Identification of CECR2-containing chromatin-remodeling complexes and
their Chromatin-binding sites in mice

by

Farshad Hassanzadeh Niri

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

Eukaryotic nuclear DNA is packaged into chromatin, a complex nucleoprotein structure. This has functional consequences by controlling the accessibility of DNA to binding factors responsible for many important cellular processes such as transcription, DNA replication and DNA repair. ATP-dependent chromatin remodeling complexes such as the ISWI family can regulate these cellular processes by altering the chromatin structure. CECR2 is a chromatin remodeling factor that forms a complex with ISWI proteins SNF2H and SNF2L. Loss-of-function mutations in *Cecr2* result in the perinatal lethal neural tube defect, exencephaly. Nonpenetrant animals that survive to adulthood exhibit subfertility. CECR2 loss affects transcription of multiple genes and is also involved in γ-H2AX formation and DSB repair. The mutant phenotypes indicate that CECR2 plays an important role in neural tube development and reproduction, but the mechanism of its function is not known. I therefore have investigated the components of the CECR2 complex and its chromatin binding sites in ES cells and testis. I hypothesized that the CECR2 complexes contain tissue-specific components that may correspond to tissue-specific functions in ES cells and testis. I also hypothesized that the CECR2 containing complexes occupy different chromatin binding sites. This work first required the development of a highly specific CECR2-specific antibody.

I confirmed that CECR2 forms a complex with SNF2H and SNF2L both in mouse ES cells and in testes. I showed that the CECR2-containing complex in mouse ES cells and testes has a size of approximately 2 MDa, suggesting the presence of additional components in the complex. Mass spectrometric analysis of CECR2-containing complexes revealed novel binding partners of CECR2 in ES cells and adult testis. I identified CCAR2 as a new member of the CECR2-containing complex in ES cells and possibly testis. CCAR2 has been shown to be

involved in DNA damage response, which is also a known function of CECR2. Strikingly, there is a difference in the composition of complexes isolated from ES cells and adult testis. LUZP1 (leucine zipper protein 1) was confirmed to be a binding partner of CECR2 only in mouse ES cells and not in testes. *Luzp1* mutant mice display exencephaly in 42% of embryos, indicating its role in neural tube closure. LUZP1 appears to play a role in stabilizing the CECR2 complex. I showed that the CECR2 complexes have different components, which opens the door toward understanding the multiple functions of CECR2.

Disruption of *Cecr2* results in dysregulation of expression of many genes in mouse embryos, however the direct transcriptional targets of the CECR2 complexes are unknown. Therefore, to find the direct binding sites of the CECR2 complex, chromatin immunoprecipitation followed by sequencing (ChIP-seq) analysis was performed in ES cells and adult testis. Looking at the overlapping binding sites of CECR2 and SNF2H allowed for a more powerful analysis and investigating the additional overlap with the binding sites of LUZP1 in ES cells allowed me to look for the ES cell-specific binding sites. Little overlap of CECR2 binding sites between ES cells and testes was observed, suggesting tissue-specific transcriptional regulation. Analysis of the ChIP-seq data revealed that the CECR2 complex occupies the promoter and cis-regulatory regions of many genes.

The identified genes in ES cells are involved in different aspects of embryonic development including brain, heart and kidney development. The *Cecr2* mutants exhibit abnormalities in these three organs. Gene ontology (GO) analyses of the genes associated with the binding sites of the CECR2 complex in ES cells showed that this complex modulates important molecular pathways including Shh and Wnt signaling. The genes identified in testes are involved in different aspects of reproduction and development. I identified many candidate

genes that can be used to investigate the CECR2 function in neurulation and fertility. *Hsd17b2*, *Lpar1*, *Nf, Lrp6* and *Phactr4* are possibly directly regulated by the CECR2 complex, all of which cause exencephaly when mutated. *Elmo1*, *Fgfr4*, *Ggt1*, *Insr*, *Itgb3* and *Schip1* from the ES cell dataset and *Cdc14b, Nfia*, *Pcsk1* and *Styx* from the testis dataset are candidate genes involved in reproduction.

My findings revealed that there are ES and testis-specific CECR2 complexes in mice. I showed that these complexes have different compositions and chromatin binding sites, which will facilitate understanding the multiple functions of CECR2 during development and reproduction.

Preface

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

Some of the research for this thesis has been completed in collaboration.

Section 2.21.1: Aligning ChIP-seq reads and detecting peaks were performed in collaboration with Dr. Paul Stothard.

Figure 3-12: Kacie Norton (graduate student) performed the IF staining and contributed the data to my thesis.

Figures 4-1 & 4-2: Alaina Tripstra (graduate student) performed IF and contributed the results as the figures to my thesis.

Section 3.10.1.3: Nhu Trieu (undergraduate student under my supervision) performed this part of the experiments and contributed the data to my thesis.

Section 3.10.4.1 Kenji Rowel Lim (undergraduate student under my supervision) performed part of the experiments in this section and contributed the data to my thesis.

Figure 3-39: Justin Elliott (undergraduate student under my supervision) performed part of the apoptosis experiment and contributed the figure to my thesis.

Acknowledgements

First and foremost, I am deeply grateful to my supervisor, Dr. Heather McDermid, for her great supervision, patience and intellectual and financial support, which made the completion of this research possible. She provided a creative and open environment for us to work in. She is not only a remarkable scientist but also a great role model to us.

I would also like to thank the rest of my thesis advisory committee, Dr. Kirst King- Jones and Dr. Roseline Godbout, for their guidance and help.

I am also grateful to the present and former members of the McDermid lab. They are all my friends and I enjoyed working with them. Thank to Renee Leduc, Kacie Norton, Alaina Trepstra and Parmveer Singh for being such great colleagues. Thanks to all undergraduate students who worked with me in the lab during my PhD program.

I also thank my friends Dr. Mohammad Shahrooie and Dr. Dhamo Neel for their helpful advice.

I would also like to thank Ehsan Misaghi for helping me in computational analyses.

Finally, I would like to dedicate this thesis to my family. They always support me with their endless love.

Table of Contents

List of tables

List of figures

List of abbreviations

AA	Amino acid
ACF	ATP-utilizing Chromatin assembly and remodeling Factor
Alx1	ALX homeobox 1
Amp	Ampicillin
AT-hook	Adenosine thymine DNA binding hook
ATP	Adenosine triphosphate
BAF	BRG1-Associated Factor
BAF155	BRG1-Associated Factor 155
BAM	Binary format for storing sequence data
BAZ	Bromodomain adjacent to zinc finger
BER	Base excision repair
BMP	Bone morphogenetic protein
bp	Base pair
Bptf/BRTF	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
Caco-2	Caucasian colon adenocarcinoma cell line (human)
Caspase-3	Cysteine-aspartic protease 3
cDNA	complementary DNA
Cecr2/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
CESCR	Cat eye syndrome critical region
CHD	Chromodomain helicase DNA-binding
ChIP-Seq	Chromatin immunoprecipitation with massively parallel DNA sequencing
CHRAC	CHRomatin Accessibility Complex
CHRAC-15	Chromatin accessibility complex, 15 kDa subunit
CHRAC-17	Chromatin accessibility complex, 17 kDa subunit
DAPI	4',6-diamidino-2-phenylindole
DDT	DNA binding homeobox and Different Transcription factors
DEPC	Diethylpyrocarbonate
DExx	Asp-Glu-xx(two indeterminate amino acids) box helicase domain
DLHP	Dorso-lateral hinge point
DMSO	Dimethyl sulfoxide

DNA	Deoxyribonucleic acid
DNMBP	Dynamin binding protein
dNTP	Deoxyribonucleotide triphosphate
DPX	Dibutyl phthalate, xylene
DSB	Double strand break
DVE	Distal visceral endoderm
E	Embryonic day
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
Elmo2	Engulfment and cell motility 2
ES	Embryonic stem
FACT	Facilitates Chromatin Transcription
FGF	Fibroblast growth factor
g	Gram
GT	Genetrap
H1	Histone H1
H2A	Histone H2A
H2A.Z	Histone H2A.Z (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
HDAC1	Histone deacetylase 1
HEK293	Human embryonic kidney 293 cell line
HEK293T	HEK293 cell line containing SV40 large T antigen
HELICc	Helicase superfamily c-terminal domain
HEPES	4-(2-hydroxyethyl)-l - piperazineethanesulfonic acid
Hh	Hedgehog
HMGA2	High mobility group A2
HR	Homologous recombination
HRP	Horseradish peroxidase
HSA	Helicase-SANT-associated domain
HSS	HAND–SANT–SLIDE
IgG	Immunoglobulin G
IGV	Integrative genomics viewer
INDEL	Insertion/deletion
INF2	Inverted formin, FH2 and WH2 domain containing
INO80	Inositol requiring 80
IR	Ionizing radiation
ISWI	Imitation SWItch
JAK/STAT	Janus Kinase/Signal transducers and activators of transcription
JNK	C-Jun NH2-terminal kinase

kb	Kilobase
kDa	Kilodalton
L	Litre
LB	Lysogeny broth
M	Molar
MBD	Methylated DNA binding domain
MBSU	Molecular biology services unit
MDa	Megadalton
MeCP2	Methyl-CpG-binding protein 2
MgCl2	Magnesium chloride
MHP	Medial hinge point
mL	Millilitre
mM	Millimolar
mm9	UCSC mouse genome assembly build 9
mRNA	Messenger Ribonucleic acid
NaCl	Sodium chloride
NaOAc	Sodium acetate
NCoR	Nuclear receptor co-repressor
NFR	Nucleosome-free regions
NGS	Next generation sequencing
NHEJ	Non-homologous end-joining
NoRC	Nucleolar Remodeling Complex
NR	nuclear hormone receptor
NTD	Neural tube defect
NURD	Nucleosome Remodeling and Deacetylase complex
NURF	NUcleosome Remodeling Factor
P	P-value
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PHD	Plant homeodomain
PIT-1	POU domain class 1 transcription factor 1
PMSF	Phenyl methyl sulfonyl fluoride
pRNA	promoter RNA
PVDF	Polyvinylidene fluoride
PWWP	Pro-Trp-Trp-Pro
qRT-PCR	Quantitative real-time PCR
RbAp46/48	Retinal blastoma Associated proteins, 46 and 48 kDa
rDNA	Ribosomal Deoxyribonucleic acid
RefSeq	Reference Sequence database
RNA	Ribonucleic acid
RNA Pol II	RNA polymerase II
RNAi	RNA interference
RSF	Remodeling and Spacing Factor

Rsf-l/RSF-1	Remodeling and spacing factor- 1
Rxn	Reaction
SAM	Sequence Alignment/Map format
SANT	Yeast SWI3, yeast ADA2, human ncor, human TFIIIB
SAP	Shrimp alkaline phosphatase
SASS	Sciences animal support services
SDS	Sodium dodecyl sulfate
Shh	Sonic hedgehog
SLIDE	SANT-like ISWI domain
SMAD	Sma and Mad related protein
Smarca1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily A, member 1 (aka SNF2L)
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)
SMRT	Silencing mediator for retinoid and thyroid hormonereceptors
SNF2	Sucrose non-fermenting 2
Snf2h/SNF2H	Sucrose non-fermenting 2-homolog
Snf2l/ SNF2L	Sucrose non-fermenting 2-like
Sry	Sex determining region Y
SWI/SNF	Switching defective/Sucrose non-fermenting
TBP	TATA-binding protein
TBS	Tris-buffered saline
TBST	TBS with 0.1% Tween-20
TC-NER	Transcription-coupled nucleotide excision repair
TE	Tris EDTA
TEMED	Tetramethylethylenediamine
TIP5	TTF-1 interacting protein 5
ToRC	Toutatis containing chromatin Remodeling Complex
Tris	Tris (hydroxymethyl)ethylaminomethane
Tris-Cl	Tris base hcl
TTF-1	Transcription termination factor-1
WCRF135	Williams Syndrome Transcription Factor-Related Chromatin-Remodeling Factor
WICH	WSTF-ISWI chromatin remodeling complex
Wnt	Wingless type
WSTF	Williams syndrome transcription factor, also known as BAZ1B
X-gal	5-bromo-4-chloro-indolyl-%-D-galactopyranoside
μM	Molar concentration in micromoles/litre

1. Introduction

1.1. Chromatin structure

In eukaryotic cells, the genome is located within the nucleus and mitochondria. In the nucleus, it is packaged into a highly organized and compact structure termed chromatin, which consists of DNA and associated proteins including histones and non-histone proteins. The structural repeating unit of chromatin is the nucleosome, which consists of 146 base pairs (bp) of DNA wrapped around an octamer of histone proteins and separated from the neighboring nucleosomes by a short linker DNA (10-80 bp in length) (Luger et al. 2012). This nucleosomal array is the first level of the chromatin compaction and is known as a "beads-on-a-string" conformation of DNA. A canonical histone octamer consists of a $(H3-H4)_2$ tetramer and two H2A-H2B dimers (Rando & Chang 2009). There are also other histone variants that can be part of the nucleosomes in specific regions of chromatin. H2A.Z, for instance, is one of the histone variants that incorporates into nucleosome adjacent to transcription start sites (Clapier & Cairns 2009). The four histone proteins are evolutionarily highly conserved and each contains a central domain and a highly basic N-terminal tail domain. Histone H2A also contains a C-terminal tail. The central domain is involved in histone-histone and histone-DNA interactions. The tails of the histones are extensions from the surface of the nucleosomes and have a role in the stability of the nucleosome structure (Iwasaki et al. 2013). The histone tails are also targets for post-translational modifications, which change the higher-order chromatin structure, and thereby control the accessibility of genes (Biswas et al. 2011).

