature stop-codon within the first half of the protein in the Dos2 interacting domain. As is the case with the majority of nonsense mutations, this mutation is likely a complete loss-of-function mutation due primarily to nonsense-mediated decay in combination with a severely truncated protein product. The proband containing this nonsense mutation was also a confirmed compound heterozygote for another low frequency protein-coding variant that affected a highly conserved amino acid and was predicted to be damaging. An additional two probands were confirmed compound heterozygotes and a third was possibly a compound heterozygote,

with one *MMS19* allele containing a rare, highly conserved variant predicted to be damaging and the other allele containing a common variant that is highly conserved for all three probands. Yet another variant identified in *MMS19* demonstrated an ~2.5-fold enrichment in the human NTD cohort, although this variant is not highly conserved and was not predicted to be damaging. These results taken alongside what is currently known about the importance of *MMS19* during development makes *MMS19* a strong candidate NTD gene. Future functional assays for *MMS19* could be performed in cell culture to assess protein-protein interactions with known binding partners, sub-cellular localization, and nucleotide excision repair capabilities. Endogenous *MMS19* will likely be present in most cell culture lines as *MMS19* is expressed at moderate to high levels in the vast majority of human tissues with the exception of liver, pneumocytes, and hematopoietic cells (www.proteinatlas.org). Therefore, a tagged *MMS19* gene construct can be used to transfect cells for protein interaction and sub-cellular localization assays. A similar strategy as described above for *DNMBP* involving the stable integration of the *MMS19* construct followed by knockout of the endogenous gene can be employed for nucleotide excision repair assays or other phenotypic output assays. We could also perform *in vivo* zebrafish assays in the future.

The third candidate gene identified in the human NTD cohort, *TJP2*, encodes for the tight junction protein TJP2 (also known as ZO-2) that functions in the assembly and maintenance of epithelial and endothelial tight junctions. Epithelial junctions are crucial for maintaining the integrity of the neural epithelium, and later in neurulation epithelial junctions become highly dynamic structures due to apical constriction, which occurs during hinge point formation [17]. We identified two sub-regions in the mouse chromosome 19 modifier region shown to contain at least one modifier gene (Figure 3.1.3) [165]. Where *Dnmbp* and *Mms19* are located within modifier region 3 in mouse, *Tjp2* is located within modifier region 1. It is interesting to note that DNMBP and TJP2 proteins have been shown by co-immunoprecipitation to interact in human Caco-2 epithelial cell culture [208]. This is consistent with the hypothesis that the two modifiers in mouse are involved in the same or related pathways since the modifiers are not additive [165]. TJP2 has also been shown to interact with the protein product of *SCRIB* [269], a gene known to function within the planar cell polarity pathway and is involved in neurulation [270,271]. It has already been established with immunofluorescence that TJP2 protein is present in the apical region of the neural epithelium of neurulating mouse embryos [213]. Homozygous knockout of *Tjp2* in mice

results in early embryonic lethality prior to implantation due to arrest during gastrulation [272]. Knockout embryos displayed reduced proliferation, increased apoptosis, and alterations in the architecture of apical junctions [272].

Tjp2 was included in the candidate modifier gene list mainly due to what was known about protein function and expression. We identified only one protein-coding variant that differed between BALB/cCrl and FVB/N, which was not predicted to be damaging. However, we identified six protein-coding variants of interest in the human NTD cohort (Table 3.10.6, Figure 3.10.5). One variant was novel, affected a highly conserved nucleotide, was predicted to be damaging by both SIFT and Polyphen2, and resulted in the alteration of an isoleucine to a threonine within a functional SH3 protein domain. Future functional assays could be performed in cell culture to analyze protein-protein interactions and sub-cellular localization. There were two probands, 20-0620476 and 20-0521274, with variants identified in both *DNMBP* and *TJP2* (Table 3.10.2). It would be particularly interesting to determine if any of the variants in TJP2 interfere with protein binding to DNMBP, and if any of the variants identified in DNMBP interfere with protein binding to TJP2. We could also perform *in vivo* zebrafish assays in the future.

Once a variant within a gene is shown to have a functional impact, we can further explore the contribution of genetic variants within that gene to human NTD etiology by expanding our sequencing efforts. This would involve sequencing the gene of interest in additional NTD affected individuals. An association of a more common variant and/or a higher burden of rare or novel variants would provide further evidence that this gene contributes to NTD etiology. We would then go back to the mouse model to determine if the gene of interest is involved in neurulation. As the genes sequenced in this study are candidate modifiers of NTDs, performing a BAC transgenic rescue of the BALB/cCrl gene variant by introducing the FVB/N gene variant would test whether or not exencephaly penetrance can be reduced in BALB/cCrl by introducing the putatitive resistant allele. In the event that a mouse knock-out model of the gene of interest has not yet been produced, future work could involve generating such a model to explore the phenotypic effects of a complete knock-out on the process of neurulation in mouse. A combination of functional analyses of the identified variants in the human cranial NTD cohort, an association or burden of variants within the gene of interest in additional human NTD cases, and

establishment that the gene of interest functions in mouse neurulation would provide a strong case for the gene of interests' involvement in human NTD etiology.

Conclusions

The complex, multifactorial nature of NTDs is exemplified in our *Cecr2* mutant mouse model, which demonstrates a reduced penetrance that appears to be changing over time (due to possible genetic background changes or environmental influences) and that varies between mouse strains. This complexity is further demonstrated by the fact that a region on chromosome 19 was shown to contain not just one, but at least two, modifier genes that are capable of rescuing the exencephaly phenotype from ~54% penetrance to ~33% penetrance. Also, there are additional modifier loci outside of the chromosome 19 region as the chromosome 19 modifier loci are only capable of a partial rescue from ~54% to ~33%, and additional suggestive modifier regions on chromosome 2 and the X chromosome have previously been identified by whole genome linkage analysis [164]. Sequencing of *CECR2* and 24 candidate modifier genes of *CECR2* in one of the worlds largest human cranial NTD cohorts identified protein-coding variants of interest; however, more work needs to be done in order to determine if these variants are NTD susceptibility factors. Confirmation of the functional impact of these variants will help to illuminate the role of *CECR2* and candidate modifiers of *CECR2* in the etiology of cranial NTDs in humans. As none of the 24 candidate modifier genes are known NTD genes, determining whether any of these candidates are indeed NTD modifier genes will lead to the identification of a novel NTD gene and aid in unraveling the complex genetic architecture of NTDs both in mice and in humans.

Chapter 4

dikar, the *Drosophila melanogaster* homologue of *Cecr2*, is dispensable for normal development

Preface

Work in this chapter was done in collaboration with Dr. John Locke, who provided the *Drosophila* stocks, reagents, lab space, and assistance with experimental design. Kathryn Wilfong-Pritchard (Biology 499 honors project student under my partial supervision) performed the irradiation experiments on a series of embryo/larva ages with varying irradiation doses (0 gray, 15 gray and 30 gray), as well as aided with data analyses for these experiments. I was responsible for overall study design, all other experiments, and data analyses.

4.1: Introduction

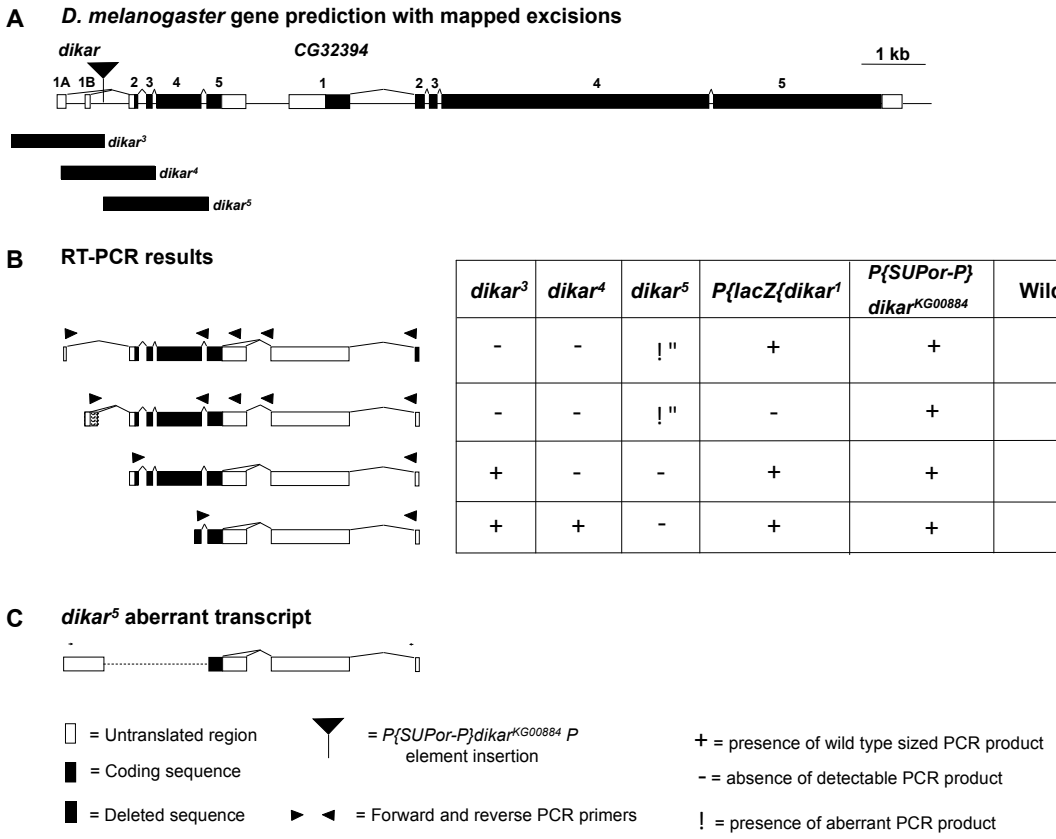
In mice, homozygous mutation in *Cecr2* results in the neural tube defect exencephaly. A homologue of mouse *Cecr2*, named *dikar*, is present in the fruitfly *Drosophila melanogaster*. However, unlike mice, *Drosophila* do not possess a neural tube, therefore making *Drosophila* not ideal for studying the actual process of neurulation. On the other hand, *Drosophila* have proven to be a valuable genetic model for studying the basic molecular underpinnings of conserved developmental genetic pathways involved in mammalian neurulation, including but not limited to Hedgehog signaling [118] and planar cell polarity (PCP) [119]. Here, phenotypic analyses of the *Drosophila Cecr2* homologue, *dikar*, were undertaken in an effort to advance our understanding of the basic molecular function of *Cecr2* (*dikar*). While preliminary results were promising, unfortunately more detailed analysis revealed that a probable second-site mutation prevented useful conclusions regarding *dikar* function in *Drosophila*. However, this research indicates that *dikar* is unlikely involved in PCP or double-strand break (DSB) repair, and thus is presented here.

Keuling et al. (2007) has previously characterized the *Cecr2* homologue, *dikar*, in *Drosophila melanogaster* [186]. What was originally annotated to be two separate genes (*dikar* and *CG32394*) in FlyBase was shown to produce a single transcript, herein referred to as *dikar* (Figure 4.1.1). Keuling et al. (2007) also produced and characterized three *dikar* deletion mutants by imprecise excision of the P element *P{SUPor-P}dikar*^{KG00884} in the first intron of *dikar*. The three mutant alleles were (a) *dikar*^{3}, which contains a deletion of non-coding exons 1A and 1B as well as upstream sequence, (b) *dikar*^{4}, which contains a deletion of non-coding exons 1A and 1B as well as coding exons 2 and 3, and (c) *dikar*^{5}, which contains a deletion of coding exons 2, 3, 4, and a portion of 5 (Figure 4.1.1A). Phenotypic analyses of homozygous mutant flies demonstrated that these flies are homozygous viable and appear morphologically normal.

As human CECR2 has been shown to physically interact with SNF2L, Keuling et al. (2007) also looked for a dominant genetic interaction between *dikar*^{5} and an *Iswi*^{1} single point mutation [273], which are mutant alleles of the fly homologues of *CECR2* and *SNF2L* respectively. No genetic interaction was observed, although this did not exclude the possibility of a physical interaction. A genetic screen by Burgio et al. (2008) in *Drosophila* of ISWI binding proteins identified 255 factors that functionally interacted with ISWI. However, none of these interacting

factors were *dikar* or *CG32394* [274], therefore making the possibility of a physical interaction less likely. According to the online database PFAM (pfam.xfam.org), *dikar* has a predicted DDT domain, and although the prediction E-value does not reach significance, it is closer to consensus than the human *CECR2* DDT domain. Also, *dikar* has a predicted WHIM1 (WSTF, HB1, Itc1p, MBD9 motif 1) domain, which is normally found in conjunction with WHIM2, WHIM3, and DDT domains and interacts with the SLIDE domain of ISWI proteins and nucleosomal linker DNA [275,276].

In summary, phenotypic analyses of the three *dikar* deletion mutations performed by Keuling et al. (2007) demonstrated no observable mutant phenotype and therefore gave no indication as to what the role of *dikar* is in *Drosophila*. The authors suggested that this might be due to compensation by other chromatin remodeling proteins, such as Acf1, which has been shown to interact with ISWI in the chromatin remodeling complexes ACF [155] and CHRAC [277]. Alternatively, reverse transcriptase PCR experiments demonstrated that all three *dikar* intragenic deletion alleles did produce some transcript, with $dikar^{3}$ and $dikar^{4}$ lacking 3'UTR sequence and $dikar^{5}$ displaying a largely aberrant transcript product that contains very little coding region (Figure 4.1.1B,C). The authors also suggested that the aberrant transcripts produced in the mutants might retain enough function for normal development. Based on studies on *Cecr2* in mammals, it is possible that *dikar* plays a role in PCP and/or DNA double-strand break repair, both of which are discussed briefly below.



dikar^{3}	*dikar*^{4}	*dikar*^{5}	*P{lacZ{dikar*^{1}	*P{SUPor-P} dikar*^{KG00884}	Wildtype
-	-	! "	+	+	+
-	-	! "	-	+	+
+	-	-	+	+	+
+	+	-	+	+	+

Figure 4.1.1: Three *dikar* deletion alleles generated by imprecise excision of the P element *P{SUPor-P}dikar*^{KG00884}. *CG32394* has been re-annotated as the C-terminal end of *dikar* (www.ensembl.org). Three independent imprecise excision events generated the *dikar*^{3}, *dikar*^{4}, and *dikar*^{5} mutant alleles (**A**). All five *dikar* alleles (*dikar*^{3}, *dikar*^{4}, *dikar*^{5}, *P{LacZ}dikar*^{1}, *P{SUPor-P}dikar*^{KG00884}) produce transcript (**B**), with *dikar*^{5} producing a largely aberrant transcript (**C**). Image modified from Keuling et al. (2007) [186].

The planar cell polarity (PCP) pathway was mainly characterized in *Drosophila*, but since then, it has been well established that PCP plays an important role in vertebrate neurulation [119]. The PCP pathway is a non-canonical Wnt signaling pathway that functions to orient cells within the plane of an epithelium and to determine the cellular orientation of actin hairs and cilia [119]. Although mutations in the PCP pathway affect all tissues in *Drosophila*, one of the major phenotypic outputs is the misplacement of actin hairs/bristles on the wing, thorax, or abdomen, which results in a swirled pattern of hair growth rather than the normal distally-pointing orientation [278]. There are 31 genes implicated in PCP in *Drosophila* [279]; however, only the six

core PCP proteins will be discussed here. Planar polarity is achieved by the asymmetric distribution of five core PCP proteins, with the three proteins, frizzled (fz), dishevelled (dsh), and diego (dgo, ANKRD6 in mammals), forming a complex at the distal end of the cell, and the two proteins, Van Gogh (Vang) and prickle (pk), forming a complex at the proximal end of the cell (Figure 4.1.2). The sixth core protein, flamingo (fmi, CELSR1 in mammals), is not asymmetrically distributed but is rather implicated in intercellular signal communication between opposing fz and Vang complexes, thereby establishing local alignment of cell neighbors within the same plane [280]. Vertebrate homologues exist for all six core PCP genes, some of which have more than one paralogue, such as for *dsh* (*Dvl1*, *Dvl2*, and *Dvl3*), *fz* (*Fzd3* and *Fzd6*), *Vang* (*Vangl1* and *Vangl2*), and *pk* (*Prickle1* and *Prickle2*) in mouse. *Cecr2* has not been directly linked to PCP in mice as evidenced by no significant misregulation of PCP genes in *Cecr2* mutant embryos [143]; however, mislocalization of PCP proteins in *Cecr2* mutants has not been tested. There is suggestive evidence that *Cecr2* may be indirectly involved in PCP. Spina bifida was an additional phenotype seen in mouse embryos homozygous mutant for the more severe *Cecr2*$^{tm1.1Hemc}$ allele that also contained a heterozygous mutation in the PCP gene *Vangl2*, thereby demonstrating a genetic interaction as spina bifida was not seen in embryos containing only one of these genotypes [143]. Also, mice homozygous mutant for *Cecr2* have misoriented inner ear stereocilia bundles [143], a common phenotype in planar cell polarity mutants.

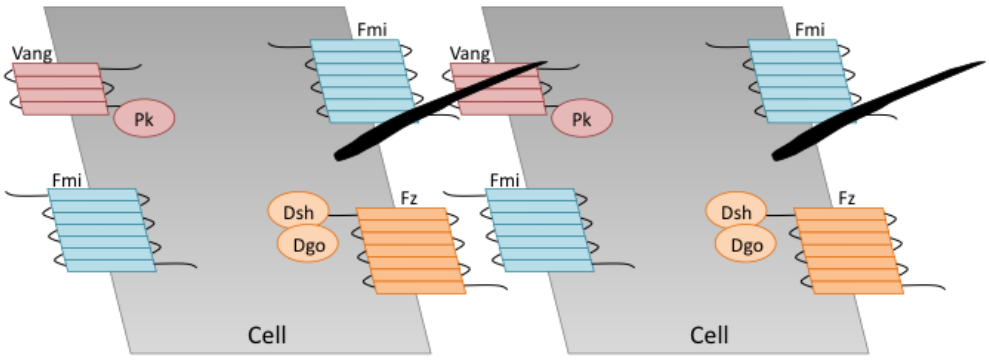


Figure 4.1.2: Distribution of the six core PCP proteins in drosophila wing cells. The Vang-prickle(Pk) complex is located to the proximal side of the cell and the frizzled(Fz)-dishevelled(Dsh)-diego(Dgo) complex is located on the distal side of the cell. Flamingo (Fmi) is located on both sides of the cell. The black actin hairs are pointing distally. Information for this figure was obtained from Seifert & Mlodzik (2007) [281].

Impaired DNA double-strand break (DSB) repair has also been implicated in the etiology of NTDs [282-286], as well as various other disorders including cancer, neurodegeneration, and immune dysfunction (reviewed in [287]). Endogenous DNA DSBs have an estimated rate of one break per 10^8 bp per cell cycle [288]. The introduction of endogenous DSBs arises due to an accumulation of single-stranded lesions during the cell cycle [289], some of which are then converted to DSBs during DNA replication [288]. There are two well-known exogenous sources of DNA double-strand breaks: chemotherapeutic drugs (DNA-alkylating agents, cross-linking agents, radio-mimetic compounds, topoisomerase inhibitors) and ionizing radiation (IR) (reviewed in [290]). Repair of DSBs begin with the recruitment of MRN complex, which then activates the ATM kinase [291]. Upon activation, ATM phosphorylates several substrates, including the variant histone H2AX (H2Av in *Drosophila*), which is called γH2AX (γH2Av) when phosphorylated. This phosphorylation event is one of the earliest markers of the presence of DSBs and the repair response [126]. γH2AX recruits MDC1 [292], which then recruits additional DNA damage response proteins, ultimately resulting in checkpoint activation and temporary cell cycle arrest to allow time for repair. DSBs are normally repaired by the cell via one of two mechanisms, non-homologous end-joining (NHEJ) or homologous recombination (HR). NHEJ functions throughout the cell cycle [293], and involves the processing of broken ends to allow for ligation of the break [294], which can be prone to introducing errors such as base mismatches and small insertions/deletions [295]. HR utilizes the DNA sequence of the sister chromatid as a template for correction of DSBs; however, HR only functions during the S and G2 phase of the cell cycle due to the requirement of a sister chromatid [293]. Decreased cell proliferation as well as increased apoptosis have been shown to result in NTDs (reviewed in [203]). Defective DNA DSB repair results in prolonged cell cycle arrest, which prevents proliferation, as well as increased rates of apoptosis. Homozygous mutation in the homologous recombination repair gene, *Xrcc2*, in mouse results in embryonic lethality with a fraction of mutant embryos displaying neural tube defects [282]. Folic acid deficiency and the antiepileptic drug valproic acid, both of which can cause NTDs, have been shown to result in higher levels of DSBs [285,286]. A study conducted at the Ewha Womans University in Korea demonstrated that siRNA knockdown of human *CECR2* followed by ionizing radiation (IR) in HEK293T cell culture resulted in an increase in unrepaired DSBs, as indicated by comet assay, and a reduced cell survival [144]. The authors also demonstrated a

reduction of γH2AX and 53BP1 foci, which occur in response to the presence of DSBs. These results led the authors to conclude that *CECR2* plays a role in DNA DSB repair.

Elucidating *dikar* function in *Drosophila* and determining if this function is conserved between *Drosophila* and humans would present with a more tractable genetic model for studying the basic molecular function of *dikar* (*Cecr2*). The following experiments will show that *dikar* does not appear to be involved in PCP and that an IR sensitivity phenotype present in homozygous mutants is not due to the $dikar^{3}$ and $dikar^{5}$ mutant alleles.

RESULTS

4.2: RNAi-mediated targeted gene silencing of *dikar* does not result in a planar cell polarity or other overt phenotype

As $dikar^{3}$ and $dikar^{5}$ homozygous mutants appear phenotypically normal and may have residual gene function, an independent method of gene knockdown, RNAi-mediated gene silencing, was utilized to assess phenotypic effects. RNAi gene knockdown can be achieved in a temporal or tissue specific manner via the GAL4/UAS system (Chapter 2, Figure 2.31.1). There are three available *Drosophila* RNAi lines that target *dikar,* all of which were used here to look for any overt phenotypes in adult flies. Specifically, to determine if down-regulation of *dikar* results in defects in planar cell polarity, expression of the genes *dikar*, *Vang*, and *fz* was down-regulated by RNAi-mediated targeted gene silencing. Since the phenotype is well characterized in *Drosophila Vang* and *fz* mutants, RNAi-mediated knockdown of these two genes acted as a phenotypic positive control that the GAL4/UAS RNAi system was functioning and also provided a visual of the typical PCP phenotype. Both *Vang* and *fz* are members of the *fz* signaling pathway, and therefore result in similar phenotypes in *Drosophila* when mutated, which include misorientation of wing-hairs, thorax bristles, abdominal bristles, and eye ommatidia (reviewed in [279,296]). Three different *dikar* RNAi lines, 2 different *Vang* RNAi lines, and 2 different *fz* lines, all of which were under the regulation of the UAS promoter, were each crossed to 8 different GAL4 driver lines. A summary of the expression patterns of GAL4 drivers used in this study is provided in table 4.2.1. The three planar cell polarity mutant phenotypes that were identified were whorls in the wing bristles, misorientation of the small thorax bristles, and misorientation of the abdomen bristles. A subset or all of these phenotypes were seen in the resulting adult progeny

when the four independent RNAi lines, two for *Vang* and two for *fz*, where crossed to 5 of the 8 GAL4 drivers with one exception (Table 4.2.2, Appendix F). The exception was that the two HS-GAL4 drivers did not induce a phenotype in progeny when crossed to the *Vang* RNAi CG8075 7376/GD line (Table 4.2.2). Three of the GAL4 drivers (eyeless-GAL4, T279-GAL4, and prd-GAL4/*TM3*) did not produce any observable planar cell polarity phenotypes in any adult progeny, which may be due to inappropriate spatial, temporal, or level of GAL4 expression. I did not observe any planar cell polarity mutant phenotypes in adult progeny resulting from RNAi targeted gene silencing of *dikar* for all 3 *dikar* RNAi lines and all 8 GAL4 driver lines tested, indicating that *dikar* does not play a major role in PCP. Additional data regarding the frequencies of each phenotype is summarized in Appendix F.

Table 4.2.1: GAL4 driver expression patterns. Only GAL4 drivers used in this study are listed. Two independent Hsp70 and Actin 5C GAL4 drivers were used in this study. Information acquired from FlyBase (www.flybase.org).

GAL4 driver	Spatial expression	Temporal expression
eyeless (ey)	Eye and nervous system	Later stages of embryonic development, larval development
maternal tubulin αTub67C (T279)	Ubiquitous	Throughout embryonic development
paired (prd)	Specific alternating pattern for segment development	Early embryogenesis during segment development, later restricted to embryonic central nervous system
Hsp70 (HS-2, HS-3)	Ubiquitous	Low basal expression, greatly elevated during heat shock
engrailed (en)	Specific polarized pattern in embryo to establish segment polarity and later in larva imaginal disc polarity	Early embryogenesis during segment development
Actin 5C (actin-2, actin-3)	Ubiquitous	Transcript levels vary throughout development

Lethality and male-specific pupal lethality were additional phenotypes observed in progeny from several crosses (Table 4.2.2, Appendix F). However, lethality was seen in a subset

of all crosses, regardless of what gene was knocked down, suggesting that this is an off-target effect of RNAi. It should be noted that nothing could be found in the literature regarding lethal phenotypes in *Drosophila* with a homozygous mutation in *fz* or *Vang*. One possibility is that high levels of GAL4 had a toxic effect that, in some cases, affected males to a greater degree than females.

Since no PCP phenotypes were seen in 3 *dikar* RNAi knockdowns paired with 8 GAL4 driver lines, and the PCP phenotypes were seen in all four PCP control lines, I conclude that *dikar* is not involved in the PCP pathway in *Drosophila*. In addition, no other abnormal phenotype was noted in any of the *dikar* mutants, other than the off-target lethality.

Table 4.2.2: External phenotypes seen in *Drosophila melanogaster* as a result of RNAi-mediated targeted gene silencing of *dikar*, *fz*, and *Vang*.

RNAi / GAL4	CG32394 24738/GD (*dikar*)	CG32394 100383/KK (*dikar*)	CG32393 107312/KK (*dikar*)	CG17697 43077/GD (*fz*)	CG17697 105493/KK (*fz*)	CG8075 100819/KK (*Vang*)	CG8075 7376/GD (*Vang*)
eyeless	×	×	×	×	×	×	◆
T279	×	×	×	×	×	×	×
prd/TM3	×	×	◆	×	×	×	×
HS-2	×	◆	×	◆	◆	◆◆	×
HS-3	×	×	◆	◆	◆◆	◆◆◆	×
en	×	×	◆	◆	◆	◆	◆◆
actin-2/CyO	×	◆	◆	◆◆◆◆	◆	◆◆◆	◆◆◆◆
actin-3/TM6B	◆	◆	◆	◆◆◆	◆	◆◆◆◆	◆◆◆◆

◆ Lethality (not sex specific, 100% penetrant except for RNAi CG8075 7376/GD (*Vang*))
◆ Male-specific pupal lethality (not 100% penetrant)
◆ Whorls in wing bristles (PCP, 100% penetrant in actin-2 and 3-GAL4, and en-GAL4)
◆ Misorientation of small thorax bristles (PCP, 100% penetrant in actin-2 and 3-GAL4)
◆ Misorientation of abdomen bristles (PCP, 100% penetrant in actin-2 and 3-GAL4)
× No apparent phenotype

4.3: Irradiation sensitivity of homozygous *dikar* mutants appears to be due to a second-site mutation

CECR2 function has been implicated in DNA double-strand break repair [144]. I tested two *Drosophila* mutant lines, $dikar^3$ and $dikar^5$, for ionizing radiation (IR) sensitivity. If homozygous *dikar* mutants are indeed deficient in DSB repair, we would expect to see an IR sensitive phenotype. When irradiated as embryos/larvae, this would manifest as a larger proportion of mutant flies dying prior to eclosion compared to their heterozygous siblings. All IR experiments were set up so that a proportion of the progeny were heterozygous (with a balancer chromosome) for the mutation of interest, which should be phenotypically wildtype. This balancer class acted as an internal control and allowed for statistical analyses within each cross.

A pilot IR experiment contained a range of *Drosophila* embryos/larvae (0-7 days post egg laying) and results indicated that flies homozygous mutant for $dikar^5$ were sensitive to IR at 20 gray (P = 9.2E-05), 30 gray (P = 7.1E-06), and 40 gray (P = 0.048, Figure 4.3.1). Progeny ratios for the 0 gray control cross did not show any IR sensitivity in the ratio of homozygous mutant to heterozygous wild type (P = 0.16). An undergraduate student, Kathryn Wilfong-Pritchard, performed further experiments to look at the effects that IR doses 15 gray and 30 gray had on various ages of embryos/larvae (0-2 days old, 2-4 days old, and 4-6 days old) for both $dikar^3$ and $dikar^5$ (Figure 4.3.2). Results showed that no adult flies were produced in crosses where the flies were exposed to IR doses of 15 gray or 30 gray at 0-2 days old for $dikar^3$ (P = 3.7E-35) and $dikar^5$ (P = 4.2E-20), meaning all flies were dying prior to eclosion regardless of genotype. For crosses where flies were irradiated at 2-4 days old, there was a genotype specific IR sensitivity seen for $dikar^3$ at both 15 gray (P = 3.4E-03) and 30 gray (P = 7.4E-06), and for $dikar^5$ at 15 gray (P = 6.1E-03) and 30 gray (P = 8.2E-04). This genotype specific IR sensitivity was also seen for 4-6 day old $dikar^5$ flies at both IR doses (15 gray P = 7.4E-04, 30 gray P = 3.4E-02), but not for $dikar^3$ (15 gray P = 0.72, 30 gray P = 0.20). This difference between the two alleles may be due to a milder phenotype in $dikar^3$ mutant flies. These results indicated that the best timing for IR experiments was 2-4 days, with an exposure between 15-30 gray.

I performed the remaining IR experiments on 2-4 day old larvae, as 0-2 day old larvae all died from the irradiation and 4-6 day old larvae were too resistant. The IR dosage was 22.5 gray, intermediate between the mild affect seen at 15 gray and the more generalized killing seen at 30

gray. These experiments included confirming IR sensitivity in *dikar*^{3} and *dikar*^{5} homozygotes, which was successful in showing that *dikar*^{3} (*P* = 3.7E-08) and *dikar*^{5} (*P* = 1.7E-21) were IR sensitive (Figure 4.3.3).

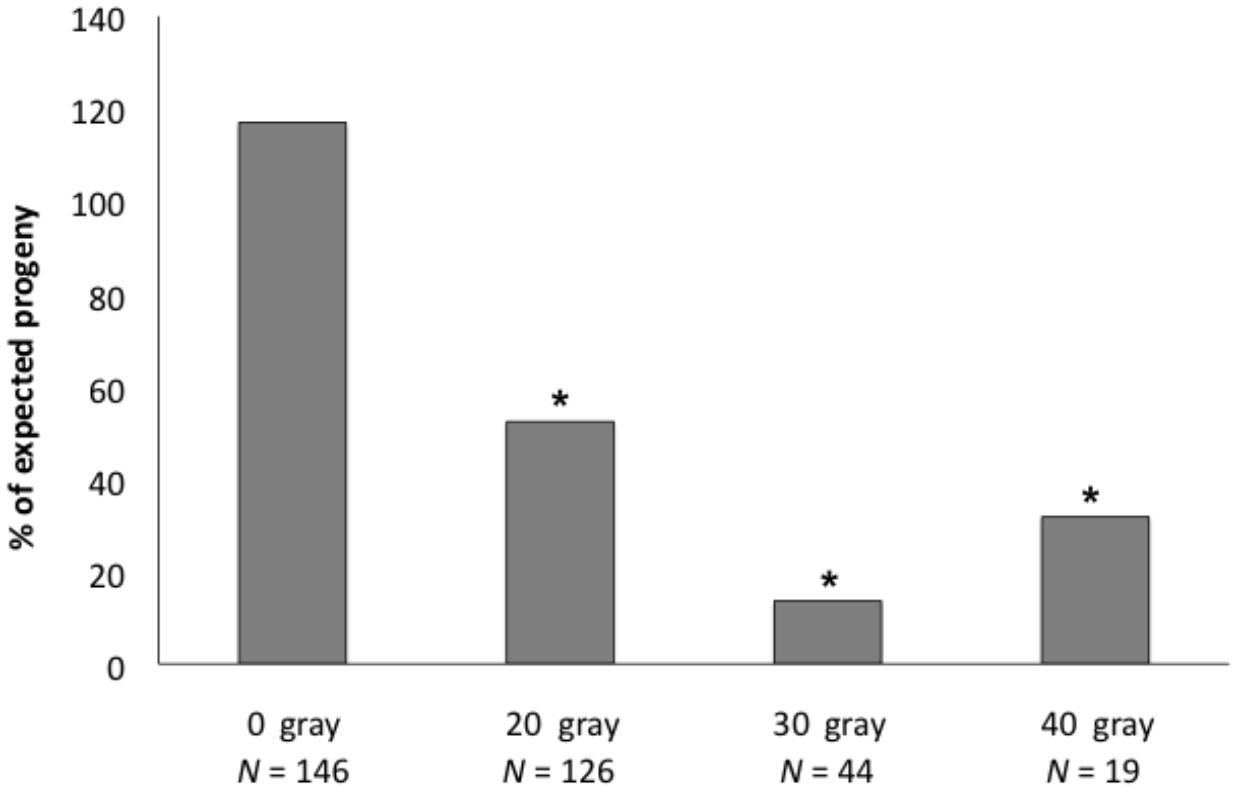


Figure 4.3.1: Flies homozygous for the *dikar*^{5} allele are sensitive to IR doses of 20 gray, 30 gray, and 40 gray. Irradiated vials contained flies ranging from 0 to 7 days old. The asterisks denote IR experiments where homozygous mutant progeny numbers were statistically different from the expected ratio (exact binomial goodness-of-fit test). Total number of flies produced for each experiment (*N*) is displayed under each graph column. Expected progeny is normalized to 100%.

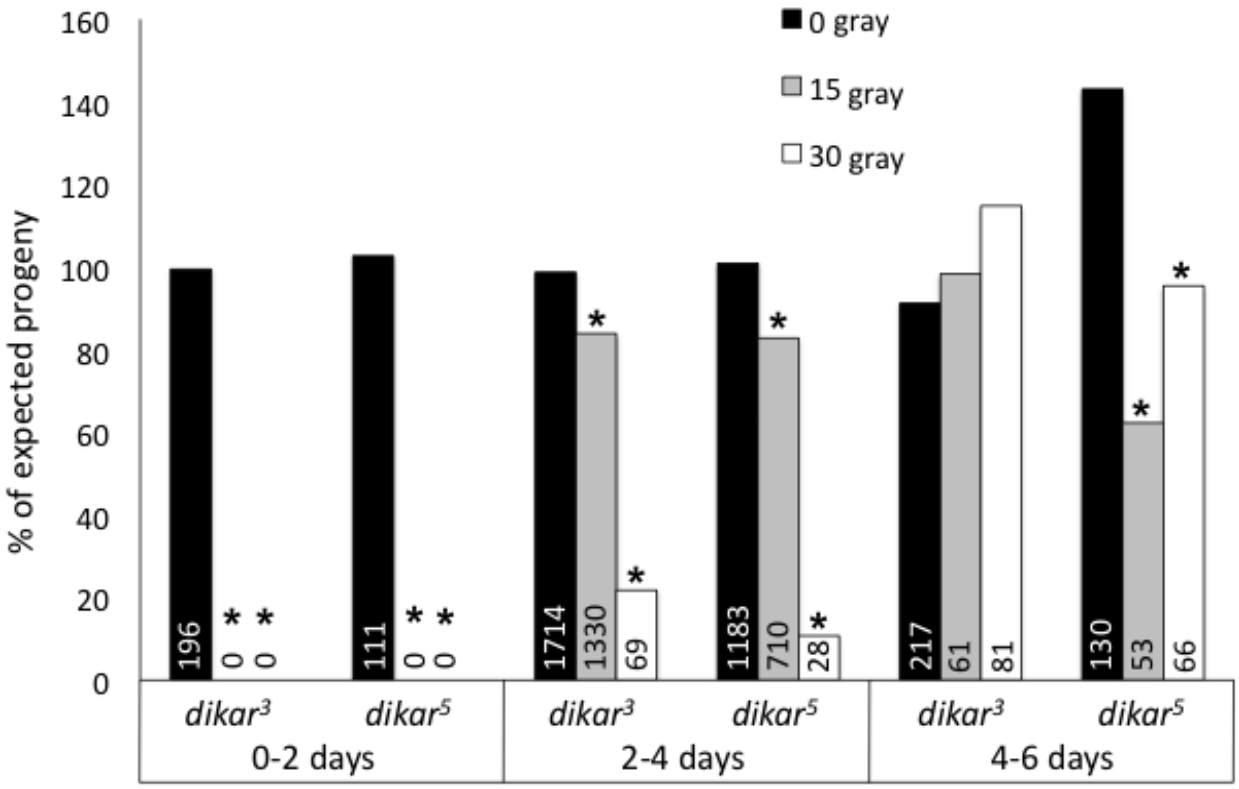


Figure 4.3.2: Irradiation sensitivity for *dikar*^{3} and *dikar*^{5} flies on three ranges of embryo/larva ages (0-2 days, 2-4 days, and 4-6 days) dosed with 0 gray, 15 gray, and 30 gray. The asterisks denote IR experiments at 15 or 30 gray where homozygous mutant progeny numbers were statistically different from the expected ratio (exact binomial goodness-of-fit test for 0-2 days) or statistically different from control experiments at 0 gray (χ^2 test-of-independence for 2-4 days and 4-6 days). Total number of flies produced for each experiment (*N*) is displayed on graph columns. Expected progeny is normalized to 100%.

Next, we tested the *dikar*^{3} and *dikar*^{5} mutations in flies that were heterozygous the *dikar*^{3} or *dikar*^{5} allele and heterozygous for a deletion of the locus. This would test each mutation with a clear null allele as well as localize the IR sensitive locus to within the deletion. The deficiency *Df(3L)ZN47* is predicted to encompass *dikar* as well as an additional 175 genes. Results showed that the *dikar* deletion alleles over the *Df(3L)ZN47* deficiency were also IR sensitive, as would be expected (*dikar*^{3}/*Df(3L)ZN47 P* = 3.8E-27, *dikar*^{5}/*Df(3L)ZN47 P* = 1.1E-11) (Figure 4.3.3). These experiments supported a role for *dikar* in DSB repair. A second deficiency, *Df(3L)ED212* was obtained from Bloomington but found to erroneously contain *dikar*.

As a negative control, the parental *P{SUPor-P}dikar*^{KG00884} line was also tested for IR sensitivity. This fly stock had to be re-ordered from Bloomington; however, the new stock from Bloomington was no longer homozygous viable for the *P{SUPor-P}dikar*^{KG00884} P element. I therefore recreated the line through a series of 3 crosses to the *dikar*^{3}/*dikar*^{3} fly line, and was able to recombine away the second-site recessive lethal. This parental line still contains the *P{SUPor-P}dikar*^{KG00884} P element, and due to the fact that it was crossed to *dikar*^{3}, the genetic background should be at least 87.5% similar to the *dikar*^{3} and *dikar*^{5} lines (even moreso considering that *dikar*^{3} and *dikar*^{5} were generated from this same parental stock several years ago). Also, the introduction of the *TM3* balancer chromosome came from the *dikar*^{3}/*TM3* fly stock, meaning that the experiments performed on the parental line, *dikar*^{3} line, and *dikar*^{5} line should be comparable. Two homozygous viable males were recovered and used to generate two independent parental *P{SUPor-P}dikar*^{KG00884} lines. Results from the two independently generated parental lines (*P{SUPor-P}dikar*^{KG00884}(1) and *P{SUPor-P}dikar*^{KG00884}(2)), which represented the same genetic background as *dikar*^{3} and *dikar*^{5} prior to imprecise excision of the P element, were also sensitive to IR (*P{SUPor-P}dikar*^{KG00884}(1) P = 1.6E-26, *P{SUPor-P}dikar*^{KG00884}(2) P = 8.0E-24). Although we were expecting the parental *P{SUPor-P}dikar*^{KG00884} line to behave similar to a wildtype *dikar* allele, these results suggested that the P element interfered with *dikar* function. Therefore, as a better control, the parental P element was precisely excised in order to re-establish a wildtype allele of *dikar*. Again, the two *dikar*^{WT} independent lines produced by precise excision of the P element showed IR sensitivity (*dikar*^{WT}(1) P = 3.1E-45, and *dikar*^{WT}(2) P = 5.6E-26). Therefore, this suggests that IR sensitivity is resulting from the genetic background itself, such as a second-site mutation, rather than the mutations in *dikar*^{3} and *dikar*^{5}.

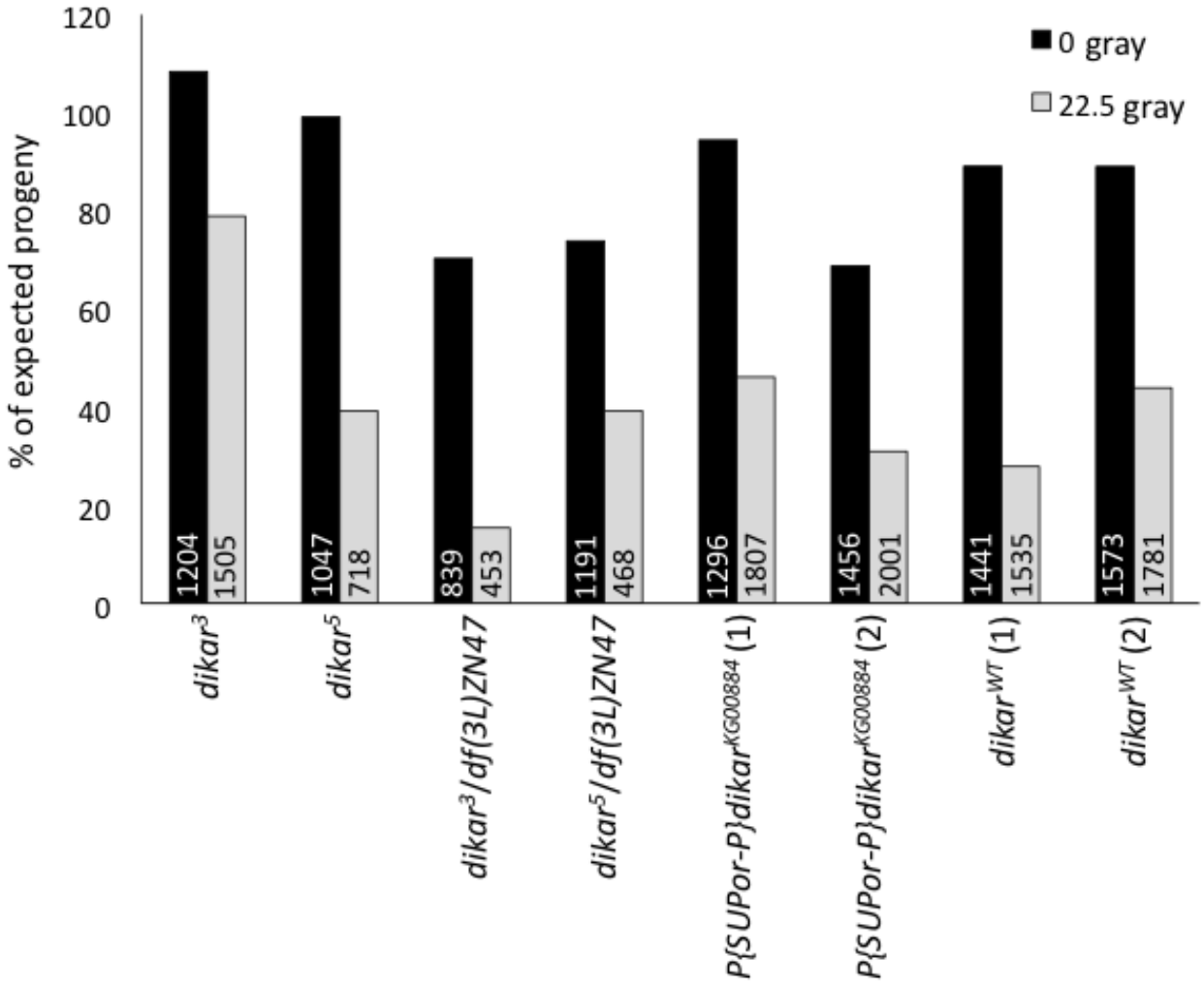


Figure 4.3.3: Irradiation sensitivity in flies homozygous for the *dikar*^{3} and *dikar*^{5} alleles, heterozygous for a *dikar* deletion over a deficiency, as well as for homozygous flies produced in control crosses (*P{SUPor-P}dikar*^{KG00884} and *dikar*^{WT}) at 22.5 gray. Progeny ratios for IR experiments at 22.5 gray were statistically different (χ^2 test-of-independence) from progeny ratios at 0 gray for all crosses. Total number of flies produced for each experiment (*N*) is displayed on graph columns. Expected progeny is normalized to 100%.

4.4: The locus conferring irradiation sensitivity is likely within the chromosomal region encompassed by the *Df(3L)ZN47* deficiency: *Gen* and *dikar* are candidate genes

Since IR sensitivity was similar in homozygous *dikar*^{3} and *dikar*^{5} flies, the *P{SUPor-P}dikar*^{KG00884} parental line, and the precisely excised P element *dikar*^{WT} flies, this suggested that

the phenotype originated from either a second-site mutation in *dikar* or a second-site mutation in another gene on chromosome 3 in this fly strain. The IR sensitivity observed in flies with *dikar*^{3} or *dikar*^{5} over the *Df(3L)ZN47* deficiency suggests that the second-site mutation lies within the bounds of the deletion. I therefore evaluated the genes that mapped within the deficiency for obvious candidates. Gene ontology (GO) analysis of the 176 genes within this region identified 13 genes annotated to bind DNA (Table 4.4.1). Three of these genes were annotated as DNA repair genes, which were *Gen*, *DNApol-ε58*, and *CG7376*. *DNApol-ε58* mainly functions in replication. *DNApol-ε58* also has roles in base excision and nucleotide excision repair (reviewed in [297]) but does not respond to DNA damage in yeast (reviewed in [297]). There is little in the literature regarding the DNA repair activity of *CG7376*. Interestingly, *Gen* is a known DSB repair gene that functions in resolving Holliday junction intermediates produced during HR [298-300]. I therefore elected to test *Gen* and *dikar* for second-site mutations.

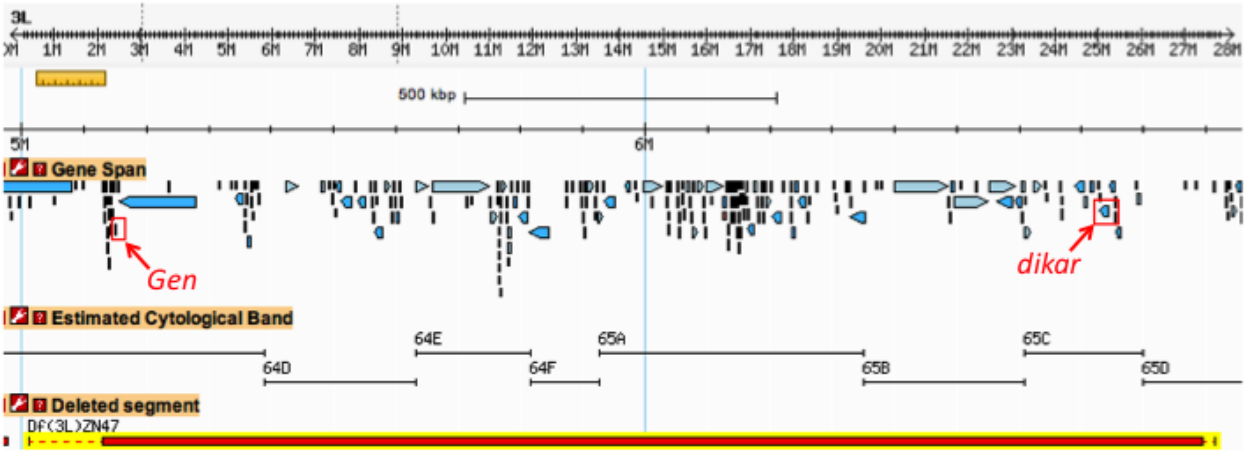


Figure 4.4.1: Genomic region encompassed by the *Df(3L)ZN47* deficiency. Image modified from FlyBase (http://flybase.org/reports/FBab0000006.html).

Table 4.4.1: Gene ontology (GO) analysis of 176 genes within the *Df(3L)ZN47* region identified 13 DNA binding genes. All genes assigned ssDNA binding, DNA repair, DNA metabolic process, or chromatin binding GO terms were also assigned the DNA binding GO term.

Gene	Flybase ID	DNA binding	ssDNA binding	DNA repair	DNA metabolic process	Chromatin binding
Gen	FBgn0263831	Yes	Yes	Yes	Yes	×
DNApol-ε58	FBgn0035644	Yes	×	Yes	Yes	×
CG7376	FBgn0035689	Yes	×	Yes	Yes	×
shep	FBgn0052423	Yes	Yes	×	Yes	×
dikar	FBgn0261934	Yes	×	×	×	Yes
Rcc1	FBgn0002638	Yes	×	×	×	Yes
Blimp-1	FBgn0035625	Yes	×	×	×	×
vvl	FBgn0086680	Yes	×	×	×	×
Ets at 65A	FBgn0005658	Yes	×	×	×	×
CG10576	FBgn0035630	Yes	×	×	×	×
CG10147	FBgn0035702	Yes	×	×	×	×
CG8398	FBgn0035708	Yes	×	×	×	×
CG13287	FBgn0035643	Yes	×	×	×	×

The coding exons and splice sites of both *dikar* and *Gen* were Sanger-sequenced in heterozygous $dikar^{WT}/TM3$ (not IR sensitive) and homozygous $dikar^{WT}/dikar^{WT}$ (IR sensitive) flies. Analyses focused on DNA sequence variants that were heterozygous and that altered the amino acid sequence in the $dikar^{WT}/TM3$ flies. Sequence of the $dikar^{WT}/dikar^{WT}$ flies elucidated what variant allele belonged to the $dikar^{WT}$ chromosome, which would be a candidate IR sensitivity variant. The other variant would therefore belong to the *TM3* balancer chromosome.

I identified a total of 11 protein-coding variants in *Gen* that differed between the $dikar^{WT}$ chromosome and the *TM3* chromosome (Figure 4.4.2). Gen contains 726 amino acids, meaning that the two alleles differ by 11/726 amino acids, or 15.15 differences per 1000 amino acids. This appears to be a particularly high number of nonsynonymous variants between two *Drosophila* lines. I compared the number of nonsynonymous variants in *Gen1*, the mouse homologue of *Gen*, in all mouse strains on the MGI database (http://www.informatics.jax.org/). In 17 strains there were a total of 7 nonsynonymous variants in mouse *Gen1* (protein product = 908 amino acids, 7.7 differences per 1000 amino acids), where no two strains contained more than 3 differences when compared to each other (3.3 differences per 1000 amino acids). It is difficult to draw any conclusions regarding the biological significance of the high number of nonsynonymous variants

that differ between the two *Drosophila Gen* alleles, as no sequence analyses between fly stocks are available.