The nucleosomal array is considered as the primary level of chromatin organization and is responsible for a 7-fold compaction of genomic DNA in eukaryotic cells (Cutter & Hayes 2015). The folding and compaction of the nucleosomal arrays leads to higher-order chromatin structures, which include secondary and tertiary structures. The 11 nm individual nucleosomal array is folded and produces a fiber with a diameter of approximately 30 nm, chromatin secondary structure, which produces a 50-fold compaction of the genomic DNA. Secondary chromatin structures interact and create chromatin tertiary structures, which are highly compacted

configurations. Linker histone H1 and the other architectural proteins including methyl-CpG-binding protein 2 (MeCP2), HMG proteins, HP1 and many others are involved in the formation of higher-order chromatin structures (Luger et al. 2012). The mitotic/meiotic chromosome is chromatin with an extreme higher-order structure which creates 10,000 to 20,000-fold compaction of DNA (Woodcock & Ghosh 2010).

1.2. Chromatin structure and gene regulation

The arrangement of genomic DNA into compact chromatin makes it inaccessible for DNA-binding factors responsible for many important cellular processes such as transcription, DNA replication and DNA repair (Groth et al. 2007). This means that the organization of genomic DNA into chromatin is not simply for packaging purposes; it also has functional consequences by affecting the accessibility of DNA (Margueron & Reinberg 2010). Therefore, eukaryotic cells have obtained a wide range of mechanisms through evolution to modulate chromatin structure in order to regulate chromatin-dependent biological processes. The alteration of chromatin structure can be achieved by one of the several general mechanisms: core histone variant replacement, covalent posttranslational modification of core histones, ATP-dependent chromatin remodeling, DNA methylation and non-coding RNAs (Felsenfeld & Groudine 2003, Magistri et al. 2012).

1.3. Chromatin remodeling

To make DNA accessible for important cellular processes, the compact structure of chromatin must be remodelled. Chromatin remodeling is divided into two main categories: covalent remodeling and ATP-dependent chromatin remodeling (Hargreaves & Crabtree 2011). Covalent remodeling is achieved by enzymes through post-translational modifications of core histone tails upon both lysine and arginine residues. The post-translational modifications of histone tails disrupt the compact structure of chromatin either over short or long distances and make genes accessible for different enzymes. There is also evidence that these post-translational modifications regulate binding sites for specific chromatin binding proteins either positively or negatively (Biswas et al. 2011). Acetylation, methylation, phosphorylation, sumoylation and ubiquitylation are among the well-known covalent histone modifications (de la Serna et al. 2006).

Most of these post-translational modifications are reversible and counterpart enzymes can remove the modifications (Bannister & Kouzarides 2011). Histone acetylation usually results in gene activation, whereas histone deacetylation generally leads to transcriptional repression. The effect of histone methylation on gene activation depends on the amino acid residues undergoing methylation. Methylation of some residues (monomethylated H3K27, H3K9, H4K20, H3K79 and H2BK5) results in gene activation; whereas methylation of some other residues (trimethylated H3K27, H3K9 and H3K79) is associated with gene repression (Barski et al. 2007). Histone modifications also exert their effect by recruiting chromatin-binding proteins including ATP-dependent chromatin remodeling complexes by creating target sites for specific protein domains (Bannister & Kouzarides 2011). The bromodomain, for instance, recognizes acetylated lysine (Dhalluin et al. 1999, Sanchez & Zhou 2009). Using this mechanism histone post-translational modifications are involved in regulation of transcription, replication, DNA repair and recombination (Lalonde et al. 2014). Figure 1.1 shows some of the well-studied histone tail post-translational modifications and protein domains recognizing these modifications.
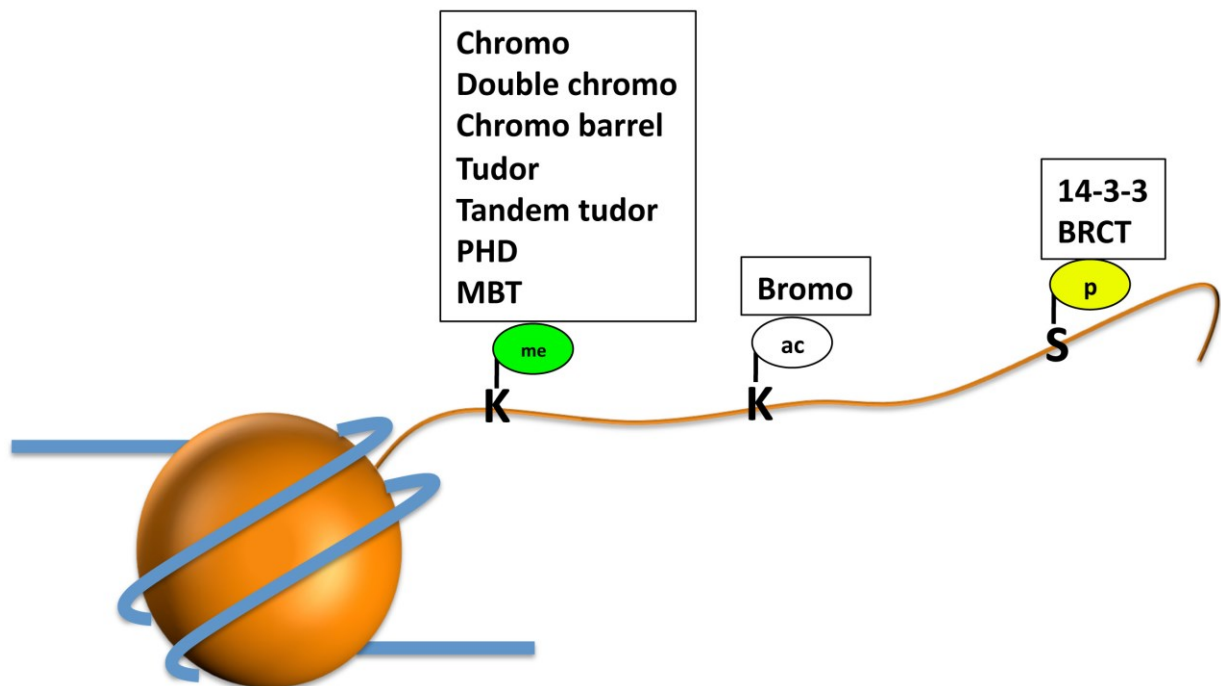
Figure 1-1 Histone post-translational modifications are binding targets for specific protein domains. Protein domains recognize methylated lysines (K-me), acetylated lysines (K-ac) and phosphorylated serines (S-p) and bind to chromatin (picture was made using information from Taverna et al. 2007).

1.4. ATP-dependent chromatin remodeling

Chromatin structure controls the accessibility of DNA to DNA-binding proteins. ATP-dependent chromatin remodeling is a process in which the energy produced from ATP hydrolysis is used to alter histone-DNA interactions within the nucleosomes making protein-binding sites accessible or inaccessible on the genome to DNA-binding proteins (Varga-Weisz 2010). ATP-dependent chromatin remodeling is the function of large multiprotein complexes. All eukaryotic cells use members of the chromatin remodeling complexes to control many crucial biological processes including replication, transcription, DNA repair, homologous recombination and chromatin assembly. All of these complexes contain a related ATPase component that belongs to the sucrose non-fermenting 2 (SNF2) family of ATPases (Piatti et al. 2011). In addition to an ATPase, these complexes contain up to 20 non-catalytic components that influence complex functions and are required for recognizing histone modifications and RNA signals to target them to specific genomic loci (Längst & Manelyte 2015). The ATPase component of these protein complexes, in addition to a related ATPase domain, contain other flanking protein domains, which are used to classify these complexes into four well-studied families: The Swi/Snf family, the Mi-2 / CHD family, the ISWI family and the Ino80 family (Figure 1-2) (Bartholomew 2014, Clapier & Cairns 2009). The Swi/Snf complex purified from *Saccharomyces cerevisiae* was the first discovered ATP-chromatin remodeling factor (Peterson & Herskowitz 1992). The Swi/Snf complexes are composed of 8-14 components and the catalytic ATPase component contains an N-terminal HSA (helicase-SANT-associated) and a C-terminal bromodomain (Clapier & Cairns 2009). These complexes remodel chromatin structure both by sliding and ejecting/inserting nucleosomes (Wilson & Roberts 2011). The ISWI complexes are multiprotein complexes and the presence of a set of HAND–SANT–SLIDE (HSS) domains at the C-terminus of the ATPase component is characteristic of this family (Toto et al. 2014). The Mi-2 / CHD complexes are composed of up to 10 components and the presence of two tandemly arranged chromodomains at the N-terminus of the ATPase component is the specific feature of this family (Marfella & Imbalzano 2007). The chromodomain domain binds to DNA, RNA and methylated histone H3 (Brehm et al. 2004). The Ino80 complexes are composed of more than 10 components and are defined by the presence of a split ATPase domain in the ATPase component (Bao & Shen 2007). This thesis focuses on the ISWI family.

1.5. The ISWI family of nucleosome remodeling complexes

This family of ATP-dependent chromatin remodeling complexes contains the ATPase ISWI (Imitation Switch) as their catalytic component (Clapier & Cairns 2009b). The ISWI protein was first characterized in *Drosophila melanogaster* and because of its sequence similarities to Brahma (the SWI2/SNF2 homolog in Drosophila) sequence, it was named ISWI
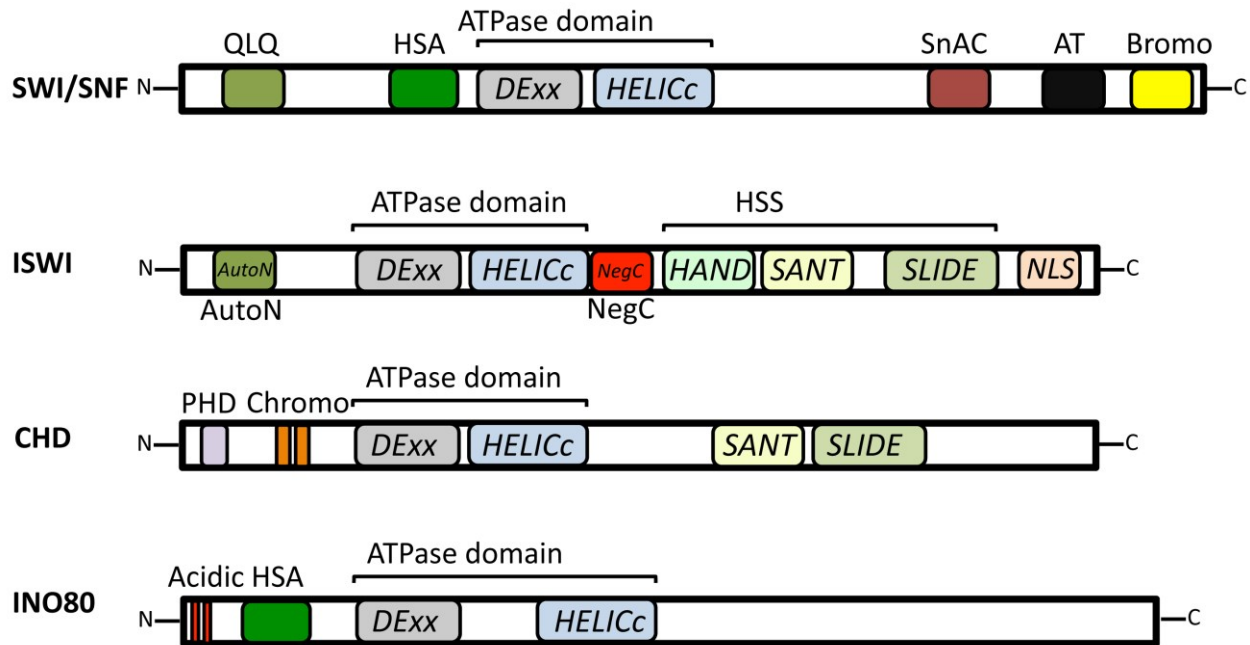
Figure 1-2 Schematic representation of the domains within the catalytic binding components for each of the four major classes of remodelers. ATPase binding components of all chromatin families share an ATPase domain composed of two parts: DExx and HELICc. Each subfamily is distinguished on the basis of the presence of the unique domains flanking the ATPase domain within their catalytic binding component: Bromodomain, HSA (helicase/SANT-associated) domain, QLQ domain and SnAC (Snf2 ATP coupling) domain for SWI/SNF family, HSS (HAND-SANT-SLIDE) module, AutoN domain, NegC domain and NLS (Nuclear Localization Signal) for ISWI family, tandem chromodomains, PHD (plant homeodomain), SANT-SLIDE module for the CHD family, HSA (helicase/SANT-associated) domain for the INO80 family (Bartholomew 2014, Clapier & Cairns 2009).

(Imitation SWItch) (Elfring et al. 1994). In addition to the conserved ATPase domain, all members of the ISWI family of ATPases have a C-terminal SANT domain (yeast SWI3, yeast ADA2, human NCoR, human TFIIIB) adjacent to a SLIDE domain (SANT-like ISWI domain), which together bind to an unmodified histone tail and DNA (Hargreaves & Crabtree 2011). Six distinct complexes containing the ISWI protein as their ATPase component and possessing chromatin-remodeling activity have been characterized in Drosophila: NURF (NUcleosome

Remodeling Factor) (Tsukiyama & Wu 1995), ACF (ATP-utilizing Chromatin assembly and remodeling Factor) (Ito et al. 1997), CHRAC (CHRomatin Accessibility Complex) (Varga-Weisz et al. 1997), RSF (Remodeling and Spacing Factor) (Hanai et al. 2008), ToRC (Toutatis containing chromatin Remodeling Complex) and NoRC (Nucleolar Remodeling Complex) (Emelyanov et al. 2012) complexes. Subsequently, ISWI remodeling complexes were identified in various species including yeast, human, mouse and Xenopus (Gangaraju & Bartholomew 2007). Figure 1-3 shows the chromatin remodeling complexes belonging to the ISWI family in different species.

1.5.1. The mammalian ISWI complexes

Mammalian homologs of Drosophila ISWI protein, SNF2L (Snf2- like) and SNF2H (Snf2 homolog), were first described in human cells (Aihara et al. 1998, Okabe et al. 1992). SNF2L is also called SMARCA1 (SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 1) and SNF2H is also called SMARCA5 (SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 5). WCRF135 (Williams Syndrome Transcription Factor-Related Chromatin-Remodeling Factor) is another name that has been used for human SNF2H as well. In this thesis, I will refer to these proteins as SNF2L and SNF2H.

Theses two ATPases are ~86% identical to each other and ~73% identical in amino acid sequence to Drosophila ISWI. Later, murine SNF2L and SNF2H were also characterized (Lazzaro & Picketts 2001). Similar to all ISWI enzymes, SNF2H and SNF2L, in addition to the conserved ATPase domain, contain a C-terminal SANT domain connected by a spacer helix to a SLIDE domain, which interacts with histones. Despite the similarity between SNF2H and SNF2L, these enzymes show obvious differences in the spatial and temporal gene expression that indicates their different functions or targets. Expression analysis has shown that *Snf2h* expression is higher in proliferating cells, whereas the expression of *Snf2l* is higher in differentiating cells (Lazzaro & Picketts 2001). During the embryonic stage (E9.5-E15.5) both *Snf2h* and *Snf2l* mRNAs have been detected throughout the embryo by Northern blot analysis and RNA *in situ* hybridization (Lazzaro & Picketts 2001). RNA *in situ* hybridization showed a higher *Snf2h* expression in the developing neocortex, cerebellum, olfactory epithelium, lungs, kidneys and gut
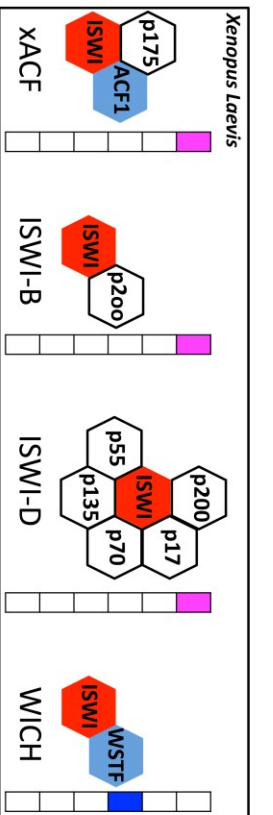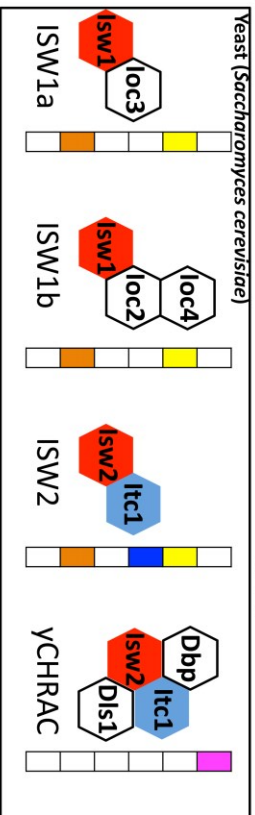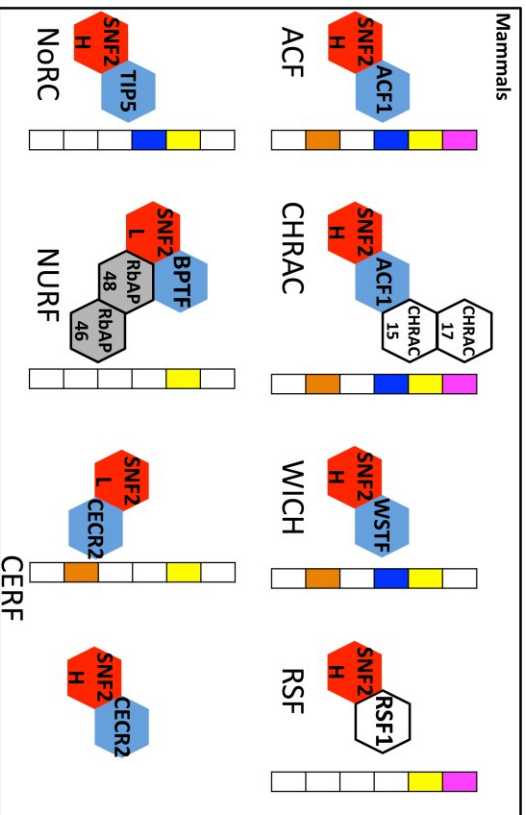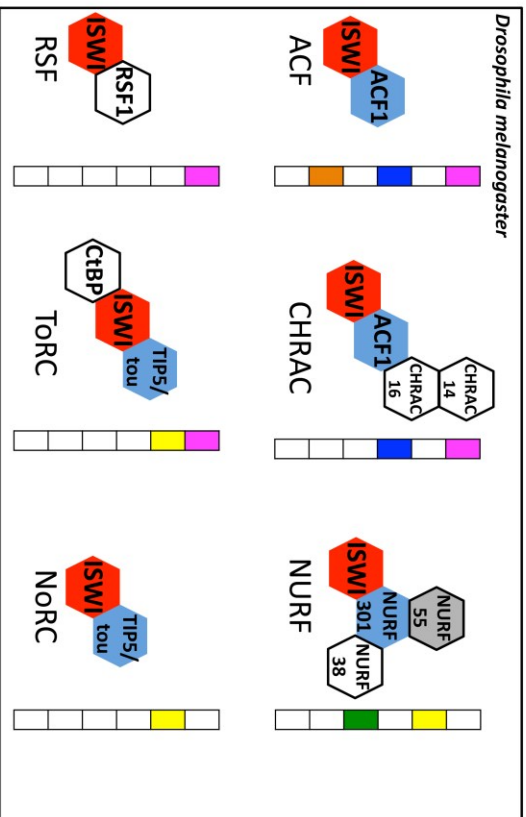
Legend:
- Chromatin assembly
- Repair
- Recombination
- Transcription
- Replication
- Chromatin structure

Shapes:
- ATPase
- BAZ and BAZ-like family
- Deacetylase complex

Drosophila melanogaster
- ACF (ISWI, ACF1)
- CHRAC (ISWI, ACF1, CHRAC 14, CHRAC 16)
- NURF (ISWI, NURF 301, NURF 55, NURF 38)
- RSF (ISWI, RSF1)
- ToRC (ISWI, TIP5/tou, CtBP)
- NoRC (ISWI, TIP5/tou)

Mammals
- ACF (SNF2H, ACF1)
- CHRAC (SNF2H, ACF1, CHRAC 15, CHRAC 17)
- WICH (SNF2H, WSTF)
- RSF (SNF2H, RSF1)
- NoRC (SNF2H, TIP5)
- NURF (SNF2L, BPTF, RbAP 48, RbAP 46)
- CERF (SNF2L, CECR2)

Yeast (Saccharomyces cerevisiae)
- ISW1a (Isw1, Ioc3)
- ISW1b (Isw1, Ioc2, Ioc4)
- ISW2 (Isw2, Itc1)
- yCHRAC (Isw2, Itc1, Dpb, Dls1)

Xenopus Laevis
- xACF (ISWI, ACF1, p175)
- ISWI-B (ISWI, p2oo)
- ISWI-D (ISWI, p55, p135, p200, p70, p17)
- WICH (ISWI, WSTF)

7

Figure 1-3 The ISWI family of ATP-dependent chromatin remodeling complexes: compositions and functions. The ISWI remodeling complexes were identified in various species including Drosophila, yeast, human, mouse and Xenopus. These complexes consist of a catalytic component (red blocks) belonging to the ISWI family of ATPases and one or more non-catalytic components. The largest non-catalytic components (blue blocks) of the ISWI complexes share several functional protein domains with the BAZ family of proteins. These complexes are involved in various biological processes including replication, transcription, DNA repair, homologous recombination and chromatin assembly (The figure is originally from Eberharter and Becker, 2004. I modified the figure and updated the information).

compared to the rest of the body. After birth, during the first two weeks there is an increase in the expression of *Snf2l* in the entire brain especially in the hippocampus and cerebellum and it continues to express during adulthood. The expression of *Snf2h* during this period is reduced throughout the brain and remains at a low level during the adulthood (Lazzaro & Picketts 2001).

In adult mice, the expression of *Snf2h* continues at low level in all tissues with a decrease in its expression in the brain and placenta whereas Snf2l expression is limited to brain, ovaries, testes, uterus and placenta (Lazzaro & Picketts 2001, Ye et al. 2009). The expression of *Snf2h* and *Snf2l* show different patterns in adult testes and ovaries as shown by *in situ* hybridization (Lazzaro & Picketts 2001). The *Snf2h* expression is detected in the periphery of the seminiferous tubules suggesting that *Snf2h* is expressed in spermatogonia and Sertoli cells. Whereas, the expression of *Snf2l* is lower than *Snf2h* and it is detected throughout the seminiferous tubules (Lazzaro & Picketts 2001). In the adult ovary, corpora lutea and granulosa cells in preovulatory follicles show a high level of *Snf2l* expression whereas the expression of *Snf2h* is higher in the rapidly proliferating, preantral follicles. *Snf2h* has also a higher level of expression in oocytes compared to Snf2l (Lazzaro et al. 2006, Lazzaro & Picketts 2001).

Analysis of the mutant mice has also shown a functional difference between *Snf2h* and *Snf2l*. A null mutation of *Snf2h* leads to the death of embryos in mice between E5.5 and E7.5 (Stopka & Skoultchi 2003), which means normal expression of *Snf2l* cannot compensate for the lack of *Snf2h* expression during early embryonic stages. On the other hand, the *Snf2l* mutation in mice causes only mild phenotypes involving brain and heart size and cell proliferation, suggesting distinct functions for *Snf2h* and *Snf2l* (Yip et al. 2012a). In the *Snf2l* mutation exon 6 encoding the ATP-binding motif has been deleted and predicted to disrupt the chromatin remodeling activity. The stable expression of *Snf2l* lacking exon 6 is detected at the transcription and translation levels in the mutant animals (Yip et al. 2012a). There is a possibility that the

protein product in the mutants has residual function and a null mutation could lead to more severe phenotypes.

In another study, increased levels of *Snf2l* expression (both at the transcription and translation levels) was detected in the cerebellum of *Snf2h* mutant mice, suggesting a temporary partial compensatory function of *Snf2l* expression in the absence of *Snf2h* expression (Alvarez-Saavedra et al. 2014). The depletion of *Snf2h* in cerebellar progenitors results in a 2.6-fold reduction in the *Snf2h* transcripts at birth based on qRT PCR. No change is detected for *Snf2l* transcripts in mutant cerebellum at this time. Interestingly, Western blot analysis shows an increase in SNF2L at postnatal (P) day 7. The depletion of *Snf2h* also leads to a reduction in EN1 level at P0 in mutant cerebella, however, the increase in *Snf2l* expression upregulated EN1 at P7, showing the compensation function of *Snf2l* although it wasn't sufficient to rescue the mutant phenotype. Also, the compensation did not occur when *Snf2h* depletion started in post-mitotic cerebellar Purkinje cells after P10 suggesting that *Snf2l* compensation occurs at a specific developmental stage.

SNF2H is part of multiple chromatin-remodeling complexes: ACF, CHRAC, WICH, NoRC, RSF and CERF (Figure 1-3). There is also evidence showing the presence of SNF2H in the mammalian ToRC complex (Emelyanov et al. 2012, Toto et al. 2014). SNF2L has been shown to be part of two complexes, human NURF and CERF (Banting 2004, Barak et al. 2003).