Figure 4.4.2: Gen protein sequence alignment shows 11 Gen amino acids that differ between the *Gen* allele on the *dikar*^{WT} chromosome and the *Gen* allele on the *TM3* chromosome. Amino acids that differ between the two chromosomes are indicated in green.

```
Gen (dikarWT)    1 MGVKELWGVLTPHCERKPINELRGKKVAIDLAGWVCESLNVVDYFVHPRHHLKNLFFRTC
Gen (TM3)        1 MGVKELWGVLTPHCERKPINELRGKKVAIDLAGWVCESLNVVDYFVHPRHHLKNLFFRTC
                   ************************************************************

Gen (dikarWT)   61 YLIWEQVTPVFVLEGVAPKLKSQVIAKRNELQFRGVKPKNSPECTQSQPSKGDKGRSRFN
Gen (TM3)       61 YLIWEQVTPVFVLEGVAPKLKSQVIAKRNELQFRGVKPKNSPECTQSQPSKGDKGRSRFN
                   ************************************************************

Gen (dikarWT)  121 HVLKQCETLLLSMGIQCVQGPGEAEAYCAFLNKHGLVDGVISQDSDCFAYGAVRVYRNFS
Gen (TM3)      121 HVLKQCETLLLSMGIQCVQGPGEAEAYCAFLNKHGLVDGVISQDSDCFAYGAVRVYRNFS
                   ************************************************************

Gen (dikarWT)  181 VSTQGAQAASGGAVDIYDMREITSRMDFGQQKIIVMALLCGCDYCPDGIGGIGKDGVLKL
Gen (TM3)      181 VSTQGAQAAAGGAVDIYDMREITSRMDFGQQKIIVMALLCGCDYCPDGIGGIGKDGVLKL
                   ********* **************************************************

Gen (dikarWT)  241 FNKYKETEILDRMRSWRGETDKYNALEIRVDDKSICSNCGHIGKTQSHTKSGCSVCRTHK
Gen (TM3)      241 FNKYKETEILDRMRSWRGETDKYNALEIRVDDKSICSNCGHIGKTQSHTKSGCSVCRTHK
                   ************************************************************

Gen (dikarWT)  301 GCDESLWKEQRLSIKSELTLRRKALLSPDFPNEEIIAEFLSEPDTIPNLNLNWRQPNLVK
Gen (TM3)      301 GCDESLWKEQRLSIKSELTLRRKALLSPDFPNEEIIAEFLSEPDTIPNLNLNWRQPNLVK
                   ************************************************************

Gen (dikarWT)  361 FIKQIGHLLQWPEIYCFQKFFPILTRWQVQQSKQEKILIQPHEIIKKRTVKGVPSLELRW
Gen (TM3)      361 FIKQIGHLLQWPEIYCFQKFFPILTRWQVQQSKQEKILIQPHEIIKKRTVKGVPSLELRW
                   ************************************************************

Gen (dikarWT)  421 HDPSGIFKGLIPDKQIAEYEAEHPKGIEELYYTIEPLDMLETAYPDLVAAFLKSKEKPAK
Gen (TM3)      421 HDPSGIFKGLIPDKQIAEYEAEHPKGIEELYYTIEPLDMLETAYSDLVAAFLKSKEKPAK
                   ****************************************** ***************

Gen (dikarWT)  481 KTTRKKKTASEEENKENEPNSKPKRVVRKIKAQPEENQPLLHQFLGRKKEGTPVKAPAPQ
Gen (TM3)      481 KTTRKKKTASEEENKENEPNSKPKRVVRKIKAQPEENQPLLHQFLGRKKEGPPVKAPAPQ
                   ************************************************** ********

Gen (dikarWT)  541 RQQCSTPITKFLPSDLESDCDAEEFDMSDIVKGIISNPNATPALTNHDGHQLHYEPMAED
Gen (TM3)      541 RQQCSTPITKFLPSDLESDCDAEEFDMSDIVKGIISNPNAKPALTNHDGHQLHYEPMAED
                   ************************************* ******************

Gen (dikarWT)  601 LSLRLAQMSLGNVYESPKVETKRDLSQVDQLPQSKRFSLEDSFDLLVKGDLQKLARTPVE
Gen (TM3)      601 LSLRLAQMSLGNVDESPKVETKRDLSQVDQLPQSKRFSLEDSFDLLVKGDLQKLARTPVE
                   ************* ********************************************

Gen (dikarWT)  661 RFKMQHRISEKIPTPVKPLDNISYFFNQSSDNADVFEELMNSSLVPQDQEDNDGEEEDDD
Gen (TM3)      661 RFKMQHRISEKIPSPVKPLANISYFFNQSSDNADVFEELMNSSLVPQDQEDNAEDEEEDD
                   ************* ***** ********************************   ** **

dikarWT        721 LVVISD
Gen (TM3)      721 LVVISD
                   ******
```

As we would predict that *Gen* variants on the IR sensitive *dikar*^{WT} chromosome would be damaging relative to *Gen* variants on the *TM3* chromosome (not IR sensitive), SIFT scores were generated to compare Gen protein variants on the *dikar*^{WT} chromosome relative to Gen protein variants on the *TM3* chromosome. SIFT predicted that 8 of the 11 variants on the *dikar*^{WT} chromosome are tolerated relative to the *TM3* chromosome (Table 4.4.2). The remaining 3 variants were predicted to be damaging; however, these predictions were low confidence due to low diversity of the sequences used at this position. SIFT scores were also generated for the reverse comparison, which was the Gen protein variant on the *TM3* chromosome relative to the Gen protein variant on the *dikar*^{WT} chromosome. Only one variant on the *TM3* balancer chromosome was predicted to be deleterious, which was the presence of a serine residue in lieu of a proline residue.

Table 4.4.2: Protein predictions of deleteriousness of the 11 variants in the Gen protein. SIFT analyses compared variants on the *dikar*^{WT} chromosome relative to the *TM3* chromosome, and vice versa.

Amino acid		***dikar*^{WT} chr. relative to *TM3* chr.**	**SIFT score**	***TM3* chr. relative to *dikar*^{WT} chr.**	**SIFT score**
***dikar*^{WT} chr.**	***TM3* chr.**	Predicted deleteriousness of…		Predicted deleteriousness of…	
S	A	S	0.40	A	1.00
P	S	P	1.00	S	**0.05**
T	P	T	1.00	P	0.09
T	K	T	0.24	K	1.00
Y	D	Y	**0.01***	D	0.53
T	S	T	0.58	S	1.00
D	A	D	**0.02***	A	1.00
D	A	D	0.63	A	0.41
G	E	G	**0.05***	E	1.00
E	D	E	1.00	D	0.42
D	E	D	0.14	E	1.00

Asterisk - Low confidence prediction due to insufficient sequence diversity

As there were 11 amino acid differences within Gen, it made sense to analyze the protein as a whole rather than analyze each variant individually. I used the web-based program I-TASSER to compare predicted protein structures between the two alleles. As part of the secondary structure prediction, I-TASSER assigned each amino acid as an α-helix, β-strand, or coil along with a number between 0-9 that denoted prediction confidence, where lower numbers

correspond to lower confidence. The secondary structure prediction also assigned each amino acid with a number ranging from 0-9 that denotes solvent accessibility, with lower numbers corresponding to buried residues and higher numbers corresponding to more exposed residues. These secondary structure analyses are displayed for the amino acids that differ in Gen between the *dikar*^{WT} chromosome and the *TM3* chromosome in Table 4.4.3. The local secondary structure prediction assignments of α-helix, β-strand, or coil assignments around all 11 amino acids (± 20 amino acids) were similar between the two protein variants, with 14 differences, all of which had low confidence predictions (≤ 4). There were some differences in the assignment of α-helix, β-strand, or coil secondary structure predictions between the two protein variants that were at least 20 amino acids away from the 11 amino acid variants; however, amino acids within these regions were also assigned low confidence values (≤ 4). Overall, there were no obvious differences in the predictions of secondary structure for the Gen protein variants.

Table 4.4.3: Secondary structure predictions for the 11 amino acids that differ in Gen between the *dikar*^{WT} chromosome and the *TM3* chromosome. Analysis was done using the web-based software I-TASSER. Confidence predictions range from 0 (low) to 9 (high). Solvent accessibility ranges from 0 (buried) to 9 (highly exposed).

Amino acid		α-helix, β-sheet, or coil		Confidence		Solvent accessibility	
dikar^{WT} chr.	*TM3* chr.	*dikar*^{WT} chr.	*TM3* chr.	*dikar*^{WT} chr.	*TM3* chr.	*dikar*^{WT} chr.	*TM3* chr.
S	A	coil	coil	4	5	4	4
P	S	coil	coil	6	7	2	2
T	P	coil	coil	8	8	5	5
T	K	coil	coil	8	8	4	4
Y	D	coil	coil	4	3	6	6
T	S	coil	coil	8	8	4	3
D	A	coil	coil	1	3	4	4
D	A	coil	coil	5	7	7	4
G	E	coil	coil	4	7	6	4
E	D	coil	coil	6	6	5	6
D	E	coil	coil	2	5	4	4

Prediction analysis of tertiary structure was also done using I-TASSER. Five models for each protein variant were generated, each with a confidence score ranging from [-5,2] where a higher score indicates higher confidence of the prediction. Model 1 for the Gen protein variant on the *dikar*^{WT} chromosome and model 2 for the Gen protein variant on the *TM3* chromosome

appeared to have a similar overall topology to each other (Figure 4.4.3); however, all tertiary structure models generated for the two Gen protein variants had confidence scores less than -3.

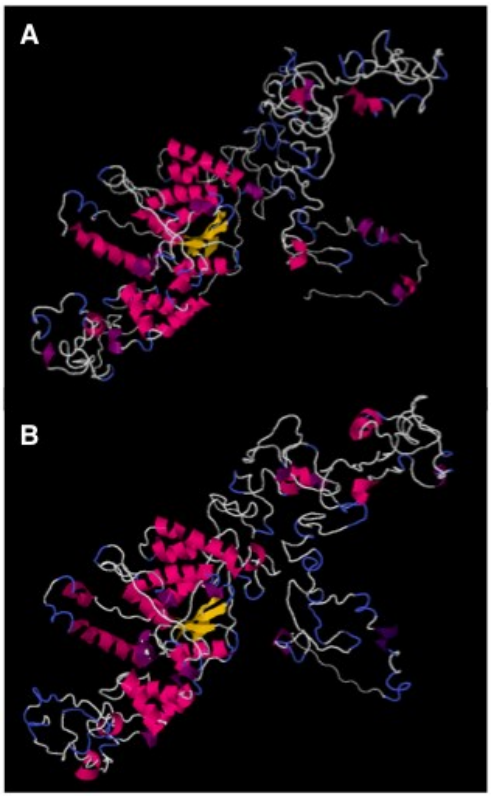


Figure 4.4.3: The most comparable Gen protein prediction models of tertiary structure between the two protein variants. Predictions of tertiary structure were generated by I-TASSER. The Gen protein variant on the *TM3* chromosome is depicted in (**A**), and the Gen protein variant on the *dikar*^{WT} chromosome is depicted in (**B**). All predictions of tertiary structure had low confidence scores (< -3).

I identified four protein-coding variants in *dikar* that differed between the *dikar*^{WT} chromosome and the *TM3* chromosome, one of which was an Asparagine insertion in a poly-Asparagine region (Figure 4.4.4). Dikar contains 3232 amino acids (3231 amino acids for the *TM3* chromosome allele), meaning that the two alleles differ by 4/3232 amino acids, or 1.2 differences per 1000 amino acids. Therefore, Dikar is much more similar at the amino acid level between the two chromosomes compared to Gen (15.15 differences per 1000 amino acids). I compared only a portion of the two Dikar protein sequences with I-TASSER, as the web-based software does not accept submissions greater than 1500 amino acids (indicated in Figure 4.4.4).

Figure 4.4.4: Dikar protein sequence alignment shows 4 Dikar amino acids that differ between the *dikar* allele on the *dikar*^{WT} chromosome and the *dikar* allele on the *TM3* chromosome. The amino acid insertion/deletion is indicated in blue and the amino acid substitutions are indicated in green. Region in bold black text was analyzed by I-TASSER.

```
dikar (dikarWT)          1 MVSTFIGLLDIQSWWEIPCISHFCSLFSSAFDLPDIDIEDLESALLSDGTSDEDQVALVP
dikar (TM3)              1 MVSTFIGLLDIQSWWEIPCISHFCSLFSSAFDLPDIDIEDLESALLSDGTSDEDQVALVP
                           ************************************************************

dikar (dikarWT)         61 ELIVRLLKGCDALQTVAKEITHSNYQMFLRRFLRQQCRLHQTENHFDTDIDFQSLPVRKR
dikar (TM3)             61 ELIVRLLKGCDALQTVAKEITHSNYQMFLRRFLRQQCRLHQTENHFDTDIDFQSLPVRKR
                           ************************************************************

dikar (dikarWT)        121 LHILHDLCHFRLDSVDVQVILSNLEADSLRVEPLGYDAKNSGYWYFYGTRLYREDKATGG
dikar (TM3)            121 LHILHDLCHFRLDSVDVQVILSNLEADSLRVEPLGYDAKNSGYWYFYGTRLYREDKATGG
                           ************************************************************

dikar (dikarWT)        181 SASSSSGSGSGGAGASTGGASNSKGIGSNGTVWQVICFTEEDWQNLAAKFKTSTNAKERE
dikar (TM3)            181 SASSSSGSGSGGAGASTGGASNSKGIGSNGTVWQVICFTEEDWQNLAAKFKTSTNAKERE
                           ************************************************************

dikar (dikarWT)        241 LFNILDDNFLPKLPQLFRERERLRRRKLLQVRNSSRIRNIAELKARQEEERQRRERENLY
dikar (TM3)            241 LFNILDDNFLPKLPQLFRERERLRRRKLLQVRNSSRIRNIAELKARQEEERQRRERENLY
                           ************************************************************

dikar (dikarWT)        301 QDRQNAHWLAPPQQQQQQQQQQQQQQQQSQQRTRRRSQSTSTASGISTYASSLRRTTTTN
dikar (TM3)            301 QDRQNAHWLAPPQQQQQQQQQQQQQQQQSQQRTRRRSQSTSTASGISTYASSLRRTTTTN
                           ************************************************************

dikar (dikarWT)        361 SSEFLCHRETFCLSPENDEDDDDDEALDAEQQSDQPEQQRQQEEQPPEEEQRQEPEELL
dikar (TM3)            361 SSEFLCHRETFCLSPENDEDDDDDEALDAEQQSDQPEQQRQQEEQPPEEEQRQEPEELL
                           ******************************************************************

dikar (dikarWT)        421 AVSATDDDFRSPEAPKSQTPAELKSKSSHHHHHHHHKKKKSSKKQKKRHHRDKGRERDRE
dikar (TM3)            421 AVSATDDDFRSPEAPKSQTPAELKSKSSHHHHHHHHKKKKSSKKQKKRHHRDKGRERDRE
                           ************************************************************

dikar (dikarWT)        481 HHHHHRRRHKHASEADEGEDDNNNHNSNSNSNSNTNSISSNNSVQPKRRRSGQQQQHYQQ
dikar (TM3)            481 HHHHHRRRHKHASEADEGEDDNNNHNSNSNSNSNTNSISSNNSVQPKRRRSGQQQQHYQQ
                           ************************************************************

dikar (dikarWT)        541 AETTEPSTSLSPEDKENFCRQLQLLDTNSAAIAVATSIADLSLPSPVAHESETEQDEQVD
dikar (TM3)            541 AETTEPSTSLSPEDKENFCRQLQLLDTNSAAIAVATSIADLSLPSPVAHESETEQDEQVD
                           ************************************************************

dikar (dikarWT)        601 QPAVVEAQPPALTSAMAPAANVKLPGRQTNNSLSSLTGNIFIPSGSAQQQQQQSQEPPSS
dikar (TM3)            601 QPAVVEAQPPALTSAMAPAANVKLPGRQTNNSLSSLTGNIFIPSGSAQQQQQQSQEPPSS
                           ************************************************************

dikar (dikarWT)        661 SSSSQSGTTTTASTVRSKKQKAVLNSSGSSSTSKMADKADRNLSKTETTGKSKKSAAGSS
dikar (TM3)            661 SSSSQSGTTTTASTVRSKKQKAVLNSSGSSSTSKMADKADRNLSKTETTGKSKKSAAGSS
                           ************************************************************

dikar (dikarWT)        721 ASSSSSKAERISGAYSDKSGDDHVNSSNRSTTNNNSSLNNNNNHKHAPHHSHNNNNNNSH
dikar (TM3)            721 ASSSSSKAERISGAYSDKSGDDHVNSSNRSTTNNNSSLNNNNNHKHAPHHSHNNNNN-SH
                           ********************************************************** **

dikar (dikarWT)        781 SMTTSATTTATTTTSQGYHYNSNHNHNQNHQNSHSHHTHHHHHHHHHTTNNHSNKHKTKS
dikar (TM3)            780 SMTTSATTTATTTTSQGYHYNSNHNHNQNHQNSHSHHTHHHHHHHHHTTNNHSNKHKTKS
                           ************************************************************
```

```
dikar (dikarWT)      841 HHSLAHNNNNNNNPTTNPTSTSLLHPVTTNNNSTYNSSNHANILSFTETEEVLQIGMHKV
dikar (TM3)          840 HHSLAHNNNNNNNSPTTNPTSTSLLHPVTTNNNSTYNSSNHANILSFTETEEVLQIGMHKV
                         ************ ********************************************

dikar (dikarWT)      901 LVYVKNHRDAWPFVDPVEEDIAPRYYSIIRRPMDLLKMEDKLDSGEYHKFSEFRNDFRLI
dikar (TM3)          900 LVYVKNHRDAWPFVDPVEEDIAPRYYSIIRRPMDLLKMEDKLDSGEYHKFSEFRNDFRLI
                         ************************************************************

dikar (dikarWT)      961 VNNCRLYNGHNNEYTEMVNNLQDAFEKATKKYFDNLSDDEDDDPNLSYPAADSKMNVFRE
dikar (TM3)          960 VNNCRLYNGHNNEYTEMVNNLQDAFEKATKKYFDNLSDDEDDDPNLSYPAADSKMNVFRE
                         ************************************************************

dikar (dikarWT)     1021 KYFSKKAKEETEKDAPGRPAVSSAEEDLSEIEAEAPQKAQKRKRKEKDKRRKKKTKSKAD
dikar (TM3)         1020 KYFSKKAKEETEKDVPGRPAVSSAEEDLSEIEAEAPQKAQKRKRKEKDKRRKKKTKSKAD
                         ************** ********************************************

dikar (dikarWT)     1081 VETDDEDMEAEREPTPPPPPPTSKKSKTSKTGKEKEKDKEKEKDKDKEKDKETSSFKRGR
dikar (TM3)         1080 VETDDEDMEAEREPTPPPPPPTSKKSKTSKTGKEKEKDKEKEKDKDKEKDKETSSFKRGR
                         ************************************************************

dikar (dikarWT)     1141 KTKSDKSASKSSKKTKKGAKKSSADSDPESDPSDSRESEDYSDDDHISLAKTKSLVKPTA
dikar (TM3)         1140 KTKSDKSASKSSKKTKKGAKKSSADSDPESDPSDSRESEDYSDDDHISLAKTKSLVKPTA
                         ************************************************************

dikar (dikarWT)     1201 RTIAAQKKKSVPAESKVKMPTPVKRQVKGKGKGGRKAKDDSLDSDQSDVNVKKQLPPTAA
dikar (TM3)         1200 RTIAAQKKKSVPAESKVKMPTPVKRQVKGKGKGGRKAKDDSLDSDQSDVNVKKQLPPTAA
                         ************************************************************

dikar (dikarWT)     1261 AALAESAAELEEDPDDPPPDEDEDEDSSRSRSMSPFKVDLHKKYSKSALNDDLSELLTTV
dikar (TM3)         1260 AALAESAAELEEDPDDPPPDEDEDEDSSRSRSMSPFKVDLHKKYSKSALNDDLSELLTTV
                         ************************************************************

dikar (dikarWT)     1321 KKVPTAETTKLSARHQDEADEERSSRESDGDFKSLSNSRASSEERPPVAKKGKKAESSKK
dikar (TM3)         1320 KKVPTAETTKLSARHQDEADEERSSRESDGDFKSLSNSRASSEERPPVAKKGKKAESSKK
                         ************************************************************

dikar (dikarWT)     1381 EKEKKGRDKDRDRDKEKDKDKDKSRSAKDKKSKDESPVLVAAAEAAAVVQAELDMLLPFM
dikar (TM3)         1380 EKEKKGRDKDRDRDKEKDKDKDKSRSAKDKKSKDESPVLVAAAEAAAVVQAELDMLLPFM
                         ************************************************************

dikar (dikarWT)     1441 DKYDVIKYRRSRAAFSGSSASNSLAPSEDSKSASTKSNRENKKASAKREKSPDAVENKRG
dikar (TM3)         1440 DKYDVIKYRRSRAAFSGSSASNSLAPSEDSKSASTKSNRENKKASAKREKSPDAVENKRG
                         ************************************************************

dikar (dikarWT)     1501 RKSKDQKRSKESDKAEKSDKASKADTEKHSEKSKKKEEPPKVVEKAPREKSPAPVSALET
dikar (TM3)         1500 RKSKDQKRSKESDKAEKSDKASKADTEKHSEKSKKKEEPPKVVEKAPREKSPAPVSVLET
                         ********************************************************* ***

dikar (dikarWT)     1561 IKEPPAPTPIATSASGKVKEGPAKKEPKKRPDKQMPPPPKPADKSSEKGSGKKSSSKKAA
dikar (TM3)         1560 IKEPPAPTPIATSASGKVKEGPAKKEPKKRPDKQMPPPPKPADKSSEKGSGKKSSSKKAA
                         ************************************************************

dikar (dikarWT)     1621 QKAGQPQTNNNTNLEALDVETEQTLKDINRWLEHTPRFTEFSSASNSPSRYNLLDDFDSG
dikar (TM3)         1620 QKAGQPQTNNNTNLEALDVETEQTLKDINRWLEHTPRFTEFSSASNSPSRYNLLDDFDSG
                         ************************************************************

dikar (dikarWT)     1681 IGSKLDAADFRRPVALAAPKAELVPTKLAEALNELVSEPKEAVSKSESESVAPTPSSVSS
dikar (TM3)         1680 IGSKLDAADFRRPVALAAPKAELVPTKLAEALNELVSEPKEAVSKSESESVAPTPSSVSS
                         ************************************************************

dikar (dikarWT)     1741 SCGTPPHSMHSGNSIGSTSATAASSSNCSNNSMPTPLPVAVTPTPTPAPPPLLPIPKPKE
dikar (TM3)         1740 SCGTPPHSMHSGNSIGSTSATAASSSNCSNNSMPTPLPVAVTPTPTPAPPPLLPIPKPKE
                         ************************************************************
```

```
dikar (dikarWT)     1801 PSTTQLILNPPPPPHIKQQLAKEAKRKSLKEKQAAQAAQAQQVKAKANVMRTIDRLQPGK
dikar (TM3)         1800 PSTTQLILNPPPPPHIKQQLAKEAKRKSLKEKQAAQAAQAQQVKAKANVMRTIDRLQPGK
                         ************************************************************

dikar (dikarWT)     1861 AKGNLLQNVVAVKSTEEGGDSHAGVNPVTTKVKELKNALITETCEGAPKLSLGTVLKTQD
dikar (TM3)         1860 AKGNLLQNVVAVKSTEEGGDSHAGVNPVTTKVKELKNALITETCEGAPKLSLGTVLKTQD
                         ************************************************************

dikar (dikarWT)     1921 FSLGKSLEEMSGKKDANEDDRPNEEASPKNSSPPTTPNTEAPKPFEALHELSKRGKSSEP
dikar (TM3)         1920 FSLGKSLEEMSGKKDANEDDRPNEEASPKNSSPPTTPNTEAPKPFEALHELSKRGKSSEP
                         ************************************************************

dikar (dikarWT)     1981 SKSEASQKEKPNLSAWLKAFGGPKVSKKSEDEEKQQTPVQDLQGDSKVAPPAHSPAGDNF
dikar (TM3)         1980 SKSEASQKEKPNLSAWLKAFGGPKVSKKSEDEEKQQTPVQDLQGDSKVAPPAHSPAGDNF
                         ************************************************************

dikar (dikarWT)     2041 SLPTVMRQRKPSTGSTNSERSSFSQDPDSPRIAIDERYGSYAAGSYTSPIGASPIGASPI
dikar (TM3)         2040 SLPTVMRQRKPSTGSTNSERSSFSQDPDSPRIAIDERYGSYAAGSYTSPIGASPIGASPI
                         ************************************************************

dikar (dikarWT)     2101 MVSPKPNDDMGKPASPYPLNGAIKVGFYQDTTTKSSPDKSCSPREMNSPYPQYSQHIYSS
dikar (TM3)         2100 MVSPKPNDDMGKPASPYPLNGAIKVGFYQDTTTKSSPDKSCSPREMNSPYPQYSQHIYSS
                         ************************************************************

dikar (dikarWT)     2161 ASSPNVSTPDMSGTSPYGGGNSYNPSGSEASKTPAYSSTSPLPIYDQYKQPRSQESDYNS
dikar (TM3)         2160 ASSPNVSTPDMSGTSPYGGGNSYNPSGSEASKTPAYSSTSPLPIYDQYKQPRSQESDYNS
                         ************************************************************

dikar (dikarWT)     2221 SMSPSTPNPHSPYQQPQSSPYTTPQQSQSTHPSPYHNQSPYHQQQHSPYHPPAAQQQQQS
dikar (TM3)         2220 SMSPSTPNPHSPYQQPQSSPYTTPQQSQSTHPSPYHNQSPYHQQQHSPYHPPAAQQQQQS
                         ************************************************************

dikar (dikarWT)     2281 SQPSHSPAAHQQALSPMHSVESPASSAATQPPTPLAQSPAEQQHSPYQQPVLSPYQQPQQ
dikar (TM3)         2280 SQPSHSPAAHQQALSPMHSVESPASSAATQPPTPLAQSPAEQQHSPYQQPVLSPYQQPQQ
                         ************************************************************

dikar (dikarWT)     2341 QVQPPVVPPVQPSAGAQASTALGNNGYAPTHDSYQQLQQQQRSLYNPATLINPLSTAASS
dikar (TM3)         2340 QVQPPVVPPVQPSAGAQASTALGNNGYAPTHDSYQQLQQQQRSLYNPATLINPLSTAASS
                         ************************************************************

dikar (dikarWT)     2401 TASITKPNNDWSLNRSLLPPSITNTVNQAAQQQQQQQQQQAGQLQQQPQMQAAQQQLQST
dikar (TM3)         2400 TASITKPNNDWSLNRSLLPPSITNTVNQAAQQQQQQQQQQAGQLQQQPQMQAAQQQLQST
                         ************************************************************

dikar (dikarWT)     2461 PQQQQQQQQQLQGTPQQQQLQGSQQQQQQQQQQQQQLQTSQQQQQQVLGHQQQHKPKYPTY
dikar (TM3)         2460 PQQQQQQQQQLQGTPQQQQLQGSQQQQQQQQQQQQQLQTSQQQQQQVLGHQQQHKPKYPTY
                         ************************************************************

dikar (dikarWT)     2521 AQYQSTNAAANAAAAADAVDNLQQQQQKIAPPTYGSDMATFMQHQMQQPPKTDTLINPLK
dikar (TM3)         2520 AQYQSTNAAANAAAAADAVDNLQQQQQKIAPPTYGSDMATFMQHQMQQPPKTDTLINPLK
                         ************************************************************

dikar (dikarWT)     2581 RPGEEIGMDYSGNAANKMQKREETPAQQQQQQQQQQPTQNQTQSLLNKQQQMFNSFLGTM
dikar (TM3)         2580 RPGEEIGMDYSGNAANKMQKREETPAQQQQQQQQQQPTQNQTQSLLNKQQQMFNSFLGTM
                         ************************************************************

dikar (dikarWT)     2641 AFGKPIGNIAPDKAFEMYNRAAAMGFPKDFAKDNSCQLQQQQQQQQSPQVTTNKQQANQQ
dikar (TM3)         2640 AFGKPIGNIAPDKAFEMYNRAAAMGFPKDFAKDNSCQLQQQQQQQQSPQVTTNKQQANQQ
                         ************************************************************

dikar (dikarWT)     2701 TQQQPQQQSQQSQPHLTQLQTALSQNYQQQQQQTQAQKLPQQQQQQPPQQLNYQQQQAQL
dikar (TM3)         2700 TQQQPQQQSQQSQPHLTQLQTALSQNYQQQQQQTQAQKLPQQQQQQPPQQLNYQQQQAQL
                         ************************************************************
```

```
dikar (dikarWT)     2761 NHNYTAQQQATAVPDKPPAAQSTVAAAVSESSSNMMNLPSTAHQHHLSQTHHLAAYNKPT
dikar (TM3)         2760 NHNYTAQQQATAVPDKPPAAQSTVAAAVSESSSNMMNLPSTAHQHHLSQTHHLAAYNKPT
                         ************************************************************

dikar (dikarWT)     2821 PPPPQTYSNPLMQSMLGYAGNYFDKTMPPAAHMYSASSSAASAYGNPAQQLPGNYVPGNN
dikar (TM3)         2820 PPPPQTYSNPLMQSMLGYAGNYFDKTMPPAAHMYSASSSAASAYGNPAQQLPGNYVPGNN
                         ************************************************************

dikar (dikarWT)     2881 NPAHQQQQQPQQQQQQQQQAPAVPPAEVKAPAKRGRKKKAATIAAEAAAAAAKQQQQQQQ
dikar (TM3)         2880 NPAHQQQQQPQQQQQQQQQAPAVPPAEVKAPAKRGRKKKAATIAAEAAAAAAKQQQQQQQ
                         ************************************************************

dikar (dikarWT)     2941 QQQQQQQQQQQAQQQQQVAGQQQHQQQQVSAQQQQQQHQQVAAQQQQHQAMPQYNMPQSMA
dikar (TM3)         2940 QQQQQQQQQQQAQQQQQVAGQQQHQQQQVSAQQQQQQHQQVAAQQQQHQAMPQYNMPQSMA
                         ************************************************************

dikar (dikarWT)     3001 SAAAATNNQLQAHAQALQQGFQLYAGLKSGGVSSPVGSSAATPVNSGSTTNQTAADAAAI
dikar (TM3)         3000 SAAAATNNQLQAHAQALQQGFQLYAGLKSGGVSSPVGSSAATPVNSGSTTNQTAADAAAI
                         ************************************************************

dikar (dikarWT)     3061 SLKTSTGGMVPGSAFNFAPTPGTLGLYGDQAAAASSYLDQFRDAPNPYYMPPAPAHSGTA
dikar (TM3)         3060 SLKTSTGGMVPGSAFNFAPTPGTLGLYGDQAAAASSYLDQFRDAPNPYYMPPAPAHSGTA
                         ************************************************************

dikar (dikarWT)     3121 ANPSGNAADKGQNPLNTAAGSYPFLAAAHPSSRAAAAAAAYPFADPNSQLYQQYLRRDDF
dikar (TM3)         3120 ANPSGNAADKGQNPLNTAAGSYPFLAAAHPSSRAAAAAAAYPFADPNSQLYQQYLRRDDF
                         ************************************************************

dikar (dikarWT)     3181 HTRMIFNQSLLGGPAAAAAAAGYGQPPPPPSAYQRAALGMPKPYDINRQSWF
dikar (TM3)         3180 HTRMIFNQSLLGGPAAAAAAAGYGQPPPPPSAYQRAALGMPKPYDINRQSWF
                         ***************************************************
```

As was done for Gen, SIFT scores were generated to compare the three amino acid variants in Dikar. Unfortunately, all three predictions had low confidence due to low numbers of sequences used by the SIFT algorithm and low sequence diversity between these sequences. Therefore, SIFT scores were not helpful for analyses of these three variants and are not displayed here.

Secondary and tertiary structure predictions were made for Dikar using the I-TASSER web-based software. Secondary structure predictions placed all three variants in coil regions of the protein, rather than α-helix or β-sheet. There was low confidence in these predictions for the second variant (alanine to valine) and moderate confidence for the third variant (alanine to valine) (Table 4.4.4). Solvent accessibility for all three variants were similar to each other, with all variants displaying a moderate solvent accessibility (score of 4-6 on a scale of 0-9) (Table 4.4.4).

There were very few differences between the secondary structure assignments of α-helix, β-sheet, or coil between the two Dikar protein variants, with any differences having low

confidence scores. A total of five tertiary structure model predictions were generated for each Dikar protein variant. Model 1 for each protein variant demonstrated a similar overall topology when compared to each other (confidence C-score = 0.96 for Dikar variant on *dikar*^{WT} chr., and C-score = 0.94 for Dikar variant on *TM3* chr.) (Figure 4.4.5). Model 2 for each protein variant also demonstrated a similar overall topology when compared to each other (confidence C-score = -1.24 for Dikar variant on *dikar*^{WT} chr., and C-score = -1.51 for Dikar variant on *TM3* chr.). However, Model 2 was very different from Model 1 (Figure 4.4.5). None of the remaining 3 models for both Dikar variants appeared to be very similar to each other, all of which had low confidence scores (C-score < -3).

Table 4.4.4: Secondary structure predictions for the 3 amino acids that differ in Dikar between the *dikar*^{WT} chromosome and the *TM3* chromosome. Analysis was done using the web-based software I-TASSER. Confidence predictions range from 0 (low) to 9 (high). Solvent accessibility ranges from 0 (buried) to 9 (highly exposed).

Amino acid		α-helix, β-sheet, or coil		Confidence		Solvent accessibility	
dikar^{WT} chr.	TM3 chr.	*dikar*^{WT} chr.	TM3 chr.	*dikar*^{WT} chr.	TM3 chr.	*dikar*^{WT} chr.	TM3 chr.
N	S	coil	coil	8	7	4	4
A	V	coil	coil	2	1	4	4
A	V	coil	coil	4	5	6	5

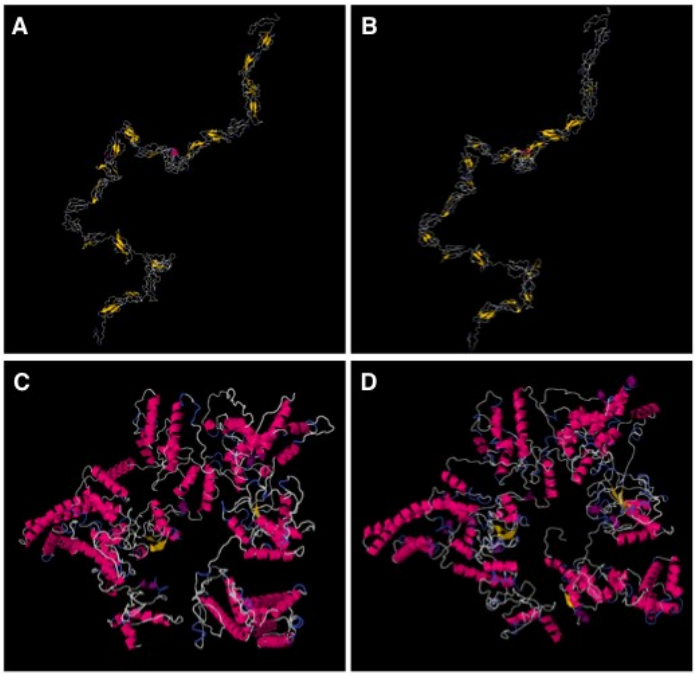


Figure 4.4.5: The most comparable Dikar protein prediction models of tertiary structure between the two protein variants. Predictions of tertiary structure were generated by I-TASSER. Model 1 predictions for Dikar on the *dikar*^{WT} chromosome (**A**, C-score = 0.96) and the *TM3* chromosome (**B**, C-score = 0.92). Model 2 predictions on the *dikar*^{WT} chromosome (**C**, C-score = -1.24) and the *TM3* chromosome (**D**, C-score = -1.51).

4.5: Discussion

Drosophila melanogaster is an amenable genetic model organism for studying the molecular function of genes and genetic pathways. Therefore, characterization of the *Drosophila Cecr2* homologue, *dikar*, was undertaken in an effort to better elucidate the basic molecular function of *dikar* (*Cecr2*). Unlike in mice, where a homozygous mutation in *Cecr2* results in a myriad of mutant phenotypes, including the NTD exencephaly, mutation or down-regulation of *dikar* failed to produce an observable mutant phenotype affecting planar cell polarity or IR sensitivity. Original characterization by Keuling et al. (2007) of two different mutant alleles, $dikar^{3}$ and $dikar^{5}$, produced no observable mutant phenotype [186].

In this thesis, an independent method of gene knockdown involving RNAi was utilized to confirm or refute that *dikar* is a dispensable gene for normal fly development. While control RNAi crosses involving the PCP genes *fz* and *Vang* both demonstrated expected PCP phenotypes when RNAi constructs were driven by a variety of promoters, none of the three RNAi knockdown of *dikar* under the same GAL4 drivers produced PCP phenotypes. No other obvious abnormal phenotype specific to *dikar* knockdown was seen in these flies either. Off-target pupal lethality, male specific in some cases, was seen to varying extents in all lines. Therefore, I found no evidence that *dikar* is involved in the PCP pathway. A caveat to this experiment is that all three *dikar* RNAi constructs may have failed to sufficiently knockdown *dikar*, although it is unlikely that all three constructs driven by 8 different GAL4 promoters failed.

A study conducted by Lee et al. (2012) concluded that *CECR2* functioned in DSB repair in HEK293T cell culture [144]. Therefore, we hypothesized that the absence of a phenotype in flies homozygous mutant for the $dikar^{3}$ or $dikar^{5}$ alleles may be due to the absence of sufficient DSBs to produce an observable phenotype. Experiments that exposed developing fly larvae to ionizing gamma radiation (IR) were undertaken in order to introduce enough DSBs to kill a fly with impaired DSB repair. Sensitivity to IR was observed in $dikar^{3}$ and $dikar^{5}$ homozygous mutant flies compared to heterozygous flies carrying a *TM3* balancer chromosome. However, the IR sensitivity was also seen in control progeny homozygous for the parental P element $P\{SUPor\text{-}P\}dikar^{KG00884}$ allele as well as progeny homozygous for the precisely excised P element ($dikar^{WT}$) allele (Figure 4.3.3). This suggested the presence of a second-site mutation causing the

IR sensitivity. As *dikar*^{3} or *dikar*^{5} over the *Df(3L)ZN47* deficiency also showed IR sensitivity, the most likely location of this second-site mutation is within the deficiency region *Df(3L)ZN47*.

Gene ontology analyses of the 147 genes within the deficiency region identified two genes, *Gen* and *dikar*, that were the best candidates for carrying second-site mutations. *Gen* was chosen due to its known function in resolving Holliday junctions [301]. Holliday junctions are DNA intermediates produced during homologous recombination (HR), an event that can occur during mitosis, meiosis, and DNA repair including but not limited to DSB repair. Homozygous mutation in *Drosophila Gen* has previously been shown to result in hypersensitivity to exogenously induced DNA damage by the crosslinking agent nitrogen mustard (HN2) and the alkylating agent methyl methanesulfonate (MMS) [299,302]. However, Sanger sequencing of both *Gen* and *dikar* failed to identify any obvious functional mutations. Analysis of 11 missense mutations in *Gen* and 3 missense mutations in *dikar* did not predict any obvious functional consequences. Additional experimentation would be required to confirm this. Although my focus was restricted to the genes within the *Df(3L)ZN47* deficiency region, it is still possible that the IR sensitivity locus lies on chromosome 3 outside this region. If this were the case, then the IR sensitivity locus would be the same on the *dikar*^{WT} containing chromosome and on the *Df(3L)ZN47* containing chromosome, but different on the *TM3* chromosome. Based on the results presented here, it is no longer worthwhile to follow up the IR phenotype in the *dikar*^{3} and *dikar*^{5} *Drosophila* lines.

It is still possible that *dikar* may cause IR sensitivity that would be epistatic to the second-site mutation. The best way to test this would be to generate a new *dikar* mutant allele on a different genetic background that does not contain the second-site IR sensitivity locus. One suggestion to identify a 3rd chromosome that is not IR sensitive relative to the *TM3* chromosome would be simply to test various 3rd chromosomes in a homozygous state compared to heterozygous over *TM3*. Unfortunately, the *TM3* chromosome would not be good for generating a new *dikar* allele in as it is not homozygous viable. The CRISPR/cas9 technology could be used to introduce a targeted germline mutation in *Drosophila* in a suitable genetic background (reviewed in [303]). One would then look for a phenotype that was present in the *dikar* mutant but not the wildtype parent.

We have yet to identify a mutant phenotype for *dikar* in *Drosophila*. Previous analyses by Keuling et al. (2007) demonstrated that the memory defect originally seen in P*{LacZ}dikar*^{1} mutant flies was due to a second-site mutation. Also, *dikar*^{3} or *dikar*^{5} mutations did not display

any obvious mutant phenotypes or interact with the *Iswi*^{1} mutation [186]. In this thesis, RNAi failed to produce any obvious phenotype specific to *dikar*. Also, IR sensitivity in *dikar*^{3} and *dikar*^{5} flies was not due to the *dikar*^{3} or *dikar*^{5} alleles, but rather to an additional second-site mutation. Keuling et al. (2007) suggested that the lack of a mutant phenotype for *dikar* may be because of functional redundancy with another chromatin modeling protein, and they proposed *Acf1* as a good candidate. ACF1 has been shown to be a component of two different ISWI complexes, ACF and CHRAC [155,277]. Fyodorov et al. (2004) previously generated and characterized *Acf1* deletion mutations by imprecise excision of the P elements EP(3)1181 and EY08629 (parental stocks available from Bloomington) [304]. Their study demonstrated that *Acf1* mutant flies had chromatin remodeling defects and displayed partial viability [304]. The authors suggested that the partial viability in *Acf1* mutant flies might be due to the presence of other ATP-dependent chromatin remodelers [304]. Future experiments could include the characterization of *dikar* and *Acf1* double-mutants to look for an exacerbation of the *Acf1* phenotype in double-mutant flies, or to attempt to rescue the *Acf1* phenotype in *Acf1* mutant flies by over-expression of *dikar*. If such an experiment were to be performed, it would be ideal if a different *dikar* mutant allele was used in lieu of or in addition to the *dikar*^{3} and *dikar*^{5} mutant alleles.

Unlike *Cecr2* in mice, *dikar* is dispensible for normal fly development. To date, no obvious phenotype has been seen in flies deficient in *dikar*. We suspect that the pupal lethality seen in RNAi-knockdown of *dikar*, *Vang*, and *fz* to be due to non-specific off-target effects. However, we are unable to conclude whether or not mutation in *dikar* results in IR sensitivity. Also, functional redundancy is another possible explanation as to why *dikar* is dispensible in flies. Therefore, we are still unable to conclude that *dikar* deficiency does not produce a mutant phenotype. Identification and characterization of a *dikar* mutant phenotype could lead to a more tractable genetic model for studying the basic molecular function of *Cecr2*.

Chapter 5

Circling behavior and stereotypic route-tracing in a genetically distinct cohort of male mice

Preface

Work in this chapter was done as part of a collaboration. I was responsible for study design, data collection, and data analyses. I collected tissue samples for the high performance liquid chromatography (HPLC) experiment, and Gail Rauw conducted the HPLC experiment. Dr. Glen Baker aided in HPLC experimental design and HPLC data analysis. A manuscript is in preparation for a version of this chapter with authorship as follows: R.Y.M. Leduc, G. Rauw, G. Baker, and H.E. McDermid.

5.1: Introduction

The gene *Arhgap19* was initially the top candidate modifier of *Cecr2*-associated exencephaly (Chapter 3). Very little is known about *Arhgap19* specifically, although it belongs to the family of GTPase activating proteins (GAPs), which are responsible for the negative regulation of Rho proteins and are involved in a myriad of cellular processes including but not limited to cell adhesion, cell migration, gene expression, cytoskeletal dynamics, and planar cell polarity [200-202]. In an effort to better understand the specific function of *Arghap19*, we set out to characterize the phenotypic effects of the $Arhgap19^{GT(YHD020)Byg}$ mutant allele (see Figure 5.5.1 for additional information on this allele). As part of our analyses, we established a mouse line homozygous for the $Arhgap19^{GT(YHD020)Byg}$ allele. We observed that male mice within this line displayed stereotypical route-tracing and circling behavior, and therefore proceeded to characterize this abnormal phenotype. Although this behavior does not appear to be associated with the *Arhgap19* mutation, an important finding was that circling was dependent on an environmental trigger, as mice housed under different conditions displayed differences in penetrance of the circling phenotype.

Implementation of social housing for laboratory mice has become standard, and increasing environmental complexity by providing environmental enrichments is becoming more common in order to improve animal welfare. It has been well established that housing male mice individually greatly increases their propensity to fight when introduced to a new male. However, male mice that are socially housed together from a young age and are not separated are less aggressive than individually housed males or males housed with a female [305]. In fact, male mice prefer social contact with another male cage-mate over environmental enrichment such as nesting material, indicating that social housing is beneficial to the welfare of male mice [306], although no differences between socially housed and individually housed males were evident when stress behaviors and stress hormones were measured [307].

However, in some instances, the provision of environmental enrichments, such as running-wheels, to socially housed male mice may adversely affect welfare due to increased aggression caused by competition for what is considered a valuable resource [308]. Increased aggression and stereotypic behavior resulting from competition over environmental enrichments, which included a running-wheel, has also been reported in socially housed female mice [309].

Running-wheels are commonly used as an environmental enrichment, as voluntary wheel-running has been widely observed in several species of wild, domestic, and laboratory animals. However, it remains controversial as to whether voluntary wheel-running is beneficial or maladaptive (reviewed in [310]). Wheel-running has been shown to be beneficial to many physiological systems including but not limited to the brain, heart, and digestive system, and has a positive effect on metabolism and aging (reviewed in [311]). Nevertheless, under certain circumstances, wheel-running can also lead to body changes (tail hyperflexion and arching of the back), increased aggression and possible addiction-like behavior (reviewed in [311]).

A commonly reported benefit of environmental enrichment for mice is the attenuation of stereotypic behavior [312,313], defined as a behavior that is highly repetitive, invariant, and has no obvious function [314]. Common stereotypic behaviors in mice include bar-chewing, cage-top twirling, circling and route-tracing. Voluntary wheel-running has been shown to significantly reduce stereotypic behavior in mice [308,315]. Howerton et al. (2008) demonstrated that the presence of a running-wheel significantly reduced the stereotypic behaviors bar mouthing, repetitive clawing, gnawing the water bottle, and route-tracing in male CD-1 mice [308]. The authors suggested that the reduction in stereotypic behaviors may be due to re-direction of these stereotypic behaviors towards stereotypic wheel-running, or that the wheel-running decreased frustration resulting from limited exploration opportunities. A consensus has yet to be reached as to whether or not wheel-running itself is a form of stereotypic behavior (reviewed in [310]). However, a recent study demonstrates that voluntary wheel-running can be considered an elective behavior rather than a stereotypic behavior since wild mice have been shown to utilize running-wheels, when provided in a natural setting, to approximately the same extent as laboratory mice [316].

Circling behavior in mice is one type of stereotypic behavior, but is a behavior that is also often attributed to defects in vestibular function of the inner ear, affecting the ability to sense gravity. Inner ear defects are seen in mice deficient in SLC26A4 [317], caspase-3 [318] and *Dmp1* [319], as well as in the stargazer, turning, and waltzing rat strains [320], all of which display circling behavior. Some circling rodent models with inner ear defects also show defects in the dopamine (DA) signaling pathway [321,322] and induced vestibular dysfunction results in alterations in the DA system [323-325]. This suggests that circling behavior due to vestibular dysfunction may partially act through secondary changes in the DA system. DA primarily acts as a neurotransmitter in the

brain, and is involved in several behavioral and cognitive functions such as voluntary movement, reward-driven learning, memory, and attention. In humans, abnormalities in the DA system are associated with neuropsychiatric disorders, which include autism [326], attention-deficit/hyperactivity disorder (ADHD) [327], and schizophrenia [328]. Dopamine transporter (DAT) knockout mice have circling behavior and are used as an ADHD model (reviewed in [329]), and polymorphisms in DAT have been associated with ADHD in humans [327,330]. The *chakragati* mouse, which contains a transgenic insertional mutation in an unknown gene, demonstrates hyperactivity and circling behavior, and is used as a model for schizophrenia [331,332]. *Chakragati* mice have asymmetry in dopamine D2 receptors and circle away from the side of the brain with higher D2 receptor levels [333], consistent with the theory that mice exhibiting unilateral circling behavior will turn away from the side of the brain hemisphere with higher DA levels [334].

Like DA, serotonin (5-hydroxytryptamine, 5-HT) is another biogenic amine involved in neurotransmission; however, 5-HT is synthesized and metabolized in a different metabolic pathway than DA. Alterations in brain 5-HT levels have been reported in partial bilateral striatal 6-hydroxydopamine-lesioned rats, which demonstrate unilateral circling behavior, however, these rats also have alterations in brain DA levels [335]. DA and 5-HT have been proposed to be agonistically linked in rodents [336], meaning that reduction in DA levels leads to reduction in 5-HT levels.

We observed an increase in stereotypic behavior in a specific cohort of male mice (herein referred to as FVB/129P2) housed with running-wheels, where pronounced circling with route-tracing was the manifesting stereotypic behavior. Circling behavior developed more often in socially housed male mice that were in cages requiring separation of inhabitants due to fighting compared to socially housed males that did not require separation. Circling male mice also show significant alterations in whole brain levels of the biogenic amines DA and 5-HT and their metabolites.

RESULTS

5.2: Circling behavior in FVB/129P2 male mice

We introduced running-wheels at the time of weaning (~21 days old) to all non-breeding mice in our colony as environmental enrichment for a period spanning 11 months. Our colony

contains BALB/cCrl, FVB/N and C57BL/6 strains, some of which carry mutations in one of the genes *Cecr2*, *Snf2l*, *Snf2h,* or *Arhgap19*. Both sexes of all mouse strains were observed to utilize the running-wheels, although wheel utilization was not recorded for individual mice. During this time, unidirectional circling behavior manifested at 2-3 months of age in a proportion of males of a particular cohort of mice (FVB/129P2). This cohort is referred to as FVB/129P2 since they are 81.25% FVB/N and 18.75% 129P2 genetic background (Figure 5.2.1). During this time, we did not see circling behavior in any male mice of other mouse lines in our colony. We did see a total of three female mice not of the FVB/129P2 line develop circling behavior. One of these mice was from the wildtype FVB/N strain, the other two were from the FVB/N mouse line that carried the *Cecr2*$^{tm1.1Hemc}$ allele; however, both mice were wildtype for the *Cecr2*$^{tm1.1Hemc}$ allele. All three female mice were housed with running-wheels since they were weaned. To look for circling behavior in FVB/129P2 mice, cages were removed from the ventilated rack and the mice were observed in their cages. Circling behavior was first noted in short bursts and a tendency to turn the same direction. Once the behavior was noted, within approximately one week the mouse adopted obvious circling behavior that consisted of running unidirectional circles in rapid succession. This was easily observed when the mice were active, and circular tracks in the sawdust bedding within cages containing circling mice were often apparent. The majority of circling mice did not utilize the wheel during bouts of circling, and we considered the circling behavior stereotypic because route-tracing was evident. For example, one circling mouse would almost always run through the doors of his house as part of his circling route, whereas another mouse would run over top of the wheel as part of his route for two circuits, and on the third circuit would briefly run on the wheel then jump off of it, and would continue this pattern with little deviation. Once a mouse adopted circling behavior, it exhibited this behavior for the rest of its life, regardless of whether the mouse was housed with a running-wheel or not, and regardless of whether the mouse was housed with other mice or alone. Circling was more pronounced when the mouse was more alert and/or aggravated, such as during a cage change. The circling mice appeared healthy; however, eating and sleeping behaviors were not measured and therefore a subtle disruption may be present. Two circling males in breeding pairs lived for 75 and 294 days, and continued to circle throughout their adult life. The male breeder mouse that lived for 75 days was separated from his female shortly after being introduced to her due to his circling behavior; therefore, he never had a chance to sire any pups. The other male breeder mouse remained with

his female breeder mouse despite his circling behavior and sired 11 litters over a period 6 months. Five of these litters were lost; however, one of these five litters was lost after the circling male was separated from the female.