1.5.2. Functions of ISWI complexes

The chromatin architecture represents a barrier for the protein machinery carrying out different biological processes such as transcription, replication and DNA repair. ATP-chromatin remodellers play an important role in those biological processes by controlling the structure of chromatin (Figure 1-3).

Transcription:

ISWI-type remodeling complexes have been connected to different aspects of transcriptional regulation including transcription activation and repression of both coding and non-coding genes (Clapier & Cairns 2009). Most ISWI complexes act by promoting nucleosome assembly and organizing chromatin to promote transcriptional repression. ACF, CHRAC and NoRC regulate transcription through catalyzing nucleosome spacing, which leads to assembly of chromatin into higher-order structure and transcriptional repression (Längst et al. 1999, Li et al.

2006, Yang et al. 2006). The majority of ISW2, which is one of the ATP-chromatin remodeling complexes of the ISWI family in yeast, occupies sequences adjacent to gene promoter regions (Whitehouse et al. 2007). ISW2 is involved in the suppression of the transcription of both coding and non-coding genes by nucleosome repositioning in the vicinity of transcription start site. Another study has shown that the yeast ISW2 complex negatively regulates the size of nucleosome-free regions (NFRs) by repositioning nucleosomes toward the middle of NFRs (Yadon et al. 2010). The resulting restricted size of the NFRs leads to repression of non-coding genes at NFRs *in vivo*. Two other yeast ISWI complexes, ISW1a and ISW1b, also cause transcriptional repression (Moreau et al. 2003). These yeast ISW1 complexes repress the basal expression of the *PHO8* gene by removing TBP (TATA-binding protein) from the promoter. This ability of the yeast ISW1 complexes is a gene-specific event and is regulated by recruiting the yeast ISW1 complexes to the *PHO8* gene by Cbf1p, a sequence-specific DNA binding protein (Moreau et al. 2003). Similar ISWI-Cbf1p gene repression has not been reported for other promoters or other organisms.

ACF1, a component of the ACF complex in human, is involved in repression of several nuclear hormone receptor (NR)-regulated genes. Knockdown of *Acf1* alters the histone occupancy in the target gene promoters and subsequently leads to activation of these genes (Ewing et al. 2007). Mouse NoRC complexes reposition nucleosomes at the rDNA promoter resulting in transcription repression (Li et al. 2006). In addition to being the consequence of direct remodeling activity, the repression function of the NoRC complex is also mediated by DNA methyltransferases and histone deacetylases recruited to the rDNA promoters by the NoRC complex to establish transcriptionally inactive heterochromatin structure (Santoro et al. 2002). Recruiting the histone deacetylase-containing complex (Sin3A/Rpd3 complex) to promote transcriptional repression has been also reported for Drosophila ISWI protein (Burgio et al. 2008).

Although most of the ISWI complexes are involved in transcriptional repression, the NURF complex exhibits the opposite function by disturbing nucleosome spacing (Alkhatib & Landry 2011) and leading to transcriptional activation (Mizuguchi et al. 1997). It has been shown in Drosophila that NURF promotes transcription by interacting with transcription factors such as GAGA, HSF, the ecdysone receptor and the Ken repressor (Alkhatib & Landry 2011, Clapier & Cairns 2009). In humans NURF localizes to *engrailed* promoters and promotes transcription of

engrailed-1. In fact, knockdown of SNF2L and BPTF (components of the NURF complex) using siRNAs leads to a significant reduction in *engrailed-1* transcripts (Barak et al. 2003). The mouse NURF complex leads to activation of Smad-regulated genes that are required for the distal visceral endoderm (DVE) function and cell proliferation (Landry et al. 2008). In another study, the role of the human RSF complex (SNF2H and RSF1) in transcription regulation of ovarian cancer cells has been reported (Choi et al. 2009). The B-WICH complex has been isolated from HeLa cells that has a role in transcription (Cavellan et al. 2006). The B-WICH complex is a high molecular weight protein assembly (3 MDa) that is formed during active transcription by interaction of WICH complex with six nuclear proteins proteins: Sf3b155/SAP155, RNA helicase II/Gu□, Myb-binding protein 1a (Myb-bp1a), Cockayne syndrome protein B (CSB), the proto-oncogene Dek and nuclear myosin 1 (NM1). The 45S rRNA, the 5S rRNA, and the 7SL RNA also have been detected in the B-WICH complex. The interaction of RNA helicase II/Gu, Myb-bp1a and Sf3b155/SAP155 with the WICH complex is mediated by RNA and the inhibition of transcription results in disassociation of these proteins from the WICH complex. Only the protein-protein interaction in the core complex (WSTF and SNF2H) is independent of transcription (Cavellan et al. 2006). The B-WICH protein assembly comprised of SNF2H, WSTF and NM1 remodel the chromatin in the 5s rRNA and 7SL RNA loci and facilitates the binding of transcription factors (Sadeghifar et al. 2015). B-WICH is involved in transcription of rRNA by RNA polymerase I and RNA polymerase III. It has been shown that nuclear myosin 1 links chromatin remodeling of the WICH complex to the transcription machinery at rRNA genes (Cavellan et al. 2006, Percipalle et al. 2006). The B-WICH complex also activates rRNA transcription by remodeling the chromatin structure at the promoter region and consequently recruiting specific factors including certain histone acetyl-transferases (bearing H3K9-Ac-activity) to the DNA, resulting in activation of rRNA genes (Vintermist et al. 2011). Knock down of WSTF using siRNA leads to disassociation of several histone acetyl-transferases at the rRNA genes and reduction in the level of acetylated H3K9, which has been connected with active genes. The NoRC complex promotes H4 deacetylation, which results in rRNA transcriptional silencing, suggesting that B-WICH functions as the counterpart of NoRC at the rDNA loci (Vintermist et al. 2011).

Replication:

There is much evidence showing that ISWI complexes (ACF and CHRAC and WICH) participate in DNA replication by their ability to create regularly spaced nucleosomal arrays (Figure 1-3) (Corona & Tamkun 2004). In Drosophila ACF/CHRAC complexes facilitate chromatin assembly following DNA replication by their nucleosomal spacing activity (Fyodorov et al. 2003). There is also evidence that ISWI complexes make DNA accessible during initiation of replication for the replication machinery. The Drosophila CHRAC complex can change the nucleosomal structure at the origin of replication *in vitro*, promoting replication initiation (Alexiadis et al. 1998). The yeast ISWI complex, ISW2, is enriched at the active replication sites and has been connected to replication of late-replicating regions throughout the yeast genome (Vincent et al. 2008). The ISW2 complex is also involved in facilitating progression of the replication fork, especially during replication stress, as there was a 25% decrease in replication fork progression in *isw2* mutants compared to wild-types.

The mammalian ISWI complexes ACF, CHRAC and WICH are also involved in replication. Depletion of SNF2H or ACF1 proteins (components of both the ACF and CHRAC complexes) by RNAi decreased the progression speed of pericentromeric heterochromatin replication (Collins et al. 2002). Interestingly, decondensation of heterochromatin induced by 5-aza-2-deoxycytidine (which inhibits DNA methylation) blocked the effects of the depletion of ACF1 and SNF2H on replication, indicating that the efficient replication of condensed pericentromeric heterochromatin needs chromatin remodeling activity of ACF/SNF2H-containing complexes (ACF and CHRAC). Another ISWI complex, WICH, consisting of SNF2H and WSTF (Williams syndrome transcription factor, also known as BAZ1B) is involved in replication both in mammals and Xenopus (Bozhenok et al. 2002). Another study showed that the human WICH complex is recruited to replication foci by PCNA (a DNA replication factor) (Poot et al. 2004). Knock down of SNF2H or WSTF leads to a compact chromatin structure of newly synthesized chromatin. The authors suggested that the WICH complex keeps newly synthesized chromatin open for binding of the factors that copy and maintain the epigenetic state after the replication fork passes and prevents formation of abnormal heterochromatin (Poot et al. 2005). The human NoRC complex also has been connected to DNA replication of late-replicating rDNA (Li et al. 2005). Overexpression of TIP5 leads to early replication of late-replicating rDNA and results in a decrease in the fraction of active rDNA transcription units, indicating the important

role of the NoRC complex in establishing the epigenetic pattern of silent chromatin of rDNA after progression of the replication fork (Li et al. 2005). Taken together, by altering the positioning and structure of nucleosomes on both sides of the replication fork and changing the accessibility of DNA for other factors, ISWI complexes play critical roles in DNA replication (Erdel & Rippe 2011, Falbo & Shen 2006).

DNA repair:

ISWI complexes have also been linked to the DNA damage response, giving them an important role in integrity of DNA. By regulating the chromatin structure, ISWI complexes facilitate DNA access to repair proteins in at least four DNA repair pathways including homologous recombination, non-homologous end-joining, Base Excision Repair (BER) and nucleotide excision repair (NER). These complexes also regulate DNA repair by recruitment of DNA repair factors to DNA-damage sites (Aydin et al. 2014, Erdel & Rippe 2011). Involvement of ISWI complexes in Base Excision Repair (BER) has been shown for yeast ISW1 and ISW2 complexes *in vitro* (Nakanishi et al. 2007). BER is involved in repairing small nucleobase lesions in single DNA strands derived from alkylation or oxidation caused by endogenous chemicals (Kim & M Wilson III 2012). ISW1 and ISW2 complexes make damaged nucleosome core DNA accessible for DNA polymerase β (Nakanishi et al. 2007). Human WICH and ACF complexes facilitate transcription-coupled nucleotide excision repair (TC-NER) (Aydin et al. 2014). TC-NER removes different types of helix distorting DNA lesions in the template strand, which are induced by ultraviolet radiation and lead to RNA polymerase II (RNA Pol II) stalling during transcription elongation (Marteijn et al. 2014). It has been shown that human WICH and ACF complex components (SNF2H, ACF1 and WSTF) are necessary for successful binding of CSB (the protein responsible for TC-NER) to lesion-stalled RNA Pol II and recovery of transcription (Aydin et al. 2014). Another major DNA repair mechanism associated with ISWI chromatin remodeling is Double Strand Break (DSB) repair, which is accomplished by homologous recombination and non-homologous end-joining. Suppression of the expression of either ACF1 or SNF2H (components of ACF and CHRAC complexes) makes human cells extremely sensitive to X-ray and chemical inducing DNA double-strand breaks (DSBs) (Lan et al. 2010). WICH is involved in DSB repairs (Aydin et al. 2014). The WSTF protein (component of WICH complex) is recruited to DSBs and phosphorylates H2AX at the DNA damage site, which leads to recruitment of downstream factors involved in DSB repair (Xiao et al. 2009). Interestingly,

depletion of SNF2L (ATPase component of NURF and CERF complexes) leads to activation of the DNA damage response in cancer lines (Ye et al. 2009). Taken together, the evidence suggests that ISWI complexes play an important role in DNA repair and are involved in different DNA repair pathways by regulating the chromatin structure and recruitment of DNA repair factors to DNA-damage sites.

The Swi/Snf family, the Mi-2 / CHD family, and the Ino80 family have overlapping roles with the ISWI family in the biological processes. The Swi/Snf family has been connected to transcriptional repression and activation (Wilson & Roberts 2011). The Swi/Snf complexes in mammalian cells play an important role in transcriptional activation of genes that are located within structurally repressive chromatin regions (Ramirez-Carrozzi et al. 2009). The Swi/Snf family is also recruited to DSB regions indicating their role in DNA repair. In human cells BAF complexes interact with acetylated H3 of γ-H2AX nucleosomes by bromodomain of BRG1 (the catalytic component of BAF complex) (Lee et al. 2010). Investigating 282 DNA replication regions in HeLa cells showed that 90 (32%) co-occur with SWI/SNF region suggesting a role in replication for Swi/Snf complexes (Euskirchen et al. 2011).

The Mi-2/CHD complexes have been implicated in the activation and repression of transcription by acting at initiation, elongation and termination stage of the transcription process (Murawska & Brehm 2011). CDH1 has been also shown to be involved in replication in yeast (Biswas et al. 2008).

Ino80 family is involved in various biological processes. The yeast Ino80 complexes bind to replication origin and stalled replication forks indicating their role in replication (Shimada et al. 2008). The yeast Ino80 complexes regulate transcription of ~20% of genes negatively or positively. The yeast complexes interact with the phosphorylated H2AX during DNA double-strand break repair (Morrison et al. 2004).

The overlapping roles of the different ATP-dependent chromatin remodeling families in biological processes such as replication, transcription and DNA repair indicate the importance of the chromatin structure in the regulation of these processes.

1.6. ISWI targeting mechanism

ISWI complexes remodel chromatin by repositioning nucleosomes along the DNA while maintaining their intact structure and not disrupting the histone octamers. This type of nucleosome repositioning is called "nucleosome sliding" (Schwanbeck et al. 2004). ISWI complexes move the entire nucleosome on the DNA and as a result the nucleosomal DNA moves into the region between adjacent nucleosomes. Alternatively, other families of ATP-dependent chromatin remodellers such as SWI/SNF can also change histone composition and even eject histone octamers from the DNA (Clapier & Cairns 2009, Deindl et al. 2013, Gangaraju & Bartholomew 2007). According to the hypothesized "continuous sampling" mechanism, ISWI complexes continuously interact with nucleosomes in transient binding reactions without accomplishing any remodeling activity unless encountering a targeting signal, which then leads to an increased binding affinity to chromatin and consequent remodeling activity (Erdel et al. 2010). Chromatin signals that mark sites of high affinity binding and activity for ISWI complexes are classified into four groups: (1) specific DNA sequences, (2) histones and DNA sequences containing post-translational modifications, (3) nucleosomes containing different histone variants substituted for their canonical histones, (4) other chromatin associated proteins that recruit remodellers to the specific loci (Erdel & Rippe 2011). It has also been shown that non-coding RNAs have a significant role in recruiting chromatin remodellers to their target in the genome (Längst & Manelyte 2015).

1.6.1. Protein domains and targeting ISWI complexes to chromatin

Protein domains are evolutionary structural units that can function independent of the rest of the protein. The domains of a multidomain protein can also function in a combinatorial mode. These structural units are also used to determine the evolutionary relationships of proteins (Vogel et al. 2004). Chromatin signals that mark the binding targets for ISWI chromatin remodeling complexes are recognized by multiple structural domains identified in both the catalytic and non-catalytic components.

Protein domains in the catalytic components

Figure 1-2 shows the domains residing in the catalytic components of the ISWI complexes. ISWI proteins use these domains to bind to DNA and unmodified H4 tails (Längst & Manelyte 2015). As an example, Drosophila ISWI protein (the catalytic component of the ISWI

complexes) needs the presence of extranucleosomal DNA and the basic patch of the histone 4 tail to slide nucleosomes (Clapier et al. 2002, Schwanbeck et al. 2004). As can be seen in figure 1-2, there is an ATPase core domain closer to the N-terminus of the protein and this domain is highly conserved in eukaryotes. The ATPase domain consists of two tandem RecA-like folds (DExx and HELICc) that are characteristic of proteins belonging to the helicase-like superfamily 2 (SF2) (Flaus 2006). In addition to the ATPase domain, which has autonomous nucleosome remodeling activity (Clapier & Cairns 2012), there are other domains within ISWI proteins that bind nucleosomal epitopes and have regulatory functions. The C-terminal half of the ISWI proteins contains a HAND-SANT-SLIDE (HSS) domain which binds the DNA fragment between nucleosomes (extranucleosomal DNA) (Clapier & Cairns 2012). The AutoN and NegC domains are in a close functional interaction with ATPase core domain. The AutoN domain is located close to N-terminal end of ATPase domain and the NegC domain is located in the C-terminal end between the ATPase core domain and the HSS domain (Clapier & Cairns 2012). The AutoN domain inhibits the intrinsic DNA-dependent ATPase activity of ISWI and the presence of the H4 tail removes the inhibition and activates the ATPase activity of the ISWI protein. Mutation of the AutoN domain of ISWI results in the removal of its inhibition and activates the ATPase activity of ISWI independent of the H4 tail. But the ISWI protein with mutated AutoN still needs H4 tail to slide the nucleosome, which is the result of the NegC domain inhibitory function. The NegC domain inhibits the coupling of the ATPase activity of the ATPase core domain to DNA translocation, which in turn prevents sliding activity. The ATPase/translocation uncoupling caused by the NegC domain is removed when the HSS domain binds the extranucleosomal DNA. With deletion of the HSS domain, the ATPase/translocation uncoupling caused by the NegC domain is not removed and the sliding activity is prevented (Clapier & Cairns 2012, Hwang et al. 2014). The combinatorial function of the ATPase domain core and the flanking domains during nucleosome sliding is explained by the "inhibition of inhibition" model. According to this model, the AutoN and NegC domains inhibit the intrinsic DNA translocase activity of the ATPase domain where AutoN inhibits its ATPase activity and NegC inhibits its coupling to DNA translocation. The HSS domain binding to extranucleosomal DNA with sufficient length relieves the NegC inhibition and the presence of the unmodified H4 tail removes the AutoN inhibition resulting in chromatin remodeling activity of the ISWI protein (Clapier & Cairns 2012). The 'inhibition of inhibition' model of ISWI regulation has been used to explain how the human ACF

complex (consisting of SNF2H and ACF1) performs chromatin remodeling. By binding to the unmodified H4 tail, ACF1 causes the inhibition of the ATPase activity of SNF2H. When it is near a longer extranucleosomal DNA, ACF1 preferentially binds to the linker DNA. The release of the H4 tail leads to removal of the AutoN inhibition and activation of chromatin remodeling. Sensing the length of the extranucleosomal DNA by ISWI complexes is important for uniform nucleosome spacing during heterochromatin formation, whereas histone tails have an important role in marking the chromatin regions to be silenced by chromatin remodeling activity of ISWI complexes (Hwang et al. 2014). This is consistent with the fact that acetylated histone H4 tails are characteristic of the promoter regions of active genes (Gonzales-Cope et al. 2016, Goudarzi et al. 2016). There is also a specific nuclear localization signal (NLS) at the very C-terminal end of the ISWI family proteins with an important role in their nuclear localization (Figure 1-3) (Vasicova et al. 2013).

Domains in non-catalytic components

The ISWI complexes are multi-protein complexes, and in addition to an ISWI-type ATPase, they contain non-catalytic components. While the ATPase component performs the catalytic function of the complex, these non-catalytic associated components are important in determining and regulating the function of ISWI complexes (Stanne et al. 2015). The importance of the non-catalytic components has been shown for all ATP-dependent chromatin remodeling families. The non-catalytic proteins are involved in the modulation of nucleosome remodeling activity of the ATPase and also have a critical role in targeting the complex to specific chromatin regions. IES2 (Ino Eighty Subunit 2) activates the ATPase activity of the Ino80 ATPase of the Ino80 complex in human cells and IES6 (Ino Eighty Subunit 6) and ARP5, two other components, are involved in binding the complex to nucleosomes (Chen et al. 2013a). The presences of the tissue-specific non-catalytic components results in the unique characteristics of these complexes in each tissue. These components are important for binding the complexes to tissue-specific genomic loci. For instance, isoforms of BAF60 can be exchanged in BAF complexes that leads to tissue-specificity of the Swi/Snf complexes (Forcales et al. 2012). BAF proteins are components of the BAF complexes. The BAF60C isoform is a component of this complex in mammalian myoblasts (Forcales et al. 2012) and by binding to muscle genes marks the binding sites of the BAF core complex. Whereas, the embryonic stem cell BAF (esBAF) complex contains the BAF60A/B isoforms and BAF170 (Vogel-Ciernia & Wood 2014) and it

plays an important role in self-renewal and pluripotency of the ES cells (Ho et al. 2009a). A good example on the effect of the non-catalytic components on the function of ISWI complex has been reported for *Saccharomyces cerevisiae*. The catalytic protein Isw1 in combination with 3 non-catalytic proteins forms 2 distinct complexes with different functions. Isw1a complex consists of Isw1 and Ioc3, whereas Isw1b complex consists of Isw1 and Ioc2 and Ioc4 (Stanne et al. 2015). The non-catalytic components, Ioc2, Ioc3 and Ioc4, are not related proteins and contain different protein domains (Vary et al. 2003). First, there is a difference in the nucleosome spacing and sliding activities of these two complexes (Krajewski 2013, Vary et al. 2003). Isw1a chromatin remodeling activity leads to uniformly spaced nucleosome array, whereas, Isw1b activity exhibits little spacing activity. Secondly, different phenotypes have been detected after the deletion of components of the Isw1a and Isw1b complexes (Vary et al. 2003). Isw1a also blocks transcription initiation of genes by removing the basal transcription machinery whereas Isw1b regulates the elongation and termination steps of transcription, which activates gene expression (Mellor & Morillon 2004). Interestingly, the Isw1b complex, but not the Isw1a complex, binds to H3K36me3 at the midregions and 3' ends of highly transcribed genes (Maltby et al. 2012). The PWWP (Pro-Trp-Trp-Pro) domain of Ioc4, the non-catalytic component of the complex, mediates the binding of the Isw1b complex to the methylated H3K36. The mutation in the Ioc4 PWWP domain disrupts the interaction of the complex with methylated H3K36 *in vitro*. This result shows the importance of the non-catalytic components of the ATP-remodeling complexes in determining its function. Another example of the functional difference caused by the non-catalytic components is the difference in the directionality of the nucleosome movement *in vitro* between the *Drosophila* ISWI ATPase protein when it is part of the CHRAC or ACF complexes. When isolated, the ISWI protein slides nucleosomes located in the centre of a DNA fragment into the end of a fragment. Conversely, when combined with non-catalytic proteins within CHRAC and ACF complexes, ISWI ATPase activity results in sliding nucleosomes from the end of the DNA segment towards the centre (Längst et al. 1999, Saha et al. 2006). The *Drosophila* CHRAC complex contains two more non-catalytic components, CHRAC14 and CHRAC16, in comparison to the ACF complex (Figure 1-3). The presence of the CHRAC14-CHRAC16 heterodimer results in increased nucleosome sliding activity in the CHRAC complex compared to the ACF complex in *Drosophila* (Hartlepp et al. 2005). The authors suggested that the CHRAC14-CHRAC16 dimer functions as a DNA chaperone that provides a DNA binding surface that leads to the disruption

of DNA and histone interactions during the remodeling activity. A similar function has been reported for human homologues of these proteins where CHRAC15 and CHRAC17 facilitate the nucleosome sliding activity of ACF (Kukimoto et al. 2004).

ISWI ATPase in *Drosophila melanogaster* forms six identified ISWI complexes (CHRAC, ACF, NURF, RSF, ToRC, and NoRC) in combination with one or more of nine known non-catalytic proteins (Figure 1-3). Although these complexes share the same catalytic component, they have distinct functions indicating the importance of the effect of non-catalytic components. In mammals, eight ATP-dependent chromatin remodeling complexes have been identified belonging to the ISWI family. In these complexes in addition to the catalytic proteins, SNF2L and SNF2H, multiple non- catalytic proteins have been detected which affect the function of the ATPase protein and also play an important role in targeting the complexes to specific genomic regions (Stanne et al. 2015, Thompson et al. 2012). Specialized large non-catalytic proteins contain many domains that explain their specific function in the context of the complex during chromatin remodeling, especially their affinity to their target sites. Interestingly, the domain structure of these non-catalytic proteins is very similar. Three of the large proteins that bind to SNF2H have been classified as the BAZ (bromodomain adjacent zinc finger) family of proteins including ACF1 (component of the ACF and CHRAC complexes), WSTF (component of the WICH complex) and TIP5 (component of the NoRC complex) (Jones et al. 2000). The BAZ family of proteins share multiple conserved motifs including a C-terminal bromodomain, an adjacent PHD (Plant homeodomain) finger, WAC (WSTF, Acf1, cbp146p), WAKZ (WSTF, Acf1, KIAA0314, ZK783.4), DDT (DNA binding homeobox and Different Transcription factors) and BAZ1 (bromodomain adjacent zinc finger) and BAZ2 (Figure 1-4) (Jones et al. 2000). Interestingly, proteins containing some similar motif organization to the BAZ family of proteins are associated with ISWI homolog proteins in various species including *Drosophila melanogaster*, mammals, Xenopus and *Saccharomyces cerevisiae* (Vary et al. 2003). BPTF, one of the non-catalytic components of the NURF complex, shares the domains PHD, DDT and Bromodomain with the BAZ family proteins (Tallant et al. 2015). RSF-1, the non-catalytic protein partner of SNF2H in the RSF complex, shares DDT and PHD finger domains with these proteins (Pessina & Lowndes 2014). The CERF component CECR2 also shares the DDT domain and bromodomain with these proteins (Banting et al. 2005). I will refer to these non-catalytic components that contain similar structural domain architecture as BAZ-like proteins. These

domains play important roles in mediating interaction with other components in the complexes and also in recognizing the targeting signals including post-translational modifications of histones, histone variants, DNA sequence/structure and RNA molecules (Längst & Manelyte 2015).