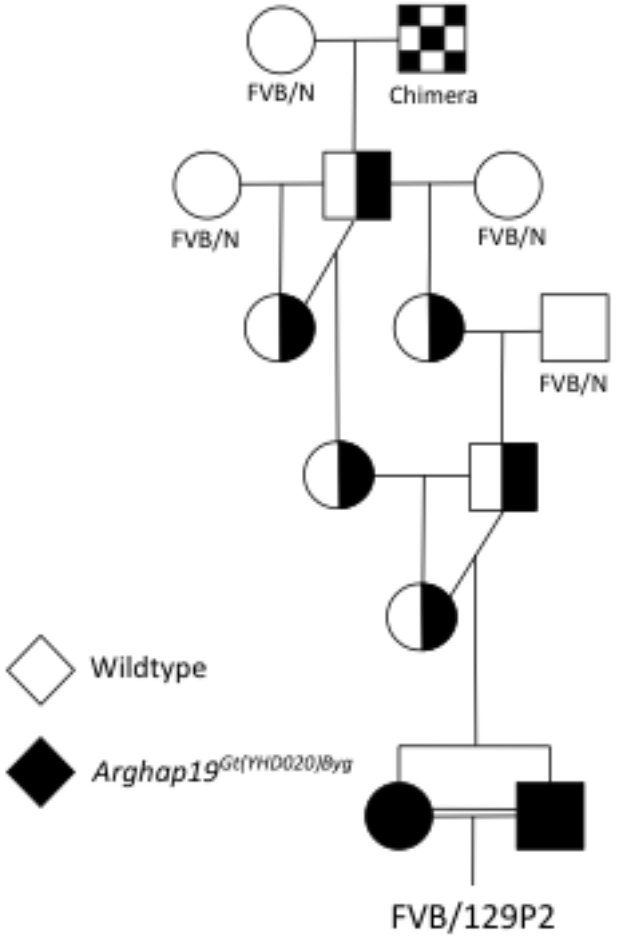


Figure 5.2.1: Mouse pedigree showing the ancestry of the FVB/129P2 line. The male chimera (checkered box) was a mosaic of 129P2 and DBA/2J:C57BL/6J hybrid cells. 129P2 germ cells were heterozygous for the *Arhgap19*$^{Gt(YHD020)Byg}$ mutation, therefore, *Arhgap19*$^{Gt(YHD020)Byg}$ was transmitted to progeny 50% of the time, and all of the chimera's progeny were 50% FVB and 50% 129P2.

Once it was clear that this severe circling behavior was limited to one genetic cohort of mice in the colony, we performed experiments to characterize the phenotype and its relationship to the presence of the running wheels. A total of 75 FVB/129P2 male mice were weaned in groups of 2-5 sibs (depending on the number of males weaned at any one time) with either a running-wheel dome house or a cardboard house. An additional 7 male mice with no brother littermates were weaned into individual housing with either a running-wheel (6 mice) or a cardboard house (1 mouse). Socially housed males were separated and housed singly when any sign of fighting was observed to avoid injury. Original observations were of wounds, particularly to the hind trunk and/or tail; however, later observations involved diligent monitoring by the technicians and separation prior to any sign of injury (detailed in Chapter 2.27). We continued to

house separated mice with the same house-type as when they were weaned. Only FVB/129P2 mice housed with running-wheels showed circling behavior (30/67, 44.8%). None of the FVB/129P2 mice housed without running-wheels (standard housing) showed circling behavior (0/15). This difference associated with the presence or absence of the running-wheel is statistically significant, as indicated by a 2 x 4 χ^2 test-of-independence ($P < 0.01$, Figure 5.2.2A).

We first noticed circling behavior in male mice approximately 4-11 weeks after receiving running-wheels at ~3 weeks old (weaning). When wheels were introduced to 9 adult male mice that ranged from 6 to 12 weeks old, six of these mice required separation due to fighting, two of which developed circling behavior approximately 6-7 weeks after wheel introduction. The 6-7 week timeframe falls within the timeframe that was seen with recently weaned mice. Therefore, this implies that circling behavior can develop in mice exposed to running-wheels later in life, and not only in mice exposed to wheels at weaning age.

It was noted that a large proportion of socially housed male mice exposed to running-wheels required separation due to fighting within a cage. In 2-5 male groupings, 72.1% (44/61) of males with running-wheels were in cages where fighting was noted, while none (0/14) of the males without wheels were associated with observed fighting ($P < 0.0001$, Figure 5.2.2B). Circling behavior in cages with running-wheels manifested in FVB/129P2 male mice in cages with fighting (27/44, 61.4%) significantly more often than in males that were not in cages with fighting (3/23, 13.0%) ($P < 0.001$, Figure 5.2.2B). One of these three circling mice was individually housed with a wheel since weaning (Figure 5.2.2C). On average, mice were separated due to fighting prior to circling behavior being observed, although appearance of fighting ranged from 51 days prior to circling to 30 days after circling (mean = 15.5 days prior, median = 14 days prior, n = 24). Of the 24 mice where the date of circling and separation due to fighting was noted, 19 mice developed circling behavior after being exposed to fighting, 1 mouse developed circling behavior on the same day that the brothers were separated due to fighting, and 4 mice exhibited circling behavior before having to be separated due to fighting within the cage. We did not measure the level of fighting in these experiments, as discussed in the materials and methods (Chapter 2.27); therefore, we were not able to decipher which animals were aggressors, victims, or bystanders. Even though mice were separated at the first signs of fighting, it is still possible that mice were exposed to stressful social interactions prior to observed fighting, which

later developed into the observed fighting. It is also possible that these stressful social interactions were occurring in cages where mice were never separated.

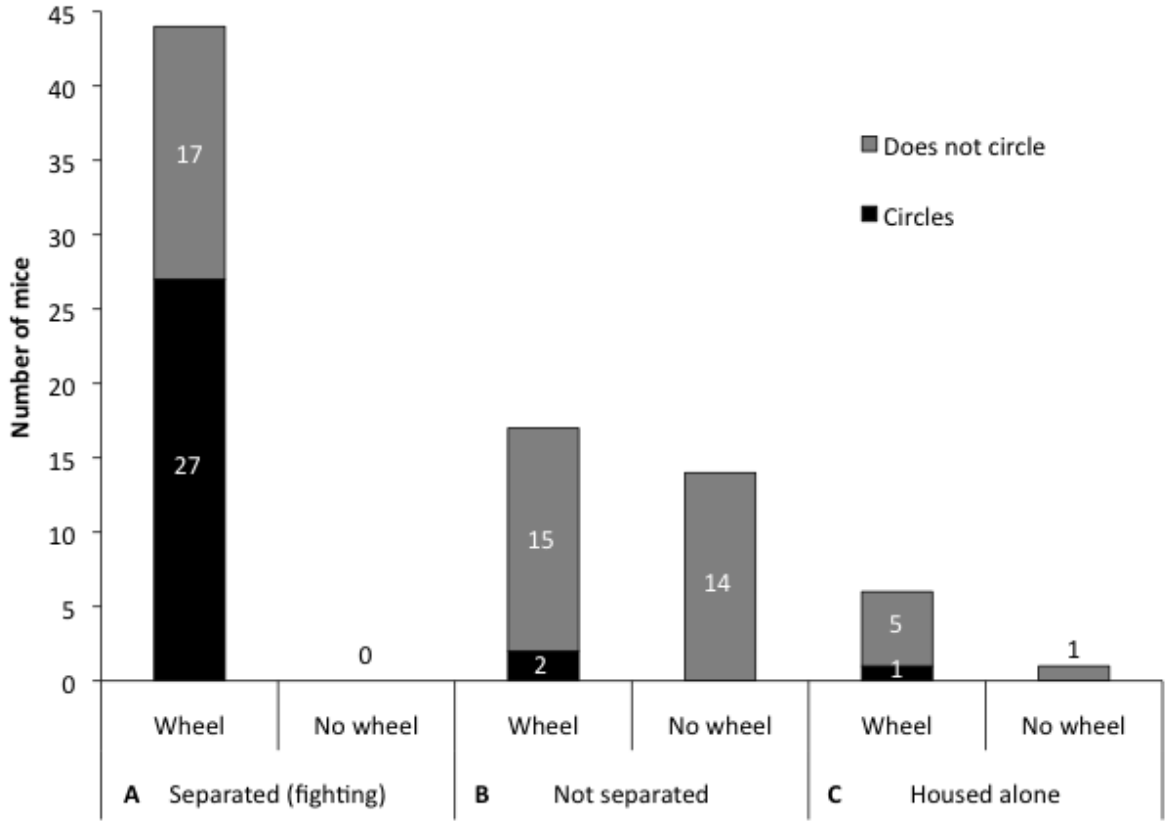


Figure 5.2.2: Circling in FVB/129P2 male mice. Circling behavior was observed only in males where a running-wheel was present in the cage (**A-C**). Mice represented under "Separated (fighting)" were initially weaned into social housing with 2 to 5 males per cage and then separated into individual housing when signs of fighting, such as wounding and fighting after cage change, were observed (**A**). Mice represented under "Not separated" were weaned into social housing with 2 to 5 males per cage and were never separated due to fighting (**B**). Mice represented under "Housed alone" were housed alone since weaning (**C**). Mice were either housed with running-wheels (Wheel), or housed with cardboard houses (No wheel). Circling behavior was seen significantly more often in male mice exposed to fighting. Overt signs of fighting were exclusively observed in mice housed with wheels. A total of 75 socially housed mice and 7 individually housed mice are represented in this graph. Numbers of mice for each category are represented on graph columns.

Both fighting and individual housing can be stressful for mice. As circling mice, all of which were housed with wheels, were often exposed to aggression and then subsequently to individual housing, it is difficult to decipher whether it was the running-wheel itself, the ensuing fighting, or the resulting separation that triggered the circling behavior in these mice. Therefore, circling behavior of the 55 male mice in which the date was recorded for circling as well as separation due to aggression was addressed (Figure 5.2.3). Prior to separation, a total of 8/55 mice developed circling behavior, which was not significantly different from the 0/14 mice that circled in the absence of a running-wheel according to a χ^2 test-of-independence ($P = 0.12909$). Of these 8 mice that developed circling behavior, 2 were never separated from their siblings, 4 were separated after the circling behavior was observed, and 2 were separated on the same day that circling behavior was observed. After separation, an additional 15 of the 55 mice went on to develop circling behavior, resulting in a total of 23/55 mice developing circling behavior. The 15/55 mice that went on to develop circling behavior post-separation was close to being significantly different from the 0/14 mice that circled in the absence of a running-wheel ($P = 0.05560$); however, the total 23/55 mice that eventually circled was significantly different from the 0/14 mice that circled in the absence of a running-wheel ($P = 0.00304$). As mice were exposed to both stressors, fighting and individual housing, we cannot eliminate individual housing as a possible stressor. However, for the 8 mice that developed circling behavior prior to separation, we can conclude that individual housing did not contribute to the development of circling behavior in these mice.

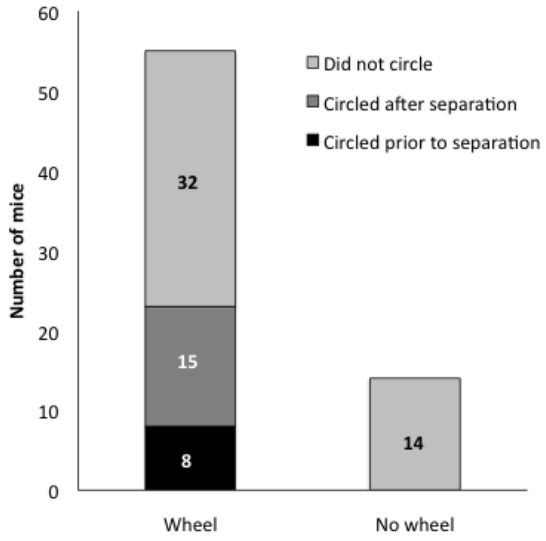


Figure 5.2.3: Socially housed mice can develop circling behavior prior to separation as well as after separation due to fighting. Only socially housed male mice where the date was recorded for observed circling behavior as well as separation is included. Numbers of mice for each category are displayed on graph columns.

5.3: Circling behavior is not associated with the *Arhgap19*$^{GT(YDH020)Byg}$ mutation

The FVB/129P2 cohort of mice are all homozygous for a genetrap mutation in *Arhgap19* (*Arhgap19*$^{GT(YDH020)Byg}$). The *Arhgap19*$^{GT(YDH020)Byg}$ allele is described in greater detail in Chapter 3, with expression analyses demonstrating that *Arhgap19* is present in the developing brain from embryonic days E9.5 to E13.5, but is no longer expressed in E14.5 brain (Figures 3.8.2, 3.8.3). We therefore performed a genetic analysis to see if circling behavior is associated with *Arhgap19*$^{GT(YDH020)Byg}$ homozygosity. We crossed the FVB/129P2 mice to FVB/N mice to create heterozygotes, which were then crossed to each other, thereby creating all 3 *Arhgap19*$^{GT(YDH020)Byg}$ genotypes in their progeny: 25% homozygous mutant, 50% heterozygous for the mutation, and 25% wildtype. If the *Arhgap19*$^{GT(YDH020)Byg}$ mutation is associated with circling, then the phenotype should only have been seen in the homozygous mutant progeny if the *Arhgap19*$^{GT(YDH020)Byg}$ allele is homozygous recessive, or in both heterozygous and homozygous mutant progeny if the *Arhgap19*$^{GT(YDH020)Byg}$ allele is dominant. All progeny were given running-wheels at weaning. Circling was seen in male mice of all 3 genotypes, including homozygous wildtype, thereby eliminating the *Arhgap19*$^{GT(YDH020)Byg}$ mutation as the cause of the circling behavior (Table 5.3.1). This suggests that the presence of an independent second-site mutation in a gene other than *Arhgap19* is responsible for circling. If this second-site mutation was also homozygous in the FVB/129P2 mice but not present in FVB/N mice, then the progeny examined above would also have been ~25% homozygous for this mutation. The penetrance of circling behavior in the FVB/129P2 mice exposed to running wheels is only 50.8% (30/59). Therefore, a circling penetrance of 6/53 (11.3%) in the overall progeny of the heterozygous males is approximately the expected incidence of 12.5% (~50% of the ~25% of homozygotes). This supports a genetic basis for the circling behavior in the presence of a running-wheel, but in a gene not linked to *Arhgap19*.

Table 5.3.1: Circling behavior is not associated with the *Arhgap19*$^{Gt(YHD020)Byg}$ mutation. Heterozygous crosses that produced all 3 genotypes showed that circling did not follow the mutant genotype in male mice. +/+ homozygous wildtype, +/- heterozygous, -/- homozygous mutant.

# of male mice	**+/+**	**+/-**	**-/-**
Circling behavior	1	3	2
Total	15	24	14

Concurrently, we also backcrossed the *Arhgap19*$^{Gt(YHD020)Byg}$ heterozygotes to the FVB/N strain for 10 generations to create a congenic line for *Arhgap19*$^{Gt(YHD020)Byg}$ on the FVB strain. When congenic *Arhgap19*$^{Gt(YHD020)Byg}$FVB males were provided with running-wheels, 0/51 males developed circling behavior. Since we selected mice at each backcross generation for the presence of *Arhgap19*$^{Gt(YHD020)Byg}$, an independent mutation would most likely be lost after 10 generations, explaining the loss of the circling behavior. We also noted that fighting in *Arhgap19*$^{Gt(YHD020)Byg}$FVB males with running-wheels remained high, with 20/48 (41.7%) males in cages requiring separation due to fighting compared to 0/15 males with no wheel. As mice of different genotypes were housed together and we did not measure fighting, we could not determine if fighting was associated with the *Arhgap19*$^{Gt(YHD020)Byg}$ mutation.

5.4: Circling behavior in female mice

Female FVB/129P2 mice housed with running wheels show a very low penetrance of circling behavior (1/33, 3.0%). This is also true of the female progeny of the *Arhgap19*$^{Gt(YHD020)Byg}$ heterozygous crosses in the previous section, with a circling incidence of 2/28 (7.1%). None of the socially housed FVB/129P2 females showed any noticeable fighting that would result in their separation. Although three out of 28 (10.7%) female progeny from the heterozygous cross were exposed to fighting and separated, none of them developed circling behavior. Unlike male congenic *Arhgap19*$^{Gt(YHD020)Byg}$FVB mice, of which none circled, there were 3/46 (6.5%) female congenic mice housed with wheels that acquired circling behavior, none of which were exposed to fighting. Of female congenic *Arhgap19*$^{Gt(YHD020)Byg}$FVB mice not housed with wheels, 1/13 developed circling behavior. These data demonstrate that the circling behavior phenotype manifests differently in female mice compared to male mice. As female

circling behavior is more sporadic and has been seen in three females from other mouse lines in our colony (mentioned in section 5.2), it possible that the cause of female circling behavior is different from the cause of the circling behavior seen in FVB/129P2 male mice in this study.

5.5: Circling mice do not have obvious behaviors typical of inner ear defects

Circling behavior in rodents is most often associated with inner ear defects. We therefore tested three of the most severe circling males with a series of behavioral tests designed to look for the inability to sense gravity due to vestibular dysfunction and compared to four non-circling control mice. The three tests consisted of 1) the trunk-curl test, where mice with inner ear dysfunction will preferentially curl their trunk rather than reach for a presented surface when suspended by their tail, 2) the contact-righting test, where mice with inner ear dysfunction will not right themselves when turned upside down while inside a tube that applies contact to their feet, sides, and back, and 3) the forced-swim test, where mice with inner ear dysfunction will display difficulties swimming and underwater tumbling. There were no obvious differences in behavior for the trunk-curl test as both circling mice and non-circling mice would occasionally curl their trunk as well as reach for a presented surface. Circling mice did spend more time curling their trunk forward, but 2 of the 3 circling mice also reached for presented surfaces (Table 5.5.1). Circling mice were capable of righting themselves in the contact-righting test, and were often quicker to do so compared to non-circling mice. Circling mice also spent more time with their heads sideways; this was because they were righting themselves more often (Table 5.5.1). All mice were capable of swimming and no consistent differences were seen between the two groups. The only abnormal swimming behavior seen in circling mice was that one of them spent 32 out of the 60 seconds swimming in circles; however, two non-circling mice also swam in circles (10.4 sec and 2.5 sec each, Table 5.5.1). None of the mice displayed underwater tumbling at any time during the two minutes of swimming. These severe circlers did not show any indication of an obvious inner ear defect; therefore no further circling mice were tested.

Table 5.5.1: Times for individual mice spent engaging in each behavior scored in the behavioral tests for inner ear function. Three male FVB/129P2 males with severe circling behavior underwent the trunk-curl, contact-righting, and forced-swim test in order to determine if they were able to sense gravity, and were compared to four non-circling FVB/129P2 male mice. The tests were video recorded and time spent in each behavior was calculated.

Trunk-curl				
	Reach (sec)	**Side curl (sec)**	**Forward curl (sec)**	**Total (sec)**
Circling 1	18.0	0.0	0.0	18.0
Circling 2	7.8	10.2	0.0	18.0
Circling 3	0.0	0.0	18.0	18.0
Average	**8.6**	**3.4**	**6.0**	**18.0**
Non-circling 1	18.0	0.0	0.0	18.0
Non-circling 2	11.6	0.0	6.4	18.0
Non-circling 3	5.7	7.7	4.6	18.0
Non-circling 4	18.0	0.0	0.0	18.0
Average	**13.3**	**1.9**	**2.8**	**18.0**
Contact-righting				
	Right side up (sec)	**Sideways (sec)**	**Upside down (sec)**	**Total (sec)**
Circling 1	1.2	9.2	9.6	20.0
Circling 2	9.8	8.2	2.0	20.0
Circling 3	3.5	9.1	7.4	20.0
Average	**4.8**	**8.8**	**6.3**	**20.0**
Non-circling 1	12.5	3.5	4.0	20.0
Non-circling 2	13.1	2.0	4.9	20.0
Non-circling 3	4.2	8.7	7.1	20.0
Non-circling 4	0.0	2.6	17.4	20.0
Average	**7.5**	**4.2**	**8.3**	**20.0**
Forced-swim				
	Regular swim (sec)	**Circular swim (sec)**	**Immobile float (sec)**	**Total (sec)**
Circling 1	60.0	0.0	0.0	60.0
Circling 2	28.0	32.0	0.0	60.0
Circling 3	24.9	0.0	35.1	60.0
Average	**37.6**	**10.7**	**11.7**	**60.0**
Non-circling 1	57.5	2.5	0.0	60.0
Non-circling 2	45.0	10.4	4.6	60.0
Non-circling 3	56.5	0.0	3.5	60.0
Non-circling 4	60.0	0.0	0.0	60.0
Average	**54.8**	**3.2**	**2.0**	**60.0**

5.6: Dopamine and serotonin biochemistry is altered in circling FVB/129P2 male mice

Changes in some amino acids and biogenic amines have been shown to be associated with circling in mice [321,329,333]. We therefore analyzed brain levels of biogenic amines, amino acids and metabolites. Compounds were measured in left and right brain hemispheres of male FVB/129P2 mice housed with wheels that did or did not develop circling behavior and compared to male FVB/129P2 mice in standard housing conditions. Although some asymmetries between brain hemispheres were noted, these asymmetries were seen in both circling and non-circling mice and no definitive pattern in circling mice could be discerned (Appendix E). Therefore, left and right hemispheres were treated as technical replicates and averaged to obtain concentrations in whole brain. Levels of the biogenic amines dopamine (DA), noradrenaline (NA), and the metabolite 3,4-dihydroxyphenylacetic acid (DOPAC) were significantly altered as indicated by one-way ANOVA (DA [$P < 0.01$, $F = 6.5$], NA [$P < 0.05$, $F = 5.9$], DOPAC [$P < 0.05$, $F = 4.5$]). Bonferroni corrected pairwise tests identified a significant drop in circling mice housed with wheels compared to males housed under standard conditions for DA, NA, and DOPAC ($P < 0.01$, $P < 0.05$, and $P < 0.05$ respectively), and non-circling mice housed with wheels displayed an intermediate but not significant drop (Figure 5.6.1A). The precursor amino acid, tyrosine (TYR), and the metabolite, homovanillic acid (HVA), demonstrated a similar trend, although differences did not reach significance. The metabolite 3-methoxytyramine (3-MT) did not appear to be different between the three groups and did not follow the same trend. The serotonin (5-HT) biogenic amine metabolic pathway was also affected ($P < 0.05$, $F = 4.6$): circling mice housed with wheels showed a significant drop in 5-HT when compared to standardly housed males ($P < 0.05$), and non-circling males housed with wheels displayed an intermediate drop that did not reach significance (Figure 5.6.1B). The precursor amino acid, tryptophan (TRYP), and the metabolite, 5-hydroxyindoleacetic acid (5-HIAA), showed similar trends although they did not reach significance. Of all other amino acids, biogenic amines, and their metabolites tested, none showed a significant difference between any of the three groups (Tables 5.6.1 and 5.6.2).

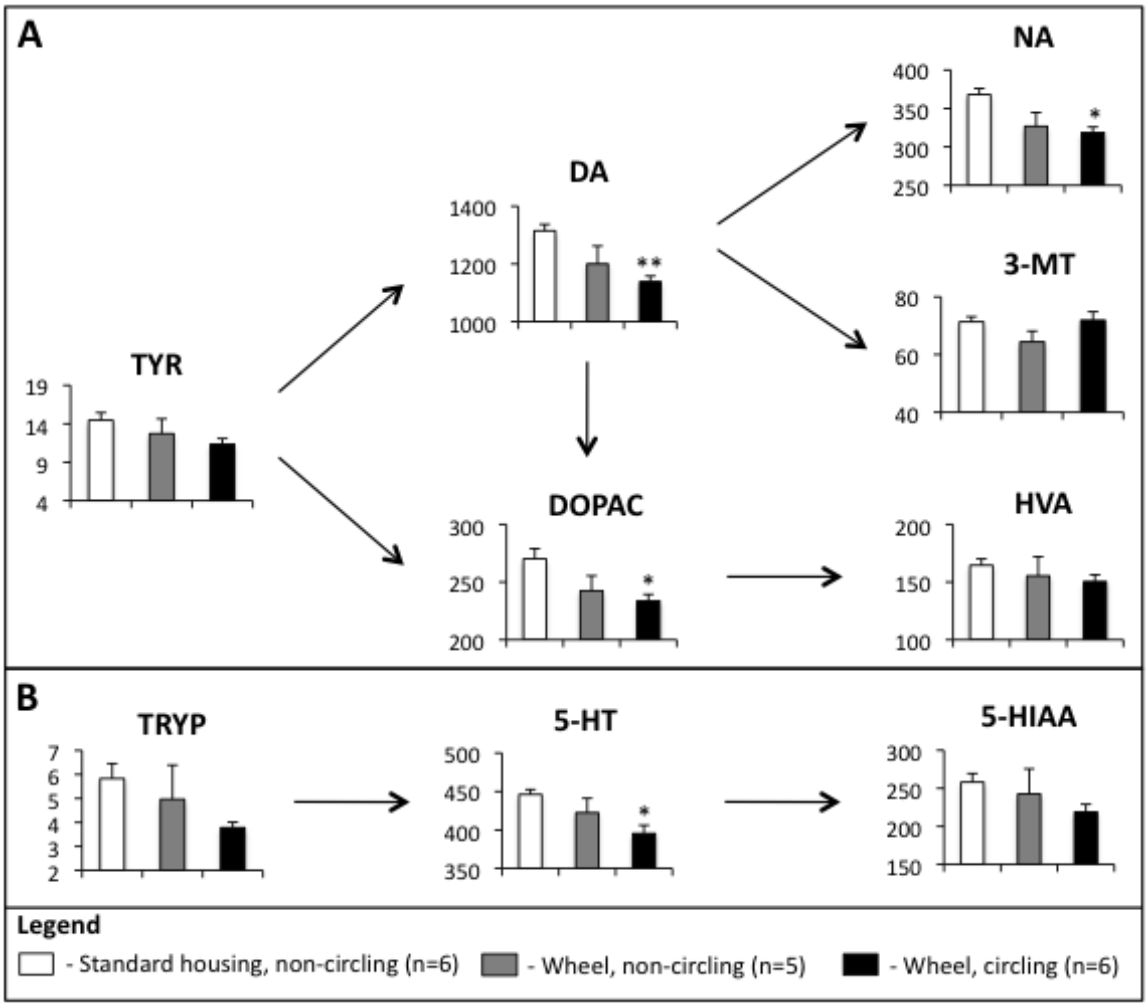


Figure 5.6.1: Male FVB/129P2 mice that circle have altered brain biochemistry. Male circling mice have significantly lower DA, NA, and DOPAC concentrations compared to male mice with standard housing, all of which did not circle (**A**). Male circling mice have significantly lower 5-HT compared to male mice with standard housing (**B**). Arrows denote the metabolic pathway, with only intermediates measured by HPLC shown. Single asterisk denotes $P < 0.05$ and double asterisk denotes $P < 0.01$ when compared to mice with standard housing. Y-axis displays total brain concentration with TYR, and TRYP as µg/g of tissue and the rest as ng/g of tissue. Error bars represent standard error of the mean. Abbreviations: tyrosine (TYR), dopamine (DA), 3,4-dihydroxyphenylacetic acid (DOPAC), noradrenaline (NA), 3-methoxytyramine (3-MT), homovanillic acid (HVA), tryptophan (TRYP), serotonin (5-HT), 5-hydroxyindoleacetic acid (5-HIAA).

Table 5.6.1: Analyses of concentrations of biogenic amines, their metabolites, and their precursor amino acids in male FVB/129P2 comparing circling mice housed with wheels, non-circling mice housed with wheels, and non-circling mice housed with cardboard housing (standard). Samples from left and right hemispheres from each mouse were treated as technical replicates. Values represent concentration ± standard error of the mean. TYR and TRYP are μg/g of tissue and the rest are ng/g of tissue. Abbreviations: noradrenaline (NA), tyrosine (TYR), 3,4-dihydroxyphenylacetic acid (DOPAC), dopamine (DA), 5-hydroxyindoleacetic acid (5-HIAA), homovanillic acid (HVA), 3-methoxytyramine (3-MT), serotonin (5-HT), tryptophan (TRYP).

	NA	TYR	DOPAC	DA	5-HIAA	HVA	3-MT	5-HT	TRYP
Standard housing (n=6)	368 ± 8	14 ± 1	270 ± 9	1315 ± 22	258 ± 11	165 ± 6	71 ± 2	446 ± 6	5.8 ± 0.6
Wheel, non-circling (n=5)	327 ± 18	13 ± 2	242 ± 13	1200 ± 62	243 ± 33	155 ± 17	64 ± 4	422 ± 19	5.0 ± 1.4
Wheel, circling (n=6)	319 ± 7*	11 ± 1	234 ± 6*	1139 ± 20**	219 ± 10	150 ± 6	72 ± 3	396 ± 11*	3.8 ± 0.2

*$P < 0.05$

**$P < 0.01$

Table 5.6.2: Analyses of amino acid concentrations in male FVB/129P2 comparing circling mice housed with wheels, non-circling mice housed with wheels, and non-circling mice housed with cardboard housing (standard). Samples from left and right hemispheres from each mouse were treated as technical replicates. Values represent concentration (μg/g of tissue) ± standard error of the mean. Abbreviations: aspartate (ASP), glutamate (GLU), L-serine (L-SER), D-serine (D-SER), glutamine (GLN), glycine (GLY), arginine (ARG), taurine (TAUR), alanine (ALA), gamma-aminobutyric acid (GABA).

	ASP	GLU	L-SER	D-SER	GLN	GLY	ARG	TAUR	ALA	GABA
Standard housing (n=6)	553±24	2073±114	77 ±4	24±1	807±42	121±6	32±1	1821 ± 74	80±2	389±15
Wheel, non-circling (n=5)	512±35	1958±15	81±6	23±1	1094±173	126±8	28±5	1671 ± 57	87±4	433±21
Wheel, circling (n=6)	596±65	2118±39	84±5	25±1	894±37	123±3	24±2	1763 ± 77	91±5	422±18

5.7: Discussion

Social housing and running-wheels are both commonly used as environmental enrichments to improve welfare of laboratory mice. However, a study by Howerton et al. (2008)

demonstrated that in certain strains of mice, social housing coupled with the provision of running-wheels can induce aggressive behavior in male mice, which was measured by video-recording of the first 15 minutes of each hour over a 48 hour period and behavior analyses of the accumulated 720 minutes of video-recordings for each cage [308]. Aggression with stereotypic behaviors (including route-tracing) has also been shown to occur in female C57BL/6J mice in clustered environmental enrichments which included running-wheels [309].

As the gene *Arhgap19* was a possible modifier of *Cecr2*-associated exencephaly, we set out to characterize any resulting phenotypes induced by the genetrap allele. As part of this analysis, we generated a line of mice homozygous for the $Arhgap19^{Gt(YHD020)Byg}$ genetrap allele on an 81.25% FVB/N and 18.75% 129P2 genetic background, herein referred to as FVB/129P2. In the current study, we have shown that this mouse line, FVB/129P2, has an underlying genetic predisposition to the development of permanent circling behavior with route-tracing that only manifested in the presence of an environmental trigger. We hypothesized that the underlying genetic predisposition was the homozygous mutation in *Arhgap19*. It should be noted that the endogenous mutant allele, $Arhgap19^{Ex6non}$, did not produce an observable phenotype in BALB/cCrl (Chapter 3). Two possible explanations for this is that the $Arhgap19^{Ex6non}$ allele may be less severe than the $Arhgap19^{Gt(YHD020)Byg}$ genetrap allele, or that manifestation of a phenotype due to mutation in *Arhgap19* differs between the FVB/129P2 strain and the BALB/cCrl strain. The possible environmental triggers were the presence of the running-wheel, the ensuing aggression resulting from the presence of the running-wheel, and individual housing as a result of separation due to fighting. As circling behavior failed to develop in the absence of the running-wheel, it is most likely that the running-wheel played a major role in triggering the circling phenotype. However, as circling behavior was seen more often in mice exposed to fighting, it is possible that the wheel indirectly triggered the circling behavior by introducing a stressful social situation, which might have been instigated by competition over the running-wheel. It is also possible that individual housing was a stressor; however, 8 male mice were observed to develop circling behavior prior to separation (Figure 5.2.3). Since aggression was not measured and could have been present in unseparated males at a low level, it is also possible that multiple stressors (wheel, aggression and individual housing) could combine to induce circling. Small numbers of mice individually housed since weaning were included in this study (6 with wheel, 1 without), thereby making it difficult to draw conclusions as to whether the stress of individual housing

alone can cause circling behavior. Phenotypic analyses of individually housed FVB/129P2 male mice with or without a running-wheel are currently underway.

This predisposition to circling behavior in males is specific to the FVB/129P2 cohort in our colony, since BALB/cCrl, FVB/N and C57BL/6 mice were also housed with wheels and none displayed circling behavior. The high level of fighting was also specific to the FVB/129P2 cohort and the later generations that were a part of the congenic *Arhgap19*$^{Gt(YHD020)Byg}$FVB cohort, with little to no fighting observed in the other strains in our colony. Although aggression and circling has been reported in female mice housed with running-wheels [309], fighting and circling was minimal and demonstrated no correlation to each other in female mice in the FVB/129P2 and related cohorts. Also, three female mice outside of the FVB/129P2 and related cohorts but with a FVB/N genetic background developed circling behavior. Taken together, these observations may indicate a basic susceptibility within the FVB/N background to circling behaviors that manifests differently in males and females.

FVB/129P2 mice are homozygous for a genetrap mutation in *Arhgap19*; however, our results from the heterozygous cross indicated that circling behavior is not a phenotype caused by this mutation, since circling does not follow the inheritance of the *Arhgap19* mutant allele. Further evidence supporting that a mutation in *Arhgap19* is not associated with circling is that BALB/cCrl mice in our colony were also housed with running-wheels during the same 11-month period. Although the BALB/cCrl substrain contains a homozygous nonsense mutation in *Arhgap19* [165], no fighting or circling behavior was observed in these mice in our colony. The circling phenotype was also no longer evident in congenic *Arhgap19*$^{Gt(YHD020)Byg}$FVB mice, suggesting that circling is caused by a second-site mutation that was unselected for and therefore lost. However, fighting levels remained high in congenic *Arhgap19*$^{Gt(YHD020)Byg}$FVB males housed with wheels. Therefore, it is possible that the mutation in *Arhgap19* increases levels of fighting in male FVB/N mice, while a different *Arhgap19* mutation in a different strain (BALB/cCrl) does not. Another possibility is that other genetic factors present in the FVB/129P2 line led to a predisposition to fighting in the presence of competition over a valued resource like a running-wheel. The putative second-site mutation in the FVB/129P2 background then predisposes to developing circling behavior in male mice housed with wheels and exposed to fighting. The second-site mutation may have originated in the 129P2 parental strain, or may be due to a spontaneous mutation that occurred in one of the mice used to generate the FVB/129P2 line.

Second-site genetrap insertions in the chimeric founder are also a possibility, and may have affected a gene resulting in the circling phenotype. Results from the heterozygous cross indicated that the circling phenotype, and therefore the second-site mutation, is not linked to *Arhgap19*. Additionally, *Arhgap19*$^{Gt(YHD020)Byg}$ genotyping used primers within the genetrap, which therefore excluded the possibility of a full-length second-site insertion that is not linked to *Arghap19*. It is still possible that a fragment of the genetrap inserted elsewhere in the genome is responsible for the circling phenotype, and backcrossing to FVB/N resulted in the loss of this second-site mutation.

Circling behavior often manifests in mice with vestibular dysfunction. If the FVB/129P2 line had vestibular defects, it would be more likely that circling behavior would manifest regardless of housing conditions; however, only male mice housed with wheels developed circling behavior. Forced-swim, contact-righting, and trunk-curl tests demonstrated that FVB/129P2 circling mice were able to sense gravity, although it is still possible that their inner ears and ability to sense gravity are subtly affected. It has been shown that vestibular dysfunction leads to secondary changes in DA signaling [323-325], and FVB/129P2 circling mice have significantly lower levels of DA and its metabolites, NA and DOPAC. This lends credence to the hypothesis that circling behavior in mice with vestibular dysfunction is more likely due to secondary alterations in DA signaling rather than to the primary vestibular lesion. This study did not afford us the ability to determine if the alterations in biogenic amines are causative or an effect of the circling behavior; we were only able to conclude that there was a significant association. Additional experiments to determine whether alterations in biogenic amines were causative or an effect of circling behavior could involve longitudinal measurements of dopamine in mice to determine if changes in DA occur prior to or after circling behavior. Methods to measure DOPA uptake, DA synthesis, and DA turnover multiple times in a single live animal have been developed for the rat and involves the use of positron emission tomography (PET) [337]. Also, fighting and separation were confounding variables for biogenic amine measurements. All of the circling and the majority of the non-circling mice housed with wheels that had biogenic amines measured were separated due to fighting, but none of the mice housed with cardboard houses that had biogenic amines measured required separation. It is important to note that several studies have linked serotonin (5-HT) hypofunction and dopamine (DA) hyperfunction to aggressive behaviors in both humans and animal models (reviewed in [338]). Currently, it is unclear

whether individual housing can alter dopaminergic or serotonergic functioning in mice; however, this could easily be tested by performing HPLC to measure DA and 5-HT in individually housed mice and compare to socially housed mice. In this study, circling mice did show lower levels of biogenic amines than non-circling mice housed with wheels (although not significant), even though both of these groups were housed under the same conditions and required separation due to fighting.

Male FVB/129P2 mice were much more likely to develop circling behavior than female FVB/129P2 mice. The sex differences in circling phenotype in our mice may be due to higher levels of fighting observed in males compared to females [339], which then acts as a stressor that triggers the abnormal circling behavior. These sex differences may also be due to inherent sex differences in DA and 5-HT systems. In humans, it has been established that DA neurobiology differs between the sexes [340-346]. Also, males are more likely to be diagnosed with autism [347], ADHD [348] and schizophrenia [349]. DA neurobiology is believed to play an important role in the etiology of these three psychiatric diseases. Sex differences in DA neurobiology have also been established in animal models, with more than one study demonstrating that sex steroids modulate DA signaling [350-354]. Velisek et al. (2005) showed sex differences in circling as well as synaptic activity in the dopaminergic substantia nigra in mice [355]. Sex differences in serotonin (5-HT) neurobiology have also been established in both humans [356-358] and animal models [359-361].

As mice housed with wheels that do not circle show intermediate levels of these neurotransmitters, it is possible that the wheel acts as an environmental trigger to induce changes in these neurotransmitter systems, which then results in behavioral abnormalities only in mice where a certain threshold is breached. As male mice, on average, were observed fighting before they developed circling behavior, the circling behavior may not have manifested directly because of the presence of the running-wheel, but may have been an abnormal behavioral response to the fighting. Combinations of stressors also may trigger circling. Stressors might include being denied access to the valuable running-wheel resource by cage-mates, aggressive behaviors that were not detected by technicians or researchers, and/or being housed alone. Our small sample size of seven singly housed male mice with wheels made it impossible to accurately determine the proportion of mice singly housed with wheels that would go on to develop circling behavior. An experiment involving the phenotypic analysis of additional FVB/129P2 male mice individually housed with wheels and comparing to FVB/129P2 male mice individually housed

with cardboard houses is currently underway. If circling behavior is only seen in mice individually housed with wheels, then it can be concluded that the wheel is sufficient to trigger circling behavior in this cohort, directly or indirectly. If circling behavior is seen in mice individually housed with either a wheel or a cardboard house, then individual housing in the absence of a running-wheel may also be a trigger of the circling behavior. If circling is not observed in any of the mice regardless of the presence or absence of a wheel, then this suggests that the wheel in conjunction with social housing is necessary to trigger circling behavior in these mice. Although this study will not tease apart the potentially combined contributions of the wheel and aggression, it will better illuminate whether or not individual housing or the wheel can act independently to trigger the circling behavior. Experiments to determine if aggression in the absence of a wheel could induce circling behavior would require strategies to provoke inter-male aggression. Aggression can be provoked between males by re-introducing individually housed males into a social setting. An individually housed male mouse would be re-introduced in an unfamiliar cage or arena (isolation-induced aggression paradigm) or introduced to another male mouse cage (resident-intruder paradigm) and specific aggressive behaviors would be scored both by live observation and video for a set period of time. Mice would have to undergo exposure to aggression on a regular basis, for example, twice a day, until circling behavior developed or a chosen end-point was reached. Control mice would have to also be individually housed and exposed to the unfamiliar arena or another male mouse cage; however, they would be put into the arena or cage alone.

The purpose of this study was to link the abnormal circling phenotype in this cohort with genetic and/or environmental contributors. The circling behavior in these mice was very obvious and permanent, meaning it was easily observed without the requirement of any special behavioral analysis equipment. We did not, however, fully characterize the circling behavior, which would involve more intensive behavioral studies in order to accurately quantify circling as well as assess any additional subtle behavioral abnormalities. Future studies could accomplish this by performing open-field tests, which quantitatively and qualitatively measures locomotion by placing a mouse in a cage apparatus that contains a video camera and/or infrared beams to aid in measuring locomotion. Employment of 24 hour video cameras could also help illuminate if other behaviors, such as feeding, grooming and reproduction, are affected in circling mice. A more complete understanding of the behavior in these mice would require a series of methodical

behavioral tests that measure general health, anxiety, hyperactivity, motor function, schizophrenia symptoms, depression, and learning and memory.

Fighting within cages and the resulting separation into individual housing are confounding variables in this study. Determining if a large proportion of males still develop circling behavior when individually housed at weaning with wheels and following up with biogenic amine measurements would aid in separating out these confounding variables. Maintaining social housing for males that are in cages with fighting so that they can be appropriately compared to non-separated control mice would be unethical. Fighting mice could obtain serious wounds and/or be killed by their cage-mates. In retrospect, an improved study design would have included a cage containing an equal number of socially housed age-matched male FVB/129P2 mice with cardboard housing for every cage of socially housed male FVB/129P2 male mice with wheels, and if either cage required separation due to fighting, the other cage would also be separated regardless of whether fighting was observed or not. Although these confounding variables, fighting and separation, do exist in combination with the presence of the running-wheel, we were still able to conclude that the circling behavior was the result of a genetic predisposition combined with an environmental trigger. Circling behavior was not observed in any FVB/129P2 male mice housed with cardboard houses. It is likely that the running-wheel is not directly triggering the behavior, but is rather contributing to a stressful environment that leads to this abnormal behavioral response in these particular mice, which may include stressful social situations including aggression and/or individual housing.

Chapter 6

General discussion and future directions

Neural tube defects (NTDs) demonstrate a complex, multifactorial etiology in humans. An elaborate interplay of hundreds of genes in several genetic pathways, along with the appropriate environmental conditions and epigenetic regulation, are required for the highly dynamic processes of neural plate induction, convergent extension, folding and fusion that ultimately results in the formation of the neural tube. Of all commonly used animal models of NTDs, mouse models best approximate human NTDs as mice are mammalian, mouse embryonic development occurs *in utero*, and mouse NTDs are also subject to a complex, multifactorial etiology. This thesis has shown that our mouse model of the cranial NTD exencephaly, which contains a homozygous mutation in *Cecr2*, demonstrates a complex etiology with genetic modifiers and possible environmental influences contributing to the penetrance of exencephaly (Figure 6.1.1). Furthermore, we have explored the potential contributions of human *CECR2* and human homologues of the mouse *Cecr2* candidate modifier genes to human cranial NTDs. An unrelated genetic predisposition to circling behavior in FVB/129P2 mice illustrated the need for the presence of an environmental trigger in order for the phenotype to develop, demonstrating a strong gene-environment interaction in these mice. Additionally, knockdown of the *Drosophila Cecr2* homologue, *dikar*, did not produce an observable phenotype; however, it is possible that mutations in additional, functionally redundant genes would be required in order for phenotype to manifest in *Drosophila*.

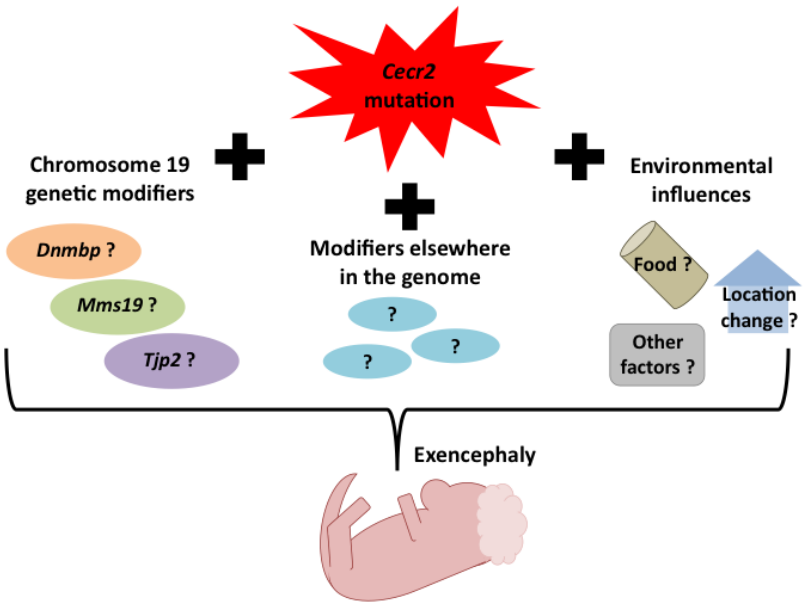


Figure 6.1.1: The *Cecr2* mutant mouse is a multifactorial model of exencephaly. Mutation in *Cecr2* requires additional factors in order for exencephaly to manifest. Modifier genes within the chromosome 19 modifier region, as well as elsewhere in the genome, contribute to *Cecr2*-associated exencephaly penetrance as mutation in *Cecr2* alone is insufficient to induce exencephaly in congenic $Cecr2^{GT45bic}$ FVB/N mice. Reduced penetrance over time in four mouse lines within our colony suggests environmental influences may contribute to *Cecr2*-associated exencephaly penetrance.

6.1: Analyzing exencephaly penetrance changes over time in *Cecr2* mutant mice

Incomplete penetrance of a disease phenotype in the presence of a genetic mutation is indicative of additional factors contributing to the manifestation of the disease. In other words, incomplete penetrance can be considered a feature of multifactorial diseases. Homozygous mutation in mouse *Cecr2* results in exencephaly with reduced penetrance. Previous penetrance analyses were performed on $Cecr2^{tm1.1Hemc}$ BALB/cCrl at generation N3-N4, with a penetrance of ~96% [142], and on $Cecr2^{GT45bic}$ FVB/N at generation N5-N6, with a penetrance of 0% [141]. Updated exencephaly penetrance analyses were completed for four different mouse lines (congenic $Cecr2^{GT45bic}$ BALB/cCrl, congenic $Cecr2^{tm1.1Hemc}$ FVB/N, $Cecr2^{GT45bic}$ MOD 5, and $Cecr2^{GT45bic}$ MOD 31), all of which demonstrated an exencephaly penetrance that dropped by ~15-20%. The $Cecr2^{tm1.1Hemc}$ allele has since been moved into the BALB/cCrl genetic background until congenic

(N10), and it is currently assumed that exencephaly penetrance remains very close to 100% due to the absence of homozygous mutant progeny at the time of genotyping (~2 weeks old). It is possible that the lack of viable homozygous mutant progeny in congenic *Cecr2*$^{tm1.1Hemc}$ BALB/cCrl may also be due to other causes in embryos non-penetrant for exencephaly. Also, the recent breeding strategy for maintaining this line involves a heterozygous parent mated with a homozygous wildtype parent. Therefore, no homozygous mutant progeny are currently being produced. Since the exencephaly penetrance in congenic *Cecr2*$^{tm1.1Hemc}$ BALB/cCrl is currently unknown, an updated penetrance analysis in these mice is called for in order to determine if exencephaly penetrance has also dropped or if it remained the same in this line.

Since the original penetrance analysis performed on incipient congenic *Cecr2*^{GT45bic} FVB/N, the *Cecr2*^{GT45bic} allele has been moved into the FVB/N background until congenic. Given that the previous analysis demonstrated 0% exencephaly at generation N6, mice within this line have been bred and maintained as homozygous mutants for *Cecr2*. However, it would be interesting to perform an updated exencephaly penetrance analysis on congenic *Cecr2*^{GT45bic} FVB/N, as changes may have occurred over time within this line as well. In fact, a single homozygous mutant *Cecr2*^{GT45bic} FVB/N pup was born with an encephalocele, which is a type of cranial NTD that is less severe than exencephaly and, in this case, was compatible with life at least up until ~2 weeks (Figure 6.1.2). Upon further investigation of the lesion, I discovered that the cranium failed to develop within a radius of ~3 mm around the lesion (data not shown), which provided evidence supporting that the lesion was in fact an encephalocele. Although the presence of a cranial NTD in this pup may be due to random chance, other cases could also have been missed if cannibalized by the mother perinatally. I therefore think it is still worthwhile to determine if cranial NTDs of any type are affecting the congenic *Cecr2*^{GT45bic} FVB/N mouse line by doing an appropriate penetrance analysis. I would recommend looking at a minimum of 100 mutant and 100 control embryos in order to pick up low frequency occurrences.

Figure 6.1.2: Encephalocele in a homozygous mutant congenic *Cecr2*^{GT45bic} FVB/N mouse pup (generation N10). This mouse strain was previously established to have an exencephaly penetrance of 0% at generation N5-N6 [141].

6.2: Identifying NTD modifier genes in susceptible genetic backgrounds in mice

The large difference in exencephaly penetrance between BALB/cCrl and FVB/N suggests the presence of modifier genes that predisposes BALB/cCrl to the development of exencephaly and protects FVB/N from exencephaly. This further exemplifies the validity of the *Cecr2* mutant mouse as a multifactorial model of NTDs. Strain differences in NTD penetrance is not unique to *Cecr2* mutant mice. A study performed by Korstanje et al. (2008) demonstrated differences in NTD penetrance in various mouse strains homozygous for the *vacuolated lens* (*vl*) mutation, and mapped five different modifier regions via whole genome linkage analysis [362]. Previous work in our lab has provided evidence that FVB/N harbors a general resistance to the development of NTDs, as demonstrated by a lower penetrance of NTDs when homozygous mutant for the NTD genes *Cecr2*, *Pax3* and *shroom* [164]. FVB/N is not necessarily resistant to the development of all types of NTDs. For example, Pani et al. (2002) showed that FVB/N mice are susceptible to NTDs that occur in maternal diabetic pregnancies relative to C57BL/6J [363]. Decades of inbreeding along with highly controlled environmental settings have produced individuals within a mouse strain that possess very little variability between each other. Therefore, it could be argued that studying an NTD causing mutation in multiple mice within a single mouse strain is analogous to studying the same NTD factor in several clones of a single human being. Although such studies in mouse are invaluable for characterizing the molecular contributions of the NTD factor under question, other genetic factors contributing to the NTD phenotype in these mice cannot be accounted for or addressed. Differences in penetrance between mouse strains, along with the responsible modifying factors, have yet to be studied for many mouse NTD genes. Future endeavors in this direction would shed light on the complex genetic architecture contributing to NTDs in mouse as well as more appropriately model the multifactorial nature of NTDs in a highly variable human population.

An effort to identify genetic factors that modify *Cecr2*-associated exencephaly and the potential role that *Cecr2* and its candidate modifiers play in human NTD etiology was a major focus of this thesis. A combination of whole genome microarray and whole exome sequence analyses aided in producing a list of candidate modifier genes within the chromosome 19 region, and genetic analyses in mouse revealed that the top candidate modifier gene, *Arhgap19*, was not a modifier of *Cecr2*-associated exencephaly. Targeted sequencing of human *CECR2* and human

homologues of the remaining candidate modifier genes aided in identifying strong candidate NTD genes in both mice and humans. I identified genetic variants within *CECR2*, *DNMBP*, *MMS19*, *TJP2*, as well as additional candidate modifier genes, which may have contributed to the development of a cranial NTD in the probands containing these variants. All variants of interest identified in probands within the human cranial NTD cohort have been seen in the normal population of the exome sequencing project (esp6500) and/or at least one normal parent. Not surprisingly, this means that none of the variants of interest within the NTD cohort are capable of causing NTDs with 100% penetrance on their own. If any of these variants are indeed located within NTD genes, they require additional susceptibility factors in order for phenotype to manifest. It is difficult to know whether only two or three factors are all that is required for a NTD to manifest, or if each proband was exposed to tens or even hundreds of susceptibility factors. Based on the complexity of the process of neurulation, which includes the involvement of hundreds of genes and can be influenced by multiple environmental factors, it is probable that each NTD case has a unique set of susceptibility factors, with some cases being exposed to a few factors with high impact and other cases being exposed to many factors that each on their own contribute a small impact. It is also likely that each proband contained more than one genetic factor that contributed to the development of exencephaly. There were several probands within the cranial NTD cohort that contained protein-coding variants of interest in more than one of the 25 genes sequenced (*CECR2* and 24 candidate modifier genes). Of 156 probands sequenced, 74 probands contained variants of interest within at least one of the 25 genes sequenced. 24 of these 74 probands contained variants of interest in more than one gene sequenced, 6 of which contained variants of interest in three or more genes. When looking specifically at the 17 probands that contained variants in *CECR2*, 6 of these probands also contained variants in at least one of the candidate modifier genes. Future analyses could involve accessing additional information in the large esp6500 database to determine the frequency that two variants within different genes co-exist within the same person (trans-heterozygosity). If two variants in two different genes were co-inherited with a frequency lower than expected, this would suggest that this combination of variants is deleterious and therefore selected against. Experiments to be performed in the near future include *in vitro* and *in vivo* functional assays in order to test the impact of identified variants on protein function. In the event that a variant is shown to be functionally deleterious, that variant can be genotyped or the gene in which that variant was

identified can be sequenced in additional human samples. The 156 probands sequenced in this study are part of a larger cranial NTD cohort that contains an additional 1050 sampled individuals, which include affected and unaffected family members. The entire Duke NTD cohort consists of 5232 sampled individuals, which is inclusive of the cranial NTD cohort discussed above as well as patients diagnosed with other NTD subtypes and their unaffected family members. Genotyping or sequencing of additional human NTD samples may identify an enrichment of the variant of interest and/or lead to the discovery of additional variants within the gene of interest. Identifying an association with NTD cases alongside characterizing a functional impact would ultimately lead to the discovery of a novel NTD gene in humans.

All genetic analyses and experiments directed towards identification and characterization of candidate modifier genes were restricted to candidate genes located within the chromosome 19 modifier region, which was responsible for only a partial rescue of the exencephaly phenotype from ~54% in BALB/cCrl to the ~33% in the MOD lines. This means that other modifiers that provide NTD resistance exist elsewhere within the FVB/N genetic background. Whole genome linkage analysis identified only one other modifier region, which was on the X chromosome and had an LOD score above the suggestive threshold but below the significance threshold (LOD = 3.04) [164]. Whole genome microarray identified four genes of interest within this region that significantly differ in expression by at least 1.5 fold between BALB/cCrl and FVB/N. These four genes are *Fmr1nb*, *Rab39b*, *Chm*, and *Col4a6*. If efforts shift to identify modifiers outside of the chromosome 19 region, future studies could include the validation and analysis of these four genes, as was done for candidate genes within the chromosome 19 modifier region. It is also possible that additional small-effect modifier genes are distributed throughout the genome. For example, while investigating the cholesterol gene *Lipa* as a potential modifier (later excluded due to the discovery of an artifact in the microarray results), I discovered differences in expression of several genes in the cholesterol synthesis pathway between BALB/cCrl and FVB/N as indicated by microarray. These differences revealed that FVB/N embryos at the time of neurulation appear to have an overall down-regulation of the cholesterol synthesis pathway relative to BALB/cCrl (Figure 6.2.1). Therefore, the NTD resistance seen in FVB/N may be partially due to reduced levels of cholesterol in the neurulating embryo. One theory as to how reduced cholesterol levels protect against the development of NTDs involves the role of cholesterol in Sonic hedgehog (Shh) signaling. Excessive amounts of Shh signaling causes exencephaly, most likely due to the

impairment of dorso-lateral hinge point (DHLP) formation (reviewed in [364]). Impaired Shh signaling due to homozygous mutation in *Shh* does not cause NTDs in mice, but rather results in cyclopia, which is a severe form of holoprosencephaly, a condition where the cerebral hemispheres fail to separate [365]. Although *Shh* itself was shown not to differ in expression levels between BALB/cCrl and FVB/N in the microarray analysis, Shh is known to undergo post-translational modifications that include the addition of a cholesterol moiety [366]. This cholesterol moiety affects the ability of Shh to diffuse to neighboring target cells, with the presence of the hydrophobic cholesterol moiety reducing the range of diffusion compared to the absence of the moiety [367]. This means that Shh with the cholesterol moiety would have a higher concentration closer to the source of diffusion, whereas Shh with the absence of the moiety would allow for a broader range of diffusion. Therefore, it is possible that embryos with lower cholesterol levels (FVB/N) have a reduction in the amount of Shh that contains the cholesterol moiety, which would lead to wider ranges of Shh diffusion that could result in a lower concentration of Shh at the target DLHPs. Therefore, Shh signaling would be reduced at the DLHPs, which would promote DLHP formation and ultimately protect against the development of exencephaly. It is important to note that the premise of this hypothesis is based on a reduction and not a complete abolition of cholesterol synthesis having a protective effect in FVB/N. In fact, human inborn errors of cholesterol synthesis result in severe phenotypic anomolies including holoprosencephaly, which suggests that Shh signaling is affected by the cholesterol synthesis pathway (reviewed in [368]). A study by Cooper et al. (2003) demonstrated that a reduction in cholesterol inhibited the response to Hh signaling in cell culture, and therefore postulated that it is an inhibition of Shh response, rather than the lack of a cholesterol moiety, that results in the phenotypes observed in patients with inborn errors of cholesterol synthesis [369]. Either way, the literature seems to agree that impaired cholesterol synthesis results in a reduction in Shh signaling. This fits with my hypothesis, which states that a mild reduction in cholesterol synthesis in FVB/N can lower Shh signaling, which then can act as a protective effect by allowing for better DLHP formation. One way to test this hypothesis would be to compare Shh diffusion patterns using immunofluorescence between BALB/cCrl and FVB/N embryos at the time of neurulation. Also, if this were the case, reducing cholesterol levels in homozygous mutant congenic *Cecr2*^{GT45bic} BALB/cCrl embryos should partially rescue the exencephaly phenotype. Increasing cholesterol levels in homozygous mutant congenic *Cecr2*^{GT45bic} FVB/N embryos may

predispose to the development of exencephaly; however, increased cholesterol levels would need to be sufficient to breach the 0% exencephaly penetrance threshold in these mice.

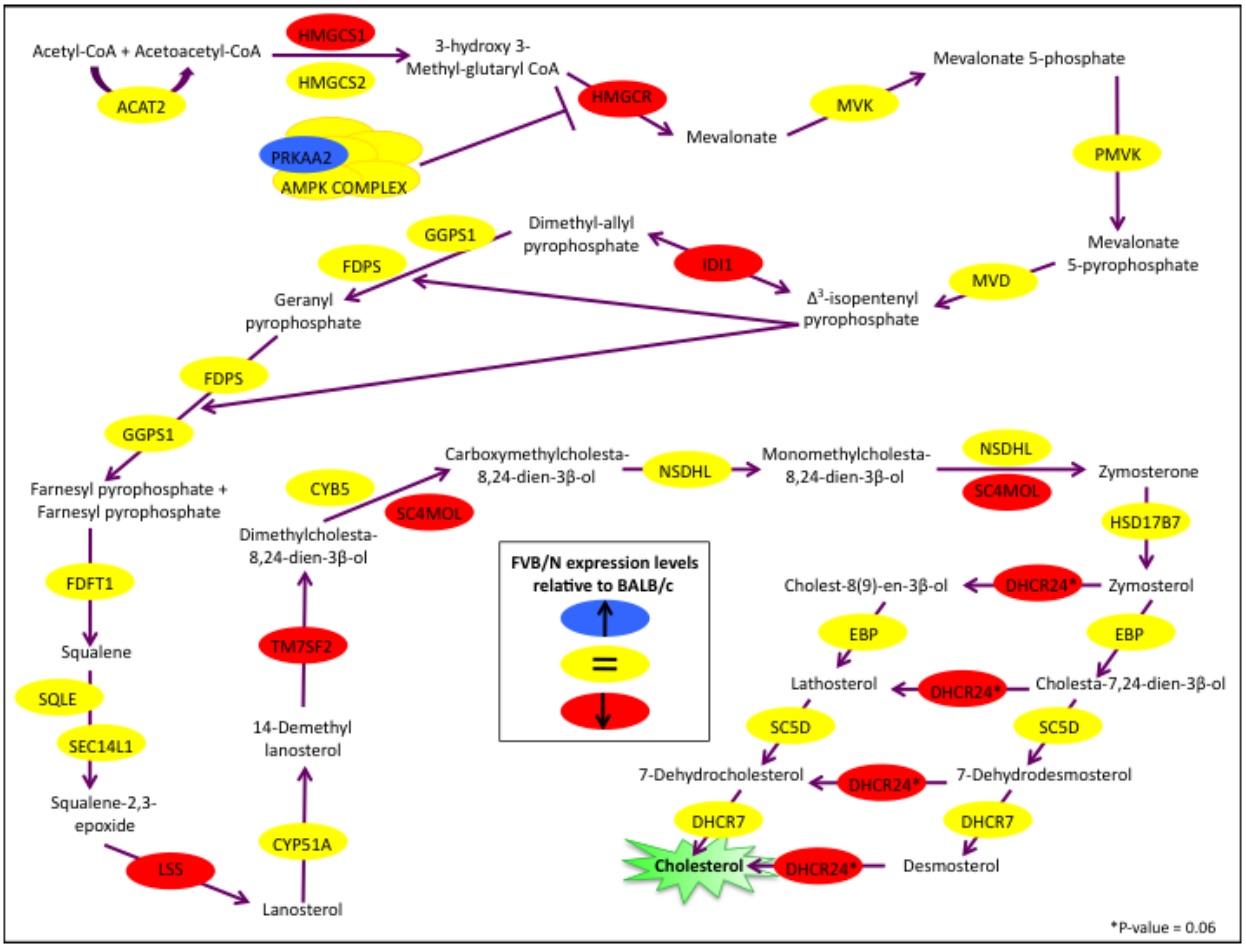


Figure 6.2.1: The cholesterol synthesis pathway appears to be down-regulated in FVB/N relative to BALB/cCrl neurulating embryos. Enzymes in which expression levels are similar between strains at E8.5 are indicated in yellow. Enzymes downregulated by at least 1.5 fold (*P*-value < 0.05, exception DHCR24 *P*-value = 0.06*) in E8.5 FVB/N relative to E8.5 BALB/c are indicated in red. The only enzyme that is up-regulated by at least 1.5 fold (*P*-value < 0.05) in E8.5 FVB/N relative to E8.5 BALB/c is indicated in blue and is part of a complex that inhibits the cholesterol synthesis pathway. Gene expression data was acquired from the whole genome microarray analysis performed by Fairbridge et al. (2010) [145]. Cholesterol pathway information was obtained from Wilcox et al. (2007) [370].

6.3: Gene-environment interactions contributing to abnormal mouse behavior

The development of abnormal circling behavior specifically in the FVB/129P2 cohort of male mice exemplified a gene-environment interaction within mice in our colony, as the exposure to an environmental trigger or combination of triggers in these genetically predisposed mice was required for phenotype to manifest. It is possible that provision of running-wheels alone or in conjunction with adverse social interactions (aggression and/or isolation) can act as an environmental trigger, which then allows for these behavioral and brain biochemistry abnormalities to manifest. Based on our results and similar cases reported in the literature [308,309], running-wheels should be introduced with caution to mice. If not for the obvious circling behavior in FVB/129P2 male mice, we would not have known that wheels could adversely affect mouse behavior. This implies that running-wheels could also be having low-level effects in other strains, which could ultimately result in unnoticed but significant changes in an experiment. We have therefore removed the wheels from all mice in our colony. We recommend opting for environmental enrichments such as nesting material and cardboard housing, which did not cause aggression or circling behaviour in either sex of all strains of mice in our colony (Balb/cCrl, FVB/N, C57BL/6J). In a broader sense, this finding taken together with the drop in exencephaly penetrance in four independent mouse lines, demonstrates the impact that environmental changes within a lab setting can have on phenotype manifestation in genetic mouse models. For example, psychiatric disorders such as ADHD and schizophrenia are hypothesized to have both genetic factors and environmental conditions that work in concert to promote the manifestation of these diseases. Research exploring the genetics or gene-environment interactions in psychiatric disorders often involves the use of mouse models. Alterations in brain biogenic amines, along with circling behavior, occurs in rodent models of ADHD [329] and schizophrenia [331,332]. However, when attempting to generate mouse models and analyze phenotype for psychiatric diseases, environmental conditions should be taken in to consideration, as the absence of a phenotype in a rodent genetic model may simply be due to the lack of an environmental trigger.

6.4: Characterizing *dikar*, the *Drosophila* homologue of *Cecr2*

Drosophila melanogaster is an invaluable genetic model for the characterization of developmental genes and genetic pathways. Therefore, I performed phenotypic analyses on flies where the *Drosophila Cecr2* homologue, *dikar*, was knocked down via RNAi or homozygous for a deletion mutation. No clear, reliable phenotype was seen associated with *dikar* loss or reduction in any experiments performed. We were unable to rule out IR sensitivity as a mutant phenotype due to a possible second-site IR sensitivity locus, which may confer IR sensitive phenotype on the chromosome bearing the *dikar* mutations. Although no phenotype was observed, it is still possible that *dikar* plays an important albeit functionally redundant role in development. It is interesting that a *dikar* mutant phenotype may only manifest in the presence of another genetic mutation, for example, in the presence of an *Acf1* mutation, or the appropriate environmental insult, such as IR. If this were the case, it would be yet another example of the complex contributions of genetic and environmental interactions on resulting phenotypes.

6.5: Concluding remarks

Exencephaly penetrance in our *Cecr2* mutant mouse model of NTDs is drastically affected by more than one genetic modifier within the chromosome 19 modifier region, additional genetic modifiers throughout the genome, as well as possible environmental influences. This establishes our *Cecr2* mutant mouse as a good model for exploring the multifactorial nature of NTDs. Further characterization of the chromosome 19 modifier region eliminated the top candidate modifier gene, *Arhgap19*, as a modifier, as well as identified additional candidate modifier genes. Sequencing of human homologues of *Cecr2* and candidate modifiers of *Cecr2* in a cranial NTD cohort identified variants of interest, particularly within *CECR2*, as well as *DNMBP*, *MMS19*, and *TJP2*. Future investigation of the functional impact of these variants could lead to the identification of one or more modifier genes within the *Cecr2* mutant mouse model, as well as a novel NTD gene in humans.

Characterization of the circling phenotype in FVB/129P2 male mice, which demonstrated a genetic predisposition that required an environmental trigger in order for phenotype to manifest, further underscored the importance of taking genomic and environmental influences into account

when investigating genetic disease. Minimizing genetic and environmental variation is important for studying the effects of a gene mutation; however, recognizing and characterizing how these variables impact a phenotype in a genetic model will provide invaluable insight into the complex underpinnings of human diseases such as psychiatric disorders and NTDs.

Bibliography

1. Greene, N.D. & Copp, A.J. Neural tube defects. *Annu Rev Neurosci* **37**, 221-42 (2014).
2. Groves, A.K. & LaBonne, C. Setting appropriate boundaries: fate, patterning and competence at the neural plate border. *Dev Biol* **389**, 2-12 (2014).
3. Xu, X. *et al.* Murine fibroblast growth factor receptor 1alpha isoforms mediate node regression and are essential for posterior mesoderm development. *Dev Biol* **208**, 293-306 (1999).
4. Greco, T.L., Sussman, D.J. & Camper, S.A. Dishevelled-2 maps to human chromosome 17 and distal to Wnt3a and vestigial tail (vt) on mouse chromosome 11. *Mamm Genome* **7**, 475-6 (1996).
5. Yamaguchi, T.P., Bradley, A., McMahon, A.P. & Jones, S. A Wnt5a pathway underlies outgrowth of multiple structures in the vertebrate embryo. *Development* **126**, 1211-23 (1999).
6. Qian, D. *et al.* Wnt5a functions in planar cell polarity regulation in mice. *Dev Biol* **306**, 121-33 (2007).
7. Castranio, T. & Mishina, Y. Bmp2 is required for cephalic neural tube closure in the mouse. *Dev Dyn* **238**, 110-22 (2009).
8. Solloway, M.J. & Robertson, E.J. Early embryonic lethality in Bmp5;Bmp7 double mutant mice suggests functional redundancy within the 60A subgroup. *Development* **126**, 1753-68 (1999).
9. Schorle, H., Meier, P., Buchert, M., Jaenisch, R. & Mitchell, P.J. Transcription factor AP-2 essential for cranial closure and craniofacial development. *Nature* **381**, 235-8 (1996).
10. Depew, M.J. *et al.* Dlx5 regulates regional development of the branchial arches and sensory capsules. *Development* **126**, 3831-46 (1999).
11. Machado, A.F., Martin, L.J. & Collins, M.D. Pax3 and the splotch mutations: structure, function, and relationship to teratogenesis, including gene-chemical interactions. *Curr Pharm Des* **7**, 751-85 (2001).
12. Ishii, M. *et al.* Combined deficiencies of Msx1 and Msx2 cause impaired patterning and survival of the cranial neural crest. *Development* **132**, 4937-50 (2005).
13. Lallemand, Y. *et al.* Analysis of Msx1; Msx2 double mutants reveals multiple roles for Msx genes in limb development. *Development* **132**, 3003-14 (2005).
14. Keller, R. Shaping the vertebrate body plan by polarized embryonic cell movements. *Science* **298**, 1950-4 (2002).
15. Wilde, J.J., Petersen, J.R. & Niswander, L. Genetic, epigenetic, and environmental contributions to neural tube closure. *Annu Rev Genet* **48**, 583-611 (2014).
16. Margolis, B. & Borg, J.P. Apicobasal polarity complexes. *J Cell Sci* **118**, 5157-9 (2005).
17. Eom, D.S., Amarnath, S., Fogel, J.L. & Agarwala, S. Bone morphogenetic proteins regulate neural tube closure by interacting with the apicobasal polarity pathway. *Development* **138**, 3179-88 (2011).
18. Ybot-Gonzalez, P., Cogram, P., Gerrelli, D. & Copp, A.J. Sonic hedgehog and the molecular regulation of mouse neural tube closure. *Development* **129**, 2507-17 (2002).

19. Ybot-Gonzalez, P. *et al.* Neural plate morphogenesis during mouse neurulation is regulated by antagonism of Bmp signalling. *Development* **134**, 3203-11 (2007).
20. Haigo, S.L., Hildebrand, J.D., Harland, R.M. & Wallingford, J.B. Shroom induces apical constriction and is required for hingepoint formation during neural tube closure. *Curr Biol* **13**, 2125-37 (2003).
21. Geelen, J.A. & Langman, J. Ultrastructural observations on closure of the neural tube in the mouse. *Anat Embryol (Berl)* **156**, 73-88 (1979).
22. Pyrgaki, C., Trainor, P., Hadjantonakis, A.K. & Niswander, L. Dynamic imaging of mammalian neural tube closure. *Dev Biol* **344**, 941-7 (2010).
23. Hall, A. & Nobes, C.D. Rho GTPases: molecular switches that control the organization and dynamics of the actin cytoskeleton. *Philos Trans R Soc Lond B Biol Sci* **355**, 965-70 (2000).
24. Holmberg, J., Clarke, D.L. & Frisen, J. Regulation of repulsion versus adhesion by different splice forms of an Eph receptor. *Nature* **408**, 203-6 (2000).
25. Abdul-Aziz, N.M., Turmaine, M., Greene, N.D. & Copp, A.J. EphrinA-EphA receptor interactions in mouse spinal neurulation: implications for neural fold fusion. *Int J Dev Biol* **53**, 559-68 (2009).
26. O'Rahilly, R. & Muller, F. The two sites of fusion of the neural folds and the two neuropores in the human embryo. *Teratology* **65**, 162-70 (2002).
27. Juriloff, D.M., Harris, M.J., Tom, C. & MacDonald, K.B. Normal mouse strains differ in the site of initiation of closure of the cranial neural tube. *Teratology* **44**, 225-33 (1991).
28. Macdonald, K.B., Juriloff, D.M. & Harris, M.J. Developmental study of neural tube closure in a mouse stock with a high incidence of exencephaly. *Teratology* **39**, 195-213 (1989).
29. Copp, A.J. Neurulation in the cranial region--normal and abnormal. *J Anat* **207**, 623-35 (2005).
30. Van Allen, M.I. *et al.* Evidence for multi-site closure of the neural tube in humans. *Am J Med Genet* **47**, 723-43 (1993).
31. Nakatsu, T., Uwabe, C. & Shiota, K. Neural tube closure in humans initiates at multiple sites: evidence from human embryos and implications for the pathogenesis of neural tube defects. *Anat Embryol (Berl)* **201**, 455-66 (2000).
32. Sulik, K.K. *et al.* Normal patterns of neural tube closure differ in the human and the mouse. *Proc Greenwood Genet Center* **18**, 129-130 (1999).
33. Pai, Y.J. *et al.* Epithelial fusion during neural tube morphogenesis. *Birth Defects Res A Clin Mol Teratol* **94**, 817-23 (2012).
34. Massa, V. *et al.* Apoptosis is not required for mammalian neural tube closure. *Proc Natl Acad Sci U S A* **106**, 8233-8 (2009).
35. Copp, A.J. & Brook, F.A. Does lumbosacral spina bifida arise by failure of neural folding or by defective canalisation? *J Med Genet* **26**, 160-6 (1989).
36. Schoenwolf, G.C. Histological and ultrastructural studies of secondary neurulation in mouse embryos. *Am J Anat* **169**, 361-76 (1984).
37. Shimokita, E. & Takahashi, Y. Secondary neurulation: Fate-mapping and gene manipulation of the neural tube in tail bud. *Dev Growth Differ* **53**, 401-10 (2011).
38. Lew, S.M. & Kothbauer, K.F. Tethered cord syndrome: an updated review. *Pediatr Neurosurg* **43**, 236-48 (2007).

39. Greene, N.D., Stanier, P. & Copp, A.J. Genetics of human neural tube defects. *Hum Mol Genet* **18**, R113-29 (2009).
40. Botto, L.D., Moore, C.A., Khoury, M.J. & Erickson, J.D. Neural-tube defects. *N Engl J Med* **341**, 1509-19 (1999).
41. Melvin, E.C. *et al.* Genetic studies in neural tube defects. NTD Collaborative Group. *Pediatr Neurosurg* **32**, 1-9 (2000).
42. Au, K.S., Ashley-Koch, A. & Northrup, H. Epidemiologic and genetic aspects of spina bifida and other neural tube defects. *Dev Disabil Res Rev* **16**, 6-15 (2010).
43. Johnson, K.M., Suarez, L., Felkner, M.M. & Hendricks, K. Prevalence of craniorachischisis in a Texas-Mexico border population. *Birth Defects Res A Clin Mol Teratol* **70**, 92-4 (2004).
44. Moore, C.A. *et al.* Elevated rates of severe neural tube defects in a high-prevalence area in northern China. *Am J Med Genet* **73**, 113-8 (1997).
45. Detrait, E.R. *et al.* Human neural tube defects: developmental biology, epidemiology, and genetics. *Neurotoxicol Teratol* **27**, 515-24 (2005).
46. Rampersaud, E. *et al.* Whole genomewide linkage screen for neural tube defects reveals regions of interest on chromosomes 7 and 10. *J Med Genet* **42**, 940-6 (2005).
47. Yan, L. *et al.* Association of the maternal MTHFR C677T polymorphism with susceptibility to neural tube defects in offsprings: evidence from 25 case-control studies. *PLoS One* **7**, e41689 (2012).
48. Frosst, P. *et al.* A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nat Genet* **10**, 111-3 (1995).
49. De Marco, P. *et al.* Reduced folate carrier polymorphism (80A-->G) and neural tube defects. *Eur J Hum Genet* **11**, 245-52 (2003).
50. Pei, L., Liu, J., Zhang, Y., Zhu, H. & Ren, A. Association of reduced folate carrier gene polymorphism and maternal folic acid use with neural tube defects. *Am J Med Genet B Neuropsychiatr Genet* **150B**, 874-8 (2009).
51. Franke, B. *et al.* An association study of 45 folate-related genes in spina bifida: Involvement of cubilin (CUBN) and tRNA aspartic acid methyltransferase 1 (TRDMT1). *Birth Defects Res A Clin Mol Teratol* **85**, 216-26 (2009).
52. Shaw, G.M. *et al.* 118 SNPs of folate-related genes and risks of spina bifida and conotruncal heart defects. *BMC Med Genet* **10**, 49 (2009).
53. Martinez, C.A. *et al.* Genetic association study of putative functional single nucleotide polymorphisms of genes in folate metabolism and spina bifida. *Am J Obstet Gynecol* **201**, 394 e1-11 (2009).
54. Parle-McDermott, A. *et al.* A common variant in MTHFD1L is associated with neural tube defects and mRNA splicing efficiency. *Hum Mutat* **30**, 1650-6 (2009).
55. Enaw, J.O. *et al.* CHKA and PCYT1A gene polymorphisms, choline intake and spina bifida risk in a California population. *BMC Med* **4**, 36 (2006).
56. Jensen, L.E., Hoess, K., Whitehead, A.S. & Mitchell, L.E. The NAT1 C1095A polymorphism, maternal multivitamin use and smoking, and the risk of spina bifida. *Birth Defects Res A Clin Mol Teratol* **73**, 512-6 (2005).
57. Zhu, H. *et al.* A known functional polymorphism (Ile120Val) of the human PCMT1 gene and risk of spina bifida. *Mol Genet Metab* **87**, 66-70 (2006).
58. Davidson, C.M. *et al.* Genes in glucose metabolism and association with spina bifida. *Reprod Sci* **15**, 51-8 (2008).

59. Wen, S. *et al.* Genetic polymorphisms in the thioredoxin 2 (TXN2) gene and risk for spina bifida. *Am J Med Genet A* **149A**, 155-60 (2009).
60. Olshan, A.F., Shaw, G.M., Millikan, R.C., Laurent, C. & Finnell, R.H. Polymorphisms in DNA repair genes as risk factors for spina bifida and orofacial clefts. *Am J Med Genet A* **135**, 268-73 (2005).
61. Volcik, K.A. *et al.* Testing for genetic associations with the PAX gene family in a spina bifida population. *Am J Med Genet* **110**, 195-202 (2002).
62. Lu, W. *et al.* Screening for novel PAX3 polymorphisms and risks of spina bifida. *Birth Defects Res A Clin Mol Teratol* **79**, 45-9 (2007).
63. Shields, D.C. *et al.* Association between historically high frequencies of neural tube defects and the human T homologue of mouse T (Brachyury). *Am J Med Genet* **92**, 206-11 (2000).
64. Deak, K.L. *et al.* SNPs in the neural cell adhesion molecule 1 gene (NCAM1) may be associated with human neural tube defects. *Hum Genet* **117**, 133-42 (2005).
65. Gibson, G. Rare and common variants: twenty arguments. *Nat Rev Genet* **13**, 135-45 (2011).
66. Lemay, P. *et al.* Loss-of-function de novo mutations play an important role in severe human neural tube defects. *J Med Genet* (2015).
67. Kibar, Z. *et al.* Mutations in VANGL1 associated with neural-tube defects. *N Engl J Med* **356**, 1432-7 (2007).
68. Lei, Y.P. *et al.* VANGL2 mutations in human cranial neural-tube defects. *N Engl J Med* **362**, 2232-5 (2010).
69. Robinson, A. *et al.* Mutations in the planar cell polarity genes CELSR1 and SCRIB are associated with the severe neural tube defect craniorachischisis. *Hum Mutat* **33**, 440-7 (2012).
70. Seo, J.H. *et al.* Mutations in the planar cell polarity gene, Fuzzy, are associated with neural tube defects in humans. *Hum Mol Genet* **20**, 4324-33 (2011).
71. Shi, Y. *et al.* Identification of novel rare mutations of DACT1 in human neural tube defects. *Hum Mutat* **33**, 1450-5 (2012).
72. Allache, R. *et al.* Novel mutations in Lrp6 orthologs in mouse and human neural tube defects affect a highly dosage-sensitive Wnt non-canonical planar cell polarity pathway. *Hum Mol Genet* **23**, 1687-99 (2014).
73. Bosoi, C.M. *et al.* Identification and characterization of novel rare mutations in the planar cell polarity gene PRICKLE1 in human neural tube defects. *Hum Mutat* **32**, 1371-5 (2011).
74. Allache, R., De Marco, P., Merello, E., Capra, V. & Kibar, Z. Role of the planar cell polarity gene CELSR1 in neural tube defects and caudal agenesis. *Birth Defects Res A Clin Mol Teratol* **94**, 176-81 (2012).
75. De Marco, P. *et al.* FZD6 is a novel gene for human neural tube defects. *Hum Mutat* **33**, 384-90 (2012).
76. Kibar, Z. *et al.* Novel mutations in VANGL1 in neural tube defects. *Hum Mutat* **30**, E706-15 (2009).
77. Kibar, Z. *et al.* Contribution of VANGL2 mutations to isolated neural tube defects. *Clin Genet* **80**, 76-82 (2011).
78. De Marco, P. *et al.* Genetic analysis of disheveled 2 and disheveled 3 in human neural tube defects. *J Mol Neurosci* **49**, 582-8 (2013).

79. Lei, Y. *et al.* Rare LRP6 variants identified in spina bifida patients. *Hum Mutat* **36**, 342-9 (2015).
80. Chen, X. *et al.* Detection of copy number variants reveals association of cilia genes with neural tube defects. *PLoS One* **8**, e54492 (2013).
81. Thiersch, J.B. Therapeutic abortions with a folic acid antagonist, 4-aminopteroylglutamic acid (4-amino P.G.A) administered by the oral route. *Am J Obstet Gynecol* **63**, 1298-304 (1952).
82. Hibbard, B.M., Hibbard, E.D. & Jeffcoate, T.N. Folic acid and reproduction. *Acta Obstet Gynecol Scand* **44**, 375-400 (1965).
83. Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. MRC Vitamin Study Research Group. *Lancet* **338**, 131-7 (1991).
84. Czeizel, A.E. & Dudas, I. Prevention of the first occurrence of neural-tube defects by periconceptional vitamin supplementation. *N Engl J Med* **327**, 1832-5 (1992).
85. Wald, N.J., Law, M.R., Morris, J.K. & Wald, D.S. Quantifying the effect of folic acid. *Lancet* **358**, 2069-73 (2001).
86. Canada Gazette part II: regulatory impact analysis statement, SOR/98-550. **132(24)**(1998).
87. De Wals, P. *et al.* Reduction in neural-tube defects after folic acid fortification in Canada. *N Engl J Med* **357**, 135-42 (2007).
88. FDA. Food Standards: Amendment of Standards of Identity for Enriched Grain Products to Require Addition of Folic Acid. *Vol. 61. Washington, DC: FDA* (1996).
89. Obican, S.G., Finnell, R.H., Mills, J.L., Shaw, G.M. & Scialli, A.R. Folic acid in early pregnancy: a public health success story. *FASEB J* **24**, 4167-74 (2010).
90. Youngblood, M.E. *et al.* 2012 Update on global prevention of folic acid-preventable spina bifida and anencephaly. *Birth Defects Res A Clin Mol Teratol* **97**, 658-63 (2013).
91. Fleischman, A.R. & Oinuma, M. Fortification of corn masa flour with folic acid in the United States. *Am J Public Health* **101**, 1360-4 (2011).
92. Zabihi, S. & Loeken, M.R. Understanding diabetic teratogenesis: where are we now and where are we going? *Birth Defects Res A Clin Mol Teratol* **88**, 779-90 (2010).
93. Stothard, K.J., Tennant, P.W., Bell, R. & Rankin, J. Maternal overweight and obesity and the risk of congenital anomalies: a systematic review and meta-analysis. *JAMA* **301**, 636-50 (2009).
94. Dreier, J.W., Andersen, A.M. & Berg-Beckhoff, G. Systematic review and meta-analyses: fever in pregnancy and health impacts in the offspring. *Pediatrics* **133**, e674-88 (2014).
95. Chambers, C.D. Risks of hyperthermia associated with hot tub or spa use by pregnant women. *Birth Defects Res A Clin Mol Teratol* **76**, 569-73 (2006).
96. Correa, A. *et al.* Lack of periconceptional vitamins or supplements that contain folic acid and diabetes mellitus-associated birth defects. *Am J Obstet Gynecol* **206**, 218 e1-13 (2012).
97. Parker, S.E., Yazdy, M.M., Tinker, S.C., Mitchell, A.A. & Werler, M.M. The impact of folic acid intake on the association among diabetes mellitus, obesity, and spina bifida. *Am J Obstet Gynecol* **209**, 239 e1-8 (2013).
98. Lacasana, M. *et al.* Effect on risk of anencephaly of gene-nutrient interactions between methylenetetrahydrofolate reductase C677T polymorphism and maternal

folate, vitamin B12 and homocysteine profile. *Public Health Nutr* **15**, 1419-28 (2012).
99. Miranda, T.B. & Jones, P.A. DNA methylation: the nuts and bolts of repression. *J Cell Physiol* **213**, 384-90 (2007).
100. Beaudin, A.E. & Stover, P.J. Folate-mediated one-carbon metabolism and neural tube defects: balancing genome synthesis and gene expression. *Birth Defects Res C Embryo Today* **81**, 183-203 (2007).
101. Barua, S. *et al.* Single-base resolution of mouse offspring brain methylome reveals epigenome modifications caused by gestational folic acid. *Epigenetics Chromatin* **7**, 3 (2014).
102. Bannister, A.J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res* **21**, 381-95 (2011).
103. Shyamasundar, S. *et al.* Analysis of epigenetic factors in mouse embryonic neural stem cells exposed to hyperglycemia. *PLoS One* **8**, e65945 (2013).
104. Kappen, C. & Salbaum, J.M. Gene expression in teratogenic exposures: a new approach to understanding individual risk. *Reprod Toxicol* **45**, 94-104 (2014).
105. Gu, H. *et al.* Diagnostic role of microRNA expression profile in the serum of pregnant women with fetuses with neural tube defects. *J Neurochem* **122**, 641-9 (2012).
106. Juriloff, D.M. & Harris, M.J. Hypothesis: the female excess in cranial neural tube defects reflects an epigenetic drag of the inactivating X chromosome on the molecular mechanisms of neural fold elevation. *Birth Defects Res A Clin Mol Teratol* **94**, 849-55 (2012).
107. Brook, F.A., Estibeiro, J.P. & Copp, A.J. Female predisposition to cranial neural tube defects is not because of a difference between the sexes in the rate of embryonic growth or development during neurulation. *J Med Genet* **31**, 383-7 (1994).
108. Wallace, M.E., Knights, P.J. & Anderson, J.R. Inheritance and morphology of exencephaly, a neonatal lethal recessive with partial penetrance, in the house mouse. *Genet Res* **32**, 135-49 (1978).
109. Chen, X. *et al.* Sex difference in neural tube defects in p53-null mice is caused by differences in the complement of X not Y genes. *Dev Neurobiol* **68**, 265-73 (2008).
110. Carter, M., Ulrich, S., Oofuji, Y., Williams, D.A. & Ross, M.E. Crooked tail (Cd) models human folate-responsive neural tube defects. *Hum Mol Genet* **8**, 2199-204 (1999).
111. Harris, M.J. Insights into prevention of human neural tube defects by folic acid arising from consideration of mouse mutants. *Birth Defects Res A Clin Mol Teratol* **85**, 331-9 (2009).
112. Gray, J.D. *et al.* Functional interactions between the LRP6 WNT co-receptor and folate supplementation. *Hum Mol Genet* **19**, 4560-72 (2010).
113. Marean, A., Graf, A., Zhang, Y. & Niswander, L. Folic acid supplementation can adversely affect murine neural tube closure and embryonic survival. *Hum Mol Genet* **20**, 3678-83 (2011).
114. Oyama, K. *et al.* Folic acid prevents congenital malformations in the offspring of diabetic mice. *Endocr J* **56**, 29-37 (2009).
115. Yuan, Q. *et al.* Folic acid supplementation changes the fate of neural progenitors in mouse embryos of hyperglycemic and diabetic pregnancy. *J Nutr Biochem* **24**, 1202-12 (2013).

116. Zohn, I.E. Mouse as a model for multifactorial inheritance of neural tube defects. *Birth Defects Res C Embryo Today* **96**, 193-205 (2012).
117. Nusslein-Volhard, C. & Wieschaus, E. Mutations affecting segment number and polarity in Drosophila. *Nature* **287**, 795-801 (1980).
118. Varjosalo, M. & Taipale, J. Hedgehog: functions and mechanisms. *Genes Dev* **22**, 2454-72 (2008).
119. Simons, M. & Mlodzik, M. Planar cell polarity signaling: from fly development to human disease. *Annu Rev Genet* **42**, 517-40 (2008).
120. Van Straaten, H.W.M., Jansen, H.C.J.P., Peeters, M.C.E., Copp, A.J. & Hekking, J.W.M. Neural tube closure in the chick embryo is multiphasic. *Developmental Dynamics* **207**, 309-318 (1996).
121. Keller, R. & Shook, D. Gastrulation in amphibians. in *Gastrulation from Cells to Embryos* (ed. Stern, C.O.) 171-203 (Cold Spring Harbor Press, Cold Spring Harbor, NY, 2004).
122. Langland, J.A. & Kimmel, C.B. Fishes. in *Embryology Constructing the Organism* (eds. Gilbert, S.F. & Rauno, A.M.) 383-407 (Sinauer Associates, Sunderland, MA, 1997).
123. Niederriter, A.R. *et al.* In vivo modeling of the morbid human genome using Danio rerio. *J Vis Exp*, e50338 (2013).
124. Kornberg, R.D. Chromatin structure: a repeating unit of histones and DNA. *Science* **184**, 868-71 (1974).
125. Clapier, C.R. & Cairns, B.R. The biology of chromatin remodeling complexes. *Annu Rev Biochem* **78**, 273-304 (2009).
126. Rogakou, E.P., Pilch, D.R., Orr, A.H., Ivanova, V.S. & Bonner, W.M. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem* **273**, 5858-68 (1998).
127. Paull, T.T. *et al.* A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. *Curr Biol* **10**, 886-95 (2000).
128. Szerlong, H. *et al.* The HSA domain binds nuclear actin-related proteins to regulate chromatin-remodeling ATPases. *Nat Struct Mol Biol* **15**, 469-76 (2008).
129. Zeng, L. & Zhou, M.M. Bromodomain: an acetyl-lysine binding domain. *FEBS Lett* **513**, 124-8 (2002).
130. Boyer, L.A., Latek, R.R. & Peterson, C.L. The SANT domain: a unique histone-tail-binding module? *Nat Rev Mol Cell Biol* **5**, 158-63 (2004).
131. Grune, T. *et al.* Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. *Mol Cell* **12**, 449-60 (2003).
132. Bultman, S. *et al.* A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. *Mol Cell* **6**, 1287-95 (2000).
133. Kim, J.K. *et al.* Srg3, a mouse homolog of yeast SWI3, is essential for early embryogenesis and involved in brain development. *Mol Cell Biol* **21**, 7787-95 (2001).
134. Yoo, A.S. & Crabtree, G.R. ATP-dependent chromatin remodeling in neural development. *Curr Opin Neurobiol* **19**, 120-6 (2009).
135. Matsumoto, S. *et al.* Brg1 is required for murine neural stem cell maintenance and gliogenesis. *Dev Biol* **289**, 372-83 (2006).
136. Ho, L. *et al.* An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proc Natl Acad Sci U S A* **106**, 5187-91 (2009).

137. Merello, E. *et al.* Expanding the mutational spectrum associated to neural tube defects: literature revision and description of novel VANGL1 mutations. *Birth Defects Res A Clin Mol Teratol* **103**, 51-61 (2015).
138. Cai, C. *et al.* Association between VANGL1 gene polymorphisms and neural tube defects. *Neuropediatrics* **45**, 234-9 (2014).
139. Iliescu, A., Gravel, M., Horth, C. & Gros, P. Independent mutations at Arg181 and Arg274 of Vangl proteins that are associated with neural tube defects in humans decrease protein stability and impair membrane targeting. *Biochemistry* **53**, 5356-64 (2014).
140. Torban, E. *et al.* Genetic interaction between members of the Vangl family causes neural tube defects in mice. *Proc Natl Acad Sci U S A* **105**, 3449-54 (2008).
141. Banting, G.S. *et al.* CECR2, a protein involved in neurulation, forms a novel chromatin remodeling complex with SNF2L. *Hum Mol Genet* **14**, 513-24 (2005).
142. Fairbridge, N.A. *et al.* Cecr2 mutations causing exencephaly trigger misregulation of mesenchymal/ectodermal transcription factors. *Birth Defects Res A Clin Mol Teratol* **88**, 619-25 (2010).
143. Dawe, C.E., Kooistra, M.K., Fairbridge, N.A., Pisio, A.C. & McDermid, H.E. Role of chromatin remodeling gene Cecr2 in neurulation and inner ear development. *Dev Dyn* **240**, 372-83 (2011).
144. Lee, S.K., Park, E.J., Lee, H.S., Lee, Y.S. & Kwon, J. Genome-wide screen of human bromodomain-containing proteins identifies Cecr2 as a novel DNA damage response protein. *Mol Cells* **34**, 85-91 (2012).
145. Fairbridge, N.A. University of Alberta (2013).
146. Thompson, P.J., Norton, K.A., Niri, F.H., Dawe, C.E. & McDermid, H.E. CECR2 is involved in spermatogenesis and forms a complex with SNF2H in the testis. *J Mol Biol* **415**, 793-806 (2012).
147. Lazzaro, M.A. & Picketts, D.J. Cloning and characterization of the murine Imitation Switch (ISWI) genes: differential expression patterns suggest distinct developmental roles for Snf2h and Snf2l. *J Neurochem* **77**, 1145-56 (2001).
148. Lazzaro, M.A. *et al.* The imitation switch protein SNF2L regulates steroidogenic acute regulatory protein expression during terminal differentiation of ovarian granulosa cells. *Mol Endocrinol* **20**, 2406-17 (2006).
149. Pepin, D., Vanderhyden, B.C., Picketts, D.J. & Murphy, B.D. ISWI chromatin remodeling in ovarian somatic and germ cells: revenge of the NURFs. *Trends Endocrinol Metab* **18**, 215-24 (2007).
150. Bochar, D.A. *et al.* A family of chromatin remodeling factors related to Williams syndrome transcription factor. *Proc Natl Acad Sci U S A* **97**, 1038-43 (2000).
151. Jones, M.H., Hamana, N., Nezu, J. & Shimane, M. A novel family of bromodomain genes. *Genomics* **63**, 40-5 (2000).
152. LeRoy, G., Loyola, A., Lane, W.S. & Reinberg, D. Purification and characterization of a human factor that assembles and remodels chromatin. *J Biol Chem* **275**, 14787-90 (2000).
153. Bozhenok, L., Wade, P.A. & Varga-Weisz, P. WSTF-ISWI chromatin remodeling complex targets heterochromatic replication foci. *EMBO J* **21**, 2231-41 (2002).
154. He, X., Fan, H.Y., Narlikar, G.J. & Kingston, R.E. Human ACF1 alters the remodeling strategy of SNF2h. *J Biol Chem* **281**, 28636-47 (2006).

155. Ito, T. *et al.* ACF consists of two subunits, Acf1 and ISWI, that function cooperatively in the ATP-dependent catalysis of chromatin assembly. *Genes Dev* **13**, 1529-39 (1999).
156. Fyodorov, D.V. & Kadonaga, J.T. Binding of Acf1 to DNA involves a WAC motif and is important for ACF-mediated chromatin assembly. *Mol Cell Biol* **22**, 6344-53 (2002).
157. Sheu, J.J. *et al.* The roles of human sucrose nonfermenting protein 2 homologue in the tumor-promoting functions of Rsf-1. *Cancer Res* **68**, 4050-7 (2008).
158. Aravind, L. & Landsman, D. AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Res* **26**, 4413-21 (1998).
159. Haynes, S.R. *et al.* The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins. *Nucleic Acids Res* **20**, 2603 (1992).
160. Kasten, M. *et al.* Tandem bromodomains in the chromatin remodeler RSC recognize acetylated histone H3 Lys14. *EMBO J* **23**, 1348-59 (2004).
161. Yang, X.J. Lysine acetylation and the bromodomain: a new partnership for signaling. *Bioessays* **26**, 1076-87 (2004).
162. Bienz, M. The PHD finger, a nuclear protein-interaction domain. *Trends Biochem Sci* **31**, 35-40 (2006).
163. Wysocka, J. *et al.* A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**, 86-90 (2006).
164. Davidson, C.E., Li, Q., Churchill, G.A., Osborne, L.R. & McDermid, H.E. Modifier locus for exencephaly in Cecr2 mutant mice is syntenic to the 10q25.3 region associated with neural tube defects in humans. *Physiol Genomics* **31**, 244-51 (2007).
165. Kooistra, M.K. *et al.* Strain-specific modifier genes of Cecr2-associated exencephaly in mice: genetic analysis and identification of differentially expressed candidate genes. *Physiol Genomics* **44**, 35-46 (2012).
166. Footz, T.K. *et al.* 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 Res* **11**, 1053-70 (2001).
167. Knijnenburg, J. *et al.* A 600 kb triplication in the cat eye syndrome critical region causes anorectal, renal and preauricular anomalies in a three-generation family. *Eur J Hum Genet* **20**, 986-9 (2012).
168. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).
169. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).
170. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303 (2010).
171. Robinson, J.T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-6 (2011).
172. Kumar, P., Henikoff, S. & Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**, 1073-81 (2009).
173. Adzhubei, I.A. *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248-9 (2010).
174. Cooper, G.M. *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res* **15**, 901-13 (2005).
175. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).

176. Pruitt, K.D. *et al.* RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res* **42**, D756-63 (2014).
177. Sherry, S.T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308-11 (2001).
178. Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA. **http://evs.gs.washington.edu/EVS/**, [Accessed January 2014].
179. Banting, G.S. University of Alberta (2003).
180. Guo, F., Chiang, M.Y., Wang, Y. & Zhang, Y.Z. An in vitro recombination method to convert restriction- and ligation-independent expression vectors. *Biotechnol J* **3**, 370-7 (2008).
181. Villefranc, J.A., Amigo, J. & Lawson, N.D. Gateway compatible vectors for analysis of gene function in the zebrafish. *Dev Dyn* **236**, 3077-87 (2007).
182. Laemmli, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-5 (1970).
183. Hardisty-Hughes, R.E., Parker, A. & Brown, S.D. A hearing and vestibular phenotyping pipeline to identify mouse mutants with hearing impairment. *Nat Protoc* **5**, 177-90 (2010).
184. Parent, M. *et al.* Analysis of amino acids and catecholamines, 5-hydroxytryptamine and their metabolites in brain areas in the rat using in vivo microdialysis. *Methods* **23**, 11-20 (2001).
185. Grant, S.L., Shulman, Y., Tibbo, P., Hampson, D.R. & Baker, G.B. Determination of d-serine and related neuroactive amino acids in human plasma by high-performance liquid chromatography with fluorimetric detection. *J Chromatogr B Analyt Technol Biomed Life Sci* **844**, 278-82 (2006).
186. Keuling, A. *et al.* Mutation analysis of Drosophila dikar/CG32394, homologue of the chromatin-remodelling gene CECR2. *Genome* **50**, 767-77 (2007).
187. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40 (2008).
188. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* **5**, 725-38 (2010).
189. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat Methods* **12**, 7-8 (2015).
190. Harris, M.J. & Juriloff, D.M. Mouse mutants with neural tube closure defects and their role in understanding human neural tube defects. *Birth Defects Res A Clin Mol Teratol* **79**, 187-210 (2007).
191. Zohn, I.E. & Sarkar, A.A. Modeling neural tube defects in the mouse. *Curr Top Dev Biol* **84**, 1-35 (2008).
192. McManus, S. Neural tube defects: identification of 'high risk' women. *Ir Med J* **80**, 166-8 (1987).
193. Strong, L.C. & Hollander, W.F. Hereditary loop-tail in the house mouse accompanied by inperforate vagina and with lethal craniorachischisis when homozygous. *J Hered* **40**, 329-334 (1949).
194. Stiefel, D., Shibata, T., Meuli, M., Duffy, P.G. & Copp, A.J. Tethering of the spinal cord in mouse fetuses and neonates with spina bifida. *J Neurosurg* **99**, 206-13 (2003).
195. Gustavsson, P. *et al.* Increased expression of Grainyhead-like-3 rescues spina bifida in a folate-resistant mouse model. *Hum Mol Genet* **16**, 2640-6 (2007).

196. Zhao, Q., Behringer, R.R. & de Crombrugghe, B. Prenatal folic acid treatment suppresses acrania and meroanencephaly in mice mutant for the Cart1 homeobox gene. *Nat Genet* **13**, 275-83 (1996).
197. De Castro, S.C. *et al.* Lamin b1 polymorphism influences morphology of the nuclear envelope, cell cycle progression, and risk of neural tube defects in mice. *PLoS Genet* **8**, e1003059 (2012).
198. Torres, M., Gomez-Pardo, E. & Gruss, P. Pax2 contributes to inner ear patterning and optic nerve trajectory. *Development* **122**, 3381-91 (1996).
199. Kooistra, M.K. University of Alberta (2010).
200. Zhang, Z. *et al.* Cloning and characterization of ARHGAP12, a novel human rhoGAP gene. *Int J Biochem Cell Biol* **34**, 325-31 (2002).
201. Tcherkezian, J. & Lamarche-Vane, N. Current knowledge of the large RhoGAP family of proteins. *Biol Cell* **99**, 67-86 (2007).
202. Bayly, R. & Axelrod, J.D. Pointing in the right direction: new developments in the field of planar cell polarity. *Nat Rev Genet* **12**, 385-91 (2011).
203. Copp, A.J. & Greene, N.D. Genetics and development of neural tube defects. *J Pathol* **220**, 217-30 (2010).
204. Neilson, K.M. *et al.* Specific domains of FoxD4/5 activate and repress neural transcription factor genes to control the progression of immature neural ectoderm to differentiating neural plate. *Dev Biol* **365**, 363-75 (2012).
205. Nagai, T. *et al.* Zic2 regulates the kinetics of neurulation. *Proc Natl Acad Sci U S A* **97**, 1618-23 (2000).
206. Wang, G.S. & Cooper, T.A. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* **8**, 749-61 (2007).
207. Kodani, A., Kristensen, I., Huang, L. & Sutterlin, C. GM130-dependent control of Cdc42 activity at the Golgi regulates centrosome organization. *Mol Biol Cell* **20**, 1192-200 (2009).
208. Otani, T., Ichii, T., Aono, S. & Takeichi, M. Cdc42 GEF Tuba regulates the junctional configuration of simple epithelial cells. *J Cell Biol* **175**, 135-46 (2006).
209. Zuo, X., Fogelgren, B. & Lipschutz, J.H. The small GTPase Cdc42 is necessary for primary ciliogenesis in renal tubular epithelial cells. *J Biol Chem* **286**, 22469-77 (2011).
210. Stehling, O. *et al.* MMS19 assembles iron-sulfur proteins required for DNA metabolism and genomic integrity. *Science* **337**, 195-9 (2012).
211. Lauder, S. *et al.* Dual requirement for the yeast MMS19 gene in DNA repair and RNA polymerase II transcription. *Mol Cell Biol* **16**, 6783-93 (1996).
212. Gonzalez-Mariscal, L., Betanzos, A., Nava, P. & Jaramillo, B.E. Tight junction proteins. *Progress in Biophysics and Molecular Biology* **81**, 1-44 (2003).
213. Lee, J.D., Silva-Gagliardi, N.F., Tepass, U., McGlade, C.J. & Anderson, K.V. The FERM protein Epb4.1l5 is required for organization of the neural plate and for the epithelial-mesenchymal transition at the primitive streak of the mouse embryo. *Development* **134**, 2007-16 (2007).
214. Thomas, S. *et al.* TCTN3 mutations cause Mohr-Majewski syndrome. *Am J Hum Genet* **91**, 372-8 (2012).
215. Garcia-Gonzalo, F.R. *et al.* A transition zone complex regulates mammalian ciliogenesis and ciliary membrane composition. *Nat Genet* **43**, 776-84 (2011).