One of the domains found in all of the BAZ-like components of ISWI complexes is the DDT (DNA binding homeobox and Different Transcription factors) domain (Figure 1-4). The DDT domain was discovered in a search for a DNA-binding motif for BPTF and the BAZ family of proteins using a bioinformatics method (Doerks et al. 2001). The DNA-binding function of the
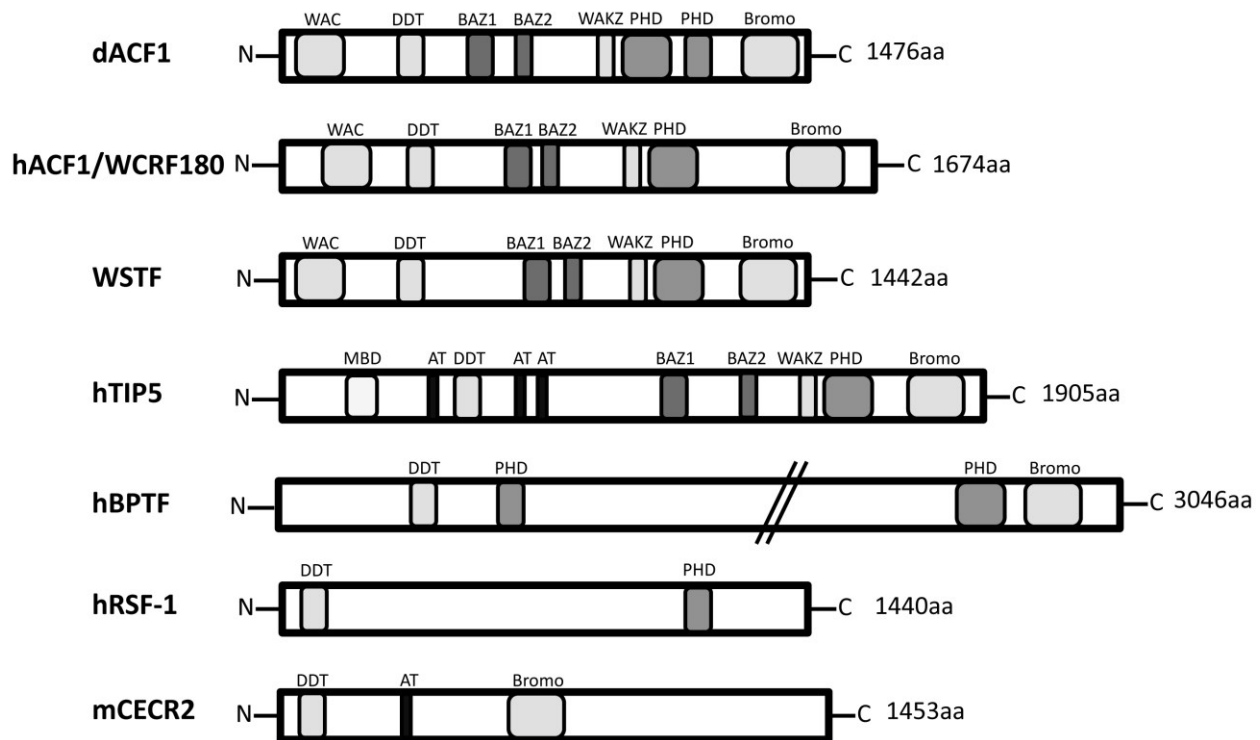
Figure 1-4 Schematic representation of the domains within the large non-catalytic components of the ISWI family of ATP-chromatin remodeling complexes. The double slashes across the hBPTF indicate that the size is not proportional to the other proteins. (Banting 2004, Barnett & Krebs 2011, Eberharter et al. 2004, Hwang et al. 2014, Jones et al. 2000, Pessina & Lowndes 2014, Tallant et al. 2015, Toto et al. 2014)

DDT domain has been shown for Swi1 in *Schizosaccharomyces pombe*. The region containing the DDT domain of Swi1 binds to DNA in vitro and the mutations in this domain blocks the interaction of Swi1 with DNA (Noguchi et al. 2012). Interestingly, the DDT domain of Swi1 is

also involved in protein-protein interaction and the Swi1 protein with mutated residues in its DDT domain loses its ability to form a complex with Swi3. Other studies have shown the importance of the DDT domain in mediating protein-protein interactions. The ACF1 protein (component of the ACF complex in Drosophila) with a deleted DDT domain does not form a complex with ISWI protein (Eberharter et al. 2004). The SLIDE domain of CHR11 (*Arabidopsis thaliana* ISWI protein) and human SNF2H interact with the DDT domain of DDT- domain-containing proteins of *Arabidopsis thaliana* (Dong et al. 2013) suggesting that the interaction between the SLIDE domain of the ISWI proteins and the DDT domain is conserved in eukaryotes.

The DNA-binding motifs within both the catalytic and BAZ-like proteins of the remodeling complexes recognize DNA sequences and conformation. WAC motifs (found in ACF1, WSTF) or AT hooks (found on TIP5 and CECR2) recognize DNA sequences and affect the outcome of the remodeling activity (Längst & Manelyte 2015). Deletion of the WAC domain of Drosophila Acf1 leads to a significant reduction in the amount of DNA-bound ACF complexes and consequently a reduction in ATPase activity of the ACF complex (Fyodorov & Kadonaga 2002). AT hooks are small protein motifs found in many DNA-binding proteins in various organisms and bind the minor groove of AT rich DNA (Aravind & Landsman 1998).

Some of the domains are able to recognize post-translational modifications on histone tails and as a result are important in anchoring the chromatin remodeling complexes on their specific binding sites. These domains are called epigenetic readers and are found in the non-catalytic components of the ATP-dependent chromatin remodeling complexes. All of the BAZ-like proteins of ISWI complexes except for RSF-1 contain bromodomains (Figure 1-4). Bromodomains preferentially bind acetylated lysines on the histone tails (Taverna et al. 2007). The bromodomain of TIP5 interacts with acetylated H4K16 and acetylated H3K14. A point mutation inside the TIP5 bromodomain negatively affects the association of the NoRC complex with chromatin and its remodeling function (Tallant et al. 2015, Zhou & Grummt 2005). The PHD finger, which is found in all the BAZ-like proteins except for CECR2 recognizes and binds trimethylated H3K4, H3K9 and H3K36 and unmodified H3K4 histone tails (Tallant et al. 2015). The presence of multiple domains and their ability to recognize histone marks on one hand and their cooperative functions on the other hand indicates the complexity of targeting these chromatin remodeling complexes.

1.6.2. Targeting by Sequence Specific Binding Proteins

Chromatin signals that are recognized by domains in catalytic and non-catalytic components of chromatin remodellers include specific DNA sequences, histones/DNA sequences containing post-translational modifications and nucleosomes containing non-canonical histone variants. In addition to protein components of chromatin remodeling complexes depicted in Figure 1-3, there are other chromatin-binding proteins that transiently interact with these complexes to recruit them to specific DNA sequences. SIRT6 is one of the factors recruited to DNA damage sites in the very early stages of repair (Toiber et al. 2013). Human SNF2H-containing complexes are recruited to the DNA damage site by their interaction with SIRT6 protein *in vivo*. SIRT6 is able to significantly increase binding affinity of SNF2H to nucleosomes *in vitro*, without affecting its chromatin remodeling activity (Toiber et al. 2013). NuMA is a protein that interacts with the WICH complex during DNA double-strand break repair (Vidi et al. 2014). A significant increase in the physical interaction between SNF2H and NuMA has been detected using FRET following exposure of the cells to gamma irradiation. Laser microirradiation causes DNA damage and leads to accumulation of NuMA at the damage sites. Knock down of NuMA using siRNA reduces SNF2H recruitment to laser microirradiation-induced DNA damage sites and leads to a reduced chromatin decompaction after DNA damage. These results show that the SNF2H complex is targeted to damaged site as a result of its interaction with NuMA.

1.6.3. Targeting to the binding sites by RNA

Non-coding RNA molecules have been also linked to ATP-dependent chromatin remodeling by their targeting function to recruit chromatin remodellers to specific genomic loci (Han & Chang 2015). An example is promoter RNA (pRNA), which is an intergenic non-coding RNA with a short life-span (Mayer et al. 2006). It is complementary in sequence to the rDNA promoter and binds the rDNA promoter. The depletion of pRNA leads to displacement of the NoRC complex from rDNA loci. This results in removal of the transcriptional repression of rDNA genes established by remodeling activity of NoRC (Mayer et al. 2006). The NoRC complex binds to pRNA by the interaction between TIP5 and a specific stem-loop structure within pRNA (Mayer et al. 2008). Disrupting the stem-loop structure of pRNA, which is conserved across mammals, negatively affects the binding of TIP5 and results in translocation of the NoRC complex from the nucleoli (Mayer et al. 2008). These data indicate that RNA

molecules can control chromatin structure by recruiting ATP-chromatin remodeling complexes to specific target sites.

In addition to targeting functions of RNAs, inhibition of ATP-dependent chromatin remodeling has been also reported for them. *Mhrt* RNA (cardiomyocyte-specific non-coding RNA) binds to the ATPase domain of BRG1 (the catalytic component of BAF complex) with high affinity and prevents its binding to its genomic targets and consequently inhibits its chromatin remodeling activity (Han et al. 2014).

1.7. Developmental importance of BAZ-like proteins

BAZ-like proteins of the ISWI chromatin remodeling complexes can play an important role during development. A loss of function mutation in *Drosophila Acf1* results in the death of 75% of offspring in the third larval stage because of a failure in larvae-pupae transition and delayed development (Fyodorov et al. 2003). The 25% surviving mutants develop to adulthood and do not show any detectable defects (Fyodorov et al. 2003). ACF1 depletion in germline cells by cell type-specific RNA interference in early phases of oogenesis results in defects in egg chambers, which are subsequently removed by apoptosis (Börner et al. 2016). The same phenotype has been seen in flies with ISWI protein depletion in adult germline cells with much higher penetrance indicating that the phenotype is caused by compromising the function of the ISWI complex (Börner et al. 2016). Mammalian ACF1 (also known as BAZ1A) forms 2 complexes (ACF and CHRAC) in combination with SNF2H. Deletion of ACF1 does not affect viability of mice (Dowdle et al. 2013). However, lack of ACF1 leads to abnormal spermiogenesis causing infertility in males. Mutant phenotypes include oligospermia, lack of motility (asthenospermia), abnormal sperm morphologies, multinucleate spermatids and the complete absence of sperm in the cauda epididymides. Intracytoplasmic injection using the ACF1 mutant sperm showed that they were also not capable of activating oocytes (Dowdle et al. 2013). The BAZ-like proteins of the NURF complex are NURF301 and BPTF in *Drosophila* and mammals, respectively (Figure 1-3). NURF301 is needed for the assembly of the NURF complex and NURF55 and NURF38 (other components of the NURF complex) bind ISWI through their binding to NURF301 (Xiao et al. 2001). NURF301 loss of function in *Drosophila* results in developmental delay and the larvae continue to survive without transforming to pupae

(Badenhorst et al. 2005). With another mutation that leads to expression of truncated NURF301 lacking the N-terminal end of the protein (containing the PHD and bromodomain), larvae successfully form pupae and reach adulthood but show developmental abnormalities and infertility phenotypes (Xiao et al. 2001). In the mammalian NURF complex, all the homozygous BPTF mutants die during fetal development between E7.5 and E8.5 indicating that BPTF is essential for normal development in mice (Goller et al. 2008, Landry et al. 2008). The *Drosophila* ortholog of TIP5, Toutatis, is a component of the NoRC and ToRC complexes (Emelyanov et al. 2012, Fauvarque et al. 2001). Depletion of Toutatis leads to wing and neural defects in Drosophila (Emelyanov et al. 2012, Vanolst 2005). Lack of TIP5 (also known as BAZ2A) in mammalian cell lines leads to upregulation of rDNA transcription and an increase in ribosome production and cell proliferation (Guetg et al. 2010). WSTF also known as BAZ1B, is the BAZ-like protein of the WICH complex in vertebrates (Bozhenok et al. 2002). This protein is the product of the WBSCR9 gene, one of the ~20 genes deleted in Williams Syndrome in human (Barnett & Krebs 2011). Patients with Williams Syndrome may exhibit intellectual deficits, growth deficiency, cardiovascular disease and facial dysmorphology (Martens et al. 2008). The *MommeD10* mutation resulted in an amino acid change, L733R, within the WSTF protein (Ashe et al. 2008). Western blot analysis showed reduced levels of the protein in mutant mice caused by the instability of the protein product. Heterozygous intercrosses produced 14 homozygotes, 187 heterozygotes and 108 wild-types reaching the weaning age (3 week old) indicating the reduced viability of the homozygous *MommeD10* mutants. The *MommeD10* mutation causes death in the first week after birth. Analysis of heterozygote mice indicated that WSTF has a critical role in craniofacial development as they exhibit protruding foreheads, shorter snouts and flattened nasal bone. The similarity of these phenotypes to those of Williams Syndrome patients suggests that reduced expression of WSTF may be responsible for facial dysmorphology. Investigating Rsf1 showed that knockout of *Rsf1* in flies only causes lethality in 10% of the homozygous pupae with a melanotic tumor phenotype (Hanai et al. 2008). The surviving mutant flies develop to adulthood and do not show any abnormalities. RSF1 in human cells has a role in DNA damage response since knockdown of RSF1 in U2OS cells leads to increased sensitivity of the mutant cells to Methyl methanesulfonate (MMS) and ionizing radiation (IR) (Pessina & Lowndes 2014).

Taken together, loss of BAZ-like proteins shows variability. Some are lethal; some appear to be dispensable and only produce mild phenotypes.

1.8. Biological and functional roles of CECR2

CECR2 (cat eye syndrome critical region, candidate 2) was identified in the human cat eye syndrome critical region (CESCR) on chromosome 22q (Footz et al. 2001). Duplication and triplication of the CESCR causes cat eye syndrome in human. The CESCR contains 14 genes, of which 10 genes are present in the homologous region on chromosome 6 in mouse including CECR2 (Footz et al. 2001). Associated symptoms of cat eye syndrome include ocular coloboma, heart, anal and kidney defects and mental retardation (Schinzel et al. 1981). The human and mouse *Cecr2* genes consist of 19 exons and produce 1464 and 1453 amino acid respectively. Both human and mouse CECR2 proteins have a very similar structural domain organization to the BAZ-like proteins (ACF1, WSTF, TIP5, BPTF and RSF1) (Figure 1-4). CECR2 shares the DDT, AT hook and bromodomain with these proteins (Banting et al. 2005). CECR2 is also present in *Drosophila melanogaster* and *Xenopus laevis*, although its function is not clear.

1.8.1. *Cecr2* mutations

There are two mouse mutations that have been used in the study of *Cecr2*. One is a genetrap mutation named *Cecr2^{Gt45Bic}*. In this mutation a β-splicetrap vector, pGT1, is located in intron 7 of *Cecr2*. When the gene is expressed, exon 7 is spliced to a βgeo cassette, which produces CECR2-βgalactosidase-neomycin-phosphotransferase fusion polypeptide (Figure 1-5) (Banting et al. 2005). This splicing event removes exons 8-19 of *Cecr2* from the transcript. The removed fragment codes for 1163 of the 1453 amino acids of the wild-type CECR2 protein. The CECR2^{Gt45Bic} fusion protein is detectable by X-gal staining and was used to examine the expression of *Cecr2* in mouse (Banting et al. 2004, Dawe et al. 2011, Thompson et al. 2012). The fusion protein is also detectable by Western blot analysis using an antibody against β-galactosidase, making biochemical analysis possible by studying the CECR2 fusion protein (Thompson et al. 2012). However, using a fusion protein may lead to erroneous conclusions if the fusion protein acts differently from the wildtype protein. *Cecr2^{Gt45Bic}* homozygous mutant embryos show an ~14-fold reduction in wild-type *Cecr2* mRNA as judged by qRT-PCR, suggesting this allele is a hypomorph and that some normal *Cecr2* transcripts are produced through splicing around the βgeo cassette (Fairbridge et al. 2010). The removed C-terminal portion of the CECR2 protein contains the conserved bromodomain, which means that the CECR2^{Gt45Bic} fusion protein should lack function. The loss of the bromodomain in the

CECR2[Gt45Bic] fusion protein could affect the targeting function of CECR2 in the CERF complex. It has been shown that a point mutation inside the bromodomain of TIP5 reduces the binding affinity of the complex to chromatin (Kumar et al. 2016, Ruthenburg et al. 2011, Tallant et al. 2015, Zhou & Grummt 2005). The CECR2[Gt45Bic] fusion protein still contains the AT hook, the DDT domain and the nuclear localization signal (NLS), which suggests that the fusion protein still can localize in the nucleus and form protein-protein interactions with other components of the CERF complex. In fact, it has been shown that the CECR2[Gt45Bic] fusion protein is able to localize to the nucleus and successfully interact with SNF2L and SNF2H in human and mouse cells (Banting et al. 2005, Tate et al. 1998, Thompson et al. 2012). Together this information suggests despite the protein exhibiting some wild-type properties, the *Cecr2*[Gt45Bic] mutation likely disrupts some CECR2 functions, because it causes mutant phenotypes.

Figure 1-5 The gene structure of wildtype *Cecr2* and mutations *Cecr2*[Gt45Bic] and *Cecr2*[tm1.1Hemc]. Cecr2[Gt45Bic] is a gene trap mutation located in intron 7. A βgeo cassette is spliced to exon 7 and leads to a CECR2-βgalactosidase-neomycin phosphotransferase fusion polypeptide (CECR2[Gt]). In the *Cecr2*[tm1.1Hemc] mutation (Deletion) exon 1 and ~1 kb upstream have been deleted, resulting in a loss of *Cecr2* expression in the *Cecr2*[tm1.1Hemc] mutants. Part of the DDT domain is also deleted in this mutation.

The second mutation, *Cecr2^tm1.1Hemc*, is a presumptive loss-of-function mutation in which exon 1 and 1 kb upstream of it have been deleted through LoxP-Cre recombination (Fairbridge et al. 2010). This mutation removes the start codon, half of the DDT domain located in exon 1, the promoter and probably the nearby regulatory regions. qRT-PCR shows an ~200 fold reduction in *Cecr2* expression in homozygous mutant embryo heads (Fairbridge et al. 2010). The increased severity of phenotype (~100% penetrance of exencephaly) caused by the *Cecr2^tm1.1Hemc* mutation compared to the *Cecr2^Gt45Bic* mutation suggests that *Cecr2^tm1.1Hemc* is probably a null mutation (Dawe et al. 2011, Fairbridge et al. 2010, Thompson et al. 2012). Lack of a working CECR2 antibody prevented the lab from investigating the *Cecr2* expression in *Cecr2^tm1.1Hemc* mutants at the protein level.

1.8.2. CECR2 mutant phenotypes

Homozygous *Cecr2^Gt45Bic* mutants exhibit exencephaly at 74% penetrance (Banting et al. 2005) resulting in perinatal death of the embryos in the BALB/c strain. Over time the penetrance has dropped to 54% in the BALB/c congenic line (Leduc et al. 2016, in press). Exencephaly is caused by the failure of neural tube closure during embryogenesis. Because of the absence of the skull vault, the exencephalic brain is exposed to the amniotic fluid and undergos degeneration later in the development (Copp et al. 2003). *Cecr2* genetrap mutants also have open eyelids, which are also seen in some of the homozygotes (Banting et al. 2005). Smaller cochleae and disorganization of sensory cells in the inner ear of *Cecr2* genetrap mutants also suggests that the *Cecr2* mutation may affect the planar cell polarity (PCP) pathway, although qRT-PCR analysis of 12 PCP genes showed no change in transcript abundance during neural tube closure (Dawe et al. 2011). There is also evidence indicating a genetic interaction between *Cecr2* and *Vangl2*, making the PCP pathway involvement controversial. Since the testis is an adult tissue with strong expression of *Cecr2*, fertility was examined in non-penetrant BALB/c *Cecr2^Gt45Bic* mutants, revealing that the homozygous adults have smaller testes compared to wild-type animals (Thompson et al. 2012). The mutant males are subfertile, producing smaller litters than their wild-type siblings. The subfertility is caused by the compromised ability of mutant sperm to successfully fertilize wild-type oocytes (Thompson et al. 2012).

The *Cecr2^tm1.1Hemc* deletion mutation produces the same phenotypes with increased severity compared to the genetrap mutation, as expected. The *Cecr2^tm1.1Hemc* mutation leads to the

same neural tube defect, exencephaly, with increased penetrance to ~100% on BALB/c background (Fairbridge et al. 2010). The lower penetrance of the $Cecr2^{Gt45Bic}$ allele compared to $Cecr2^{tm1.1Hemc}$ can be explained by the presence of wild-type $Cecr2$ expression in the $Cecr2^{Gt45Bic}$ mutants, presumably due to limited splicing around the genetrap. The $Cecr2^{tm1.1Hemc}$ mutants (BALB/c background) also exhibit disorganization of sensory cells in the inner ear (Dawe et al. 2011). Because of the perinatal death of all $Cecr2^{tm1.1Hemc}$ BALB/c mutants, there is no data about the fertility of these mutants.

qRT-PCR analysis showed that the genetrap mutation leads to downregulation of four genes (*Alx1*, *Dlx5*, *Ncapd2*, and *Six1*) in the head region of the 10-14 somite stage in homozygous BALB/c embryos (Fairbridge et al. 2010). Transcripts of three genes (*Epha7*, *Eya1*, and *Lix1*) also showed significant downregulation in the head region of the 18-20 somite stage (after cranial neural tube closure). From these genes *Alx1*, *Dlx5* and *Epha7* result in exencephaly when mutated. *Alx1* and *Dlx5* are mesenchymal/ectodermal transcription factors suggesting that *Cecr2* can affect the downstream genes regulated by these transcription factors.

Interestingly, the exencephaly caused by a *Cecr2* mutation is strain dependent. While the $Cecr2^{tm1.1Hemc}$ mutation leads to exencephaly in 54% of the embryos on the BALB/c strain, it does not result in exencephaly when congenic on the FVB/N strain, suggesting the presence of modifier genes that affect susceptibility to exencephaly (Kooistra et al. 2012).

1.8.3. CECR2 expression pattern

X-gal staining of the CECR2^{Gt45Bic} fusion protein in mutant mice during E10.5-E14.5 showed a weak expression of *Cecr2* throughout the embryo with a strong expression in the central nervous system and adjacent spinal ganglia, the limb mesenchyme, nasal epithelium, the lens and neuroretina of the forming eye and in the inter-costal mesenchyme (Banting et al. 2005). There was not detectable *Cecr2* expression in the heart and liver by Northern blot analysis. The other fetal tissues expressing the CECR2^{Gt45Bic} fusion protein include the foregut/midgut at 5 somite stage, hindgut, the pharyngeal region of the foregut, the first branchial arches and the cochlear floor with the strongest expression being in the regions with the most immature cochlear hair cells at the 16–17 somite stage (Dawe et al. 2011). The expression decreases closer to birth in almost all tissues. One of the tissues with persistent expression of the CECR2^{Gt45Bic} fusion protein during adulthood is the testis. In developing testis, the fusion protein is detectable as early

as E16.6 embryos when the sex cords contain differentiating gonocytes. The expression continues to the adult testis with the highest expression of the fusion protein in spermatogonia and weaker expression in spermatocytes (Dawe et al. 2011). Western blot analysis using an anti-β-gal antibody against the fusion protein also confirmed the expression of *Cecr2* in adult testis (Thompson et al. 2012). Analysis of chicken CECR2 showed that it consists of 1473 amino acids with a conserved AT hook and a bromodomain (Chen et al. 2010). The chicken *Cecr2* has a very similar expression pattern to mouse *Cecr2*. It is predominantly expressed in the neural fold and neural tube during neurulation. There is also strong expression of *Cecr2* in the developing somites and the spinal cord at later developmental stages. Similar to mouse *Cecr2*, there is a noticeable expression of *Cecr2* in the limbs during the embryonic stages. The mesonephric duct and pharyngeal arches also express *Cecr2* in developing chicken embryos.

1.9. CECR2-containing complexes

The CECR2-containing complex was first isolated from human embryonic kidney cells (HEK293) in which SNF2L binds to CECR2 to form CECR2-containing remodeling factor (CERF) (Banting et al. 2005). Similar to NURF, another ISWI complex, CERF exhibits ATPase activity in the presence of nucleosomes and ATP-dependent chromatin remodeling activity *in vitro* (Banting et al. 2005, Barak et al. 2003). CERF and NURF both contain SNF2L as their catalytic ATPase. As reviewed in Wiechens et al., 2016, other mammalian ISWI complexes (ACF, CHRAC, WICH, RSF and NoRC) contain SNF2H as the ATPase component. There has been no evidence of interaction between CECR2 and SNF2H in human cells so far, however CECR2 biochemical analyses using the CECR2[Gt45Bic] fusion protein showed interaction between CECR2[Gt45Bic] and SNF2H in mouse ES cells and adult mouse testis (Thompson et al. 2012). An antibody against β-galactosidase was able to immunoprecipitate a protein complex containing CECR2[Gt45Bic] and SNF2H. The reciprocal approach using an antibody for SNF2H was able to immunoprecipitate the same complex in mouse ES cells and the adult mouse testis. The fact that the anti-β-galactosidase antibody isolated complex did not contain other SNF2H associated proteins including ACF1, WSTF, RSF1 or TIP5 indicated that CECR2[Gt45Bic] interacts with SNF2H in an independent complex. The study indicates that using the same technique they were not able to show an interaction between CECR2[Gt45Bic] and SNF2L and concluded that

CECR2$^{\text{Gt45Bic}}$ probably preferentially binds to SNF2H rather than SNF2L in both the adult testis and ES cells (Thompson et al. 2012). The human CERF complex isolated from HEK293 cells has a size of approximately 0.6 MDa. Interestingly, CECR2$^{\text{Gt45Bic}}$-containing complexes isolated from mouse ES cells and adult testes have different sizes. The ES cell complex is ~300-400 kDa, consistent with the size of a CECR2/SNF2H heterodimer, whereas the testis complex size is ~0.9-1 MDa (Thompson et al. 2012). The size of CECR2$^{\text{Gt45Bic}}$-containing complexes isolated from adult mouse testis suggests that these complexes may have included other unidentified components or they consisted of repeats of heterodimers of CECR2$^{\text{Gt45Bic}}$ and SNF2H. The evidence indicates that there must be a tissue-specific CERF. I refer here to tissue-specific CERF complexes by adding letters the (first letter of the tissue) in front of CERF: tCERF for the testis comple, esCERF for ES cell complex and nCERF for neural stem cells CERF.