216. Paton, C.M. & Ntambi, J.M. Biochemical and physiological function of stearoyl-CoA desaturase. *Am J Physiol Endocrinol Metab* **297**, E28-37 (2009).
217. Tamplin, O.J. *et al.* Microarray analysis of Foxa2 mutant mouse embryos reveals novel gene expression and inductive roles for the gastrula organizer and its derivatives. *BMC Genomics* **9**, 511 (2008).
218. Wang, Y., Kakinuma, N., Zhu, Y. & Kiyama, R. Nucleo-cytoplasmic shuttling of human Kank protein accompanies intracellular translocation of beta-catenin. *J Cell Sci* **119**, 4002-10 (2006).
219. Kang, H.S. *et al.* Transcription factor Glis3, a novel critical player in the regulation of pancreatic beta-cell development and insulin gene expression. *Mol Cell Biol* **29**, 6366-79 (2009).
220. Watanabe, N. *et al.* A murine model of neonatal diabetes mellitus in Glis3-deficient mice. *FEBS Lett* **583**, 2108-13 (2009).
221. Lichti-Kaiser, K., ZeRuth, G., Kang, H.S., Vasanth, S. & Jetten, A.M. Gli-similar proteins: their mechanisms of action, physiological functions, and roles in disease. *Vitam Horm* **88**, 141-71 (2012).
222. Kim, Y.S., Nakanishi, G., Lewandoski, M. & Jetten, A.M. GLIS3, a novel member of the GLIS subfamily of Kruppel-like zinc finger proteins with repressor and activation functions. *Nucleic Acids Res* **31**, 5513-25 (2003).
223. Waring, P. & Mullbacher, A. Cell death induced by the Fas/Fas ligand pathway and its role in pathology. *Immunology and Cell Biology* **77**, 312-317 (1999).
224. Hao, Z., Hampel, B., Yagita, H. & Rajewsky, K. T cell-specific ablation of Fas leads to Fas ligand-mediated lymphocyte depletion and inflammatory pulmonary fibrosis. *The Journal of Experimental Medicine* **199**, 1355-1365 (2004).
225. Li, X. *et al.* Developmental expression of sideroflexin family genes in Xenopus embryos. *Dev Dyn* **239**, 2742-7 (2010).
226. Yao, C. *et al.* Overlapping and distinct functions of CstF64 and CstF64tau in mammalian mRNA 3' processing. *RNA* **19**, 1781-90 (2013).
227. Dass, B. *et al.* Loss of polyadenylation protein tauCstF-64 causes spermatogenic defects and male infertility. *Proc Natl Acad Sci U S A* **104**, 20374-9 (2007).
228. Pavlopoulos, E. *et al.* Neuralized1 activates CPEB3: a function for nonproteolytic ubiquitin in synaptic plasticity and memory storage. *Cell* **147**, 1369-83 (2011).
229. Richter, J.D. CPEB: a life in translation. *Trends Biochem Sci* **32**, 279-85 (2007).
230. Tamplin, O.J., Cox, B.J. & Rossant, J. Integrated microarray and ChIP analysis identifies multiple Foxa2 dependent target genes in the notochord. *Dev Biol* **360**, 415-25 (2011).
231. Raijmakers, R., Noordman, Y.E., van Venrooij, W.J. & Pruijn, G.J. Protein-protein interactions of hCsl4p with other human exosome subunits. *J Mol Biol* **315**, 809-18 (2002).
232. Di Pietro, S.M., Falcon-Perez, J.M. & Dell'Angelica, E.C. Characterization of BLOC-2, a complex containing the Hermansky-Pudlak syndrome proteins HPS3, HPS5 and HPS6. *Traffic* **5**, 276-83 (2004).
233. Huizing, M. *et al.* Clinical and cellular characterization of Hermansky-Pudlak syndrome type 6. *The Journal of Medical Genetics* **46**, 803-810 (2009).
234. Beaupre, B.A., Hoag, M.R. & Moran, G.R. Renalase does not catalyze the oxidation of catecholamines. *Archives of Biochemistry and Biophysics* **579**, 62-66 (2015).

235. Wu, Y. *et al.* Renalase deficiency aggravates ischemic myocardial damage. *Kidney Int* **79**, 853-60 (2011).
236. D'Agostino, D. & Lowe, M.E. Pancreatic lipase-related protein 2 is the major colipase-dependent pancreatic lipase in suckling mice. *J Nutr* **134**, 132-4 (2004).
237. Lee, H.K., Lee, H.S. & Moody, S.A. Neural transcription factors: from embryos to neural stem cells. *Mol Cells* **37**, 705-12 (2014).
238. Suda, Y., Nakabayashi, J., Matsuo, I. & Aizawa, S. Functional equivalency between Otx2 and Otx1 in development of the rostral head. *Development* **126**, 743-57 (1999).
239. Tian, E., Kimura, C., Takeda, N., Aizawa, S. & Matsuo, I. Otx2 is required to respond to signals from anterior neural ridge for forebrain specification. *Dev Biol* **242**, 204-23 (2002).
240. Klejman, M.P. *et al.* NC2alpha interacts with BTAF1 and stimulates its ATP-dependent association with TATA-binding protein. *Mol Cell Biol* **24**, 10072-82 (2004).
241. Wansleeben, C. *et al.* An ENU-induced point mutation in the mouse Btaf1 gene causes post-gastrulation embryonic lethality and protein instability. *Mech Dev* **128**, 279-88 (2011).
242. Dann, C.E.r., Bruick, R.K. & Deisenhofer, J. Structure of factor-inhibiting hypoxia-inducible factor 1: An asparaginyl hydroxylase involved in the hypoxic response pathway. *Proc Natl Acad Sci U S A* **99**, 15351-15356 (2002).
243. Lai, F., Zhou, Y., Luo, X., Fox, J. & King, M.L. Nanos1 functions as a translational repressor in the Xenopus germline. *Mech Dev* **128**, 153-163 (2011).
244. Zucchini, C. *et al.* The human TruB family of pseudouridine synthase genes, including the Dyskeratosis Congenita 1 gene and the novel member TRUB1. *Int J Mol Med* **11**, 697-704 (2003).
245. Salazar, M.A. *et al.* Tuba, a novel protein containing bin/amphiphysin/Rvs and Dbl homology domains, links dynamin to regulation of the actin cytoskeleton. *J Biol Chem* **278**, 49031-43 (2003).
246. Li, F., Martienssen, R. & Cande, W.Z. Coordination of DNA replication and histone modification by the Rik1-Dos2 complex. *Nature* **475**, 244-8 (2011).
247. Queimado, L. *et al.* Cloning the human and mouse MMS19 genes and functional complementation of a yeast mms19 deletion mutant. *Nucleic Acids Res* **29**, 1884-91 (2001).
248. Cingolani, G., Petosa, C., Weis, K. & Muller, C.W. Structure of importin-beta bound to the IBB domain of importin-alpha. *Nature* **399**, 221-9 (1999).
249. Chook, Y.M. & Blobel, G. Structure of the nuclear transport complex karyopherin-beta2-Ran x GppNHp. *Nature* **399**, 230-7 (1999).
250. Vetter, I.R., Nowak, C., Nishimoto, T., Kuhlmann, J. & Wittinghofer, A. Structure of a Ran-binding domain complexed with Ran bound to a GTP analogue: implications for nuclear transport. *Nature* **398**, 39-46 (1999).
251. Lee, H.J. & Zheng, J.J. PDZ domains and their binding partners: structure, specificity, and modification. *Cell Commun Signal* **8**, 8 (2010).
252. Zhu, J. *et al.* Guanylate kinase domains of the MAGUK family scaffold proteins as specific phospho-protein-binding modules. *EMBO J* **30**, 4986-97 (2011).

253. Johnston, C.A., Hirono, K., Prehoda, K.E. & Doe, C.Q. Identification of an Aurora-A/PinsLINKER/Dlg spindle orientation pathway using induced cell polarity in S2 cells. *Cell* **138**, 1150-63 (2009).
254. Kim, E. *et al.* GKAP, a novel synaptic protein that interacts with the guanylate kinase-like domain of the PSD-95/SAP90 family of channel clustering molecules. *J Cell Biol* **136**, 669-78 (1997).
255. Brenman, J.E. *et al.* Localization of postsynaptic density-93 to dendritic microtubules and interaction with microtubule-associated protein 1A. *J Neurosci* **18**, 8805-13 (1998).
256. Hanada, T., Lin, L., Tibaldi, E.V., Reinherz, E.L. & Chishti, A.H. GAKIN, a novel kinesin-like protein associates with the human homologue of the Drosophila discs large tumor suppressor in T lymphocytes. *J Biol Chem* **275**, 28774-84 (2000).
257. Pak, D.T., Yang, S., Rudolph-Correia, S., Kim, E. & Sheng, M. Regulation of dendritic spine morphology by SPAR, a PSD-95-associated RapGAP. *Neuron* **31**, 289-303 (2001).
258. Yang, X.Y. *et al.* Mutations in the COPII vesicle component gene SEC24B are associated with human neural tube defects. *Hum Mutat* **34**, 1094-101 (2013).
259. Flanagan, S.E., Patch, A.M. & Ellard, S. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genet Test Mol Biomarkers* **14**, 533-7 (2010).
260. Jordan, D.M. *et al.* Identification of cis-suppression of human disease mutations by comparative genomics. *Nature* (2015).
261. Juriloff, D.M. & Harris, M.J. A consideration of the evidence that genetic defects in planar cell polarity contribute to the etiology of human neural tube defects. *Birth Defects Res A Clin Mol Teratol* **94**, 824-40 (2012).
262. Moody, S.A., Klein, S.L., Karpinski, B.A., Maynard, T.M. & Lamantia, A.S. On becoming neural: what the embryo can tell us about differentiating neural stem cells. *Am J Stem Cells* **2**, 74-94 (2013).
263. Wong, A. *et al.* Diverse fates of paralogs following segmental duplication of telomeric genes. *Genomics* **84**, 239-47 (2004).
264. Yoshizawa, M. *et al.* Dynamic and coordinated expression profile of dbl-family guanine nucleotide exchange factors in the developing mouse brain. *Gene Expr Patterns* **3**, 375-81 (2003).
265. Schlessinger, K., McManus, E.J. & Hall, A. Cdc42 and noncanonical Wnt signal transduction pathways cooperate to promote cell polarity. *J Cell Biol* **178**, 355-61 (2007).
266. Menzies, A.S. *et al.* Mena and vasodilator-stimulated phosphoprotein are required for multiple actin-dependent processes that shape the vertebrate nervous system. *J Neurosci* **24**, 8029-38 (2004).
267. Lanier, L.M. *et al.* Mena is required for neurulation and commissure formation. *Neuron* **22**, 313-25 (1999).
268. Gari, K. *et al.* MMS19 links cytoplasmic iron-sulfur cluster assembly to DNA metabolism. *Science* **337**, 243-5 (2012).
269. Metais, J.Y., Navarro, C., Santoni, M.J., Audebert, S. & Borg, J.P. hScrib interacts with ZO-2 at the cell-cell junctions of epithelial cells. *FEBS Lett* **579**, 3725-30 (2005).

270. Murdoch, J.N. *et al.* Genetic interactions between planar cell polarity genes cause diverse neural tube defects in mice. *Dis Model Mech* **7**, 1153-63 (2014).
271. Lei, Y. *et al.* Mutations in planar cell polarity gene SCRIB are associated with spina bifida. *PLoS One* **8**, e69262 (2013).
272. Xu, J. *et al.* Early embryonic lethality of mice lacking ZO-2, but Not ZO-3, reveals critical and nonredundant roles for individual zonula occludens proteins in mammalian development. *Mol Cell Biol* **28**, 1669-78 (2008).
273. Deuring, R. *et al.* The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure in vivo. *Mol Cell* **5**, 355-65 (2000).
274. Burgio, G. *et al.* Genetic identification of a network of factors that functionally interact with the nucleosome remodeling ATPase ISWI. *PLoS Genet* **4**, e1000089 (2008).
275. Yamada, K. *et al.* Structure and mechanism of the chromatin remodelling factor ISW1a. *Nature* **472**, 448-53 (2011).
276. Aravind, L. & Iyer, L.M. The HARE-HTH and associated domains: novel modules in the coordination of epigenetic DNA and protein modifications. *Cell Cycle* **11**, 119-31 (2012).
277. Eberharter, A. *et al.* Acf1, the largest subunit of CHRAC, regulates ISWI-induced nucleosome remodelling. *EMBO J* **20**, 3781-8 (2001).
278. Gubb, D. & Garcia-Bellido, A. A genetic analysis of the determination of cuticular polarity during development in Drosophila melanogaster. *J Embryol Exp Morphol* **68**, 37-57 (1982).
279. Adler, P.N. Planar signaling and morphogenesis in Drosophila. *Dev Cell* **2**, 525-35 (2002).
280. Chen, W.S. *et al.* Asymmetric homotypic interactions of the atypical cadherin flamingo mediate intercellular polarity signaling. *Cell* **133**, 1093-105 (2008).
281. Seifert, J.R. & Mlodzik, M. Frizzled/PCP signalling: a conserved mechanism regulating cell polarity and directed motility. *Nat Rev Genet* **8**, 126-38 (2007).
282. Deans, B., Griffin, C.S., Maconochie, M. & Thacker, J. Xrcc2 is required for genetic stability, embryonic neurogenesis and viability in mice. *EMBO J* **19**, 6675-85 (2000).
283. King, T.M. *et al.* The impact of BRCA1 on spina bifida meningomyelocele lesions. *Ann Hum Genet* **71**, 719-28 (2007).
284. Wang, X. *et al.* Genetic interactions between Brca1 and Gadd45a in centrosome duplication, genetic stability, and neural tube closure. *J Biol Chem* **279**, 29606-14 (2004).
285. Blount, B.C. *et al.* Folate deficiency causes uracil misincorporation into human DNA and chromosome breakage: implications for cancer and neuronal damage. *Proc Natl Acad Sci U S A* **94**, 3290-5 (1997).
286. Defoort, E.N., Kim, P.M. & Winn, L.M. Valproic acid increases conservative homologous recombination frequency and reactive oxygen species formation: a potential mechanism for valproic acid-induced neural tube defects. *Mol Pharmacol* **69**, 1304-10 (2006).
287. Bohgaki, T., Bohgaki, M. & Hakem, R. DNA double-strand break signaling and human disorders. *Genome Integr* **1**, 15 (2010).

288. Vilenchik, M.M. & Knudson, A.G. Endogenous DNA double-strand breaks: production, fidelity of repair, and induction of cancer. *Proc Natl Acad Sci U S A* **100**, 12871-6 (2003).
289. Vilenchik, M.M. & Knudson, A.G., Jr. Inverse radiation dose-rate effects on somatic and germ-line mutations and DNA damage rates. *Proc Natl Acad Sci U S A* **97**, 5381-6 (2000).
290. Mehta, A. & Haber, J.E. Sources of DNA double-strand breaks and models of recombinational DNA repair. *Cold Spring Harb Perspect Biol* **6**, a016428 (2014).
291. Uziel, T. *et al.* Requirement of the MRN complex for ATM activation by DNA damage. *EMBO J* **22**, 5612-21 (2003).
292. Stucki, M. *et al.* MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell* **123**, 1213-26 (2005).
293. Jeggo, P.A., Geuting, V. & Lobrich, M. The role of homologous recombination in radiation-induced double-strand break repair. *Radiother Oncol* **101**, 7-12 (2011).
294. Weterings, E. & Chen, D.J. The endless tale of non-homologous end-joining. *Cell Res* **18**, 114-24 (2008).
295. Kuhfittig-Kulle, S. *et al.* The mutagenic potential of non-homologous end joining in the absence of the NHEJ core factors Ku70/80, DNA-PKcs and XRCC4-LigIV. *Mutagenesis* **22**, 217-33 (2007).
296. Strutt, D. Frizzled signalling and cell polarisation in Drosophila and vertebrates. *Development* **130**, 4501-13 (2003).
297. Pospiech, H. & Syvaoja, J.E. DNA Polymerase epsilon - More Than Just a Polymerase. *TheScientificWorldJOURNAL* **3**, 87-104 (2003).
298. Wei, D.S. & Rong, Y.S. A genetic screen for DNA double-strand break repair mutations in Drosophila. *Genetics* **177**, 63-77 (2007).
299. Andersen, S.L., Kuo, H.K., Savukoski, D., Brodsky, M.H. & Sekelsky, J. Three structure-selective endonucleases are essential in the absence of BLM helicase in Drosophila. *PLoS Genet* **7**, e1002315 (2011).
300. Bailly, A.P. *et al.* The Caenorhabditis elegans homolog of Gen1/Yen1 resolvases links DNA damage signaling to DNA double-strand break repair. *PLoS Genet* **6**, e1001025 (2010).
301. Ip, S.C. *et al.* Identification of Holliday junction resolvases from humans and yeast. *Nature* **456**, 357-61 (2008).
302. Laurencon, A. *et al.* A large-scale screen for mutagen-sensitive loci in Drosophila. *Genetics* **167**, 217-31 (2004).
303. Bassett, A.R. & Liu, J.L. CRISPR/Cas9 and genome editing in Drosophila. *J Genet Genomics* **41**, 7-19 (2014).
304. Fyodorov, D.V., Blower, M.D., Karpen, G.H. & Kadonaga, J.T. Acf1 confers unique activities to ACF/CHRAC and promotes the formation rather than disruption of chromatin in vivo. *Genes Dev* **18**, 170-83 (2004).
305. Crawley, J.N., Schleidt, W.M. & Contrera, J.F. Does social environment decrease propensity to fight in male mice? *Behavioral Biology* **15**, 73-83 (1975).
306. Van Loo, P.L., Van de Weerd, H.A., Van Zutphen, L.F. & Baumans, V. Preference for social contact versus environmental enrichment in male laboratory mice. *Lab Anim* **38**, 178-88 (2004).

307. Arndt, S.S. *et al.* Individual housing of mice--impact on behaviour and stress responses. *Physiol Behav* **97**, 385-93 (2009).
308. Howerton, C.L., Garner, J.P. & Mench, J.A. Effects of a running wheel-igloo enrichment on aggression, hierarchy linearity, and stereotypy in group-housed male CD-1 (ICR) mice. *Applied Animal Behaviour Science* **115**, 90-103 (2008).
309. Akre, A.K., Bakken, M., Hovland, A.L., Palme, R. & Mason, G. Clustered environmental enrichments induce more aggression and stereotypic behaviour than do dispersed enrichments in female mice. *Appl. Anim. Behav. Sci.* **131**, 145-152 (2011).
310. Sherwin, C.M. Voluntary wheel running: a review and novel interpretation. *Anim Behav* **56**, 11-27 (1998).
311. Richter, S.H., Gass, P. & Fuss, J. Resting Is Rusting: A Critical View on Rodent Wheel-Running Behavior. *Neuroscientist* (2014).
312. Hadley, C., Hadley, B., Ephraim, S., Yang, M. & Lewis, M.H. Spontaneous stereotypy and environmental enrichment in deer mice (Peromyscus maniculatus): reversibility of experience. *Appl. Anim. Behav. Sci.* **97**, 312-322 (2006).
313. Powell, S.B., Newman, H.A., McDonald, T.A., Bugenhagen, P. & Lewis, M.H. Development of spontaneous stereotyped behavior in deer mice: effects of early and late exposure to a more complex environment. *Dev Psychobiol* **37**, 100-8 (2000).
314. Mason, G. Stereotypies: a critical review. *Animal Behaviour* **41**, 1015-1037 (1991).
315. Richter, H. *et al.* Wheel-running in a transgenic mouse model of Alzheimer's disease: protection or symptom? *Behav Brain Res* **190**, 74-84 (2008).
316. Meijer, J.H. & Robbers, Y. Wheel running in the wild. *Proc Biol Sci* **281**(2014).
317. Lu, Y.C. *et al.* Establishment of a knock-in mouse model with the SLC26A4 c.919-2A>G mutation and characterization of its pathology. *PLoS One* **6**, e22150 (2011).
318. Makishima, T. *et al.* Inner ear dysfunction in caspase-3 deficient mice. *BMC Neurosci* **12**, 102 (2011).
319. Lv, K. *et al.* Circling behavior developed in Dmp1 null mice is due to bone defects in the vestibular apparatus. *Int J Biol Sci* **6**, 537-45 (2010).
320. Loscher, W. Abnormal circling behavior in rat mutants and its relevance to model specific brain dysfunctions. *Neurosci Biobehav Rev* **34**, 31-49 (2010).
321. Kincaid, A.E. Spontaneous circling behavior and dopamine neuron loss in a genetically hypothyroid mouse. *Neuroscience* **105**, 891-8 (2001).
322. Loscher, W. *et al.* Behavioral and neurochemical dysfunction in the circling (ci) rat: a novel genetic animal model of a movement disorder. *Neuroscience* **74**, 1135-42 (1996).
323. Stiles, L., Zheng, Y., Darlington, C.L. & Smith, P.F. The D(2) dopamine receptor and locomotor hyperactivity following bilateral vestibular deafferentation in the rat. *Behav Brain Res* **227**, 150-8 (2012).
324. Schirmer, M. *et al.* Auditory and vestibular defects and behavioral alterations after neonatal administration of streptomycin to Lewis rats: Similarities and differences to the circling (ci2/ci2) Lewis rat mutant. *Brain Res* **1155**, 179-95 (2007).
325. Giardino, L., Zanni, M. & Pignataro, O. DA1 and DA2 receptor regulation in the striatum of young and old rats after peripheral vestibular lesion. *Brain Res* **736**, 111-7 (1996).
326. Nakamura, K. *et al.* Brain serotonin and dopamine transporter bindings in adults with high-functioning autism. *Arch Gen Psychiatry* **67**, 59-68 (2010).

327. Cook, E.H., Jr. *et al.* Association of attention-deficit disorder and the dopamine transporter gene. *Am J Hum Genet* **56**, 993-8 (1995).
328. Guillin, O., Abi-Dargham, A. & Laruelle, M. Neurobiology of dopamine in schizophrenia. *Int Rev Neurobiol* **78**, 1-39 (2007).
329. Russell, V.A., Sagvolden, T. & Johansen, E.B. Animal models of attention-deficit hyperactivity disorder. *Behav Brain Funct* **1**, 9 (2005).
330. Asherson, P. *et al.* Confirmation that a specific haplotype of the dopamine transporter gene is associated with combined-type ADHD. *Am J Psychiatry* **164**, 674-7 (2007).
331. Torres, G. *et al.* A neurobehavioral screening of the ckr mouse mutant: implications for an animal model of schizophrenia. *Brain Res Bull* **62**, 315-26 (2004).
332. Dawe, G.S. *et al.* Antipsychotic drugs dose-dependently suppress the spontaneous hyperactivity of the chakragati mouse. *Neuroscience* **171**, 162-72 (2010).
333. Fitzgerald, L.W. *et al.* Asymmetric elevation of striatal dopamine D2 receptors in the chakragati mouse: neurobehavioral dysfunction in a transgenic insertional mutant. *Brain Res* **580**, 18-26 (1992).
334. Miller, R. & Beninger, R.J. On the interpretation of asymmetries of posture and locomotion produced with dopamine agonists in animals with unilateral depletion of striatal dopamine. *Prog Neurobiol* **36**, 229-56 (1991).
335. Scholtissen, B. *et al.* Functional investigations into the role of dopamine and serotonin in partial bilateral striatal 6-hydroxydopamine lesioned rats. *Pharmacol Biochem Behav* **83**, 175-85 (2006).
336. Kaariainen, T.M., Garcia-Horsman, J.A., Piltonen, M., Huotari, M. & Mannisto, P.T. Serotonergic activation after 2-week intrastriatal infusion of L-dopa and slow recovery of circling in rats with unilateral nigral lesions. *Basic Clin Pharmacol Toxicol* **102**, 300-7 (2008).
337. Walker, M.D. *et al.* In-vivo measurement of LDOPA uptake, dopamine reserve and turnover in the rat brain using [18F]FDOPA PET. *J Cereb Blood Flow Metab* **33**, 59-66 (2013).
338. Seo, D., Patrick, C.J. & Kennealy, P.J. Role of Serotonin and Dopamine System Interactions in the Neurobiology of Impulsive Aggression and its Comorbidity with other Clinical Disorders. *Aggress Violent Behav* **13**, 383-395 (2008).
339. Beeman, E.A. The effect of male hormone on aggressive behavior in mice. *Physiol Zool* **20**, 373-405 (1947).
340. Konradi, C. *et al.* Variations of monoamines and their metabolites in the human brain putamen. *Brain Res* **579**, 285-90 (1992).
341. Kaasinen, V., Nagren, K., Hietala, J., Farde, L. & Rinne, J.O. Sex differences in extrastriatal dopamine d(2)-like receptors in the human brain. *Am J Psychiatry* **158**, 308-11 (2001).
342. Laakso, A. *et al.* Sex differences in striatal presynaptic dopamine synthesis capacity in healthy subjects. *Biol Psychiatry* **52**, 759-63 (2002).
343. Munro, C.A. *et al.* Sex differences in striatal dopamine release in healthy adults. *Biol Psychiatry* **59**, 966-74 (2006).
344. Pohjalainen, T., Rinne, J.O., Nagren, K., Syvalahti, E. & Hietala, J. Sex differences in the striatal dopamine D2 receptor binding characteristics in vivo. *Am J Psychiatry* **155**, 768-73 (1998).

345. Riccardi, P. *et al.* Sex differences in amphetamine-induced displacement of [(18)F]fallypride in striatal and extrastriatal regions: a PET study. *Am J Psychiatry* **163**, 1639-41 (2006).
346. Riccardi, P. *et al.* Sex differences in the relationship of regional dopamine release to affect and cognitive function in striatal and extrastriatal regions using positron emission tomography and [(1)(8)F]fallypride. *Synapse* **65**, 99-102 (2011).
347. Smalley, S.L., Asarnow, R.F. & Spence, M.A. Autism and genetics. A decade of research. *Arch Gen Psychiatry* **45**, 953-61 (1988).
348. Rucklidge, J.J. Gender differences in ADHD: implications for psychosocial treatments. *Expert Rev Neurother* **8**, 643-55 (2008).
349. McGrath, J. *et al.* A systematic review of the incidence of schizophrenia: the distribution of rates and the influence of sex, urbanicity, migrant status and methodology. *BMC Med* **2**, 13 (2004).
350. Becker, J.B. Direct effect of 17 beta-estradiol on striatum: sex differences in dopamine release. *Synapse* **5**, 157-64 (1990).
351. Becker, J.B. Estrogen rapidly potentiates amphetamine-induced striatal dopamine release and rotational behavior during microdialysis. *Neurosci Lett* **118**, 169-71 (1990).
352. Becker, J.B. Gender differences in dopaminergic function in striatum and nucleus accumbens. *Pharmacol Biochem Behav* **64**, 803-12 (1999).
353. Chiodo, L.A. & Caggiula, A.R. Alterations in basal firing rate and autoreceptor sensitivity of dopamine neurons in the substantia nigra following acute and extended exposure to estrogen. *Eur J Pharmacol* **67**, 165-6 (1980).
354. Di Paolo, T. Modulation of brain dopamine transmission by sex steroids. *Rev Neurosci* **5**, 27-41 (1994).
355. Velisek, L., Veliskova, J., Ravizza, T., Giorgi, F.S. & Moshe, S.L. Circling behavior and [14C]2-deoxyglucose mapping in rats: possible implications for autistic repetitive behaviors. *Neurobiol Dis* **18**, 346-55 (2005).
356. Nishizawa, S. *et al.* Differences between males and females in rates of serotonin synthesis in human brain. *Proc Natl Acad Sci U S A* **94**, 5308-13 (1997).
357. Maron, E. *et al.* Gender differences in brain serotonin transporter availability in panic disorder. *J Psychopharmacol* **25**, 952-9 (2011).
358. Jovanovic, H. *et al.* Sex differences in the serotonin 1A receptor and serotonin transporter binding in the human brain measured by PET. *Neuroimage* **39**, 1408-19 (2008).
359. Zhang, L., Ma, W., Barker, J.L. & Rubinow, D.R. Sex differences in expression of serotonin receptors (subtypes 1A and 2A) in rat brain: a possible role of testosterone. *Neuroscience* **94**, 251-9 (1999).
360. Goel, N., Innala, L. & Viau, V. Sex differences in serotonin (5-HT) 1A receptor regulation of HPA axis and dorsal raphe responses to acute restraint. *Psychoneuroendocrinology* **40**, 232-41 (2014).
361. Fischette, C.T., Biegon, A. & McEwen, B.S. Sex differences in serotonin 1 receptor binding in rat brain. *Science* **222**, 333-5 (1983).
362. Korstanje, R. *et al.* Quantitative trait loci affecting phenotypic variation in the vacuolated lens mouse mutant, a multigenic mouse model of neural tube defects. *Physiol Genomics* **35**, 296-304 (2008).

363. Pani, L., Horal, M. & Loeken, M.R. Polymorphic susceptibility to the molecular causes of neural tube defects during diabetic embryopathy. *Diabetes* **51**, 2871-4 (2002).
364. Murdoch, J.N. & Copp, A.J. The relationship between sonic Hedgehog signaling, cilia, and neural tube defects. *Birth Defects Res A Clin Mol Teratol* **88**, 633-52 (2010).
365. Chiang, C. *et al.* Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature* **383**, 407-13 (1996).
366. Guerrero, I. & Chiang, C. A conserved mechanism of Hedgehog gradient formation by lipid modifications. *Trends Cell Biol* **17**, 1-5 (2007).
367. Li, Y., Zhang, H., Litingtung, Y. & Chiang, C. Cholesterol modification restricts the spread of Shh gradient in the limb bud. *Proc Natl Acad Sci U S A* **103**, 6548-53 (2006).
368. Haas, D. & Muenke, M. Abnormal sterol metabolism in holoprosencephaly. *Am J Med Genet C Semin Med Genet* **154C**, 102-8 (2010).
369. Cooper, M.K. *et al.* A defective response to Hedgehog signaling in disorders of cholesterol biosynthesis. *Nat Genet* **33**, 508-13 (2003).
370. Wilcox, C.B. *et al.* Coordinate up-regulation of TMEM97 and cholesterol biosynthesis genes in normal ovarian surface epithelial cells treated with progesterone: implications for pathogenesis of ovarian cancer. *BMC Cancer* **7**, 223 (2007).

Appendices

Appendix A: List of primers

Appendix B: Cranial NTD phenotypes for 156 probands

Appendix C: Co-immunoprecipitation assay comparing human wildtype CECR2 to variant CECR2 (E32K and P632L) in the ability to complex with SNF2L in transfected HEK293 cells

Appendix D: *CECR2* ORF constructs in the pENTR^{TM}11 plasmid

Appendix E: Biogenic amine and amino acid concentrations in half brain samples

Appendix F: *Drosophila* progeny counts for RNAi crosses

Appendix A: List of primers. All primers were obtained from Integrated DNA Technologies unless otherwise noted in additional information. Primers used in Sanger sequencing reactions are labeled with an asterisk.

Name	Sequence	Additional information	Thesis chapter
Cecr2 Intron7 F4	CCCCATTTATTTGCTTGAGCTG	*Cecr2*^{GT45bic} genotyping	2.5
Cecr2 Intron7 R4	CACGAACAATGGAAGGAATGA		
pGT1R4	ACGCCATACAGTCCTCTTCACATC		
SRY FOR	GAGAGCATGGAGGGGCAT	sex genotyping	2.5
SRY REV	CCACTCCTCTGTGACACT		
Ingenious SDL2	GTAGCGCCTATTTGTAATGGTCA	*Cecr2*$^{tm1.1Hemc}$ genotyping	2.6
LoxCECR2_DEL3R	AATGGTGGCGAAATCAACTC		
IngeniousLox 1	TTAGAATAGGTGAGGGAGGAG		
CECR2_DDT F*	CCGGTTGCTCTTCTCACAGT	*CECR2* novel missense, position 22:17956663	2.9
CECR2_DDT R	TGTTAGCTGGGGTATCCAAGAG		
CECR2-E4-F*	AGTCATGACCCACAGTAGAAACA	*CECR2* rs201912432	2.9
CECR2-E4-R	TGAGACTATAGCTTGCCCTGA		
CECR2-E14-2-F	GCAGAACGTTGCAGGAAACC	*CECR2* rs199780601	2.9
CECR2-14-2-R*	TCTTTGGGACTCTCTGGCCT		
CECR2-E13-1-F*	TCTCTTTCCAGGTTAAATGTTGTTC	*CECR2* rs199565531 and rs62623401	2.9
CECR2-E13-1-R	TCCATGAGGTAGGAAGCCTG		
CECR2-E14-3-F*	CCAACCCGTATGGATGCAGT	*CECR2* rs142851999	2.9
CECR2-E14-3-R	CATCTCTAAGGCCATTTTTCTGTT		
CECR2-E15-F*	GAAGGCAGTACCCTCGTGAC	*CECR2* rs181553013	2.9
CECR2-E15-R	TGCCAAGGAAGCACGATACA		
CECR2-E5-F*	GAAGCCTTTGGCCCTAACCT	*CECR2* rs5747211	2.9
CECR2-E5-R	GACATGCCTACCATCCCAGG		
CECR2-E12-F	TGTCTCCAGCCGAGGTTTAG	*CECR2* rs1296794	2.9
CECR2-E12-R*	TGTGCCATGCTTCTCTATGG		
DNMBP-E10-F	ACCACATGCAGGTGTTTAAATCAAT	*DNMBP* novel stopgain-position 10:101654735	2.9
DNMBP-E10-R*	GGTGGAAAAAGTTGGGCTGC		
DNMBP-E12-F*	GATCTGTATGGGTTTAGCTTGTG	*DNMBP* novel missense-position 10:101646214	2.9
DNMBP-E12-R	GGGATGAACTAGCGACACATTC		
DNMBP-E15-1-F*	CACTATTGGCCTGGCTTTTG	*DNMBP* rs372003127 and rs11190305	2.9
DNMBP-E15-1-R	CATCTGCAGAGTCCCCTGAC		
DNMBP-E5-F*	GGTTGGGGTGGCAGAGAATA	*DNMBP* rs114927649 and rs17854134	2.9
DNMBP-E5-R	CCCCAAAGTAGGCTGAGACA		
DNMBP-E15-2-F*	GCCAGAAGCAGCCTCAAGAT	*DNMBP* rs147752816	2.9
DNMBP-E15-2-R	AGAGGGGTCCAGGAATAGCC		
DNMBP-E3-2-F*	AAGAGATAGGGCCGGATGAG	*DNMBP* rs35924554	2.9
DNMBP-E3-2-R	GGTACTGTTCTGAGTGCGGG		
M155-F*	ACATTGCTGTGGAGGCAAGA	*MMS19* novel stopgain-position 10:99237155, req'd Q sol'n	2.9
M155-R	TGCTGTGCTGCTGAGTCAAT		
M991-F*	TAAGTGGCAGGATGGCCTTT	*MMS19* rs200490757, req'd Q sol'n	2.9
M991-R	TTCATGATGTGGCCATCCCC		

Name	Sequence	Additional information	Thesis chapter
M456-F*	GCAGCATATCGCCCCTTTTC	*MMS19* rs36023427, req'd Q sol'n	2.9
M456-R	TTGCCTTGGATGAGCAAGGT		
M846/645-F*	AGTTACCTCTGTTCACTTGCCA	*MMS19* rs17112809 and rs12360068, req'd Q sol'n	2.9
M846/645-R	GCCTAATATCAGACGTAGGCA		
M707-F*	CACACAGGGGAGGATGGAAG	*MMS19* rs3740526, req'd Q sol'n	2.9
M707-R	TCAAATGGCAGCACCTCTGT		
T989/040-F*	CTGTTGGAGCTGTGACTCCC	*TJP2* novel missense-position 9:71850989 and rs149911553, req'd Q sol'n	2.9
T989/040-R	CAGTTCAGGCCACTCTCCAT		
T149/988-F	GCTGTTTGCCCTGCTGTATT	*TJP2* rs199892018 and rs41277907, req'd Q sol'n	2.9
T149/988-R*	CCTGGAGCTCCTGCATTCTC		
T842-F*	TCGGCAGGGATACGGTTTTC	*TJP2* rs41305539, req'd Q sol'n	2.9
T842-R	GCTCGTAGTCTGGGTCGTAG		
T877-F*	AAGGTCCCCACATTTCCCAC	*TJP2* rs34774441, req'd Q sol'n	2.9
T877-R	CTGCCTTTGGCATAACTGACA		
T235-F*	CTGGTGTCGTCCATCACCC	*TJP2* rs77236826, req'd Q sol'n	2.9
T235-R	TGGACAGACGGCACCTAAAG		
CPEB3 F*	CAGCAGAGTTGAAAGCAGCG	*CPEB3* rs140779166	2.9
CPEB3 R	ACCAAGCAGACAAAGGTGTG		
CSTF2T_nov F*	CTGCAGCGCCCATTATTGAC	*CSTF2T* novel missense-position 10:53458828	2.9
CSTF2T_nov R	GGAGTCTGCATCAGAGGAGG		
CSTF2T_tri F*	CCAGAGGGATGGAAGCAAGG	*CSTF2T* rs143644186, rs148098627, and rs142002882	2.9
CSTF2T_tri R	CCCTAATCCAAGTGTGGGGA		
EXOSC1 F*	GTGGCCAGAGTGGCTCTTAT	*EXOSC1* rs141001349	2.9
EXOSC1 R	TCCTTTTGTCCTAGATCATCCATGT		
FAM45A F*	TGGGAATTTGGGGTGTCAGG	*FAM45A* rs149569390, req'd Q sol'n	2.9
FAM45A R	GCTGTACTGCTCACCACTGT		
FAS F*	GGCTTCTGCATCCTGCCATA	*FAS* rs56006128	2.9
FAS R	TTTCTTTTCAAGGAAAGCTGATACC		
GLIS3_tri F*	AGCATGAAGCAGGAGTGGTC	*GLIS3* rs143051164, rs148572278, and rs35154632, req'd Q sol'n	2.9
GLIS3_tri R	GGCGTTCGGTCTTGAACAGG		
GLIS3_sin F*	CTGTTCAAGACCGAACGCCT	*GLIS3* rs72687988, req'd Q sol'n	2.9
GLIS3_sin R	CCTTCCTCAAGCTGAAGGGG		
HPS6 F*	CTTTCAGCCACTGTGTGTGC	*HPS6* rs36078476	2.9
HPS6 R	AGTTCCAATGTGGAGCCCAG		
PNLIPRP2_rar F*	AGTCTGACCCAGCTTTGTGG	*PNLIPRP2* rs200056143	2.9
PNLIPRP2_rar R	CTCTGGGTGTCTTCAGCCTC		
PNLIPRP2_mut F*	CAGCCCCCAGGTGATTCTTA	*PNLIPRP2* rs62623669	2.9
PNLIPRP2_mut R	TGATCCTGAGAGCTACGCCT		
PNLIPRP2_com F	TCCCTTGCTCCCACCTAGAA	*PNLIPRP2* rs4751996 and rs4751995	2.9
PNLIPRP2_com R*	GGAGTTAGCACATGACTCATTTTA		
RNLS_nov F*	AGAGCAGGCACTGATAGGGA	*RNLS* novel missense-position 10:90332762	2.9
RNLS_nov R	GACGACGGCCGTAAGTATCA		
RNLS_com F*	CAGGGAGCACGGAACCAAA	*RNLS* rs2296545	2.9
RNLS_com R	GCTGTCTCAGGTCTCCCTCT		
SCD_nov F*	CCCTCCATTGACCTGGTGTC	*SCD* novel missense-position 10:102116427	2.9
SCD_nov R	ACTGCCCCCTAATTTTATAGTGGA		

Name	Sequence	Additional information	Thesis chapter
SCD_rar F*	TGCGTCTCGGATACACCCTA	*SCD* rs150416868	2.9
SCD_rar R	GAGAAAAACCCCAACCGGGA		
SFXN2_tri F*	GAGAGCAGGCAGTCTGCAA	*SFXN2* novel missense-position 10:104486433, novel missense-position 10:104486843, and rs201068739, req'd Q sol'n	2.9
SFXN2_tri R	AGAACTGGAGCATGAAGCCC		
SFXN2_sin F*	CTGATCCATGGCAGAGCAGG	*SFXN2* rs151088981	2.9
SFXN2_sin R	AGCCTGATGCCCTCACAAAA		
TCTN3_357 F*	TTGTTGGCTGGCACTGGAA	*TCTN3* rs55859130	2.9
TCTN3_357 R	TGGTAGGGAGGATAATGGGCT		
TCTN3_261 F*	TCACAGTCTGTCACCCTCCT	*TCTN3* rs147928670	2.9
TCTN3_261 R	AGGACGTACTTGCAAAGGCA		
TMEM180_399 F*	GGGCCTCTCTTTCTCATCCG	*TMEM180* rs371861479	2.9
TMEM180_399 R	GGTTGCTGCAGTAGAGGGAG		
TMEM180_098 F*	CGGCTCCCTCTCTGTCTTTG	*TMEM180* rs149406506	2.9
TMEM180_098 R	GAAGTCCAGCCCAGCTGTAG		
KANK1_509 F*	GAGCACACCTTGCATCTCCT	*KANK1* novel missense-position 9:732509	2.9
KANK1_509 R	GGTACCAGGCCATTCTACCA		
KANK1_588 F*	TCCCAGAAGACCGAACGAGT	*KANK1* rs143775530	2.9
KANK1_588 R	CTCAGGGGCTGCTATCTGGA		
KANK1_972 F	CCCAGCTGTTGTAGATCCAGG	*KANK1* rs61737969	2.9
KANK1_972 R*	GAGTTGTGGTGAGGGAGGTG		
pCECR2-1 R*	CACCGGGTCCTCTTTGTACA	*CECR2* ORF sequence validation	2.13
CECR2-OVR F*	ATCACGCCTCAGACATTCCA	*CECR2* ORF sequence validation	2.13
pCECR2-2 F*	TTCTAAAGGGCCTGGATGCA	*CECR2* ORF sequence validation	2.13
pCECR2-3 F*	TGTGCCAGACAGAAGAGGAA	*CECR2* ORF sequence validation	2.13
pCECR2-4 F*	GAGGAGAAGGTCAAGGCAGT	*CECR2* ORF sequence validation	2.13
pCECR2-5 F*	GACCATGTTCAGGAATTGTCGA	*CECR2* ORF sequence validation	2.13
pCECR2-6 F*	CTTCCAATGGCCGAGGTTTT	*CECR2* ORF sequence validation	2.13
pCECR2-7 F*	GCAGATAAGTGGCCCAAGTC	*CECR2* ORF sequence validation	2.13
pCECR2-8 F*	CAGGACCCTCTCACCAGC	*CECR2* ORF sequence validation	2.13
pCECR2-9 F*	GCCTGAGAATGACCAAGCAG	*CECR2* ORF sequence validation	2.13
pCECR2-10 F*	GGAGTCATTGGGGAAGCATC	*CECR2* ORF sequence validation	2.13
pCECR2-11 F*	GTATTCCTACCACCCACCGC	*CECR2* ORF sequence validation	2.13
pCECR2-12 F*	GGCCAGAGAGTCCCAAAGAA	*CECR2* ORF sequence validation	2.13
pCECR2-13 F*	CTTACTCTTCCCCTGTGGCT	*CECR2* ORF sequence validation	2.13
CECR2-KpnI-F	TTTGGTACCATGTGCCCAGAGGAGGGCGG	*CECR2* ORF cloning	2.13
CECR2-XhoI-R	TTTCTCGAGCTAGCTCTGATCCAGGGGAAGTGTTGGAGG		
CECR2-Kozak F	GTCGACTGGATCCGCCACCATGTGCCCAGAG	Kozak sequence introduction	2.14
CECR2-Kozak R	CTCTGGGCACATGGTGGCGGATCCAGTCGAC		

Name	Sequence	Additional information	Thesis chapter
CECR2 DEL2 F	CTAAGGAAATACTTAACCTGGTCGACTGGATCCGCCATG	Upstream vector single bp deletion	2.14
CECR2 DEL2 R	CATGGCGGATCCAGTCGACCAGGTTAACTATTTCCTTAG		
CECR2 E32K F	CTCTTCACAGAGATGACGTGAAGTTTATCAGTGACCTGATTG	E32K novel missense, position 22:17956663	2.14
CECR2 E32K R	CAATCAGGTCACTGATAAACTTCACGTCATCTCTGTGAAGAG		
CECR2 R245Q F	GAGAGGACCTCCCTTCAAGAACGGCAGCTCTAC	R245Q, rs201912432	2.14
CECR2 R245Q R	GTAGAGCTGCCGTTCTTGAAGGGAGGTCCTCTC		
CECR2 R271H F	CCCAGAAGGGAAAACATCCACAGCGCACAAAG	R271H, rs5747211	2.14
CECR2 R271H R	CTTTGTGCGCTGTGGATGTTTTCCCTTCTGGG		
CECR2 P632L F	CTCCTCTGGAGTCCTGGAGCCACACCCC	P632L, rs1296794	2.14
CECR2 P632L R	GGGGTGTGGCTCCAGGACTCCAGAGGAG		
CECR2 M680V F	CACACCTTTCTAACGTGGGCCCACACCCTG	M680V, rs62623401	2.14
CECR2 M680V R	CAGGGTGTGGGCCCACGTTAGAAAGGTGTG		
CECR2 Y738C F	CCTCCCAGCCATATGTGTCGATCGTACAAGTAC	Y738C, rs199565531	2.14
CECR2 Y738C R	GTACTTGTACGATCGACACATATGGCTGGGAGG		
CECR2 P1210R F	CAGTCAGCCTCCCCGACCAAGGTCCCTC	P1210R, rs199780601	2.14
CECR2 P1210R R	GAGGGACCTTGGTCGGGGAGGCTGACTG		
CECR2 S1261L F	GGAAATGTACAGACCATTAGGAATGCAGATGCACC	S1261L, rs142851999	2.14
CECR2 S1261L R	GGTGCATCTGCATTCCTAATGGTCTGTACATTTCC		
CECR2 P1430L F	CACAGGAGGAGGTGCTGCCTCATAAGCCTC	P1430L, rs181553013	2.14
CECR2 P1430L R	GAGGCTTATGAGGCAGCACCTCCTCCTGTG		
T7 pGEM*	TAATACGACTCACTATAGGG	Sequence validation of pEZY3 sub-cloning	2.15
SP6 pGEM*	TATTTAGGTGACACTATAG	Sequence validation of pCSDes sub-cloning	2.15
Ar19Ex6-F*	GACATGCACTGGTGAGCACT	*Arhgap*[19] nonsense mutation genotyping	2.22
Ar19Ex6-R	ACTCTGCCCCATCTTCCTTT		
Elmo2-1 F	AGCCCTCAGCCTCATCCT	Endogenous control, Probe #13, no known SNVs under primer/probe (www.ensembl.org)	2.24
Elmo2-1 R	AAGCCTAAGAGAGCAGGACTCA		
VldlrPr102-F	TTGCAGCTCAGAAGCTGTTTT	Probe #102, no known SNVs under primer/probe (www.ensembl.org)	2.24
VldlrPr102-R	AAATGTCTACCAACCTTGTCATCA		
Scd1-34-F	TTCCCTCCTGCAAGCTCTAC	Probe #34, no known SNVs under primer/probe (www.ensembl.org)	2.24
Scd1-34-R	CAGAGCGCTGGTCATGTAGT		
Lipo1-Pr106-F	CACCTCAAGGATTTTCTTCAGTTAT	Probe #106	2.24
Lipo1-Pr106-R	ACATGAATAAGGCACATTGTTCTC		
lipo1pr106seq-F	CATGGAATTTTGAAGCCTCTG	Sequence validation of *Lipo1* primer/probe	2.24
lipo1pr106seq-R*	TCCCATGAGGAATTCGGTTA		

Name	Sequence	Additional information	Thesis chapter
Lipa-Pr102-F	TGCCAGACCCTCTTCTCAAG	Probe #102	2.24
Lipa-Pr102-R	ATGATGACGTGGGTGCAAA		
lipapr102seq-F*	GGCTGCACCATAGGTTTCAT	Sequence validation of *Lipa* primer/probe	2.24
lipapr102seq-R	CGCAGGGCAGTGTGTTGTAT		
Fam160b1-45-F	CTGGAAGGCGATTACACACTACT	Probe #45	2.24
Fam160b1-45-R	GGATGGGATGTTTGTGTCG		
Fam160b1-45-SF*	TCCAAGCTTACGTCCATCCT	Sequence validation of *Fam160b1* primer/probe	2.24
Fam160b1-45-SR	TCCCGCTCATTTTCTTCTTG		
Pnliprp2-94-F	TGGAACAAACCTTTTTCCTGA	Probe #94	2.24
Pnliprp2-94-R	GTGTGACTGATACTTTGTATCTCCAAC		
Pnliprp2-94-SF*	TCCAGCACAATGACTGCTTC	Sequence validation of *Pnliprp2* primer/probe	2.24
Pnliprp2-94-SR	GCGACCAAGATGTACCCACT		
Fam45a13-F	GCTGTCCAGGAGTTCACCAG	Probe #13	2.24
Fam45a13-R	GATGGTCCAATCCTGTCGAT		
Fam45a13-SF*	TGCTGGCTCCATCAAAGACA	Sequence validation of *Fam45a* primer/probe	2.24
Fam45a13-SR	AGAGGTCTGGTCTGTTGCTC		
Hif1an96-F	GCCTCTATCCATACCCTGTCC	Probe #96	2.24
Hif1an96-R	CAGGATTGTCAAAGTCCACCT		
Hif1an96-SF*	GGAAATGTGACACCTGCTCA	Sequence validation of *Hif1an* primer/probe	2.24
Hif1an96-SR	GCCAACCACTGTCTCATAACC		
Sfxn2-18-F	TGCCGCTAACTGTGTCAATATC	Probe #18	2.24
Sfxn2-18-R	CTTCACACAGATGCCCTGAA		
Sfxn2-18-SF*	AGCTACTACTACTGCGGTGG	Sequence validation of *Sfxn2* primer/probe	2.24
Sfxn2-18-SR	TCTGAGAATGGCCAAGCTCA		
Sfrp5-104-F	GACAACGACCTCTGCATCG	Probe #104	2.24
Sfrp5-104-R	TCAGCGCTGTGCTCCAT		
Sfrp5-104-SF*	CTCATGGAGGCCTACGGTTT	Sequence validation of *Sfrp5* primer/probe	2.24
Sfrp5-104-SR	CCACAAAGTCACTGGAGCAC		
Rnls17-F	TGACCTTGTCATCCTCACCA	Probe #17	2.24
Rnls17-R	TCCCTCTGGCGTTCACTAAT		
Rnls17-SF*	AGTCTCCCTCAAGCACTGTG	Sequence validation of *Rnls* primer/probe	2.24
Rnls17-SR	AGCATAGCGAGAGGAGTAGC		
March5-76-F	GGGTAGATGAAAAGCAAAGAGGA	Probe #76	2.24
March5-76-R	GATCCAAGACATAAACCACTGGA		
March5-76-SF*	GCTTTGCCACCGATGAAGAT	Sequence validation of *March5* primer/probe	2.24
March5-76-SR	CTGCATCACTGTCACTGCTC		
Foxd4-1 F	AAAAACAAAACAGTCTGCAATACAA	Probe #9, assay designed by M. Kooistra	2.24
Foxd4-1 R	AACACTGGCAGTTCCTAAGCA		
foxd4qRTseq-F	CACGGTTTTCCCGACTTACA	Sequence validation of *Foxd4* primer/probe	2.24
foxd4qRTseq-R*	TTTTAACACTCCCCGTTGGA		
scd2qRTseq-F*	GCTTTGTCCCTGATGACCTC	Sequence validation of M. Kooistra's *Scd2* primer/probe	2.24
scd2qRTseq-R	GTGTTGAACTTGGCCCCTTA		
pax2qRTseq-F*	CCCAAAGTGGTGGACAAGAT	Sequence validation of M. Kooistra's *Pax2* primer/probe	2.24
pax2qRTseq-R	GTTAGAGGCGCTGGAAACAG		

Name	Sequence	Additional information	Thesis chapter
exosc1qRTseq-F	GCGAACGTCTGTGTAACTTGG	Sequence validation of M. Kooistra's *Exosc1* primer/probe	2.23
exosc1qRTseq-R*	GCATTTTTGAGTGGTGTGGA		
dmrt2qRTseq-F*	GTGGTTCCAGAGAGGTCTGC	Sequence validation of M. Kooistra's *dmrt2* primer/probe	2.23
dmrt2qRTseq-R	CTGGCTCTCCTTGACCAAAC		
ar19qRTseq-F*	GCATCATCCCTGCTCTCCTA	Sequence validation of M. Kooistra's *Arhgap19* primer/probe	2.23
ar19qRTseq-R	TCCAGTTTCCATGGTGACAA		
tmem180qRTseq-F*	GGCCCAGATCCATACTCTGA	Sequence validation of M. Kooistra's *Tmem180* primer/probe	2.23
tmem180qRTseq-R	CCCTACCCTTTCTGGTCCTC		
Ar19In2WT-F	GATGTTGACTGGCTCGGTCT	*Arhgap19*$^{Gt(YHD020)Byg}$ genotyping	2.25
Ar19In2WT-R	CCCAATACTGTCAGCCTGGT		
Engeo F	AACAAACTTGGCCTCACCAG		
Engeo R	AAATTCAGACGGCAAACGAC		
DK Pelem 3F*	CGAGCGCAGAATTCAGACAA	Sequence validation of P element precise excision	2.29
DK Pelem 5R	ACCTTTCATCGCGTGCAATT		
Gen-1F*	TGTCACAACTTTGCACTTGG	Sequencing *Gen* (*Drosophila*)	2.32
Gen-1R	CCTTTTGGCTATCACCTGGC		
DK5WT3-F	CGTTTGAGGCGTAGGAAGTG	Testing for presence of wildtype *dikar* in Df(3L)ED212 deficiency region	2.30
DK5WT3-R	TGTAAGCGAAGAGTATTTTGGCT		
DK5WT5-F	AATAGTAGTGCCCATTCGCG	Testing for presence of wildtype *dikar* in Df(3L)ED212 deficiency region	2.30
DK5WT5-R	CATGTGTGTGTGTTGGGA		
Gen-2F*	ACGCTTTACCAACAGGAACC	Sequencing *Gen* (*Drosophila*)	2.32
Gen-2R	ATAGACACGAACGGCTCCAT		
Gen-3F*	CAACAAGCATGGCGTACGTA	Sequencing *Gen* (*Drosophila*)	2.32
Gen-3R	GACTCTCATCGCAGCCTTTG		
Gen-4F*	CTGCAGAATGAGAAGCTGGC	Sequencing *Gen* (*Drosophila*)	2.32
Gen-4R	GCAATTCTAAGCTGGGCACA		
Gen-5F*	AAATCTCAACTGGCGGCAAC	Sequencing *Gen* (*Drosophila*)	2.32
Gen-5R	GAGTACCTTCCTTCTTGCGC		
Gen-6F*	ACCGCATCTGAAGAGGAGAA	Sequencing *Gen* (*Drosophila*)	2.32
Gen-6R	TGCTGCATCTTGAATCGCTC		
Gen-7F*	TTCATTACGAACCCATGGCG	Sequencing *Gen* (*Drosophila*)	2.32
Gen-7R	AACAGTTGGAGCAGGCAATG		
DMCTCO-F3*	TCTGCCTTTGATCTGCCCGACAT	Sequencing *dikar*, primers designed by A. Kueling and ordered from Invitrogen	2.32
DMCTCO-R3*	CGATGTTGCGAATGCGGCTGGAGTT		
DMCTCO-F3	TCTGCCTTTGATCTGCCCGACAT	Sequencing *dikar*, primers designed by A. Kueling and ordered from Invitrogen	2.32
DMCECR2-5E-R1*	GCTCATCCTTGTTCCGTCTA		
DMCECR2-5E-F1	GGAGAATCTCTACCAGGATC	Sequencing *dikar*, primers designed by A. Kueling and ordered from Invitrogen	2.32
DMCTCO-R2*	GCTGCTGCTCTGGCTGATCGGACTG		
Drcecr2-6e-F1*	CGACAGCTAGTGGTATTTCAA	Sequencing *dikar*, primers designed by A. Kueling, forward primer ordered from Qiagen, reverse primer ordered from Invitrogen	2.32
DMCTCO-R1*	GGCGCTATATCCTCCTCCACGGG		
DMCECR2-6E-F2*	GCACAGCCACCAGCATTAAC	Sequencing *dikar*, primers designed by A. Kueling and ordered from Invitrogen	2.32
DMCECR2-9E-R3*	CGGACGCTGATGTGGCTATT		

Name	Sequence	Additional information	Thesis chapter
DMCECR2-9E-F4*	CTCCGATGACGAAGATGATG	Sequencing *dikar*, primers designed by A. Kueling and ordered from Invitrogen	2.32
DMCECR2-9E-R4*	CCTTTGCGCTGCGAGACTTA		
DMCECR2-9E-F5	GAGCGAGGATTACAGCGATG	Sequencing *dikar*, primers designed by A. Kueling and ordered from Invitrogen	2.32
DMCECR2-9E-R3*	CGGACGCTGATGTGGCTATT		
DMCECR2-9E-F5	GAGCGAGGATTACAGCGATG	Sequencing *dikar*, primers designed by A. Kueling and ordered from Invitrogen	2.32
DMCECR2-10E-R3*	CCGATTTGCTAACTGCCTCT		
DMCECR2-9E-F1*	GGCAAAGCGAAGGGTAATCT	Sequencing *dikar*, primers designed by A. Kueling and ordered from Invitrogen	2.32
DM-CECR2-10E-R1*	GGCATGGTCTTGTCGAAATA		
DMCECR2-10E-F2*	CTCTGGGCAACAATGGATAT	Sequencing *dikar*, primers designed by A. Kueling and ordered from Invitrogen	2.32
DM-CECR2-10E-R1	GGCATGGTCTTGTCGAAATA		
DMCECR2-10E-F3*	GACGCCGTGGATAATTTGCA	Sequencing *dikar*, primers designed by A. Kueling and ordered from Invitrogen	2.32
DM-CECR2-10E-R1	GGCATGGTCTTGTCGAAATA		
DM-CECR2-10E-F1*	GCACTATGGCATTCGGCAA	Sequencing *dikar*, primers designed by A. Kueling and ordered from Invitrogen	2.32
DMCECR2-10E-R2*	GGCAAACGGATAAGCAGCT		
Dikar14-F*	GACTTCCACACCAGAGAGCT	Sequencing *dikar*	2.32
Dikar14-R	ATGACCTGCACATCGACGG		
Dikar15-F	TGGAAGAATTGGTTGGATTTCGT	Sequencing *dikar*	2.32
Dikar15-R*	GATTTCGGCGCTTCTGGAG		
Dikar16-F*	GGAGCAGCAGTCCGATCA	Sequencing *dikar*	2.32
Dikar16-R*	GACTTGCCCGTCGTCTCC		
Dikar17-F*	TCGGCTCATTTCATTCACACA	Sequencing *dikar*	2.32
Dikar17-R*	GGGTGGAGTAGTGATGTGCT		
Dikar18-F*	AGCACATCACTACTCCACCC	Sequencing *dikar*	2.32
Dikar18-R	ACATCCGCCTTACTCTTAGT		
Dikar19-F	TTAATCGCTGGCTGGAACAC	Sequencing *dikar*	2.32
Dikar19-R*	ACTCAGTTTTGGTGCTCCCT		
Dikar20-F*	CTGGCAGCACAAACTCAGAG	Sequencing *dikar*	2.32
Dikar20-R*	TTCGTAATACTTGGCGGCAG		
Dikar21-F*	CCCCAGGTGACAACAAACAA	Sequencing *dikar*	2.32
Dikar21-R*	GTTGCGCCTGTTGTTGCT		
Dikar21n-F*	ATCAGCAACAGCAAGCACAA	Sequencing *dikar*, primers nested within above primer pair	2.32
Dikar21n-R*	CTTCAGCCGCAATTGTAGCT		
Dikar22-F*	CAGAGATGCACCCAATCCCT	Sequencing *dikar*	2.32
Dikar22-R	TGCATGTGGATTTGAAACCGT		

Appendix B: Cranial NTD phenotypes for 156 probands. Coding exons of 25 genes were sequenced in each proband. Probands included in Table 4.6.1 are listed in the same order here.

Proband ID	Cranial NTD type	Genes containing variants of interest
A1024167	anencephaly	*CECR2, DNMBP*
20-0212267	anencephaly	*CECR2, DNMBP*
20-0504243	anencephaly	*CECR2, DNMBP*
A1015381	anencephaly	*CECR2, MMS19*
20-0421416	craniorachischisis	*CECR2*
20-0601809	anencephaly	*CECR2*
20-0608812	acrania	*CECR2, GLIS3, SFXN2*
20-0711091	anencephaly	*CECR2*
Z0156471	anencephaly	*CECR2*
A1021271	anencephaly	*CECR2*
A1021263	anencephaly	*CECR2*
20-0714522	anencephaly	*CECR2*
20-0800216	cranial meningocele	*CECR2*
20-0515262	anencephaly	*CECR2*
20-0601092	craniorachischisis	*CECR2, PNLIPRP2*
A1024158	anencephaly	*CECR2*
20-0602035	anencephaly	*CECR2*
20-0620476	anencephaly	*DNMBP, MMS19, TJP2, FAS*
20-0521275	anencephaly	*DNMBP, MMS19, TJP2*
20-0421455	anencephaly	*DNMBP, TCTN3*
20-0504690	anencephaly	*DNMBP*
20-0801162	anencephaly	*DNMBP*
A1021253	anencephaly	*DNMBP, SFXN2*
A1024197	anencephaly	*DNMBP*
A1024256	anencephaly	*DNMBP*
20-0426714	anencephaly	*DNMBP*
20-0613589	occipital encephalocele	*DNMBP, GLIS3*
20-0410539	anencephaly	*DNMBP*
20-0512275	acrania	*DNMBP, EXOSC1*
20-0514451	anencephaly	*DNMBP*
20-0601184	anencephaly	*DNMBP, SCD*
20-0806627	anencephaly	*DNMBP*
20-027140	anencephaly	*DNMBP*
20-0302223	anencephaly	*MMS19, TJP2*
A1024176	anencephaly	*MMS19, TJP2, GLIS3*
20-0700015	anencephaly	*MMS19, GLIS3, PNLIPRP2*
20-0300929	anencephaly	*MMS19, SFXN2, TMEM180*
20-0604224	acrania	*MMS19, GLIS3*
A1030655	anencephaly	*MMS19*
Z0167301	anencephaly	*MMS19*
20-0427674	anencephaly	*MMS19*
20-0708160	occipital encephalocele	*MMS19*
20-0705228	anencephaly	*TJP2, PNLIPRP2*

Proband ID	Cranial NTD type	Genes containing variants of interest
Z0082501	anencephaly	*TJP2*
A1024159	anencephaly	*TJP2*
A1021228	anencephaly	*TJP2*
A1024188	anencephaly	*TJP2*
20-0518590	anencephaly	*TJP2*
20-0609211	anencephaly	*TJP2*
20-0613565	anencephaly	*TJP2*
A1034284	anencephaly	*TJP2*
20-0426206	anencephaly	*TJP2, HPS6*
20-0517865	anencephaly	*TJP2*
A1030628	anencephaly	*TJP2, CPEB3*
20-0422758	anencephaly	*CSTF2T, PNLIPRP2*
A1021274	anencephaly	*CSTF2T, KANK1*
A1034295	anencephaly	*FAM45A, TCTN3*
Z0082809	anencephaly	*GLIS3*
20-0608051	anencephaly	*GLIS3*
20-0424645	occipital encephalocele	*GLIS3*
20-0802143	anencephaly	*HPS6*
20-0602679	anencephaly	*PNLIPRP2*
A1030639	anencephaly	*PNLIPRP2*
Z0174261	acrania	*PNLIPRP2*
20-0404771	anencephaly	*PNLIPRP2*
A1030634	anencephaly	*RNLS*
20-0800185	anencephaly	*SCD*
A1021240	anencephaly	*SFXN2*
Z0082701	acrania	*TCTN3*
20-0300594	anencephaly	*TCTN3*
20-014072	anencephaly	*TCTN3*
A1015387	anencephaly	*TMEM180*
20-028661	encephalocele	*KANK1*
Z0156466	acrania	*KANK1*
19-960239	occipital encephalocele	-
19-973956	encephalocele	-
20-0212266	anencephaly	-
20-0300536	anencephaly	-
20-0307768	anencephaly	-
20-0409349	anencephaly	-
20-0409672	anencephaly	-
20-0409946	craniorachischisis	-
20-0410779	encephalocele	-
20-0411056	anencephaly	-
20-0411879	anencephaly	-
20-0412168	anencephaly	-
20-0424420	anencephaly	-
20-0424564	anencephaly	-
20-0425711	occipital encephalocele	-
20-0427136	anencephaly	-
20-0504382	anencephaly	-

Proband ID	Cranial NTD type	Genes containing variants of interest
20-0505812	anencephaly	-
20-0511614	anencephaly	-
20-0512010	occipital encephalocele	-
20-0513594	anencephaly	-
20-0515291	anencephaly	-
20-0515436	anencephaly	-
20-0516527	anencephaly	-
20-0518777	anencephaly	-
20-0519874	anencephaly	-
20-0521489	anencephaly	-
20-0601091	anencephaly	-
20-0601287	anencephaly	-
20-0602131	anencephaly	-
20-0606470	anencephaly	-
20-0608659	acrania	-
20-0609034	acrania	-
20-0612708	anencephaly	-
20-0700618	anencephaly	-
20-0700642	anencephaly	-
20-0700669	anencephaly	-
20-0707657	anencephaly	-
20-0708234	anencephaly	-
20-0708241	anencephaly	-
20-0711055	anencephaly	-
20-0714509	anencephaly	-
20-0800725	anencephaly	-
20-0802386	anencephaly	-
20-0802445	anencephaly	-
20-0804648	anencephaly	-
20-0805590	anencephaly	-
20-0806702	anencephaly	-
A1000707	occipital encephalocele	-
A1000764	anencephaly	-
A1013093	anencephaly	-
A1013101	anencephaly	-
A1013102	anencephaly	-
A1013566	acrania	-
A1015376	anencephaly	-
A1015378	anencephaly	-
A1015395	anencephaly	-
A1021227	anencephaly	-
A1021243	anencephaly	-
A1021267	anencephaly	-
A1021273	acrania	-
A1024152	anencephaly	-
A1024156	anencephaly	-
A1024183	anencephaly	-
A1024186	anencephaly	-

Proband ID	Cranial NTD type	Genes containing variants of interest
A1024191	acrania	-
A1024196	anencephaly	-
A1024259	anencephaly	-
A1030617	anencephaly	-
A1030621	acrania	-
A1030651	anencephaly	-
A1030654	anencephaly	-
A1034279	anencephaly	-
A1034304	acrania	-
A1034308	anencephaly	-
Z0141223	anencephaly	-
Z0141445	anencephaly	-
Z0161943	anencephaly	-
Z0164031	anencephaly	-
Z0164034	anencephaly	-
Z0174262	anencephaly	-
Z0176661	anencephaly	-

Appendix C: Co-immunoprecipitation assay comparing human wildtype CECR2 to variant CECR2 (E32K and P632L) in the ability to complex with SNF2L in transfected HEK293 cells.

A total of 9 variants of interest were identified in human *CECR2* when the coding exons were sequenced in a human cranial NTD cohort consisting of 156 probands (Chapter 3). In order to determine whether or not these variants are deleterious, functional assays need to be performed. Banting (2003) previously generated a human wildtype *CECR2* clone in the pMIB/V5-His A vector (Invitrogen) [175]. I then sequence verified this *CECR2* clone and moved it into the pENTR^{TM}11 Dual Selection Vector (Gateway® Life Technologies). The additional 9 *CECR2* constructs, each of which contain one of the 9 variants of interest, were generated from the *CECR2* clone by site-directed mutagenesis (see Appendix D for sequence information for all 10 *CECR2* constructs). Two variants of particular interest were the novel E32K variant within the CECR2 DDT domain, and the P632L variant. Although the P632L variant was common (MAF ~ 0.19), it was of particular interest because it contributed to compound heterozygosity in 4 probands, possibly contributed to compound heterozygosity in 4 additional probands, one of which was the proband containing the novel E32K variant, and contributed to homozygosity in yet another 3 probands. The *CECR2* constructs for these two variants were sub-cloned into the pEZY3 mammalian expression vector (Addgene plasmid #18672, gift from Dr. Fred Berry) [176]. As the E32K variant is within the DDT domain, which functions to interact with the ISWI binding partner SNF2L, we were interested in testing whether or not this variant (as well as the P632L variant) affected SNF2L binding. HEK293 cells were chosen as the human cell type to test this protein interaction as HEK293 cells were originally used to characterize the CERF complex (CECR2 and SNF2L), which involved the successful transfection and characterization of CECR2 in HEK293 cells [138]. Ideally, a cell type with little to no endogenous CECR2 expression in conjunction with detectable levels of the CECR2 binding partner, SNF2L, would be used. Western blot analysis indicated the presence of endogenous CECR2 protein after a long exposure (> 2 hours) and endogenous SNF2L protein after a short exposure (~5 minutes) in HEK293 cells (Figure C.1). HEK293 cells were then transfected with a wildtype-*CECR2* containing expression vector, an E32K-*CECR2* containing expression vector, or a P632L-*CECR2* containing expression vector. Then, co-immunoprecipitation (co-IP) assays utilizing the α-

CECR2 antibody were performed to determine how much SNF2L could be pulled down in CECR2 complexes. Negative controls included (1) an IgG immunoprecipitation (IP) in lieu of an α-CECR2 IP, which should not specifically bind to and pull down CECR2 containing complexes, (2) non-transfected HEK293 cells, which should only contain endogenous levels of CECR2, and (3) EGFP transfected cells, which should behave similarly to non-transfected cells in the co-IP assay. Preliminary results indicated that α-CECR2 antibody was able to pull down SNF2L in cells transfected with the wildtype-CECR2 construct as well as the two variant CECR2 constructs (Figure C.2). The ability for α-CECR2 antibody to pull down SNF2L appears to be specific because the loading-control protein, Tubulin, was present in input protein but not pulled down in IPs. In order to determine if there is a difference in SNF2L binding between wildtype and variant CECR2, we need to be able to compare the amount of CECR2 present in the transfected cells. However, α-CECR2 antibody was unable to pull down detectable levels of CECR2 in all transfected HEK293 cell lines except for HEK293 cells transfected with wildtype-CECR2 (Figure C.2A). Further experimentation is required to explain the discrepancy between the lack of CECR2 and the presence of SNF2L in the CECR2 IPs. Also, α-CECR2 antibody was able to pull down detectable levels of SNF2L in non-transfected and EGFP-transfected cells (Figure C.2B). Optimization of this protocol in HEK293 cells is required in order to be able to detect CECR2 protein from α-CECR2 IPs. Performing functional assays of the CECR2 complex in a cell line that does not express endogenous CECR2 but expresses CECR2 binding partners, SNF2H and/or SNF2L, would be ideal.

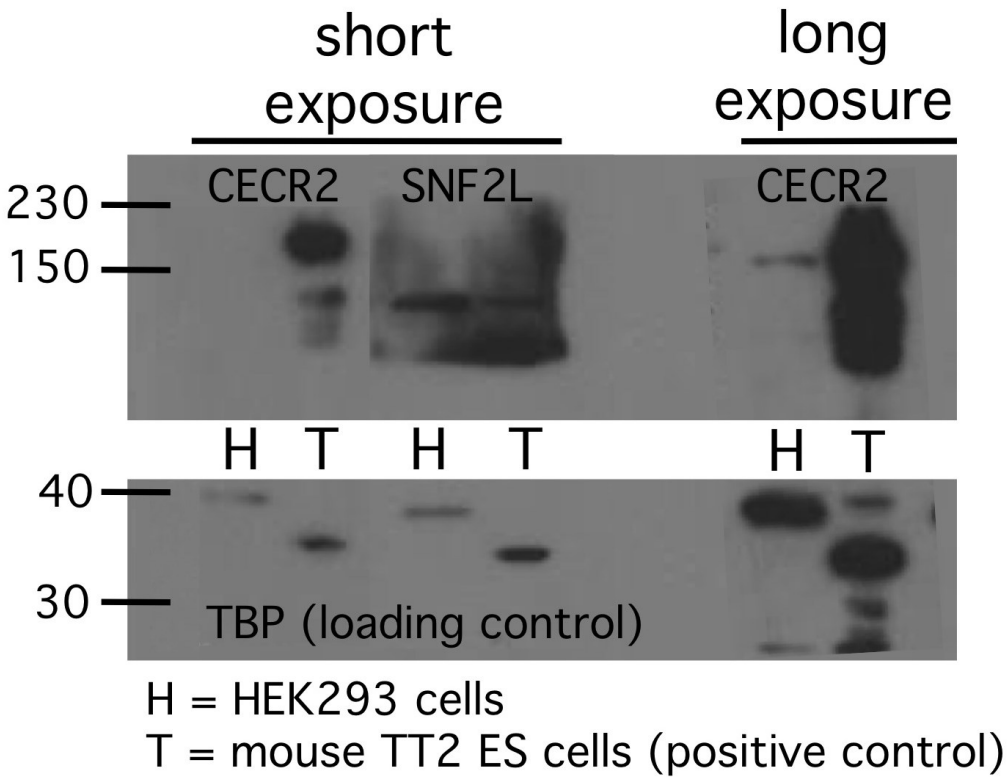


Figure C.1: HEK293 cells have low endogenous levels of CECR2 when compared to the level of CECR2 in mouse TT2 ES cells. CECR2 is easily detectable after a short exposure (~5 minutes) in TT2 ES cells, but is not detectable at all in HEK293 cells. A detectable level of CECR2 is seen after a long exposure (>2 hours) in HEK293 cells. The level of the CECR2 ISWI binding partner, SNF2L, is detectable after a short exposure and is similar to SNF2L levels in mouse TT2 ES cells. Tata-binding protein (TBP) was used as a loading control. The difference in molecular weight is because this antibody detects an ~40 kDa band for human (HEK293 cells) and an ~34 kDa band for mouse (TT2 ES cells) according to the manufacturer (ProteinTech^{TM}). Numbers represent molecular weight in kilo Daltons (kDa). H = HEK293 cells, T = mouse TT2 ES cells.

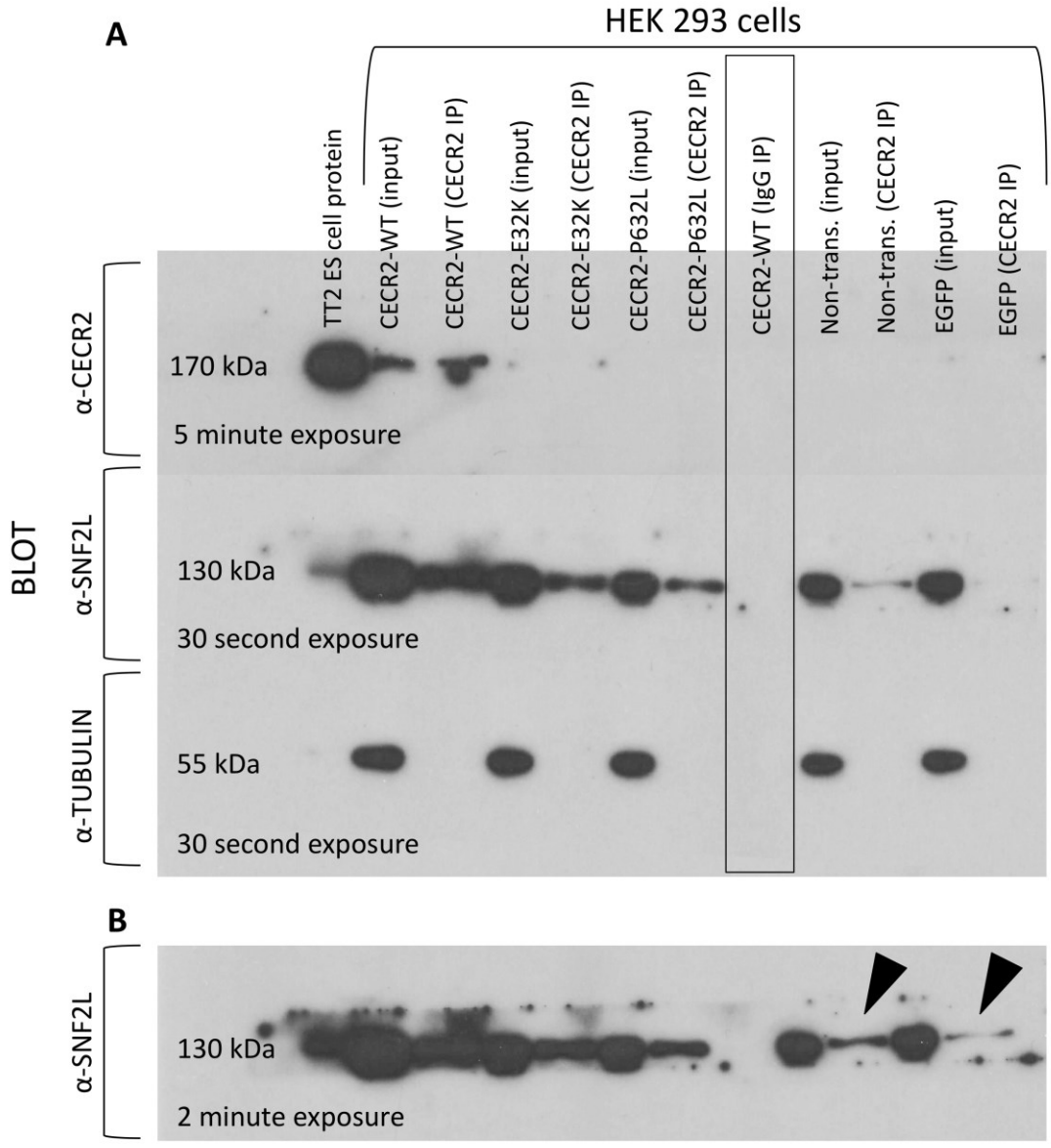


Figure C.2: α-CECR2 antibody is capable of specifically pulling down SNF2L in HEK293 cells. SNF2L was present in CECR2 IP protein from HEK293 cells transfected with wildtype-*CECR2*, E32K-*CECR2*, and P632L-*CECR2* (**A**). IgG immunoprecipitation performed on HEK293 cells transfected with wildtype-*CECR2* did not pull down CECR2, SNF2L, or the loading control protein, Tubulin, as expected (IgG lane is indicated with a black box). SNF2L was also present in CECR2 IP protein from non-transfected HEK293 cells as well as HEK293 cells transfected with *EGFP* as shown by longer exposure of the α-SNF2L blot, with bands indicated by black arrowheads (**B**).

Appendix D: *CECR2* ORF constructs in the pENTR™11 plasmid. Vector sequence is in black text and the *CECR2* ORF is in red text. Gateway® recombination sites (*attL*) are underlined. Introduced missense mutations are highlighted in red.

CECR2 – WT

CAAATAATGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATAAGCAATGCTTTTTTATAATGCCAACTTTGTAC
AAAAAAGCAGGCTTCGAAGGAGATAGAACCAATTCTCTAAGGAAATACTTAACCTGGTCGACTGGATCCGCCACC**ATGTGCCCAGAG**
GAGGGCGGCGCGGCCGGGCTGGGCGAGCTCCGCTCCTGGTGGGAGGTCCCGGCCATCGCGCACTTCTGCTCGCTCTTTCGCACCGCG
TTCCGCCTGCCCGACTTCGAGATCGAGGAGTTAGAAGCCGCTCTTCACAGAGATGACGTGGAGTTTATCAGTGACCTGATTGCCTGC
CTGCTTCAGGGCTGCTATCAACGAAGAGATATCACGCCTCAGACATTCCACAGCTACCTAGAGGACATCATCAACTACCGCTGGGAG
CTCGAAGAAGGGAAGCCCAACCCTCTGAGGGAAGCCAGTTTCCAGGACCTGCCTCTTCGCACACGGGTGGAGATCCTGCACCGACTC
TGTGATTACCGGCTGGATGCAGACGATGTCTTCGATCTTCTAAAGGGCCTGGATGCAGACAGTCTCCGTGTGGAGCCATTGGGTGAA
GACAATTCTGGGGCACTATATTGGTATTTCTATGGAACACGAATGTACAAAGAGGACCCGGTGCAAGGAAAATCCAATGGAGAACTC
TCTTTGAGCAGGGAAAGTGAAGGACAAAAAAATGTCTCAAGTATTCCTGGAAAAACGGGAAAAAGAAGAGGAAGACCCCCAAAACGG
AAGAAACTGCAGGAGGAGATTCTGTTGAGTGAAAAGCAGGAAGAAAATTCCTTGGCATCCGAGCCACAGACAAGACATGGGTCCCAA
GGGCCAGGCCAAGGTACTTGGTGGCTCCTGTGCCAGACAGAAGAGGAATGGAGACAGGTCACCGAGAGTTTTCGCGAGAGGACCTCC
CTTCGAGAACGGCAGCTCTACAAGCTCCTCAGTGAGGACTTCCTGCCTGAGATCTGCAACATGATCGCCCAGAAGGGAAAACGTCCA
CAGCGCACAAAGGCAGAGTTGCATCCTAGGTGGATGTCTGACCACCTGTCCATCAAACCCGTCAAGCAAGAGGAGACTCCTGTGCTG
ACCAGAATAGAAAAACAAAAGCGCAAAGAGGAGGAAGAAGAGCGTCAGATTCTTCTAGCAGTGCAGAAGAAGGAGCAGGAGCAGATG
CTAAAGGAAGAGAGGAAACGCGAGTTGGAGGAGAAGGTCAAGGCAGTGGAAGATCGAGCGAAGAGGAGAAAGCTCAGGGAAGAAAGG
GCATGGCTGCTGGCTCAAGGAAAGGAGCTCCCTCCAGAACTTTCCCATCTGGACCCCAATTCCCCCATGAGAGAGGAAAAAAAGACT
AAAGACCTCTTTGAGTTGGATGATGATTTCACTGCTATGTATAAAGTTCTAGACGTGGTAAAGGCTCACAAGGATTCCTGGCCCTTC
TTGGAACCTGTGGATGAATCTTATGCCCCTAACTATTATCAGATTATTAAGGCCCCCATGGATATTTCCAGCATGGAGAAGAAACTG
AATGGAGGTTTATACTGTACCAAGGAGGAATTTGTAAATGACATGAAGACCATGTTCAGGAATTGTCGAAAGTATAATGGGGAAAGT
AGTGAGTATACCAAGATGTCTGATAATTTAGAGAGGTGTTTCCATCGGGCAATGATGAAACATTTTCCTGGAGAAGATGGAGACACA
GATGAAGAATTTTGGATTCGAGAGGATGAAAAGCGGGAGAAAAGACGGAGTCGGGCTGGGCGAAGTGGTGGGAGCCATGTTTGGACC
CGCTCCAGGGACCCAGAAGGGTCCAGCAGGAAACAGCAGCCCATGGAGAATGGAGGAAAGTCGTTGCCCCCCACACGCCGAGCGCCC
TCTTCTGGGGACGATCAGAGCAGCAGCTCCACACAGCCCCCGCGGGAGGTGGGCACTTCCAATGGCCGAGGTTTTTCTCATCCCCTG
CATTGTGGTGGGACACCCAGCCAGGCACCCTTTTTAAACCAGATGAGGCCAGCAGTACCAGGAACATTTGGCCCTCTGCGAGGATCA
GATCCTGCCACCTTGTATGGCTCCTCTGGAGTCCCGGAGCCACACCCCGGGGAGCCTGTGCAGCAGCGTCAGCCTTTCACCATGCAG
CCTCCAGTTGGAATTAACAGCCTCCGAGGACCCAGGCTAGGCACACCAGAGGAGAAGCAAATGTGCGGGGGGCTGACACACCTTTCT
AACATGGGCCCACACCCTGGATCCTTGCAGCTTGGGCAGATAAGTGGCCCAAGTCAGGATGGAAGCATGTATGCTCCAGCTCAGTTC
CAGCCAGGATTCATTCCTCCCCGGCATGGGGGCGCTCCAGCCCGGCCACCAGACTTTCCTGAAAGCTCAGAAATTCCTCCCAGCCAT
ATGTATCGATCGTACAAGTACCTGAATCGAGTACACTCTGCCGTCTGGAATGGGAACCATGGTGCTACGAACCAAGGACCCTTGGGC
CCAGATGAGAAGCCCCACCTGGGGCCAGGACCCTCTCACCAGCCTCGCACTCTCGGTCACGTGATGGATTCCCGAGTCATGAGACCA
CCTGTCCCCCCCAACCAGTGGACTGAACAATCAGGCTTCCTACCTCATGGAGTTCCTTCCTCAGGGTACATGCGACCGCCCTGCAAG
TCTGCCGGACATCGGTTACAGCCACCTCCAGTGCCAGCACCCAGTTCTTTGTTTGGAGCACCTGCCCAGGCTCTTCGGGGGGTGCAG
GGAGGGGACTCCATGATGGACAGCCCAGAGATGATTGCGATGCAGCAGCTCTCCTCCCGCGTCTGCCCCCCAGGTGTGCCTTACCAC
CCCCACCAGCCTGCACACCCCCGTTTACCTGGCCCTTTTCCGCAGGTAGCTCACCCAATGTCAGTCACTGTGTCAGCCCCCAAGCCT
GCCCTGGGCAACCCTGGGAGGGCACCGGAGAACAGTGAAGCACAAGAGCCTGAGAATGACCAAGCAGAGCCGTTGCCTGGCCTTGAA
GAGAAACCACCAGGTGTTGGTACTTCAGAGGGGGTCTACCTCACACAACTACCTCACCCCACACCTCCCCTGCAGACTGACTGCACC
AGGCAGAGCTCACCACAAGAAAGGGAAACAGTGGGCCCGGAGCTCAAAAGCAGCTCCTCCGAATCTGCGGACAACTGTAAAGCAATG
AAGGGCAAGAATCCCTGGCCCTCGGATAGCAGCTACCCCGGCCCAGCCGCCCAAGGGTGCGTGAGAGACCTCTCCACGGTGGCAGAC
AGGGGCGCTCTATCCGAGAACGGAGTCATTGGGGAAGCATCTCCTTGTGGATCGGAGGGGAAGGGCCTTGGTAGCAGTGGTTCCGAA
AAGCTGCTCTGCCCCAGAGGCAGAACGTTGCAGGAAACCATGCCATGCACGGGACAGAACGCAGCGACACCGCCCAGCACAGACCCC
GGTTTGACGGGAGGCACTGTGAGCCAGTTTCCCCCGCTGTATATGCCTGGCCTAGAGTACCCGAATTCAGCTGCCCATTACCACATC
AGTCCAGGCCTGCAGGGTGTGGGCCCTGTGATGGGAGGGAAGTCCCCAGCATCCCATCCCCAGCATTTTCCCCCAAGGGGCTTTCAG
TCTAACCACCCACATTCTGGAGGCTTTCCCCGGTATCGCCCCCCACAAGGAATGAGGTATTCCTACCACCCACCGCCACAGCCTTCC
TACCACCACTATCAGCGAACTCCTTACTATGCCTGTCCACAGAGCTTTTCTGACTGGCAGAGACCTCTCCATCCCCAGGGAAGCCCA
AGCGGACCCCCAGCCAGTCAGCCTCCCCCACCAAGGTCCCTCTTCTCAGATAAGAATGCCATGGCCAGTCTGCAAGGCTGTGAGACA
CTGAATGCTGCCTTAACTTCTCCAACCCGTATGGATGCAGTGGCTGCTAAAGTCCCAAATGACGGGCAGAATCCTGGTCCAGAGGAA
GAGAAGCTGGATGAATCTATGGAGAGGCCAGAGAGTCCCAAAGAATTTTTAGACCTGGACAACCATAACGCAGCTACCAAGCGGCAG
AGCTCGTTGTCAGCCAGCGAGTATCTCTATGGAACTCCTCCGCCTCTGAGTTCAGGAATGGGATTTGGTTCATCTGCATTTCCACCC
CACAGTGTGATGCTGCAGACGGGGCCTCCCTATACCCCTCAGCGGCCGGCCAGTCACTTTCAGCCCAGGGCTTACTCTTCTCCTGTG
GCTGCCCTCCCACCTCACCACCCAGGGGCCACCCAGCCCAACGGCCTCTCTCAGGAGGGTCCCATCTATCGCTGCCAGGAAGAAGGC
CTGGGTCACTTTCAAGCTGTGATGATGGAACAAATTGGCACTAGAAGTGGAATAAGAGGACCTTTCCAGGAAATGTACAGACCATCA
GGAATGCAGATGCACCCGGTCCAGTCGCAGGCCTCGTTCCCAAAGACCCCCACAGCAGCAACATCACAGGAGGAGGTGCCGCCTCAT
AAGCCTCCAACACTTCCCCTGGATCAGAGCTAGCTCGAGATATCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCA
TTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCATTATTTG

CECR2 – E32K

CAAATAATGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATAAGCAATGCTTTTTTATAATGCCAACTTTGTAC
AAAAAAGCAGGCTTCGAAGGAGATAGAACCAATTCTCTAAGGAAATACTTAACCTGGTCGACTGGATCCGCCACC**ATGTGCCCAGAG**
GAGGGCGGCGCGGCCGGGCTGGGCGAGCTCCGCTCCTGGTGGGAGGTCCCGGCCATCGCGCACTTCTGCTCGCTCTTTCGCACCGCG
TTCCGCCTGCCCGACTTCGAGATCGAGGAGTTAGAAGCCGCTCTTCACAGAGATGACGTGAAGTTTATCAGTGACCTGATTGCCTGC
CTGCTTCAGGGCTGCTATCAACGAAGAGATATCACGCCTCAGACATTCCACAGCTACCTAGAGGACATCATCAACTACCGCTGGGAG
CTCGAAGAAGGGAAGCCCAACCCTCTGAGGGAAGCCAGTTTCCAGGACCTGCCTCTTCGCACACGGGTGGAGATCCTGCACCGACTC
TGTGATTACCGGCTGGATGCAGACGATGTCTTCGATCTTCTAAAGGGCCTGGATGCAGACAGTCTCCGTGTGGAGCCATTGGGTGAA
GACAATTCTGGGGCACTATATTGGTATTTCTATGGAACACGAATGTACAAAGAGGACCCGGTGCAAGGAAAATCCAATGGAGAACTC
TCTTTGAGCAGGGAAAGTGAAGGACAAAAAAATGTCTCAAGTATTCCTGGAAAAACGGGAAAAAGAAGAGGAAGACCCCCAAAACGG
AAGAAACTGCAGGAGGAGATTCTGTTGAGTGAAAAGCAGGAAGAAAATTCCTTGGCATCCGAGCCACAGACAAGACATGGGTCCCAA
GGGCCAGGCCAAGGTACTTGGTGGCTCCTGTGCCAGACAGAAGAGGAATGGAGACAGGTCACCGAGAGTTTTCGCGAGAGGACCTCC
CTTCGAGAACGGCAGCTCTACAAGCTCCTCAGTGAGGACTTCCTGCCTGAGATCTGCAACATGATCGCCCAGAAGGGAAAACGTCCA
CAGCGCACAAAGGCAGAGTTGCATCCTAGGTGGATGTCTGACCACCTGTCCATCAAACCCGTCAAGCAAGAGGAGACTCCTGTGCTG
ACCAGAATAGAAAAACAAAAGCGCAAAGAGGAGGAAGAAGAGCGTCAGATTCTTCTAGCAGTGCAGAAGAAGGAGCAGGAGCAGATG
CTAAAGGAAGAGAGGAAACGCGAGTTGGAGGAGAAGGTCAAGGCAGTGGAAGATCGAGCGAAGAGGAGAAAGCTCAGGGAAGAAAGG
GCATGGCTGCTGGCTCAAGGAAAGGAGCTCCCTCCAGAACTTTCCCATCTGGACCCCAATTCCCCCATGAGAGAGGAAAAAAAGACT
AAAGACCTCTTTGAGTTGGATGATGATTTCACTGCTATGTATAAAGTTCTAGACGTGGTAAAGGCTCACAAGGATTCCTGGCCCTTC
TTGGAACCTGTGGATGAATCTTATGCCCCTAACTATTATCAGATTATTAAGGCCCCCATGGATATTTCCAGCATGGAGAAGAAACTG
AATGGAGGTTTATACTGTACCAAGGAGGAATTTGTAAATGACATGAAGACCATGTTCAGGAATTGTCGAAAGTATAATGGGGAAAGT
AGTGAGTATACCAAGATGTCTGATAATTTAGAGAGGTGTTTCCATCGGGCAATGATGAAACATTTTCCTGGAGAAGATGGAGACACA
GATGAAGAATTTTGGATTCGAGAGGATGAAAAGCGGGAGAAAAGACGGAGTCGGGCTGGGCGAAGTGGTGGGAGCCATGTTTGGACC
CGCTCCAGGGACCCAGAAGGGTCCAGCAGGAAACAGCAGCCCATGGAGAATGGAGGAAAGTCGTTGCCCCCCACACGCCGAGCGCCC
TCTTCTGGGGACGATCAGAGCAGCAGCTCCACACAGCCCCCGCGGGAGGTGGGCACTTCCAATGGCCGAGGTTTTTCTCATCCCCTG
CATTGTGGTGGGACACCCAGCCAGGCACCCTTTTTAAACCAGATGAGGCCAGCAGTACCAGGAACATTTGGCCCTCTGCGAGGATCA
GATCCTGCCACCTTGTATGGCTCCTCTGGAGTCCCGGAGCCACACCCCGGGGAGCCTGTGCAGCAGCGTCAGCCTTTCACCATGCAG
CCTCCAGTTGGAATTAACAGCCTCCGAGGACCCAGGCTAGGCACACCAGAGGAGAAGCAAATGTGCGGGGGGCTGACACACCTTTCT
AACATGGGCCCACACCCTGGATCCTTGCAGCTTGGGCAGATAAGTGGCCCAAGTCAGGATGGAAGCATGTATGCTCCAGCTCAGTTC
CAGCCAGGATTCATTCCTCCCCGGCATGGGGGCGCTCCAGCCCGGCCACCAGACTTTCCTGAAAGCTCAGAAATTCCTCCCAGCCAT
ATGTATCGATCGTACAAGTACCTGAATCGAGTACACTCTGCCGTCTGGAATGGGAACCATGGTGCTACGAACCAAGGACCCTTGGGC
CCAGATGAGAAGCCCCACCTGGGGCCAGGACCCTCTCACCAGCCTCGCACTCTCGGTCACGTGATGGATTCCCGAGTCATGAGACCA
CCTGTCCCCCCCAACCAGTGGACTGAACAATCAGGCTTCCTACCTCATGGAGTTCCTTCCTCAGGGTACATGCGACCGCCCTGCAAG
TCTGCCGGACATCGGTTACAGCCACCTCCAGTGCCAGCACCCAGTTCTTTGTTTGGAGCACCTGCCCAGGCTCTTCGGGGGGTGCAG
GGAGGGGACTCCATGATGGACAGCCCAGAGATGATTGCGATGCAGCAGCTCTCCTCCCGCGTCTGCCCCCCAGGTGTGCCTTACCAC
CCCCACCAGCCTGCACACCCCCGTTTACCTGGCCCTTTTCCGCAGGTAGCTCACCCAATGTCAGTCACTGTGTCAGCCCCCAAGCCT
GCCCTGGGCAACCCTGGGAGGGCACCGGAGAACAGTGAAGCACAAGAGCCTGAGAATGACCAAGCAGAGCCGTTGCCTGGCCTTGAA
GAGAAACCACCAGGTGTTGGTACTTCAGAGGGGGTCTACCTCACACAACTACCTCACCCCACACCTCCCCTGCAGACTGACTGCACC
AGGCAGAGCTCACCACAAGAAAGGGAAACAGTGGGCCCGGAGCTCAAAAGCAGCTCCTCCGAATCTGCGGACAACTGTAAAGCAATG
AAGGGCAAGAATCCCTGGCCCTCGGATAGCAGCTACCCCGGCCCAGCCGCCCAAGGGTGCGTGAGAGACCTCTCCACGGTGGCAGAC
AGGGGCGCTCTATCCGAGAACGGAGTCATTGGGGAAGCATCTCCTTGTGGATCGGAGGGGAAGGGCCTTGGTAGCAGTGGTTCCGAA
AAGCTGCTCTGCCCCAGAGGCAGAACGTTGCAGGAAACCATGCCATGCACGGGACAGAACGCAGCGACACCGCCCAGCACAGACCCC
GGTTTGACGGGAGGCACTGTGAGCCAGTTTCCCCCGCTGTATATGCCTGGCCTAGAGTACCCGAATTCAGCTGCCCATTACCACATC
AGTCCAGGCCTGCAGGGTGTGGGCCCTGTGATGGGAGGGAAGTCCCCAGCATCCCATCCCCAGCATTTTCCCCCAAGGGGCTTTCAG
TCTAACCACCCACATTCTGGAGGCTTTCCCCGGTATCGCCCCCCACAAGGAATGAGGTATTCCTACCACCCACCGCCACAGCCTTCC
TACCACCACTATCAGCGAACTCCTTACTATGCCTGTCCACAGAGCTTTTCTGACTGGCAGAGACCTCTCCATCCCCAGGGAAGCCCA
AGCGGACCCCCAGCCAGTCAGCCTCCCCCACCAAGGTCCCTCTTCTCAGATAAGAATGCCATGGCCAGTCTGCAAGGCTGTGAGACA
CTGAATGCTGCCTTAACTTCTCCAACCCGTATGGATGCAGTGGCTGCTAAAGTCCCAAATGACGGGCAGAATCCTGGTCCAGAGGAA
GAGAAGCTGGATGAATCTATGGAGAGGCCAGAGAGTCCCAAAGAATTTTTAGACCTGGACAACCATAACGCAGCTACCAAGCGGCAG
AGCTCGTTGTCAGCCAGCGAGTATCTCTATGGAACTCCTCCGCCTCTGAGTTCAGGAATGGGATTTGGTTCATCTGCATTTCCACCC
CACAGTGTGATGCTGCAGACGGGGCCTCCCTATACCCCTCAGCGGCCGGCCAGTCACTTTCAGCCCAGGGCTTACTCTTCTCCTGTG
GCTGCCCTCCCACCTCACCACCCAGGGGCCACCCAGCCCAACGGCCTCTCTCAGGAGGGTCCCATCTATCGCTGCCAGGAAGAAGGC
CTGGGTCACTTTCAAGCTGTGATGATGGAACAAATTGGCACTAGAAGTGGAATAAGAGGACCTTTCCAGGAAATGTACAGACCATCA
GGAATGCAGATGCACCCGGTCCAGTCGCAGGCCTCGTTCCCAAAGACCCCCACAGCAGCAACATCACAGGAGGAGGTGCCGCCTCAT
AAGCCTCCAACACTTCCCCTGGATCAGAGCTAGCTCGAGATATCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCA
TTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCATTATTTG

CECR2 – R245Q

CAAATAATGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATAAGCAATGCTTTTTTATAATGCCAACTTTGTAC
AAAAAAGCAGGCTTCGAAGGAGATAGAACCAATTCTCTAAGGAAATACTTAACCTGGTCGACTGGATCCGCCACC**ATGTGCCCAGAG**
GAGGGCGGCGCGGCCGGGCTGGGCGAGCTCCGCTCCTGGTGGGAGGTCCCGGCCATCGCGCACTTCTGCTCGCTCTTTCGCACCGCG
TTCCGCCTGCCCGACTTCGAGATCGAGGAGTTAGAAGCCGCTCTTCACAGAGATGACGTGGAGTTTATCAGTGACCTGATTGCCTGC
CTGCTTCAGGGCTGCTATCAACGAAGAGATATCACGCCTCAGACATTCCACAGCTACCTAGAGGACATCATCAACTACCGCTGGGAG
CTCGAAGAAGGGAAGCCCAACCCTCTGAGGGAAGCCAGTTTCCAGGACCTGCCTCTTCGCACACGGGTGGAGATCCTGCACCGACTC
TGTGATTACCGGCTGGATGCAGACGATGTCTTCGATCTTCTAAAGGGCCTGGATGCAGACAGTCTCCGTGTGGAGCCATTGGGTGAA
GACAATTCTGGGGCACTATATTGGTATTTCTATGGAACACGAATGTACAAAGAGGACCCGGTGCAAGGAAAATCCAATGGAGAACTC
TCTTTGAGCAGGGAAAGTGAAGGACAAAAAAATGTCTCAAGTATTCCTGGAAAAACGGGAAAAAGAAGAGGAAGACCCCCAAAACGG
AAGAAACTGCAGGAGGAGATTCTGTTGAGTGAAAAGCAGGAAGAAAATTCCTTGGCATCCGAGCCACAGACAAGACATGGGTCCCAA
GGGCCAGGCCAAGGTACTTGGTGGCTCCTGTGCCAGACAGAAGAGGAATGGAGACAGGTCACCGAGAGTTTTCGCGAGAGGACCTCC
CTTCAAGAACGGCAGCTCTACAAGCTCCTCAGTGAGGACTTCCTGCCTGAGATCTGCAACATGATCGCCCAGAAGGGAAAACGTCCA
CAGCGCACAAAGGCAGAGTTGCATCCTAGGTGGATGTCTGACCACCTGTCCATCAAACCCGTCAAGCAAGAGGAGACTCCTGTGCTG
ACCAGAATAGAAAAACAAAAGCGCAAAGAGGAGGAAGAAGAGCGTCAGATTCTTCTAGCAGTGCAGAAGAAGGAGCAGGAGCAGATG
CTAAAGGAAGAGAGGAAACGCGAGTTGGAGGAGAAGGTCAAGGCAGTGGAAGATCGAGCGAAGAGGAGAAAGCTCAGGGAAGAAAGG
GCATGGCTGCTGGCTCAAGGAAAGGAGCTCCCTCCAGAACTTTCCCATCTGGACCCCAATTCCCCCATGAGAGAGGAAAAAAAGACT
AAAGACCTCTTTGAGTTGGATGATGATTTCACTGCTATGTATAAAGTTCTAGACGTGGTAAAGGCTCACAAGGATTCCTGGCCCTTC
TTGGAACCTGTGGATGAATCTTATGCCCCTAACTATTATCAGATTATTAAGGCCCCCATGGATATTTCCAGCATGGAGAAGAAACTG
AATGGAGGTTTATACTGTACCAAGGAGGAATTTGTAAATGACATGAAGACCATGTTCAGGAATTGTCGAAAGTATAATGGGGAAAGT
AGTGAGTATACCAAGATGTCTGATAATTTAGAGAGGTGTTTCCATCGGGCAATGATGAAACATTTTCCTGGAGAAGATGGAGACACA
GATGAAGAATTTTGGATTCGAGAGGATGAAAAGCGGGAGAAAAGACGGAGTCGGGCTGGGCGAAGTGGTGGGAGCCATGTTTGGACC
CGCTCCAGGGACCCAGAAGGGTCCAGCAGGAAACAGCAGCCCATGGAGAATGGAGGAAAGTCGTTGCCCCCCACACGCCGAGCGCCC
TCTTCTGGGGACGATCAGAGCAGCAGCTCCACACAGCCCCCGCGGGAGGTGGGCACTTCCAATGGCCGAGGTTTTTCTCATCCCCTG
CATTGTGGTGGGACACCCAGCCAGGCACCCTTTTTAAACCAGATGAGGCCAGCAGTACCAGGAACATTTGGCCCTCTGCGAGGATCA
GATCCTGCCACCTTGTATGGCTCCTCTGGAGTCCCGGAGCCACACCCCGGGGAGCCTGTGCAGCAGCGTCAGCCTTTCACCATGCAG
CCTCCAGTTGGAATTAACAGCCTCCGAGGACCCAGGCTAGGCACACCAGAGGAGAAGCAAATGTGCGGGGGGCTGACACACCTTTCT
AACATGGGCCCACACCCTGGATCCTTGCAGCTTGGGCAGATAAGTGGCCCAAGTCAGGATGGAAGCATGTATGCTCCAGCTCAGTTC
CAGCCAGGATTCATTCCTCCCCGGCATGGGGGCGCTCCAGCCCGGCCACCAGACTTTCCTGAAAGCTCAGAAATTCCTCCCAGCCAT
ATGTATCGATCGTACAAGTACCTGAATCGAGTACACTCTGCCGTCTGGAATGGGAACCATGGTGCTACGAACCAAGGACCCTTGGGC
CCAGATGAGAAGCCCCACCTGGGGCCAGGACCCTCTCACCAGCCTCGCACTCTCGGTCACGTGATGGATTCCCGAGTCATGAGACCA
CCTGTCCCCCCCAACCAGTGGACTGAACAATCAGGCTTCCTACCTCATGGAGTTCCTTCCTCAGGGTACATGCGACCGCCCTGCAAG
TCTGCCGGACATCGGTTACAGCCACCTCCAGTGCCAGCACCCAGTTCTTTGTTTGGAGCACCTGCCCAGGCTCTTCGGGGGGTGCAG
GGAGGGGACTCCATGATGGACAGCCCAGAGATGATTGCGATGCAGCAGCTCTCCTCCCGCGTCTGCCCCCCAGGTGTGCCTTACCAC
CCCCACCAGCCTGCACACCCCCGTTTACCTGGCCCTTTTCCGCAGGTAGCTCACCCAATGTCAGTCACTGTGTCAGCCCCCAAGCCT
GCCCTGGGCAACCCTGGGAGGGCACCGGAGAACAGTGAAGCACAAGAGCCTGAGAATGACCAAGCAGAGCCGTTGCCTGGCCTTGAA
GAGAAACCACCAGGTGTTGGTACTTCAGAGGGGGTCTACCTCACACAACTACCTCACCCCACACCTCCCCTGCAGACTGACTGCACC
AGGCAGAGCTCACCACAAGAAAGGGAAACAGTGGGCCCGGAGCTCAAAAGCAGCTCCTCCGAATCTGCGGACAACTGTAAAGCAATG
AAGGGCAAGAATCCCTGGCCCTCGGATAGCAGCTACCCCGGCCCAGCCGCCCAAGGGTGCGTGAGAGACCTCTCCACGGTGGCAGAC
AGGGGCGCTCTATCCGAGAACGGAGTCATTGGGGAAGCATCTCCTTGTGGATCGGAGGGGAAGGGCCTTGGTAGCAGTGGTTCCGAA
AAGCTGCTCTGCCCCAGAGGCAGAACGTTGCAGGAAACCATGCCATGCACGGGACAGAACGCAGCGACACCGCCCAGCACAGACCCC
GGTTTGACGGGAGGCACTGTGAGCCAGTTTCCCCCGCTGTATATGCCTGGCCTAGAGTACCCGAATTCAGCTGCCCATTACCACATC
AGTCCAGGCCTGCAGGGTGTGGGCCCTGTGATGGGAGGGAAGTCCCCAGCATCCCATCCCCAGCATTTTCCCCCAAGGGGCTTTCAG
TCTAACCACCCACATTCTGGAGGCTTTCCCCGGTATCGCCCCCCACAAGGAATGAGGTATTCCTACCACCCACCGCCACAGCCTTCC
TACCACCACTATCAGCGAACTCCTTACTATGCCTGTCCACAGAGCTTTTCTGACTGGCAGAGACCTCTCCATCCCCAGGGAAGCCCA
AGCGGACCCCCAGCCAGTCAGCCTCCCCCACCAAGGTCCCTCTTCTCAGATAAGAATGCCATGGCCAGTCTGCAAGGCTGTGAGACA
CTGAATGCTGCCTTAACTTCTCCAACCCGTATGGATGCAGTGGCTGCTAAAGTCCCAAATGACGGGCAGAATCCTGGTCCAGAGGAA
GAGAAGCTGGATGAATCTATGGAGAGGCCAGAGAGTCCCAAAGAATTTTTAGACCTGGACAACCATAACGCAGCTACCAAGCGGCAG
AGCTCGTTGTCAGCCAGCGAGTATCTCTATGGAACTCCTCCGCCTCTGAGTTCAGGAATGGGATTTGGTTCATCTGCATTTCCACCC
CACAGTGTGATGCTGCAGACGGGGCCTCCCTATACCCCTCAGCGGCCGGCCAGTCACTTTCAGCCCAGGGCTTACTCTTCTCCTGTG
GCTGCCCTCCCACCTCACCACCCAGGGGCCACCCAGCCCAACGGCCTCTCTCAGGAGGGTCCCATCTATCGCTGCCAGGAAGAAGGC
CTGGGTCACTTTCAAGCTGTGATGATGGAACAAATTGGCACTAGAAGTGGAATAAGAGGACCTTTCCAGGAAATGTACAGACCATCA
GGAATGCAGATGCACCCGGTCCAGTCGCAGGCCTCGTTCCCAAAGACCCCCACAGCAGCAACATCACAGGAGGAGGTGCCGCCTCAT
AAGCCTCCAACACTTCCCCTGGATCAGAGCTAGCTCGAGATATCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCA
TTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCATTATTTG

CECR2 – R271H

CAAATAATGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATAAGCAATGCTTTTTTATAATGCCAACTTTGTAC
AAAAAAGCAGGCTTCGAAGGAGATAGAACCAATTCTCTAAGGAAATACTTAACCTGGTCGACTGGATCCGCCACC**ATGTGCCCAGAG**
GAGGGCGGCGCGGCCGGGCTGGGCGAGCTCCGCTCCTGGTGGGAGGTCCCGGCCATCGCGCACTTCTGCTCGCTCTTTCGCACCGCG
TTCCGCCTGCCCGACTTCGAGATCGAGGAGTTAGAAGCCGCTCTTCACAGAGATGACGTGGAGTTTATCAGTGACCTGATTGCCTGC
CTGCTTCAGGGCTGCTATCAACGAAGAGATATCACGCCTCAGACATTCCACAGCTACCTAGAGGACATCATCAACTACCGCTGGGAG
CTCGAAGAAGGGAAGCCCAACCCTCTGAGGGAAGCCAGTTTCCAGGACCTGCCTCTTCGCACACGGGTGGAGATCCTGCACCGACTC
TGTGATTACCGGCTGGATGCAGACGATGTCTTCGATCTTCTAAAGGGCCTGGATGCAGACAGTCTCCGTGTGGAGCCATTGGGTGAA
GACAATTCTGGGGCACTATATTGGTATTTCTATGGAACACGAATGTACAAAGAGGACCCGGTGCAAGGAAAATCCAATGGAGAACTC
TCTTTGAGCAGGGAAAGTGAAGGACAAAAAAATGTCTCAAGTATTCCTGGAAAAACGGGAAAAAGAAGAGGAAGACCCCCAAAACGG
AAGAAACTGCAGGAGGAGATTCTGTTGAGTGAAAAGCAGGAAGAAAATTCCTTGGCATCCGAGCCACAGACAAGACATGGGTCCCAA
GGGCCAGGCCAAGGTACTTGGTGGCTCCTGTGCCAGACAGAAGAGGAATGGAGACAGGTCACCGAGAGTTTTCGCGAGAGGACCTCC
CTTCGAGAACGGCAGCTCTACAAGCTCCTCAGTGAGGACTTCCTGCCTGAGATCTGCAACATGATCGCCCAGAAGGGAAAACATCCA
CAGCGCACAAAGGCAGAGTTGCATCCTAGGTGGATGTCTGACCACCTGTCCATCAAACCCGTCAAGCAAGAGGAGACTCCTGTGCTG
ACCAGAATAGAAAAACAAAAGCGCAAAGAGGAGGAAGAAGAGCGTCAGATTCTTCTAGCAGTGCAGAAGAAGGAGCAGGAGCAGATG
CTAAAGGAAGAGAGGAAACGCGAGTTGGAGGAGAAGGTCAAGGCAGTGGAAGATCGAGCGAAGAGGAGAAAGCTCAGGGAAGAAAGG
GCATGGCTGCTGGCTCAAGGAAAGGAGCTCCCTCCAGAACTTTCCCATCTGGACCCCAATTCCCCCATGAGAGAGGAAAAAAAGACT
AAAGACCTCTTTGAGTTGGATGATGATTTCACTGCTATGTATAAAGTTCTAGACGTGGTAAAGGCTCACAAGGATTCCTGGCCCTTC
TTGGAACCTGTGGATGAATCTTATGCCCCTAACTATTATCAGATTATTAAGGCCCCCATGGATATTTCCAGCATGGAGAAGAAACTG
AATGGAGGTTTATACTGTACCAAGGAGGAATTTGTAAATGACATGAAGACCATGTTCAGGAATTGTCGAAAGTATAATGGGGAAAGT
AGTGAGTATACCAAGATGTCTGATAATTTAGAGAGGTGTTTCCATCGGGCAATGATGAAACATTTTCCTGGAGAAGATGGAGACACA
GATGAAGAATTTTGGATTCGAGAGGATGAAAAGCGGGAGAAAAGACGGAGTCGGGCTGGGCGAAGTGGTGGGAGCCATGTTTGGACC
CGCTCCAGGGACCCAGAAGGGTCCAGCAGGAAACAGCAGCCCATGGAGAATGGAGGAAAGTCGTTGCCCCCCACACGCCGAGCGCCC
TCTTCTGGGGACGATCAGAGCAGCAGCTCCACACAGCCCCCGCGGGAGGTGGGCACTTCCAATGGCCGAGGTTTTTCTCATCCCCTG
CATTGTGGTGGGACACCCAGCCAGGCACCCTTTTTAAACCAGATGAGGCCAGCAGTACCAGGAACATTTGGCCCTCTGCGAGGATCA
GATCCTGCCACCTTGTATGGCTCCTCTGGAGTCCCGGAGCCACACCCCGGGGAGCCTGTGCAGCAGCGTCAGCCTTTCACCATGCAG
CCTCCAGTTGGAATTAACAGCCTCCGAGGACCCAGGCTAGGCACACCAGAGGAGAAGCAAATGTGCGGGGGGCTGACACACCTTTCT
AACATGGGCCCACACCCTGGATCCTTGCAGCTTGGGCAGATAAGTGGCCCAAGTCAGGATGGAAGCATGTATGCTCCAGCTCAGTTC
CAGCCAGGATTCATTCCTCCCCGGCATGGGGGCGCTCCAGCCCGGCCACCAGACTTTCCTGAAAGCTCAGAAATTCCTCCCAGCCAT
ATGTATCGATCGTACAAGTACCTGAATCGAGTACACTCTGCCGTCTGGAATGGGAACCATGGTGCTACGAACCAAGGACCCTTGGGC
CCAGATGAGAAGCCCCACCTGGGGCCAGGACCCTCTCACCAGCCTCGCACTCTCGGTCACGTGATGGATTCCCGAGTCATGAGACCA
CCTGTCCCCCCCAACCAGTGGACTGAACAATCAGGCTTCCTACCTCATGGAGTTCCTTCCTCAGGGTACATGCGACCGCCCTGCAAG
TCTGCCGGACATCGGTTACAGCCACCTCCAGTGCCAGCACCCAGTTCTTTGTTTGGAGCACCTGCCCAGGCTCTTCGGGGGGTGCAG
GGAGGGGACTCCATGATGGACAGCCCAGAGATGATTGCGATGCAGCAGCTCTCCTCCCGCGTCTGCCCCCCAGGTGTGCCTTACCAC
CCCCACCAGCCTGCACACCCCCGTTTACCTGGCCCTTTTCCGCAGGTAGCTCACCCAATGTCAGTCACTGTGTCAGCCCCCAAGCCT
GCCCTGGGCAACCCTGGGAGGGCACCGGAGAACAGTGAAGCACAAGAGCCTGAGAATGACCAAGCAGAGCCGTTGCCTGGCCTTGAA
GAGAAACCACCAGGTGTTGGTACTTCAGAGGGGGTCTACCTCACACAACTACCTCACCCCACACCTCCCCTGCAGACTGACTGCACC
AGGCAGAGCTCACCACAAGAAAGGGAAACAGTGGGCCCGGAGCTCAAAAGCAGCTCCTCCGAATCTGCGGACAACTGTAAAGCAATG
AAGGGCAAGAATCCCTGGCCCTCGGATAGCAGCTACCCCGGCCCAGCCGCCCAAGGGTGCGTGAGAGACCTCTCCACGGTGGCAGAC
AGGGGCGCTCTATCCGAGAACGGAGTCATTGGGGAAGCATCTCCTTGTGGATCGGAGGGGAAGGGCCTTGGTAGCAGTGGTTCCGAA
AAGCTGCTCTGCCCCAGAGGCAGAACGTTGCAGGAAACCATGCCATGCACGGGACAGAACGCAGCGACACCGCCCAGCACAGACCCC
GGTTTGACGGGAGGCACTGTGAGCCAGTTTCCCCCGCTGTATATGCCTGGCCTAGAGTACCCGAATTCAGCTGCCCATTACCACATC
AGTCCAGGCCTGCAGGGTGTGGGCCCTGTGATGGGAGGGAAGTCCCCAGCATCCCATCCCCAGCATTTTCCCCCAAGGGGCTTTCAG
TCTAACCACCCACATTCTGGAGGCTTTCCCCGGTATCGCCCCCCACAAGGAATGAGGTATTCCTACCACCCACCGCCACAGCCTTCC
TACCACCACTATCAGCGAACTCCTTACTATGCCTGTCCACAGAGCTTTTCTGACTGGCAGAGACCTCTCCATCCCCAGGGAAGCCCA
AGCGGACCCCCAGCCAGTCAGCCTCCCCCACCAAGGTCCCTCTTCTCAGATAAGAATGCCATGGCCAGTCTGCAAGGCTGTGAGACA
CTGAATGCTGCCTTAACTTCTCCAACCCGTATGGATGCAGTGGCTGCTAAAGTCCCAAATGACGGGCAGAATCCTGGTCCAGAGGAA
GAGAAGCTGGATGAATCTATGGAGAGGCCAGAGAGTCCCAAAGAATTTTTAGACCTGGACAACCATAACGCAGCTACCAAGCGGCAG
AGCTCGTTGTCAGCCAGCGAGTATCTCTATGGAACTCCTCCGCCTCTGAGTTCAGGAATGGGATTTGGTTCATCTGCATTTCCACCC
CACAGTGTGATGCTGCAGACGGGGCCTCCCTATACCCCTCAGCGGCCGGCCAGTCACTTTCAGCCCAGGGCTTACTCTTCTCCTGTG
GCTGCCCTCCCACCTCACCACCCAGGGGCCACCCAGCCCAACGGCCTCTCTCAGGAGGGTCCCATCTATCGCTGCCAGGAAGAAGGC
CTGGGTCACTTTCAAGCTGTGATGATGGAACAAATTGGCACTAGAAGTGGAATAAGAGGACCTTTCCAGGAAATGTACAGACCATCA
GGAATGCAGATGCACCCGGTCCAGTCGCAGGCCTCGTTCCCAAAGACCCCCACAGCAGCAACATCACAGGAGGAGGTGCCGCCTCAT
AAGCCTCCAACACTTCCCCTGGATCAGAGCTAGCTCGAGATATCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCA
TTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCATTATTTG

CECR2 – P632L

CAAATAATGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATAAGCAATGCTTTTTTATAATGCCAACTTTGTAC
AAAAAAGCAGGCTTCGAAGGAGATAGAACCAATTCTCTAAGGAAATACTTAACCTGGTCGACTGGATCCGCCACC**ATGTGCCCAGAG**
GAGGGCGGCGCGGCCGGGCTGGGCGAGCTCCGCTCCTGGTGGGAGGTCCCGGCCATCGCGCACTTCTGCTCGCTCTTTCGCACCGCG
TTCCGCCTGCCCGACTTCGAGATCGAGGAGTTAGAAGCCGCTCTTCACAGAGATGACGTGGAGTTTATCAGTGACCTGATTGCCTGC
CTGCTTCAGGGCTGCTATCAACGAAGAGATATCACGCCTCAGACATTCCACAGCTACCTAGAGGACATCATCAACTACCGCTGGGAG
CTCGAAGAAGGGAAGCCCAACCCTCTGAGGGAAGCCAGTTTCCAGGACCTGCCTCTTCGCACACGGGTGGAGATCCTGCACCGACTC
TGTGATTACCGGCTGGATGCAGACGATGTCTTCGATCTTCTAAAGGGCCTGGATGCAGACAGTCTCCGTGTGGAGCCATTGGGTGAA
GACAATTCTGGGGCACTATATTGGTATTTCTATGGAACACGAATGTACAAAGAGGACCCGGTGCAAGGAAAATCCAATGGAGAACTC
TCTTTGAGCAGGGAAAGTGAAGGACAAAAAAATGTCTCAAGTATTCCTGGAAAAACGGGAAAAAGAAGAGGAAGACCCCCAAAACGG
AAGAAACTGCAGGAGGAGATTCTGTTGAGTGAAAAGCAGGAAGAAAATTCCTTGGCATCCGAGCCACAGACAAGACATGGGTCCCAA
GGGCCAGGCCAAGGTACTTGGTGGCTCCTGTGCCAGACAGAAGAGGAATGGAGACAGGTCACCGAGAGTTTTCGCGAGAGGACCTCC
CTTCGAGAACGGCAGCTCTACAAGCTCCTCAGTGAGGACTTCCTGCCTGAGATCTGCAACATGATCGCCCAGAAGGGAAAACGTCCA
CAGCGCACAAAGGCAGAGTTGCATCCTAGGTGGATGTCTGACCACCTGTCCATCAAACCCGTCAAGCAAGAGGAGACTCCTGTGCTG
ACCAGAATAGAAAAACAAAAGCGCAAAGAGGAGGAAGAAGAGCGTCAGATTCTTCTAGCAGTGCAGAAGAAGGAGCAGGAGCAGATG
CTAAAGGAAGAGAGGAAACGCGAGTTGGAGGAGAAGGTCAAGGCAGTGGAAGATCGAGCGAAGAGGAGAAAGCTCAGGGAAGAAAGG
GCATGGCTGCTGGCTCAAGGAAAGGAGCTCCCTCCAGAACTTTCCCATCTGGACCCCAATTCCCCCATGAGAGAGGAAAAAAAGACT
AAAGACCTCTTTGAGTTGGATGATGATTTCACTGCTATGTATAAAGTTCTAGACGTGGTAAAGGCTCACAAGGATTCCTGGCCCTTC
TTGGAACCTGTGGATGAATCTTATGCCCCTAACTATTATCAGATTATTAAGGCCCCCATGGATATTTCCAGCATGGAGAAGAAACTG
AATGGAGGTTTATACTGTACCAAGGAGGAATTTGTAAATGACATGAAGACCATGTTCAGGAATTGTCGAAAGTATAATGGGGAAAGT
AGTGAGTATACCAAGATGTCTGATAATTTAGAGAGGTGTTTCCATCGGGCAATGATGAAACATTTTCCTGGAGAAGATGGAGACACA
GATGAAGAATTTTGGATTCGAGAGGATGAAAAGCGGGAGAAAAGACGGAGTCGGGCTGGGCGAAGTGGTGGGAGCCATGTTTGGACC
CGCTCCAGGGACCCAGAAGGGTCCAGCAGGAAACAGCAGCCCATGGAGAATGGAGGAAAGTCGTTGCCCCCCACACGCCGAGCGCCC
TCTTCTGGGGACGATCAGAGCAGCAGCTCCACACAGCCCCCGCGGGAGGTGGGCACTTCCAATGGCCGAGGTTTTTCTCATCCCCTG
CATTGTGGTGGGACACCCAGCCAGGCACCCTTTTTAAACCAGATGAGGCCAGCAGTACCAGGAACATTTGGCCCTCTGCGAGGATCA
GATCCTGCCACCTTGTATGGCTCCTCTGGAGTCCTGGAGCCACACCCCGGGGAGCCTGTGCAGCAGCGTCAGCCTTTCACCATGCAG
CCTCCAGTTGGAATTAACAGCCTCCGAGGACCCAGGCTAGGCACACCAGAGGAGAAGCAAATGTGCGGGGGGCTGACACACCTTTCT
AACATGGGCCCACACCCTGGATCCTTGCAGCTTGNGCAGATAAGTGGCCCAAGTCAGGATGGAAGCATGTATGCTCCAGCTCAGTTC
CAGCCAGGATTCATTCCTCCCCGGCATGGGGGCGCTCCAGCCCGGCCACCAGACTTTCCTGAAAGCTCAGAAATTCCTCCCAGCCAT
ATGTATCGATCGTACAAGTACCTGAATCGAGTACACTCTGCCGTCTGGAATGGGAACCATGGTGCTACGAACCAAGGACCCTTGGGC
CCAGATGAGAAGCCCCACCTGGGGCCAGGACCCTCTCACCAGCCTCGCACTCTCGGTCACGTGATGGATTCCCGAGTCATGAGACCA
CCTGTCCCCCCCAACCAGTGGACTGAACAATCAGGCTTCCTACCTCATGGAGTTCCTTCCTCAGGGTACATGCGACCGCCCTGCAAG
TCTGCCGGACATCGGTTACAGCCACCTCCAGTGCCAGCACCCAGTTCTTTGTTTGGAGCACCTGCCCAGGCTCTTCGGGGGGTGCAG
GGAGGGGACTCCATGATGGACAGCCCAGAGATGATTGCGATGCAGCAGCTCTCCTCCCGCGTCTGCCCCCCAGGTGTGCCTTACCAC
CCCCACCAGCCTGCACACCCCCGTTTACCTGGCCCTTTTCCGCAGGTAGCTCACCCAATGTCAGTCACTGTGTCAGCCCCCAAGCCT
GCCCTGGGCAACCCTGGGAGGGCACCGGAGAACAGTGAAGCACAAGAGCCTGAGAATGACCAAGCAGAGCCGTTGCCTGGCCTTGAA
GAGAAACCACCAGGTGTTGGTACTTCAGAGGGGGTCTACCTCACACAACTACCTCACCCCACACCTCCCCTGCAGACTGACTGCACC
AGGCAGAGCTCACCACAAGAAAGGGAAACAGTGGGCCCGGAGCTCAAAAGCAGCTCCTCCGAATCTGCGGACAACTGTAAAGCAATG
AAGGGCAAGAATCCCTGGCCCTCGGATAGCAGCTACCCCGGCCCAGCCGCCCAAGGGTGCGTGAGAGACCTCTCCACGGTGGCAGAC
AGGGGCGCTCTATCCGAGAACGGAGTCATTGGGGAAGCATCTCCTTGTGGATCGGAGGGGAAGGGCCTTGGTAGCAGTGGTTCCGAA
AAGCTGCTCTGCCCCAGAGGCAGAACGTTGCAGGAAACCATGCCATGCACGGGACAGAACGCAGCGACACCGCCCAGCACAGACCCC
GGTTTGACGGGAGGCACTGTGAGCCAGTTTCCCCCGCTGTATATGCCTGGCCTAGAGTACCCGAATTCAGCTGCCCATTACCACATC
AGTCCAGGCCTGCAGGGTGTGGGCCCTGTGATGGGAGGGAAGTCCCCAGCATCCCATCCCCAGCATTTTCCCCCAAGGGGCTTTCAG
TCTAACCACCCACATTCTGGAGGCTTTCCCCGGTATCGCCCCCCACAAGGAATGAGGTATTCCTACCACCCACCGCCACAGCCTTCC
TACCACCACTATCAGCGAACTCCTTACTATGCCTGTCCACAGAGCTTTTCTGACTGGCAGAGACCTCTCCATCCCCAGGGAAGCCCA
AGCGGACCCCCAGCCAGTCAGCCTCCCCCACCAAGGTCCCTCTTCTCAGATAAGAATGCCATGGCCAGTCTGCAAGGCTGTGAGACA
CTGAATGCTGCCTTAACTTCTCCAACCCGTATGGATGCAGTGGCTGCTAAAGTCCCAAATGACGGGCAGAATCCTGGTCCAGAGGAA
GAGAAGCTGGATGAATCTATGGAGAGGCCAGAGAGTCCCAAAGAATTTTTAGACCTGGACAACCATAACGCAGCTACCAAGCGGCAG
AGCTCGTTGTCAGCCAGCGAGTATCTCTATGGAACTCCTCCGCCTCTGAGTTCAGGAATGGGATTTGGTTCATCTGCATTTCCACCC
CACAGTGTGATGCTGCAGACGGGGCCTCCCTATACCCCTCAGCGGCCGGCCAGTCACTTTCAGCCCAGGGCTTACTCTTCTCCTGTG
GCTGCCCTCCCACCTCACCACCCAGGGGCCACCCAGCCCAACGGCCTCTCTCAGGAGGGTCCCATCTATCGCTGCCAGGAAGAAGGC
CTGGGTCACTTTCAAGCTGTGATGATGGAACAAATTGGCACTAGAAGTGGAATAAGAGGACCTTTCCAGGAAATGTACAGACCATCA
GGAATGCAGATGCACCCGGTCCAGTCGCAGGCCTCGTTCCCAAAGACCCCCACAGCAGCAACATCACAGGAGGAGGTGCCGCCTCAT
AAGCCTCCAACACTTCCCCTGGATCAGAGCTAGCTCGAGATATCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCA
TTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCATTATTTG

CECR2 – M680V

CAAATAATGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATAAGCAATGCTTTTTTATAATGCCAACTTTGTAC
AAAAAAGCAGGCTTCGAAGGAGATAGAACCAATTCTCTAAGGAAATACTTAACCTGGTCGACTGGATCCGCCACC**ATGTGCCCAGAG**
GAGGGCGGCGCGGCCGGGCTGGGCGAGCTCCGCTCCTGGTGGGAGGTCCCGGCCATCGCGCACTTCTGCTCGCTCTTTCGCACCGCG
TTCCGCCTGCCCGACTTCGAGATCGAGGAGTTAGAAGCCGCTCTTCACAGAGATGACGTGGAGTTTATCAGTGACCTGATTGCCTGC
CTGCTTCAGGGCTGCTATCAACGAAGAGATATCACGCCTCAGACATTCCACAGCTACCTAGAGGACATCATCAACTACCGCTGGGAG
CTCGAAGAAGGGAAGCCCAACCCTCTGAGGGAAGCCAGTTTCCAGGACCTGCCTCTTCGCACACGGGTGGAGATCCTGCACCGACTC
TGTGATTACCGGCTGGATGCAGACGATGTCTTCGATCTTCTAAAGGGCCTGGATGCAGACAGTCTCCGTGTGGAGCCATTGGGTGAA
GACAATTCTGGGGCACTATATTGGTATTTCTATGGAACACGAATGTACAAAGAGGACCCGGTGCAAGGAAAATCCAATGGAGAACTC
TCTTTGAGCAGGGAAAGTGAAGGACAAAAAAATGTCTCAAGTATTCCTGGAAAAACGGGAAAAAGAAGAGGAAGACCCCCAAAACGG
AAGAAACTGCAGGAGGAGATTCTGTTGAGTGAAAAGCAGGAAGAAAATTCCTTGGCATCCGAGCCACAGACAAGACATGGGTCCCAA
GGGCCAGGCCAAGGTACTTGGTGGCTCCTGTGCCAGACAGAAGAGGAATGGAGACAGGTCACCGAGAGTTTTCGCGAGAGGACCTCC
CTTCGAGAACGGCAGCTCTACAAGCTCCTCAGTGAGGACTTCCTGCCTGAGATCTGCAACATGATCGCCCAGAAGGGAAAACGTCCA
CAGCGCACAAAGGCAGAGTTGCATCCTAGGTGGATGTCTGACCACCTGTCCATCAAACCCGTCAAGCAAGAGGAGACTCCTGTGCTG
ACCAGAATAGAAAAACAAAAGCGCAAAGAGGAGGAAGAAGAGCGTCAGATTCTTCTAGCAGTGCAGAAGAAGGAGCAGGAGCAGATG
CTAAAGGAAGAGAGGAAACGCGAGTTGGAGGAGAAGGTCAAGGCAGTGGAAGATCGAGCGAAGAGGAGAAAGCTCAGGGAAGAAAGG
GCATGGCTGCTGGCTCAAGGAAAGGAGCTCCCTCCAGAACTTTCCCATCTGGACCCCAATTCCCCCATGAGAGAGGAAAAAAAGACT
AAAGACCTCTTTGAGTTGGATGATGATTTCACTGCTATGTATAAAGTTCTAGACGTGGTAAAGGCTCACAAGGATTCCTGGCCCTTC
TTGGAACCTGTGGATGAATCTTATGCCCCTAACTATTATCAGATTATTAAGGCCCCCATGGATATTTCCAGCATGGAGAAGAAACTG
AATGGAGGTTTATACTGTACCAAGGAGGAATTTGTAAATGACATGAAGACCATGTTCAGGAATTGTCGAAAGTATAATGGGGAAAGT
AGTGAGTATACCAAGATGTCTGATAATTTAGAGAGGTGTTTCCATCGGGCAATGATGAAACATTTTCCTGGAGAAGATGGAGACACA
GATGAAGAATTTTGGATTCGAGAGGATGAAAAGCGGGAGAAAAGACGGAGTCGGGCTGGGCGAAGTGGTGGGAGCCATGTTTGGACC
CGCTCCAGGGACCCAGAAGGGTCCAGCAGGAAACAGCAGCCCATGGAGAATGGAGGAAAGTCGTTGCCCCCCACACGCCGAGCGCCC
TCTTCTGGGGACGATCAGAGCAGCAGCTCCACACAGCCCCCGCGGGAGGTGGGCACTTCCAATGGCCGAGGTTTTTCTCATCCCCTG
CATTGTGGTGGGACACCCAGCCAGGCACCCTTTTTAAACCAGATGAGGCCAGCAGTACCAGGAACATTTGGCCCTCTGCGAGGATCA
GATCCTGCCACCTTGTATGGCTCCTCTGGAGTCCCGGAGCCACACCCCGGGGAGCCTGTGCAGCAGCGTCAGCCTTTCACCATGCAG
CCTCCAGTTGGAATTAACAGCCTCCGAGGACCCAGGCTAGGCACACCAGAGGAGAAGCAAATGTGCGGGGGGCTGACACACCTTTCT
AACGTGGGCCCACACCCTGGATCCTTGCAGCTTGGGCAGATAAGTGGCCCAAGTCAGGATGGAAGCATGTATGCTCCAGCTCAGTTC
CAGCCAGGATTCATTCCTCCCCGGCATGGGGGCGCTCCAGCCCGGCCACCAGACTTTCCTGAAAGCTCAGAAATTCCTCCCAGCCAT
ATGTATCGATCGTACAAGTACCTGAATCGAGTACACTCTGCCGTCTGGAATGGGAACCATGGTGCTACGAACCAAGGACCCTTGGGC
CCAGATGAGAAGCCCCACCTGGGGCCAGGACCCTCTCACCAGCCTCGCACTCTCGGTCACGTGATGGATTCCCGAGTCATGAGACCA
CCTGTCCCCCCCAACCAGTGGACTGAACAATCAGGCTTCCTACCTCATGGAGTTCCTTCCTCAGGGTACATGCGACCGCCCTGCAAG
TCTGCCGGACATCGGTTACAGCCACCTCCAGTGCCAGCACCCAGTTCTTTGTTTGGAGCACCTGCCCAGGCTCTTCGGGGGGTGCAG
GGAGGGGACTCCATGATGGACAGCCCAGAGATGATTGCGATGCAGCAGCTCTCCTCCCGCGTCTGCCCCCCAGGTGTGCCTTACCAC
CCCCACCAGCCTGCACACCCCCGTTTACCTGGCCCTTTTCCGCAGGTAGCTCACCCAATGTCAGTCACTGTGTCAGCCCCCAAGCCT
GCCCTGGGCAACCCTGGGAGGGCACCGGAGAACAGTGAAGCACAAGAGCCTGAGAATGACCAAGCAGAGCCGTTGCCTGGCCTTGAA
GAGAAACCACCAGGTGTTGGTACTTCAGAGGGGGTCTACCTCACACAACTACCTCACCCCACACCTCCCCTGCAGACTGACTGCACC
AGGCAGAGCTCACCACAAGAAAGGGAAACAGTGGGCCCGGAGCTCAAAAGCAGCTCCTCCGAATCTGCGGACAACTGTAAAGCAATG
AAGGGCAAGAATCCCTGGCCCTCGGATAGCAGCTACCCCGGCCCAGCCGCCCAAGGGTGCGTGAGAGACCTCTCCACGGTGGCAGAC
AGGGGCGCTCTATCCGAGAACGGAGTCATTGGGGAAGCATCTCCTTGTGGATCGGAGGGGAAGGGCCTTGGTAGCAGTGGTTCCGAA
AAGCTGCTCTGCCCCAGAGGCAGAACGTTGCAGGAAACCATGCCATGCACGGGACAGAACGCAGCGACACCGCCCAGCACAGACCCC
GGTTTGACGGGAGGCACTGTGAGCCAGTTTCCCCCGCTGTATATGCCTGGCCTAGAGTACCCGAATTCAGCTGCCCATTACCACATC
AGTCCAGGCCTGCAGGGTGTGGGCCCTGTGATGGGAGGGAAGTCCCCAGCATCCCATCCCCAGCATTTTCCCCCAAGGGGCTTTCAG
TCTAACCACCCACATTCTGGAGGCTTTCCCCGGTATCGCCCCCCACAAGGAATGAGGTATTCCTACCACCCACCGCCACAGCCTTCC
TACCACCACTATCAGCGAACTCCTTACTATGCCTGTCCACAGAGCTTTTCTGACTGGCAGAGACCTCTCCATCCCCAGGGAAGCCCA
AGCGGACCCCCAGCCAGTCAGCCTCCCCCACCAAGGTCCCTCTTCTCAGATAAGAATGCCATGGCCAGTCTGCAAGGCTGTGAGACA
CTGAATGCTGCCTTAACTTCTCCAACCCGTATGGATGCAGTGGCTGCTAAAGTCCCAAATGACGGGCAGAATCCTGGTCCAGAGGAA
GAGAAGCTGGATGAATCTATGGAGAGGCCAGAGAGTCCCAAAGAATTTTTAGACCTGGACAACCATAACGCAGCTACCAAGCGGCAG
AGCTCGTTGTCAGCCAGCGAGTATCTCTATGGAACTCCTCCGCCTCTGAGTTCAGGAATGGGATTTGGTTCATCTGCATTTCCACCC
CACAGTGTGATGCTGCAGACGGGGCCTCCCTATACCCCTCAGCGGCCGGCCAGTCACTTTCAGCCCAGGGCTTACTCTTCTCCTGTG
GCTGCCCTCCCACCTCACCACCCAGGGGCCACCCAGCCCAACGGCCTCTCTCAGGAGGGTCCCATCTATCGCTGCCAGGAAGAAGGC
CTGGGTCACTTTCAAGCTGTGATGATGGAACAAATTGGCACTAGAAGTGGAATAAGAGGACCTTTCCAGGAAATGTACAGACCATCA
GGAATGCAGATGCACCCGGTCCAGTCGCAGGCCTCGTTCCCAAAGACCCCCACAGCAGCAACATCACAGGAGGAGGTGCCGCCTCAT
AAGCCTCCAACACTTCCCCTGGATCAGAGCTAGCTCGAGATATCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCA
TTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCATTATTTG

CECR2 – Y738C

CAAATAATGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATAAGCAATGCTTTTTTATAATGCCAACTTTGTAC
AAAAAAGCAGGCTTCGAAGGAGATAGAACCAATTCTCTAAGGAAATACTTAACCTGGTCGACTGGATCCGCCACC**ATGTGCCCAGAG**
GAGGGCGGCGCGGCCGGGCTGGGCGAGCTCCGCTCCTGGTGGGAGGTCCCGGCCATCGCGCACTTCTGCTCGCTCTTTCGCACCGCG
TTCCGCCTGCCCGACTTCGAGATCGAGGAGTTAGAAGCCGCTCTTCACAGAGATGACGTGGAGTTTATCAGTGACCTGATTGCCTGC
CTGCTTCAGGGCTGCTATCAACGAAGAGATATCACGCCTCAGACATTCCACAGCTACCTAGAGGACATCATCAACTACCGCTGGGAG
CTCGAAGAAGGGAAGCCCAACCCTCTGAGGGAAGCCAGTTTCCAGGACCTGCCTCTTCGCACACGGGTGGAGATCCTGCACCGACTC
TGTGATTACCGGCTGGATGCAGACGATGTCTTCGATCTTCTAAAGGGCCTGGATGCAGACAGTCTCCGTGTGGAGCCATTGGGTGAA
GACAATTCTGGGGCACTATATTGGTATTTCTATGGAACACGAATGTACAAAGAGGACCCGGTGCAAGGAAAATCCAATGGAGAACTC
TCTTTGAGCAGGGAAAGTGAAGGACAAAAAAATGTCTCAAGTATTCCTGGAAAAACGGGAAAAAGAAGAGGAAGACCCCCAAAACGG
AAGAAACTGCAGGAGGAGATTCTGTTGAGTGAAAAGCAGGAAGAAAATTCCTTGGCATCCGAGCCACAGACAAGACATGGGTCCCAA
GGGCCAGGCCAAGGTACTTGGTGGCTCCTGTGCCAGACAGAAGAGGAATGGAGACAGGTCACCGAGAGTTTTCGCGAGAGGACCTCC
CTTCGAGAACGGCAGCTCTACAAGCTCCTCAGTGAGGACTTCCTGCCTGAGATCTGCAACATGATCGCCCAGAAGGGAAAACGTCCA
CAGCGCACAAAGGCAGAGTTGCATCCTAGGTGGATGTCTGACCACCTGTCCATCAAACCCGTCAAGCAAGAGGAGACTCCTGTGCTG
ACCAGAATAGAAAAACAAAAGCGCAAAGAGGAGGAAGAAGAGCGTCAGATTCTTCTAGCAGTGCAGAAGAAGGAGCAGGAGCAGATG
CTAAAGGAAGAGAGGAAACGCGAGTTGGAGGAGAAGGTCAAGGCAGTGGAAGATCGAGCGAAGAGGAGAAAGCTCAGGGAAGAAAGG
GCATGGCTGCTGGCTCAAGGAAAGGAGCTCCCTCCAGAACTTTCCCATCTGGACCCCAATTCCCCCATGAGAGAGGAAAAAAAGACT
AAAGACCTCTTTGAGTTGGATGATGATTTCACTGCTATGTATAAAGTTCTAGACGTGGTAAAGGCTCACAAGGATTCCTGGCCCTTC
TTGGAACCTGTGGATGAATCTTATGCCCCTAACTATTATCAGATTATTAAGGCCCCCATGGATATTTCCAGCATGGAGAAGAAACTG
AATGGAGGTTTATACTGTACCAAGGAGGAATTTGTAAATGACATGAAGACCATGTTCAGGAATTGTCGAAAGTATAATGGGGAAAGT
AGTGAGTATACCAAGATGTCTGATAATTTAGAGAGGTGTTTCCATCGGGCAATGATGAAACATTTTCCTGGAGAAGATGGAGACACA
GATGAAGAATTTTGGATTCGAGAGGATGAAAAGCGGGAGAAAAGACGGAGTCGGGCTGGGCGAAGTGGTGGGAGCCATGTTTGGACC
CGCTCCAGGGACCCAGAAGGGTCCAGCAGGAAACAGCAGCCCATGGAGAATGGAGGAAAGTCGTTGCCCCCCACACGCCGAGCGCCC
TCTTCTGGGGACGATCAGAGCAGCAGCTCCACACAGCCCCCGCGGGAGGTGGGCACTTCCAATGGCCGAGGTTTTTCTCATCCCCTG
CATTGTGGTGGGACACCCAGCCAGGCACCCTTTTTAAACCAGATGAGGCCAGCAGTACCAGGAACATTTGGCCCTCTGCGAGGATCA
GATCCTGCCACCTTGTATGGCTCCTCTGGAGTCCCGGAGCCACACCCCGGGGAGCCTGTGCAGCAGCGTCAGCCTTTCACCATGCAG
CCTCCAGTTGGAATTAACAGCCTCCGAGGACCCAGGCTAGGCACACCAGAGGAGAAGCAAATGTGCGGGGGGCTGACACACCTTTCT
AACATGGGCCCACACCCTGGATCCTTGCAGCTTGGGCAGATAAGTGGCCCAAGTCAGGATGGAAGCATGTATGCTCCAGCTCAGTTC
CAGCCAGGATTCATTCCTCCCCGGCATGGGGGCGCTCCAGCCCGGCCACCAGACTTTCCTGAAAGCTCAGAAATTCCTCCCAGCCAT
ATGTGTCGATCGTACAAGTACCTGAATCGAGTACACTCTGCCGTCTGGAATGGGAACCATGGTGCTACGAACCAAGGACCCTTGGGC
CCAGATGAGAAGCCCCACCTGGGGCCAGGACCCTCTCACCAGCCTCGCACTCTCGGTCACGTGATGGATTCCCGAGTCATGAGACCA
CCTGTCCCCCCCAACCAGTGGACTGAACAATCAGGCTTCCTACCTCATGGAGTTCCTTCCTCAGGGTACATGCGACCGCCCTGCAAG
TCTGCCGGACATCGGTTACAGCCACCTCCAGTGCCAGCACCCAGTTCTTTGTTTGGAGCACCTGCCCAGGCTCTTCGGGGGGTGCAG
GGAGGGGACTCCATGATGGACAGCCCAGAGATGATTGCGATGCAGCAGCTCTCCTCCCGCGTCTGCCCCCCAGGTGTGCCTTACCAC
CCCCACCAGCCTGCACACCCCCGTTTACCTGGCCCTTTTCCGCAGGTAGCTCACCCAATGTCAGTCACTGTGTCAGCCCCCAAGCCT
GCCCTGGGCAACCCTGGGAGGGCACCGGAGAACAGTGAAGCACAAGAGCCTGAGAATGACCAAGCAGAGCCGTTGCCTGGCCTTGAA
GAGAAACCACCAGGTGTTGGTACTTCAGAGGGGGTCTACCTCACACAACTACCTCACCCCACACCTCCCCTGCAGACTGACTGCACC
AGGCAGAGCTCACCACAAGAAAGGGAAACAGTGGGCCCGGAGCTCAAAAGCAGCTCCTCCGAATCTGCGGACAACTGTAAAGCAATG
AAGGGCAAGAATCCCTGGCCCTCGGATAGCAGCTACCCCGGCCCAGCCGCCCAAGGGTGCGTGAGAGACCTCTCCACGGTGGCAGAC
AGGGGCGCTCTATCCGAGAACGGAGTCATTGGGGAAGCATCTCCTTGTGGATCGGAGGGGAAGGGCCTTGGTAGCAGTGGTTCCGAA
AAGCTGCTCTGCCCCAGAGGCAGAACGTTGCAGGAAACCATGCCATGCACGGGACAGAACGCAGCGACACCGCCCAGCACAGACCCC
GGTTTGACGGGAGGCACTGTGAGCCAGTTTCCCCCGCTGTATATGCCTGGCCTAGAGTACCCGAATTCAGCTGCCCATTACCACATC
AGTCCAGGCCTGCAGGGTGTGGGCCCTGTGATGGGAGGGAAGTCCCCAGCATCCCATCCCCAGCATTTTCCCCCAAGGGGCTTTCAG
TCTAACCACCCACATTCTGGAGGCTTTCCCCGGTATCGCCCCCCACAAGGAATGAGGTATTCCTACCACCCACCGCCACAGCCTTCC
TACCACCACTATCAGCGAACTCCTTACTATGCCTGTCCACAGAGCTTTTCTGACTGGCAGAGACCTCTCCATCCCCAGGGAAGCCCA
AGCGGACCCCCAGCCAGTCAGCCTCCCCCACCAAGGTCCCTCTTCTCAGATAAGAATGCCATGGCCAGTCTGCAAGGCTGTGAGACA
CTGAATGCTGCCTTAACTTCTCCAACCCGTATGGATGCAGTGGCTGCTAAAGTCCCAAATGACGGGCAGAATCCTGGTCCAGAGGAA
GAGAAGCTGGATGAATCTATGGAGAGGCCAGAGAGTCCCAAAGAATTTTTAGACCTGGACAACCATAACGCAGCTACCAAGCGGCAG
AGCTCGTTGTCAGCCAGCGAGTATCTCTATGGAACTCCTCCGCCTCTGAGTTCAGGAATGGGATTTGGTTCATCTGCATTTCCACCC
CACAGTGTGATGCTGCAGACGGGGCCTCCCTATACCCCTCAGCGGCCGGCCAGTCACTTTCAGCCCAGGGCTTACTCTTCTCCTGTG
GCTGCCCTCCCACCTCACCACCCAGGGGCCACCCAGCCCAACGGCCTCTCTCAGGAGGGTCCCATCTATCGCTGCCAGGAAGAAGGC
CTGGGTCACTTTCAAGCTGTGATGATGGAACAAATTGGCACTAGAAGTGGAATAAGAGGACCTTTCCAGGAAATGTACAGACCATCA
GGAATGCAGATGCACCCGGTCCAGTCGCAGGCCTCGTTCCCAAAGACCCCCACAGCAGCAACATCACAGGAGGAGGTGCCGCCTCAT
AAGCCTCCAACACTTCCCCTGGATCAGAGCTAGCTCGAGATATCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCA
TTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCATTATTTG

CECR2 – P1210R

CAAATAATGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATAAGCAATGCTTTTTTATAATGCCAACTTTGTAC
AAAAAAGCAGGCTTCGAAGGAGATAGAACCAATTCTCTAAGGAAATACTTAACCTGGTCGACTGGATCCGCCACC**ATGTGCCCAGAG**
GAGGGCGGCGCGGCCGGGCTGGGCGAGCTCCGCTCCTGGTGGGAGGTCCCGGCCATCGCGCACTTCTGCTCGCTCTTTCGCACCGCG
TTCCGCCTGCCCGACTTCGAGATCGAGGAGTTAGAAGCCGCTCTTCACAGAGATGACGTGGAGTTTATCAGTGACCTGATTGCCTGC
CTGCTTCAGGGCTGCTATCAACGAAGAGATATCACGCCTCAGACATTCCACAGCTACCTAGAGGACATCATCAACTACCGCTGGGAG
CTCGAAGAAGGGAAGCCCAACCCTCTGAGGGAAGCCAGTTTCCAGGACCTGCCTCTTCGCACACGGGTGGAGATCCTGCACCGACTC
TGTGATTACCGGCTGGATGCAGACGATGTCTTCGATCTTCTAAAGGGCCTGGATGCAGACAGTCTCCGTGTGGAGCCATTGGGTGAA
GACAATTCTGGGGCACTATATTGGTATTTCTATGGAACACGAATGTACAAAGAGGACCCGGTGCAAGGAAAATCCAATGGAGAACTC
TCTTTGAGCAGGGAAAGTGAAGGACAAAAAAATGTCTCAAGTATTCCTGGAAAAACGGGAAAAAGAAGAGGAAGACCCCCAAAACGG
AAGAAACTGCAGGAGGAGATTCTGTTGAGTGAAAAGCAGGAAGAAAATTCCTTGGCATCCGAGCCACAGACAAGACATGGGTCCCAA
GGGCCAGGCCAAGGTACTTGGTGGCTCCTGTGCCAGACAGAAGAGGAATGGAGACAGGTCACCGAGAGTTTTCGCGAGAGGACCTCC
CTTCGAGAACGGCAGCTCTACAAGCTCCTCAGTGAGGACTTCCTGCCTGAGATCTGCAACATGATCGCCCAGAAGGGAAAACGTCCA
CAGCGCACAAAGGCAGAGTTGCATCCTAGGTGGATGTCTGACCACCTGTCCATCAAACCCGTCAAGCAAGAGGAGACTCCTGTGCTG
ACCAGAATAGAAAAACAAAAGCGCAAAGAGGAGGAAGAAGAGCGTCAGATTCTTCTAGCAGTGCAGAAGAAGGAGCAGGAGCAGATG
CTAAAGGAAGAGAGGAAACGCGAGTTGGAGGAGAAGGTCAAGGCAGTGGAAGATCGAGCGAAGAGGAGAAAGCTCAGGGAAGAAAGG
GCATGGCTGCTGGCTCAAGGAAAGGAGCTCCCTCCAGAACTTTCCCATCTGGACCCCAATTCCCCCATGAGAGAGGAAAAAAAGACT
AAAGACCTCTTTGAGTTGGATGATGATTTCACTGCTATGTATAAAGTTCTAGACGTGGTAAAGGCTCACAAGGATTCCTGGCCCTTC
TTGGAACCTGTGGATGAATCTTATGCCCCTAACTATTATCAGATTATTAAGGCCCCCATGGATATTTCCAGCATGGAGAAGAAACTG
AATGGAGGTTTATACTGTACCAAGGAGGAATTTGTAAATGACATGAAGACCATGTTCAGGAATTGTCGAAAGTATAATGGGGAAAGT
AGTGAGTATACCAAGATGTCTGATAATTTAGAGAGGTGTTTCCATCGGGCAATGATGAAACATTTTCCTGGAGAAGATGGAGACACA
GATGAAGAATTTTGGATTCGAGAGGATGAAAAGCGGGAGAAAAGACGGAGTCGGGCTGGGCGAAGTGGTGGGAGCCATGTTTGGACC
CGCTCCAGGGACCCAGAAGGGTCCAGCAGGAAACAGCAGCCCATGGAGAATGGAGGAAAGTCGTTGCCCCCCACACGCCGAGCGCCC
TCTTCTGGGGACGATCAGAGCAGCAGCTCCACACAGCCCCCGCGGGAGGTGGGCACTTCCAATGGCCGAGGTTTTTCTCATCCCCTG
CATTGTGGTGGGACACCCAGCCAGGCACCCTTTTTAAACCAGATGAGGCCAGCAGTACCAGGAACATTTGGCCCTCTGCGAGGATCA
GATCCTGCCACCTTGTATGGCTCCTCTGGAGTCCCGGAGCCACACCCCGGGGAGCCTGTGCAGCAGCGTCAGCCTTTCACCATGCAG
CCTCCAGTTGGAATTAACAGCCTCCGAGGACCCAGGCTAGGCACACCAGAGGAGAAGCAAATGTGCGGGGGGCTGACACACCTTTCT
AACATGGGCCCACACCCTGGATCCTTGCAGCTTGGGCAGATAAGTGGCCCAAGTCAGGATGGAAGCATGTATGCTCCAGCTCAGTTC
CAGCCAGGATTCATTCCTCCCCGGCATGGGGGCGCTCCAGCCCGGCCACCAGACTTTCCTGAAAGCTCAGAAATTCCTCCCAGCCAT
ATGTATCGATCGTACAAGTACCTGAATCGAGTACACTCTGCCGTCTGGAATGGGAACCATGGTGCTACGAACCAAGGACCCTTGGGC
CCAGATGAGAAGCCCCACCTGGGGCCAGGACCCTCTCACCAGCCTCGCACTCTCGGTCACGTGATGGATTCCCGAGTCATGAGACCA
CCTGTCCCCCCCAACCAGTGGACTGAACAATCAGGCTTCCTACCTCATGGAGTTCCTTCCTCAGGGTACATGCGACCGCCCTGCAAG
TCTGCCGGACATCGGTTACAGCCACCTCCAGTGCCAGCACCCAGTTCTTTGTTTGGAGCACCTGCCCAGGCTCTTCGGGGGGTGCAG
GGAGGGGACTCCATGATGGACAGCCCAGAGATGATTGCGATGCAGCAGCTCTCCTCCCGCGTCTGCCCCCCAGGTGTGCCTTACCAC
CCCCACCAGCCTGCACACCCCCGTTTACCTGGCCCTTTTCCGCAGGTAGCTCACCCAATGTCAGTCACTGTGTCAGCCCCCAAGCCT
GCCCTGGGCAACCCTGGGAGGGCACCGGAGAACAGTGAAGCACAAGAGCCTGAGAATGACCAAGCAGAGCCGTTGCCTGGCCTTGAA
GAGAAACCACCAGGTGTTGGTACTTCAGAGGGGGTCTACCTCACACAACTACCTCACCCCACACCTCCCCTGCAGACTGACTGCACC
AGGCAGAGCTCACCACAAGAAAGGGAAACAGTGGGCCCGGAGCTCAAAAGCAGCTCCTCCGAATCTGCGGACAACTGTAAAGCAATG
AAGGGCAAGAATCCCTGGCCCTCGGATAGCAGCTACCCCGGCCCAGCCGCCCAAGGGTGCGTGAGAGACCTCTCCACGGTGGCAGAC
AGGGGCGCTCTATCCGAGAACGGAGTCATTGGGGAAGCATCTCCTTGTGGATCGGAGGGGAAGGGCCTTGGTAGCAGTGGTTCCGAA
AAGCTGCTCTGCCCCAGAGGCAGAACGTTGCAGGAAACCATGCCATGCACGGGACAGAACGCAGCGACACCGCCCAGCACAGACCCC
GGTTTGACGGGAGGCACTGTGAGCCAGTTTCCCCCGCTGTATATGCCTGGCCTAGAGTACCCGAATTCAGCTGCCCATTACCACATC
AGTCCAGGCCTGCAGGGTGTGGGCCCTGTGATGGGAGGGAAGTCCCCAGCATCCCATCCCCAGCATTTTCCCCCAAGGGGCTTTCAG
TCTAACCACCCACATTCTGGAGGCTTTCCCCGGTATCGCCCCCCACAAGGAATGAGGTATTCCTACCACCCACCGCCACAGCCTTCC
TACCACCACTATCAGCGAACTCCTTACTATGCCTGTCCACAGAGCTTTTCTGACTGGCAGAGACCTCTCCATCCCCAGGGAAGCCCA
AGCGGACCCCCAGCCAGTCAGCCTCCCCGACCAAGGTCCCTCTTCTCAGATAAGAATGCCATGGCCAGTCTGCAAGGCTGTGAGACA
CTGAATGCTGCCTTAACTTCTCCAACCCGTATGGATGCAGTGGCTGCTAAAGTCCCAAATGACGGGCAGAATCCTGGTCCAGAGGAA
GAGAAGCTGGATGAATCTATGGAGAGGCCAGAGAGTCCCAAAGAATTTTTAGACCTGGACAACCATAACGCAGCTACCAAGCGGCAG
AGCTCGTTGTCAGCCAGCGAGTATCTCTATGGAACTCCTCCGCCTCTGAGTTCAGGAATGGGATTTGGTTCATCTGCATTTCCACCC
CACAGTGTGATGCTGCAGACGGGGCCTCCCTATACCCCTCAGCGGCCGGCCAGTCACTTTCAGCCCAGGGCTTACTCTTCTCCTGTG
GCTGCCCTCCCACCTCACCACCCAGGGGCCACCCAGCCCAACGGCCTCTCTCAGGAGGGTCCCATCTATCGCTGCCAGGAAGAAGGC
CTGGGTCACTTTCAAGCTGTGATGATGGAACAAATTGGCACTAGAAGTGGAATAAGAGGACCTTTCCAGGAAATGTACAGACCATCA
GGAATGCAGATGCACCCGGTCCAGTCGCAGGCCTCGTTCCCAAAGACCCCCACAGCAGCAACATCACAGGAGGAGGTGCCGCCTCAT
AAGCCTCCAACACTTCCCCTGGATCAGAGCTAGCTCGAGATATCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCA
TTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCATTATTTG

CECR2 - S1261L

CAAATAATGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATAAGCAATGCTTTTTTATAATGCCAACTTTGTAC
AAAAAAGCAGGCTTCGAAGGAGATAGAACCAATTCTCTAAGGAAATACTTAACCTGGTCGACTGGATCCGCCACC**ATGTGCCCAGAG**
GAGGGCGGCGCGGCCGGGCTGGGCGAGCTCCGCTCCTGGTGGGAGGTCCCGGCCATCGCGCACTTCTGCTCGCTCTTTCGCACCGCG
TTCCGCCTGCCCGACTTCGAGATCGAGGAGTTAGAAGCCGCTCTTCACAGAGATGACGTGGAGTTTATCAGTGACCTGATTGCCTGC
CTGCTTCAGGGCTGCTATCAACGAAGAGATATCACGCCTCAGACATTCCACAGCTACCTAGAGGACATCATCAACTACCGCTGGGAG
CTCGAAGAAGGGAAGCCCAACCCTCTGAGGGAAGCCAGTTTCCAGGACCTGCCTCTTCGCACACGGGTGGAGATCCTGCACCGACTC
TGTGATTACCGGCTGGATGCAGACGATGTCTTCGATCTTCTAAAGGGCCTGGATGCAGACAGTCTCCGTGTGGAGCCATTGGGTGAA
GACAATTCTGGGGCACTATATTGGTATTTCTATGGAACACGAATGTACAAAGAGGACCCGGTGCAAGGAAAATCCAATGGAGAACTC
TCTTTGAGCAGGGAAAGTGAAGGACAAAAAAATGTCTCAAGTATTCCTGGAAAAACGGGAAAAAGAAGAGGAAGACCCCCAAAACGG
AAGAAACTGCAGGAGGAGATTCTGTTGAGTGAAAAGCAGGAAGAAAATTCCTTGGCATCCGAGCCACAGACAAGACATGGGTCCCAA
GGGCCAGGCCAAGGTACTTGGTGGCTCCTGTGCCAGACAGAAGAGGAATGGAGACAGGTCACCGAGAGTTTTCGCGAGAGGACCTCC
CTTCGAGAACGGCAGCTCTACAAGCTCCTCAGTGAGGACTTCCTGCCTGAGATCTGCAACATGATCGCCCAGAAGGGAAAACGTCCA
CAGCGCACAAAGGCAGAGTTGCATCCTAGGTGGATGTCTGACCACCTGTCCATCAAACCCGTCAAGCAAGAGGAGACTCCTGTGCTG
ACCAGAATAGAAAAACAAAAGCGCAAAGAGGAGGAAGAAGAGCGTCAGATTCTTCTAGCAGTGCAGAAGAAGGAGCAGGAGCAGATG
CTAAAGGAAGAGAGGAAACGCGAGTTGGAGGAGAAGGTCAAGGCAGTGGAAGATCGAGCGAAGAGGAGAAAGCTCAGGGAAGAAAGG
GCATGGCTGCTGGCTCAAGGAAAGGAGCTCCCTCCAGAACTTTCCCATCTGGACCCCAATTCCCCCATGAGAGAGGAAAAAAAGACT
AAAGACCTCTTTGAGTTGGATGATGATTTCACTGCTATGTATAAAGTTCTAGACGTGGTAAAGGCTCACAAGGATTCCTGGCCCTTC
TTGGAACCTGTGGATGAATCTTATGCCCCTAACTATTATCAGATTATTAAGGCCCCCATGGATATTTCCAGCATGGAGAAGAAACTG
AATGGAGGTTTATACTGTACCAAGGAGGAATTTGTAAATGACATGAAGACCATGTTCAGGAATTGTCGAAAGTATAATGGGGAAAGT
AGTGAGTATACCAAGATGTCTGATAATTTAGAGAGGTGTTTCCATCGGGCAATGATGAAACATTTTCCTGGAGAAGATGGAGACACA
GATGAAGAATTTTGGATTCGAGAGGATGAAAAGCGGGAGAAAAGACGGAGTCGGGCTGGGCGAAGTGGTGGGAGCCATGTTTGGACC
CGCTCCAGGGACCCAGAAGGGTCCAGCAGGAAACAGCAGCCCATGGAGAATGGAGGAAAGTCGTTGCCCCCCACACGCCGAGCGCCC
TCTTCTGGGGACGATCAGAGCAGCAGCTCCACACAGCCCCCGCGGGAGGTGGGCACTTCCAATGGCCGAGGTTTTTCTCATCCCCTG
CATTGTGGTGGGACACCCAGCCAGGCACCCTTTTTAAACCAGATGAGGCCAGCAGTACCAGGAACATTTGGCCCTCTGCGAGGATCA
GATCCTGCCACCTTGTATGGCTCCTCTGGAGTCCCGGAGCCACACCCCGGGGAGCCTGTGCAGCAGCGTCAGCCTTTCACCATGCAG
CCTCCAGTTGGAATTAACAGCCTCCGAGGACCCAGGCTAGGCACACCAGAGGAGAAGCAAATGTGCGGGGGGCTGACACACCTTTCT
AACATGGGCCCACACCCTGGATCCTTGCAGCTTGGGCAGATAAGTGGCCCAAGTCAGGATGGAAGCATGTATGCTCCAGCTCAGTTC
CAGCCAGGATTCATTCCTCCCCGGCATGGGGGCGCTCCAGCCCGGCCACCAGACTTTCCTGAAAGCTCAGAAATTCCTCCCAGCCAT
ATGTATCGATCGTACAAGTACCTGAATCGAGTACACTCTGCCGTCTGGAATGGGAACCATGGTGCTACGAACCAAGGACCCTTGGGC
CCAGATGAGAAGCCCCACCTGGGGCCAGGACCCTCTCACCAGCCTCGCACTCTCGGTCACGTGATGGATTCCCGAGTCATGAGACCA
CCTGTCCCCCCCAACCAGTGGACTGAACAATCAGGCTTCCTACCTCATGGAGTTCCTTCCTCAGGGTACATGCGACCGCCCTGCAAG
TCTGCCGGACATCGGTTACAGCCACCTCCAGTGCCAGCACCCAGTTCTTTGTTTGGAGCACCTGCCCAGGCTCTTCGGGGGGTGCAG
GGAGGGGACTCCATGATGGACAGCCCAGAGATGATTGCGATGCAGCAGCTCTCCTCCCGCGTCTGCCCCCCAGGTGTGCCTTACCAC
CCCCACCAGCCTGCACACCCCCGTTTACCTGGCCCTTTTCCGCAGGTAGCTCACCCAATGTCAGTCACTGTGTCAGCCCCCAAGCCT
GCCCTGGGCAACCCTGGGAGGGCACCGGAGAACAGTGAAGCACAAGAGCCTGAGAATGACCAAGCAGAGCCGTTGCCTGGCCTTGAA
GAGAAACCACCAGGTGTTGGTACTTCAGAGGGGGTCTACCTCACACAACTACCTCACCCCACACCTCCCCTGCAGACTGACTGCACC
AGGCAGAGCTCACCACAAGAAAGGGAAACAGTGGGCCCGGAGCTCAAAAGCAGCTCCTCCGAATCTGCGGACAACTGTAAAGCAATG
AAGGGCAAGAATCCCTGGCCCTCGGATAGCAGCTACCCCGGCCCAGCCGCCCAAGGGTGCGTGAGAGACCTCTCCACGGTGGCAGAC
AGGGGCGCTCTATCCGAGAACGGAGTCATTGGGGAAGCATCTCCTTGTGGATCGGAGGGGAAGGGCCTTGGTAGCAGTGGTTCCGAA
AAGCTGCTCTGCCCCAGAGGCAGAACGTTGCAGGAAACCATGCCATGCACGGGACAGAACGCAGCGACACCGCCCAGCACAGACCCC
GGTTTGACGGGAGGCACTGTGAGCCAGTTTCCCCCGCTGTATATGCCTGGCCTAGAGTACCCGAATTCAGCTGCCCATTACCACATC
AGTCCAGGCCTGCAGGGTGTGGGCCCTGTGATGGGAGGGAAGTCCCCAGCATCCCATCCCCAGCATTTTCCCCCAAGGGGCTTTCAG
TCTAACCACCCACATTCTGGAGGCTTTCCCCGGTATCGCCCCCCACAAGGAATGAGGTATTCCTACCACCCACCGCCACAGCCTTCC
TACCACCACTATCAGCGAACTCCTTACTATGCCTGTCCACAGAGCTTTTCTGACTGGCAGAGACCTCTCCATCCCCAGGGAAGCCCA
AGCGGACCCCCAGCCAGTCAGCCTCCCCCACCAAGGTCCCTCTTCTCAGATAAGAATGCCATGGCCAGTCTGCAAGGCTGTGAGACA
CTGAATGCTGCCTTAACTTCTCCAACCCGTATGGATGCAGTGGCTGCTAAAGTCCCAAATGACGGGCAGAATCCTGGTCCAGAGGAA
GAGAAGCTGGATGAATCTATGGAGAGGCCAGAGAGTCCCAAAGAATTTTTAGACCTGGACAACCATAACGCAGCTACCAAGCGGCAG
AGCTCGTTGTCAGCCAGCGAGTATCTCTATGGAACTCCTCCGCCTCTGAGTTCAGGAATGGGATTTGGTTCATCTGCATTTCCACCC
CACAGTGTGATGCTGCAGACGGGGCCTCCCTATACCCCTCAGCGGCCGGCCAGTCACTTTCAGCCCAGGGCTTACTCTTCTCCTGTG
GCTGCCCTCCCACCTCACCACCCAGGGGCCACCCAGCCCAACGGCCTCTCTCAGGAGGGTCCCATCTATCGCTGCCAGGAAGAAGGC
CTGGGTCACTTTCAAGCTGTGATGATGGAACAAATTGGCACTAGAAGTGGAATAAGAGGACCTTTCCAGGAAATGTACAGACCATTA
GGAATGCAGATGCACCCGGTCCAGTCGCAGGCCTCGTTCCCAAAGACCCCCACAGCAGCAACATCACAGGAGGAGGTGCCGCCTCAT
AAGCCTCCAACACTTCCCCTGGATCAGAGCTAGCTCGAGATATCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCA
TTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCATTATTTG

CECR2 - P1430L

CAAATAATGATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATAAGCAATGCTTTTTTATAATGCCAACTTTGTAC
AAAAAAGCAGGCTTCGAAGGAGATAGAACCAATTCTCTAAGGAAATACTTAACCTGGTCGACTGGATCCGCCACC**ATGTGCCCAGAG**
GAGGGCGGCGCGGCCGGGCTGGGCGAGCTCCGCTCCTGGTGGGAGGTCCCGGCCATCGCGCACTTCTGCTCGCTCTTTCGCACCGCG
TTCCGCCTGCCCGACTTCGAGATCGAGGAGTTAGAAGCCGCTCTTCACAGAGATGACGTGGAGTTTATCAGTGACCTGATTGCCTGC
CTGCTTCAGGGCTGCTATCAACGAAGAGATATCACGCCTCAGACATTCCACAGCTACCTAGAGGACATCATCAACTACCGCTGGGAG
CTCGAAGAAGGGAAGCCCAACCCTCTGAGGGAAGCCAGTTTCCAGGACCTGCCTCTTCGCACACGGGTGGAGATCCTGCACCGACTC
TGTGATTACCGGCTGGATGCAGACGATGTCTTCGATCTTCTAAAGGGCCTGGATGCAGACAGTCTCCGTGTGGAGCCATTGGGTGAA
GACAATTCTGGGGCACTATATTGGTATTTCTATGGAACACGAATGTACAAAGAGGACCCGGTGCAAGGAAAATCCAATGGAGAACTC
TCTTTGAGCAGGGAAAGTGAAGGACAAAAAAATGTCTCAAGTATTCCTGGAAAAACGGGAAAAAGAAGAGGAAGACCCCCAAAACGG
AAGAAACTGCAGGAGGAGATTCTGTTGAGTGAAAAGCAGGAAGAAAATTCCTTGGCATCCGAGCCACAGACAAGACATGGGTCCCAA
GGGCCAGGCCAAGGTACTTGGTGGCTCCTGTGCCAGACAGAAGAGGAATGGAGACAGGTCACCGAGAGTTTTCGCGAGAGGACCTCC
CTTCGAGAACGGCAGCTCTACAAGCTCCTCAGTGAGGACTTCCTGCCTGAGATCTGCAACATGATCGCCCAGAAGGGAAAACGTCCA
CAGCGCACAAAGGCAGAGTTGCATCCTAGGTGGATGTCTGACCACCTGTCCATCAAACCCGTCAAGCAAGAGGAGACTCCTGTGCTG
ACCAGAATAGAAAAACAAAAGCGCAAAGAGGAGGAAGAAGAGCGTCAGATTCTTCTAGCAGTGCAGAAGAAGGAGCAGGAGCAGATG
CTAAAGGAAGAGAGGAAACGCGAGTTGGAGGAGAAGGTCAAGGCAGTGGAAGATCGAGCGAAGAGGAGAAAGCTCAGGGAAGAAAGG
GCATGGCTGCTGGCTCAAGGAAAGGAGCTCCCTCCAGAACTTTCCCATCTGGACCCCAATTCCCCCATGAGAGAGGAAAAAAAGACT
AAAGACCTCTTTGAGTTGGATGATGATTTCACTGCTATGTATAAAGTTCTAGACGTGGTAAAGGCTCACAAGGATTCCTGGCCCTTC
TTGGAACCTGTGGATGAATCTTATGCCCCTAACTATTATCAGATTATTAAGGCCCCCATGGATATTTCCAGCATGGAGAAGAAACTG
AATGGAGGTTTATACTGTACCAAGGAGGAATTTGTAAATGACATGAAGACCATGTTCAGGAATTGTCGAAAGTATAATGGGGAAAGT
AGTGAGTATACCAAGATGTCTGATAATTTAGAGAGGTGTTTCCATCGGGCAATGATGAAACATTTTCCTGGAGAAGATGGAGACACA
GATGAAGAATTTTGGATTCGAGAGGATGAAAAGCGGGAGAAAAGACGGAGTCGGGCTGGGCGAAGTGGTGGGAGCCATGTTTGGACC
CGCTCCAGGGACCCAGAAGGGTCCAGCAGGAAACAGCAGCCCATGGAGAATGGAGGAAAGTCGTTGCCCCCCACACGCCGAGCGCCC
TCTTCTGGGGACGATCAGAGCAGCAGCTCCACACAGCCCCCGCGGGAGGTGGGCACTTCCAATGGCCGAGGTTTTTCTCATCCCCTG
CATTGTGGTGGGACACCCAGCCAGGCACCCTTTTTAAACCAGATGAGGCCAGCAGTACCAGGAACATTTGGCCCTCTGCGAGGATCA
GATCCTGCCACCTTGTATGGCTCCTCTGGAGTCCCGGAGCCACACCCCGGGGAGCCTGTGCAGCAGCGTCAGCCTTTCACCATGCAG
CCTCCAGTTGGAATTAACAGCCTCCGAGGACCCAGGCTAGGCACACCAGAGGAGAAGCAAATGTGCGGGGGGCTGACACACCTTTCT
AACATGGGCCCACACCCTGGATCCTTGCAGCTTGGGCAGATAAGTGGCCCAAGTCAGGATGGAAGCATGTATGCTCCAGCTCAGTTC
CAGCCAGGATTCATTCCTCCCCGGCATGGGGGCGCTCCAGCCCGGCCACCAGACTTTCCTGAAAGCTCAGAAATTCCTCCCAGCCAT
ATGTATCGATCGTACAAGTACCTGAATCGAGTACACTCTGCCGTCTGGAATGGGAACCATGGTGCTACGAACCAAGGACCCTTGGGC
CCAGATGAGAAGCCCCACCTGGGGCCAGGACCCTCTCACCAGCCTCGCACTCTCGGTCACGTGATGGATTCCCGAGTCATGAGACCA
CCTGTCCCCCCCAACCAGTGGACTGAACAATCAGGCTTCCTACCTCATGGAGTTCCTTCCTCAGGGTACATGCGACCGCCCTGCAAG
TCTGCCGGACATCGGTTACAGCCACCTCCAGTGCCAGCACCCAGTTCTTTGTTTGGAGCACCTGCCCAGGCTCTTCGGGGGGTGCAG
GGAGGGGACTCCATGATGGACAGCCCAGAGATGATTGCGATGCAGCAGCTCTCCTCCCGCGTCTGCCCCCCAGGTGTGCCTTACCAC
CCCCACCAGCCTGCACACCCCCGTTTACCTGGCCCTTTTCCGCAGGTAGCTCACCCAATGTCAGTCACTGTGTCAGCCCCCAAGCCT
GCCCTGGGCAACCCTGGGAGGGCACCGGAGAACAGTGAAGCACAAGAGCCTGAGAATGACCAAGCAGAGCCGTTGCCTGGCCTTGAA
GAGAAACCACCAGGTGTTGGTACTTCAGAGGGGGTCTACCTCACACAACTACCTCACCCCACACCTCCCCTGCAGACTGACTGCACC
AGGCAGAGCTCACCACAAGAAAGGGAAACAGTGGGCCCGGAGCTCAAAAGCAGCTCCTCCGAATCTGCGGACAACTGTAAAGCAATG
AAGGGCAAGAATCCCTGGCCCTCGGATAGCAGCTACCCCGGCCCAGCCGCCCAAGGGTGCGTGAGAGACCTCTCCACGGTGGCAGAC
AGGGGCGCTCTATCCGAGAACGGAGTCATTGGGGAAGCATCTCCTTGTGGATCGGAGGGGAAGGGCCTTGGTAGCAGTGGTTCCGAA
AAGCTGCTCTGCCCCAGAGGCAGAACGTTGCAGGAAACCATGCCATGCACGGGACAGAACGCAGCGACACCGCCCAGCACAGACCCC
GGTTTGACGGGAGGCACTGTGAGCCAGTTTCCCCCGCTGTATATGCCTGGCCTAGAGTACCCGAATTCAGCTGCCCATTACCACATC
AGTCCAGGCCTGCAGGGTGTGGGCCCTGTGATGGGAGGGAAGTCCCCAGCATCCCATCCCCAGCATTTTCCCCCAAGGGGCTTTCAG
TCTAACCACCCACATTCTGGAGGCTTTCCCCGGTATCGCCCCCCACAAGGAATGAGGTATTCCTACCACCCACCGCCACAGCCTTCC
TACCACCACTATCAGCGAACTCCTTACTATGCCTGTCCACAGAGCTTTTCTGACTGGCAGAGACCTCTCCATCCCCAGGGAAGCCCA
AGCGGACCCCCAGCCAGTCAGCCTCCCCCACCAAGGTCCCTCTTCTCAGATAAGAATGCCATGGCCAGTCTGCAAGGCTGTGAGACA
CTGAATGCTGCCTTAACTTCTCCAACCCGTATGGATGCAGTGGCTGCTAAAGTCCCAAATGACGGGCAGAATCCTGGTCCAGAGGAA
GAGAAGCTGGATGAATCTATGGAGAGGCCAGAGAGTCCCAAAGAATTTTTAGACCTGGACAACCATAACGCAGCTACCAAGCGGCAG
AGCTCGTTGTCAGCCAGCGAGTATCTCTATGGAACTCCTCCGCCTCTGAGTTCAGGAATGGGATTTGGTTCATCTGCATTTCCACCC
CACAGTGTGATGCTGCAGACGGGGCCTCCCTATACCCCTCAGCGGCCGGCCAGTCACTTTCAGCCCAGGGCTTACTCTTCTCCTGTG
GCTGCCCTCCCACCTCACCACCCAGGGGCCACCCAGCCCAACGGCCTCTCTCAGGAGGGTCCCATCTATCGCTGCCAGGAAGAAGGC
CTGGGTCACTTTCAAGCTGTGATGATGGAACAAATTGGCACTAGAAGTGGAATAAGAGGACCTTTCCAGGAAATGTACAGACCATCA
GGAATGCAGATGCACCCGGTCCAGTCGCAGGCCTCGTTCCCAAAGACCCCCACAGCAGCAACATCACAGGAGGAGGTGCTGCCTCAT
AAGCCTCCAACACTTCCCCTGGATCAGAGCTAGCTCGAGATATCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCA
TTGCTTATCAATTTGTTGCAACGAACAGGTCACTATCAGTCAAAATAAAATCATTATTTG

Appendix E: Biogenic amine and amino acid concentrations in half brain samples.

Table E.1: Individual sample concentrations of biogenic amines, their metabolites, and their precursor amino acids in male FVB/129P2 in circling mice housed with wheels, non-circling mice housed with wheels, and non-circling mice housed with cardboard housing (standard). TYR and TRYP are μg/g of tissue and the rest are ng/g of tissue. Abbreviations: clockwise circler (CW), counterclockwise circler (CCW), noradrenaline (NA), 3,4-dihydroxyphenylacetic acid (DOPAC), dopamine (DA), 5-hydroxyindoleacetic acid (5-HIAA), homovanillic acid (HVA), 3-methoxytyramine (3-MT), serotonin (5-HT), tyrosine (TYR), tryptophan (TRYP).

Wheel	Phenotype	Hemisphere	NA	TYR	DOPAC	DA	5HIAA	HVA	3MT	5HT	TRYP
No	Non-circler	Left	348.97	16.90	287.67	1285.77	287.67	176.43	71.75	463.08	8.46
		Right	371.31	17.04	269.25	1174.77	276.81	164.51	60.87	418.91	7.89
No	Non-circler	Left	380.09	15.12	305.56	1386.01	271.55	179.50	78.79	428.99	5.86
		Right	394.49	16.69	292.11	1331.94	272.84	172.53	75.60	444.96	6.30
No	Non-circler	Left	399.03	16.77	300.56	1411.71	277.97	186.26	63.92	472.61	6.81
		Right	395.84	16.33	272.17	1277.75	273.52	168.44	74.20	443.99	6.75
No	Non-circler	Left	342.96	13.55	262.51	1298.95	229.29	148.65	78.00	402.17	4.56
		Right	368.04	14.74	255.05	1324.47	235.88	148.01	65.58	440.28	4.77
No	Non-circler	Left	360.42	11.23	256.63	1272.29	275.15	165.20	74.18	475.53	5.74
		Right	347.86	11.40	261.47	1276.26	269.06	170.71	61.83	445.95	5.04
No	Non-circler	Left	351.30	11.45	257.71	1368.44	223.11	151.93	76.32	446.12	3.78
		Right	354.73	12.26	221.15	1372.17	204.83	141.90	75.08	471.13	3.98
Yes	Non-circler	Left	360.39	26.63	277.87	1264.22	361.21	210.14	59.75	436.79	10.16
		Right	363.62	10.29	291.39	1347.07	356.90	215.75	73.77	456.92	10.64
Yes	Non-circler	Left	366.33	13.77	238.37	1363.34	234.22	149.33	79.17	468.07	4.00
		Right	382.50	13.50	253.46	1376.70	248.70	158.37	76.09	475.83	4.24
Yes	Non-circler	Left	272.94	10.42	230.31	1080.96	195.86	135.40	67.62	355.62	2.62
		Right	292.99	9.70	230.27	1058.28	196.89	125.86	56.96	390.10	2.61
Yes	Non-circler	Left	319.52	14.03	256.81	1198.61	251.18	163.60	50.32	412.68	4.97
		Right	316.49	14.54	236.24	1194.60	251.28	163.64	67.82	457.48	4.83
Yes	Non-circler	Left	293.37	6.78	197.98	1076.42	153.79	110.96	49.94	370.41	2.60
		Right	299.90	7.52	211.87	1042.78	176.36	120.13	62.48	400.47	3.00
Yes	CW	Left	319.90	10.31	241.79	1172.94	236.84	179.90	74.63	397.98	4.50
		Right	294.61	9.92	234.71	997.83	261.44	176.45	56.34	296.04	4.83
Yes	CW	Left	331.70	10.80	243.60	1152.03	234.78	149.10	84.32	417.64	3.86
		Right	351.54	11.85	263.50	1077.28	263.55	153.16	77.89	411.56	4.24
Yes	CW	Left	295.23	9.35	217.00	1128.22	184.84	140.03	66.10	359.49	2.61
		Right	321.09	11.33	212.25	1116.39	201.98	142.73	76.43	455.12	3.32
Yes	CCW	Left	348.75	15.25	249.35	1118.78	231.64	149.27	81.13	397.47	3.71
		Right	331.89	14.26	212.30	1136.52	199.85	135.69	72.87	439.70	3.55
Yes	CCW	Left	311.69	9.63	230.05	1134.26	206.40	137.71	59.58	402.00	3.68
		Right	311.19	9.90	251.88	1179.21	211.76	138.81	63.96	392.68	3.60
Yes	CCW	Left	288.00	10.69	231.24	1248.05	189.68	146.19	76.30	351.94	3.48
		Right	318.70	12.78	217.23	1201.87	205.42	155.43	72.92	423.19	4.02

Table E.2: Individual sample amino acid concentrations in male FVB/129P2 comparing circling mice housed with wheels, non-circling mice housed with wheels, and non-circling mice housed with cardboard housing (standard). Concentration is μg/g of tissue.
Abbreviations: clockwise circler (CW), counterclockwise circler (CCW), aspartate (ASP), glutamate (GLU), L-serine (L-SER), D-serine (D-SER), glutamine (GLN), glycine (GLY), arginine (ARG), taurine (TAUR), alanine (ALA), gamma-aminobutyric acid (GABA).

Wheel	Phenotype	Hemisphere	ASP	GLU	L-SER	D-SER	GLN	GLY	ARG	TAUR	ALA	GABA
No	Non-circler	Left	544.7	1944.4	70.7	23.5	735.1	116.1	32.2	1760.5	77.1	351.6
		Right	528.3	1910.0	72.2	23.0	841.5	108.6	30.7	1586.9	74.4	335.0
No	Non-circler	Left	524.8	1888.6	66.0	22.5	710.0	104.0	28.6	1685.9	72.7	348.3
		Right	552.4	1780.2	64.6	21.7	676.8	97.7	29.6	1487.2	71.2	345.0
No	Non-circler	Left	549.0	1956.4	69.1	23.2	660.1	110.4	28.1	1668.9	71.0	337.1
		Right	653.6	2379.0	85.5	26.5	819.5	139.7	34.9	2132.7	85.5	439.7
No	Non-circler	Left	470.4	1744.1	67.5	21.3	753.4	108.4	33.3	1630.0	79.3	382.7
		Right	482.5	1799.3	74.3	21.1	763.7	122.7	38.9	1857.6	86.2	462.8
No	Non-circler	Left	615.7	2518.8	97.0	29.4	962.8	140.5	34.7	1988.7	89.3	430.2
		Right	664.9	2493.3	94.7	28.6	938.6	136.1	33.9	1931.6	87.2	408.7
No	Non-circler	Left	585.6	2434.9	87.0	27.4	1016.6	144.9	31.5	2264.2	91.2	458.0
		Right	467.9	2022.0	71.8	23.6	806.7	117.6	26.4	1855.6	76.4	365.6
Yes	Non-circler	Left	663.7	2144.0	113.9	26.5	1853.6	161.7	42.0	1626.3	98.4	495.6
		Right	580.4	1816.0	92.6	23.8	1547.3	139.7	36.6	1451.1	84.1	421.0
Yes	Non-circler	Left	557.4	1978.2	84.7	24.5	862.0	138.5	37.7	1952.3	101.6	550.0
		Right	514.1	1887.9	71.2	22.6	744.3	114.3	30.7	1691.8	84.7	410.7
Yes	Non-circler	Left	517.2	1894.5	84.5	24.8	860.0	130.4	31.9	1750.6	92.3	458.6
		Right	539.5	1932.8	87.3	24.5	878.0	136.1	29.0	1771.3	94.2	462.0
Yes	Non-circler	Left	416.9	1861.8	67.9	18.9	1242.0	107.9	14.9	1442.8	71.5	352.9
		Right	459.6	2065.0	74.8	21.1	1285.7	120.4	17.6	1647.4	80.5	401.0
Yes	Non-circler	Left	413.5	1925.4	66.6	22.2	818.1	106.3	17.0	1658.8	80.1	384.2
		Right	457.2	2072.2	69.9	22.3	849.1	108.0	18.6	1720.2	82.1	396.2
Yes	CW	Left	842.7	2285.7	101.7	26.5	1018.4	128.7	25.0	1531.7	106.9	402.3
		Right	817.1	2107.4	93.7	25.0	899.6	109.2	24.3	1313.3	97.0	350.7
Yes	CW	Left	769.0	2203.2	91.1	26.2	961.0	119.0	25.8	1822.9	89.7	397.6
		Right	746.3	2032.7	86.2	24.3	892.4	115.6	25.5	1656.5	85.0	371.8
Yes	CW	Left	587.3	2358.9	102.3	32.7	1079.5	137.9	27.0	1876.0	111.3	498.7
		Right	478.2	2120.7	89.8	29.6	967.9	123.6	22.1	1765.8	94.9	421.0
Yes	CCW	Left	499.8	1964.1	77.6	23.2	808.9	124.2	29.2	1742.8	94.3	447.8
		Right	571.2	2213.6	89.2	25.1	862.5	143.4	33.3	1932.1	104.7	514.1
Yes	CCW	Left	442.4	1957.8	63.5	20.6	719.7	111.7	18.4	1773.5	72.4	378.1
		Right	490.8	2233.8	78.6	25.2	880.9	141.2	23.6	2204.9	90.1	491.5
Yes	CCW	Left	433.3	1856.5	63.2	20.8	719.7	104.3	18.0	1672.6	69.3	369.2
		Right	475.5	2081.1	72.5	22.6	917.4	121.8	20.7	1860.5	79.0	417.3

Appendix F: *Drosophila* progeny counts for RNAi crosses.

Table F.1: *Drosophila* progeny counts for RNAi CG32394 2478/GD (*dikar*) crosses. Number of males (M), females (F) and combined sexes (T) are given for both reciprocal crosses, and the number of progeny from both crosses were summed to give the combined totals. χ^2 goodness-of-fit tests were done to compare observed to expected sex ratios and, where applicable, balancer (TM3, CyO, TM6B) to non-balancer (Non-TM3, Non-CyO, Non-TM6B) ratios. Statistical analyses were not performed to compare sex ratios in balancer classes. Significant *P*-values are bolded and marked with an asterisk. The proportion of pupae that failed to eclose was an estimate based on the observation of uneclosed pupae within the vial.

RNAi CG32394 24738/GD (*dikar*)						
GAL4 driver	**M RNAi x F GAL4**	**M GAL4 x F RNAi**	**Combined totals**	**Male to female ratio χ^2 *P*-value**	**Balancer to non-balancer ratio χ^2 *P*-value**	**Proportion of pupae that failed to eclose**
GAL4-eyeless	M-77 F-66 T-163	M-83 F-69 T-152	M-160 F-155 T-315	0.7782	NA	≤ 1/10
GAL4-T279	M-50 F-47 T-97	M-83 F-101 T-184	M-133 F-148 T-281	0.3709	NA	≤ 1/10
GAL4-prd/TM3	TM3: M-37 F-38 T-75 Non-TM3: M-31 F-38 T-69	TM3: M-26 F-29 T-55 Non-TM3: M-39 F-49 T-88	TM3: M-63 F-67 T-130 Non-TM3: M-70 F-87 T-157	0.1749	0.1110	≤ 1/10
GAL4-HS-2	M-39 F-62 T-101	M-51 F-55 T-106	M-90 F-117 T-207	0.0606	NA	≤ 1/10
GAL4-HS-3	M-65 F-69 T-134	M-66 F-69 T-135	M-131 F-138 T-269	0.6695	NA	~ 1/6
GAL4-en	M-75 F-77 T-152	M-88 F-82 T-170	M-163 F-159 T-322	0.8236	NA	~ 1/6
GAL4-actin-2/CyO	CyO: M-52 F-57 T-109 Non-CyO: M-34 F-48 T-82	CyO: M-38 F-41 T-79 Non-CyO: M-46 F-49 T-95	CyO: M-90 F-98 T-188 Non-CyO: M-80 F-97 T-177	0.2013	0.5648	≤ 1/10
GAL4-actin-3/TM6B	TM6B: M-22 F-20 T-42 Non-TM6B: M-30 F-44 T-74	TM6B: M-25 F-39 T-64 Non-TM6B: M-36 F-58 T-94	TM6B: M-47 F-59 T-106 Non-TM6B: M-66 F-102 T-168	**0.0055***	**0.0002***	≤ 1/10

Table F.2: *Drosophila* progeny counts for RNAi CG32394 100383/KK (*dikar*) crosses. Refer to Table F.1 for details.

RNAi CG32394 100383/KK (*dikar*)						
GAL4 driver	**M RNAi x F GAL4**	**M GAL4 x F RNAi**	**Combined totals**	**Male to female ratio χ^2 *P*-value**	**Balancer to non-balancer ratio χ^2 *P*-value**	**Proportion of pupae that failed to eclose**
GAL4-eyeless	M-54 F-80 T-134	M-93 F-70 T-163	M-147 F-150 T-297	0.8618	NA	~ 1/6
GAL4-T279	M-81 F-66 T-147	M-73 F-74 T-147	M-154 F-140 T-294	0.4142	NA	~ 1/6
GAL4-prd/TM3	TM3: M-38 F-36 T-74 Non-TM3: M-39 F-43 T-82	TM3: M-41 F-52 T-93 Non-TM3: M-56 F-41 T-97	TM3: M-79 F-88 T-167 Non-TM3: M-95 F-84 T-179	0.4110	0.5188	~ 1/6
GAL4-HS-2	M-26 F-38 T-64	M-40 F-85 T-125	M-66 F-123 T-189	**<0.0001***	NA	~ 1/2
GAL4-HS-3	M-38 F-42 T-80	M-40 F-36 T-76	M-78 F-78 T-156	1.000	NA	~ 1/6
GAL4-en	M-52 F-53 T-105	M-79 F-65 T-144	M-131 F-118 T-249	0.4100	NA	≤ 1/10
GAL4-actin-2/CyO	CyO: M-52 F-60 T-112 Non-CyO: M-14 F-85 T-99	CyO: M-84 F-74 T-158 Non-CyO: M-23 F-76 T-99	CyO: M-136 F-134 T-270 Non-CyO: M-37 F-161 T-198	**<0.0001***	**0.0009***	~ 1/3
GAL4-actin-3/TM6B	TM6B: M-64 F-52 T-116 Non-TM6B: M-23 F-64 T-87	TM6B: M-66 F-64 T-130 Non-TM6B: M-10 F-66 T-76	TM6B: M-130 F-116 T-246 Non-TM6B: M-33 F-130 T-163	**<0.0001***	**<0.0001***	~ 1/3

Table F.3: *Drosophila* progeny counts for RNAi CG32393 107312/KK (*dikar*) crosses. Refer to Table F.1 for details.

RNAi CG32393 107312/KK (*dikar*)						
GAL4 driver	**M RNAi x F GAL4**	**M GAL4 x F RNAi**	**Combined totals**	**Male to female ratio χ^2 *P*-value**	**Balancer to non-balancer ratio χ^2 *P*-value**	**Proportion of pupae that failed to eclose**
GAL4-eyeless	M-66 F-85 T-151	M-34 F-37 T-71	M-100 F-122 T-222	0.1398	NA	≤ 1/10
GAL4-T279	M-85 F-75 T-160	M-92 F-79 T-171	M-177 F-154 T-331	0.2062	NA	~ 1/6
GAL4-prd/TM3	TM3: M-24 F-32 T-56 Non-TM3: M-41 F-58 T-99	TM3: M-32 F-37 T-69 Non-TM3: M-30 F-42 T-72	TM3: M-56 F-69 T-125 Non-TM3: M-71 F-100 T-171	**0.0266***	**0.0075***	~ 1/6
GAL4-HS-2	M-54 F-55 T-109	M-37 F-33 T-70	M-91 F-88 T-179	0.8226	NA	~ 1/6
GAL4-HS-3	M-52 F-73 T-125	M-50 F-75 T-125	M-102 F-148 T-250	**0.0036***	NA	~ 1/6
GAL4-en	M-62 F-92 T-154	M-18 F-20 T-38	M-80 F-112 T-192	**0.0209***	NA	≤ 1/10
GAL4-actin-2/CyO	CyO: M-67 F-71 T-138 Non-CyO: M-0 F-0 T-0	CyO: M-42 F-60 T-82 Non-CyO: M-0 F-0 T-0	CyO: M-109 F-131 T-240 Non-CyO: M-0 F-0 T-0	1.000	**<0.0001***	~ 1/2
GAL4-actin-3/TM6B	TM6B: M-33 F-26 T-59 Non-TM6B: M-0 F-0 T-0	TM6B: M-38 F-36 T-74 Non-TM6B: M-0 F-0 T-0	TM6B: M-71 F-62 T-133 Non-TM6B: M-0 F-0 T-0	1.000	**<0.0001***	~ 1/2

Table F.4: *Drosophila* progeny counts for RNAi CG17697 43077/GD (*fz*) crosses. Refer to Table F.1 for details.

RNAi CG17697 43077/GD (*fz*)						
GAL4 driver	**M RNAi x F GAL4**	**M GAL4 x F RNAi**	**Combined totals**	**Male to female ratio χ^2 *P*-value**	**Balancer to non-balancer ratio χ^2 *P*-value**	**Proportion of pupae that failed to eclose**
GAL4-eyeless	M-77 F-79 T-156	M-46 F-47 T-93	M-123 F-126 T-249	0.8492	NA	≤ 1/10
GAL4-T279	M-73 F-65 T-138	M-65 F-73 T-138	M-138 F-138 T-276	1.000	NA	≤ 1/10
GAL4-prd/TM3	TM3: M-33 F-27 T-60 Non-TM3: M-37 F-32 T-69	TM3: M-33 F-42 T-75 Non-TM3: M-32 F-47 T-79	TM3: M-66 F-69 T-135 Non-TM3: M-69 F-79 T-148	0.4111	0.4397	~ 1/6
GAL4-HS-2	M-27 F-29 T-56	M-46 F-48 T-94	M-73 F-77 T-150	0.7440	NA	≤ 1/10
GAL4-HS-3	M-64 F-59 T-123	M-71 F-82 T-153	M-135 F-141 T-276	0.7180	NA	≤ 1/10
GAL4-en	M-65 F-64 T-129	M-64 F-55 T-119	M-129 F-119 T-249	0.5254	NA	~ 1/6
GAL4-actin-2/CyO	CyO: M-26 F-29 T-55 Non-CyO: M-9 F-24 T-33	CyO: M-50 F-44 T-94 Non-CyO: M-32 F-51 T-83	CyO: M-76 F-73 T-149 Non-CyO: M-41 F-75 T-116	**0.0016***	**0.0426***	~ 1/6
GAL4-actin-3/TM6B	TM6B: M-16 F-24 T-40 Non-TM6B: M-26 F-36 T-62	TM6B: M-11 F-14 T-25 Non-TM6B: M-18 F-22 T-40	TM6B: M-27 F-38 T-65 Non-TM6B: M-44 F-58 T-102	0.1657	**0.0042***	≤ 1/10

Table F.5: *Drosophila* progeny counts for RNAi CG17697 105697/KK (*fz*) crosses. Refer to Table F.1 for details.

RNAi CG17697 105493/KK (*fz*)						
GAL4 driver	**M RNAi x F GAL4**	**M GAL4 x F RNAi**	**Combined totals**	**Male to female ratio χ^2 *P*-value**	**Balancer to non-balancer ratio χ^2 *P*-value**	**Proportion of pupae that failed to eclose**
GAL4-eyeless	M-62 F-71 T-133	M-21 F-20 T-41	M-83 F-91 T-174	0.5442	NA	≤ 1/10
GAL4-T279	M-66 F-54 T-120	M-47 F-64 T-111	M-113 F-118 T-231	0.7422	NA	~ 1/6
GAL4-prd/TM3	TM3: M-22 F-34 T-56 Non-TM3: M-30 F-30 T-60	TM3: M-11 F-17 T-28 Non-TM3: M-15 F-20 T-35	TM3: M-33 F-51 T-84 Non-TM3: M-45 F-50 T-95	0.6080	0.4110	≤ 1/10
GAL4-HS-2	M-28 F-40 T-68	M-50 F-53 T-103	M-78 F-93 T-171	0.2513	NA	~ 1/6
GAL4-HS-3	M-51 F-54 T-105	M-43 F-73 T-116	M-94 F-127 T-221	**0.0264***	NA	~ 1/6
GAL4-en	M-42 F-34 T-76	M-58 F-40 T-98	M-100 F-74 T-174	**0.0487***	NA	~ 1/6
GAL4-actin-2/CyO	CyO: M-66 F-77 T-143 Non-CyO: M-0 F-0 T-0	CyO: M-26 F-30 T-56 Non-CyO: M-0 F-0 T-0	CyO: M-92 F-107 T-199 Non-CyO: M-0 F-0 T-0	1.000	**<0.0001***	~ 1/2
GAL4-actin-3/TM6B	TM6B: M-43 F-35 T-78 Non-TM6B: M-0 F-0 T-0	TM6B: M-11 F-10 T-21 Non-TM6B: M-0 F-0 T-0	TM6B: M-54 F-45 T-99 Non-TM6B: M-0 F-0 T-0	1.000	**<0.0001***	~ 1/2

Table F.6: *Drosophila* progeny counts for RNAi CG8075 100819/KK (*Vang*) crosses. Refer to Table F.1 for details.

RNAi CG8075 100819/KK (*Vang*)						
GAL4 driver	**M RNAi x F GAL4**	**M GAL4 x F RNAi**	**Combined totals**	**Male to female ratio χ^2 *P*-value**	**Balancer to non-balancer ratio χ^2 *P*-value**	**Proportion of pupae that failed to eclose**
GAL4-eyeless	M-62 F-77 T-139	M-68 F-64 T-132	M-130 F-141 T-271	0.5040	NA	~ 1/6
GAL4-T279	M-74 F-82 T-156	M-79 F-83 T-162	M-153 F-165 T-318	0.5010	NA	≤ 1/10
GAL4-prd/TM3	TM3: M-30 F-20 T-50 Non-TM3: M-24 F-32 T-56	TM3: M-34 F-42 T-76 Non-TM3: M-50 F-43 T-93	TM3: M-64 F-62 T-126 Non-TM3: M-74 F-75 T-149	0.9347	0.1655	≤ 1/10
GAL4-HS-2	M-39 F-55 T-94	M-44 F-51 T-95	M-83 F-106 T-189	0.0943	NA	~ 1/6
GAL4-HS-3	M-51 F-46 T-97	M-63 F-75 T-138	M-114 F-121 T-235	0.6479	NA	~ 1/6
GAL4-en	M-63 F-56 T-119	M-104 F-88 T-192	M-167 F-144 T-311	0.1922	NA	≤ 1/10
GAL4-actin-2/CyO	CyO: M-34 F-67 T-101 Non-CyO: M-54 F-56 T-110	CyO: M-40 F-58 T-98 Non-CyO: M-33 F-51 T-84	CyO: M-74 F-125 T-199 Non-CyO: M-87 F-107 T-194	0.1510	0.8009	≤ 1/10
GAL4-actin-3/TM6B	TM6B: M-29 F-24 T-53 Non-TM6B: M-33 F-55 T-88	TM6B: M-16 F-22 T-38 Non-TM6B: M-26 F-51 T-77	TM6B: M-45 F-46 T-91 Non-TM6B: M-59 F-106 T-165	**0.0003***	**<0.0001***	≤ 1/10

Table F.7: *Drosophila* progeny counts for RNAi CG8075 7376/GD (*Vang*) crosses. Refer to Table F.1 for details.

RNAi CG8075 7376/GD (*Vang*)						
GAL4 driver	**M RNAi x F GAL4**	**M GAL4 x F RNAi**	**Combined totals**	**Male to female ratio χ^2 *P*-value**	**Balancer to non-balancer ratio χ^2 *P*-value**	**Proportion of pupae that failed to eclose**
GAL4-eyeless	M-119 F-123 T-242	M-78 F-115 T-193	M-197 F-238 T-435	**0.0493***	NA	~ 1/3
GAL4-T279	M-54 F-33 T-87	M-61 F-64 T-125	M-115 F-97 T-212	0.2164	NA	~ 1/6
GAL4-prd/TM3	TM3: M-36 F-38 T-74 Non-TM3: M-35 F-44 T-79	TM3: M-35 F-47 T-82 Non-TM3: M-37 F-38 T-75	TM3: M-71 F-85 T-156 Non-TM3: M-72 F-82 T-154	0.4203	0.9096	≤ 1/10
GAL4-HS-2	M-25 F-21 T-46	M-69 F-69 T-138	M-94 F-90 T-184	0.7681	NA	~ 1/6
GAL4-HS-3	M-29 F-20 T-49	M-57 F-83 T-140	M-86 F-103 T-189	0.2162	NA	~ 1/6
GAL4-en	M-55 F-55 T-110	M-13 F-20 T-33	M-68 F-75 T-143	0.5583	NA	~ 1/3
GAL4-actin-2/CyO	CyO: M-82 F-76 T-158 Non-CyO: M-47 F-56 T-103	CyO: M-65 F-102 T-167 Non-CyO: M-27 F-37 T-64	CyO: M-147 F-178 T-325 Non-CyO: M-74 F-93 T-167	0.1415	**<0.0001***	≤ 1/10
GAL4-actin-3/TM6B	TM6B: M-70 F-63 T-133 Non-TM6B: M-32 F-19 T-51	TM6B: M-49 F-62 T-111 Non-TM6B: M-15 F-35 T-50	TM6B: M-119 F-125 T-244 Non-TM6B: M-47 F-54 T-101	0.4861	**<0.0001***	≤ 1/10