The size of CECR2$^{\text{Gt45Bic}}$-containing complexes in adult mouse testis indicates that there might be other binding components present in the complex. Other ISWI chromatin remodeling complexes with more than two components have previously been isolated from mammalian tissues. Human NURF, for instance, in addition to SNF2L and BPTF contains two more components, RbAP48 and RbAP46 (Barak et al. 2003). CHRAC is another ATP-chromatin remodeling complex with four binding components including SNF2H, ACF1, CHRAC15 and CHRAC17 (Poot et al. 2000). The WICH complex interaction with several nuclear proteins during active transcription leads to a 3 MDa protein complex called B-WICH with a role in transcription (Cavellan et al. 2006). The ISWI complex, ISWI-D, has been isolated from *Xenopus laevis* containing ISWI and 5 unidentified proteins with the molecular weights of 200, 135, 70, 55 and 17 kDa (Guschin et al. 2000). Understanding the function of CECR2-containing complexes in mammals will be aided by finding additional CERF componets in mouse.

1.10. Genome-wide analysis of the ATP-dependent chromatin remodellers

The ATP-dependent chromatin remodeling complexes bind DNA and nucleosomes by recognizing histone variants, histone tail post-translational modifications and specific DNA sequences. These complexes play an important role in positioning of the nucleosomes within the genome (Narlikar et al. 2013), which affects gene expression directly and indirectly. Chromatin immunoprecipitation with massively parallel DNA sequencing is used to identify the binding

sites of these complexes to identify the genes that are directly regulated by them. The genome-wide binding of different chromatin remodellers in yeast showed their distinct contribution to nucleosome positioning (Yen et al. 2012). Ino80 (the catalytic component of the Ino80 complex), Isw1, and Isw2 (the catalytic components of the ISWI complexes) proteins bind many nucleosomes within genic arrays and share many target genes, suggesting that these complexes function together. Although Ioc3 and Ioc4 (the non-catalytic components) share many binding targets with Isw1, their binding sites are largely distinct, indicating that the ISW1a and ISW1b complexes bind distinct sets of genes (Yen et al. 2012). This is consistent with the expression profile of the mutants of the components of the ISW1a and ISW1b complexes (Vary et al. 2003). Northern blotting and DNA microarray analyses showed that the ioc2 and ioc3 double mutant yeast strain has a transcriptional profile similar to the isw1 single mutant as expected, while the ioc2 and ioc3 single mutants show distinct transcriptional profiles indicating independent functions of the ISW1a and ISW1b complexes (Vary et al. 2003). The motif discovery analysis showed the Reb1 (a DNA-binding protein) motif in and around Isw2-enriched regions suggesting that both Reb1 and ISW2 bind the same set of genes close to their transcriptional start sites. Reb1 is involved in organizing nucleosomes, suggesting that ISW2 and Reb1 function together in positioning the nucleosomes at the TSS of the genes. The majority of binding targets of the ISWI and SWI/SNF do not overlap indicating the distinct function of these two complexes as well. While the ISWI complexes function within genic regions, the RSC and SWI/SNF complexes bind different genes and are limited to the 5' end of genes (Yen et al. 2012). ChIP-seq analysis also showed that ISW1a occupies both non-coding RNA and tRNA genes in yeast (Parnell et al. 2015). It has also been shown that Isw2 binds to upstream of tRNAs (Gelbart et al. 2005). Interestingly, the chromatin remodeling of the tDNA genes do not affect the transcription of the tRNA suggesting that it is involved in another process than regulating the tRNA gene transcription. Further investigation showed that the Isw2 binding is needed for appropriate Ty1 elements integration upstream of tDNAs. In *Drosophila*, loss of ISWI results in defects in global transcription and dramatic chromosome condensation defects, which could be the direct or indirect effect of the protein. Genome-wide identification of ISWI chromatin-binding sites in larvae showed that ISWI binds to 925 unique gene loci and 141 intergenic regions (Sala et al. 2011). The majority of ISWI peaks in the genomic regions were close to the TSS and the rest were in exons, introns and the 3'end of several genes. Comparing the binding sites of ISWI and

the genes showing altered expression level in the ISWI mutant larvae showed only 10% overlap indicating that the majority of the genes at the transcriptional level are indirectly regulated by the chromatin remodeling of ISWI in *Drosophila*. Motif discovery analysis using the DNA sequences of ISWI targets identified a variety of consensus sequences indicating that a variety of factors recruit ISWI to the chromatin binding sites. One of the identified motifs is the consensus sequence for the GAGA factor (Sala et al. 2011). This is consistent with the finding that NURF recruits the GAGA factor to its binding sites *in vitro* (Tsukiyama & Wu 1995). The NURF complex in *Drosophila* binds to gene enhancer elements, downstream of the transcription start site of active genes and distal insulator sites (Kwon et al. 2016). Brg1 controls the growth of mouse pro-B cells by indirectly regulating the expression of genes encoding products associated with ribosome biogenesis (Bossen et al. 2015). Brg1 binds to a distal regulatory element that modulates *Myc* expression and makes the region accessible for B lineage-specific transcription factors. In turn, c-Myc binds to the promoter of the genes that are associated with ribosome biogenesis and induces their expression to control cell growth.

Genome-wide analysis of the three chromatin-remodellers BRG1, CHD1 and SNF2H in mouse mammary epithelial cells by ChIP-seq identified 38,896, 37,525 and 46,614 binding sites for these proteins, respectively (Morris et al. 2013). Approximately 60% of the binding sites were located in the promoters and within the body of the genes and ~40% of the binding sites were located in intergenic regions. Motif discovery analysis showed that the most significantly enriched motif associated with SNF2H binding sites was CTCF. CTCF is a zinc finger DNA-binding protein and a transcriptional regulator in vertebrates that binds all insulator elements and blocks enhancer function (Bell et al. 1999). Another study confirmed this finding by showing that SNF2H depletion in HeLa cells using siRNA leads to disorganization of nucleosomes and reduction in CTCF binding and subsequent changes in the expression of many CTCF dependent genes (Wiechens et al. 2016). The same study showed that SNF2L has a minor role at remodeling the CTCF sites. SNF2L is involved in the organization of nucleosomes adjacent to CTCF sites as part of the NURF complex, since depletion of BPTF lead to similar results as seen for SNF2L depletion in disorganization of nucleosomes adjacent to CTCF binding sites.

Genome-wide analysis of human BPTF (a component of NURF) by ChIP-seq identified 26,069 binding sites on the genome (Vicent et al. 2011). 40.8% of the binding sites were located in introns, 46.7% in intergenic regions and 5.1% in promoter regions. BPTF is involved in

mammalian erythropoiesis by direct regulating transcription levels of erythroid genes upon lineage differentiation (Li et al. 2016). Comparing ChIP-seq data with global transcriptome profiles showed that strong expression of erythroid genes including *β-globin*, *NOTCH2*, *P4.2*, *GATA-1*, *ANK1* and *KLF3* in human erythroblasts during differentiation is caused by chromatin remodeling activity of BPTF in promoters of these genes. The NURF complex is recruited by SET domain containing 1A (SETD1A) and prevents heterochromatin formation at erythroid-specific promoters/enhancers during terminal erythroid differentiation and maintains the DNA accessible for transcription.

The mouse SNF2L protein is involved in neurogenesis by directly regulating the expression of *Foxg1*, a key regulator of neurogenesis (Yip et al. 2012a). ChIP experiments using cortices of E15.5 embryos showed that SNF2L binds the promoter of *Foxg1*, which results in repression of expression of the transcription factor Foxg1. The human NURF complex has also been associated with cerebellar development through regulation of the expression of EN1 and EN2 by directly binding to their promoters (Alvarez-Saavedra et al. 2014, Barak et al. 2003).

Taken together, the genome-wide analyses of the ATP-dependent chromatin remodeling complexes show that these complexes play important roles in various biological events by affecting gene expression directly or indirectly. These complexes bind to the transcription start sites or the regulatory elements and modulate the expression of the genes directly, which consequently can affect the expression of the downstream genes.

1.11. Rationale:

Previous work provides strong evidence that CECR2/SNF2H-containing complexes exist in the mouse. All the work with CECR2 and its complexes both in human and mouse were based on tagged and fusion proteins. To remove any potential artifacts related to the tag and β-geo cassette, confirmation of the data using antibodies generated against the wild-type protein is important.

The composition of CECR2-containing complexes was analyzed after isolating the complex with the CECR2 antibody. The novel components may contain histone or DNA recognition domains that can change the targeting patterns or modify the remodeling activity of the catalytic component (SNF2h and SNF2L). It has been shown that the non-catalytic components of the ATP-dependent complexes are important for targeting, altering chromatin

binding and remodeling activities of the catalytic component (Reviewed by Bartholomew 2014). The structural domain of the components has been shown to be involved in recognizing the specific histone tail modifications (Maltby et al. 2012). Revealing the structural domains of the novel components would help to hypothesize their function in the complex.

The CECR2-containing complexes are chromatin remodelers and are able to regulate expression of many genes. Microarray analysis has shown that a CECR2 mutation leads to misregulation of 114 genes in mouse embryos (Fairbridge et al. 2010), however it is unknown how many of these genes are regulated directly. Finding the specific binding targets of CERF would be very useful to understand the function of *Cecr2* and how it is involved in central nervous system and reproduction development.

My main questions were:

In my project I wanted to investigate the composition of CECR2-containing complexes and find the chromatin binding sites, to be able to understand how CECR2 is involved in neurulation and reproduction. I hypothesize that there are tissue-specific components of the CECR2-containing complexes that play important roles in tissue-specific functions of the CECR2 complexes in adult testes and mouse ES cells. I also expect to find tissue-specific binding sites for CECR2-containing complexes in adult testes and ES cells. The CECR2 complexes may control reproduction and ES cell specific functions by directly binding to key regulators of the known biological processes.

- Are there tissue-specific CECR2-containing complexes with different composition and possibly different functions?

- What are the chromatin-binding sites of wild-type CERF complexes in ES cells and adult testes and how do these direct targets associate CECR2 with reproduction and neurulation?

2. Materials and Methods

2.1. Maintaining the mouse colony

All experiments involving animals were carried out with experimental protocols and procedures reviewed and approved by the Animal Care and Use Committee of the University of Alberta (University of Alberta AUP 00000094). A sub-strain of BALB/c originated from Charles River Laboratories and FVB/N strain originated from Jackson Laboratories were bred within our colony and used for this research.

Mice were housed in the Sciences Animal Support Services (SASS) facility at the University of Alberta. Mice were kept in filter-top cages in groups of 5/cage in the same room under a 14:10 light/dark cycle at a room temperature of 22±2 °C. Breeding animals were fed a breeding diet containing 9% fat (lab diet 5001), and animals not in the breeding schedule were fed a standard diet containing 4% fat (lab diet 9F 5020).

2.1.1. Breeding

Breeding was done as harem breeding in which three females were housed with a male in each cage. Mated females were identified by visualizing vaginal plugs each subsequent morning and were separated from males. Pregnancies were timed by counting days from the morning of the plugging day. Dependent on the experiment and the need for the embryos, animals were euthanized and embryos were dissected out of the uterus.

2.1.2. Euthanasia

To meet the CCAC guidelines on rodent euthanasia, mice were euthanized with CO_2 after being anesthetized with Isoflurane in an anesthesia/CO_2 chamber located in the animal facility and then transferred to the lab. Mice also were euthanized by cervical dislocation in the lab when immediate dissections were needed. After confirmation of death mice were dissected and unused tissues were stored at -20°C until incineration.

2.2. Genotyping Cecr2 mutations

2.2.1. Genomic DNA extraction

Tissue samples for isolating genomic DNA were obtained from ear notches, embryo tails or the extraembryonic membrane of an embryo. Genomic DNA from the tissue samples was extracted using the following protocol.

The tissue samples were incubated overnight at 65°C in 400 µl DNA extraction buffer (50 mM Tris pH 8.0, 100 mM NaCl, 0.5% SDS, 0.5 µM EDTA) containing 150 µg of proteinase K (Invitrogen). After a brief vortex, an additional 75 µg of proteinase K was added and the samples were incubated at 65 °C for one hour. To precipitate proteins out of the solution, 75 µl of 8 M potassium acetate (final concentration of ~125 M) and 500µl of chloroform were added. After mixing, samples were incubated at -20°C for at least 20 minutes. The samples were thawed and centrifuged at 21,000 g for 5 minutes at room temperature in an Eppendorf tabletop microcentrifuge. The upper aqueous layer was transferred to a new tube and 800 µl (~2 volumes of the solution) of 100% ethanol were added to precipitate the genomic DNA. The samples were shaken manually and centrifuged at 21,000 g for 15 minutes and the supernatant was discarded. The DNA pellets were washed with 500 µl of 70% ethanol and centrifuged at 21000 g for 5 minutes at room temperature. After discarding the supernatant, the pellets were air-dried and resuspended TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). For DNA obtained from ear notches 75 µl of TE buffer and for embryo tails or the extraembryonic membrane 20 µl of TE buffer were used. Isolated genomic DNA samples were stored at 4°C for up to a month or at -20°C for long-term storage.

2.2.2. *Cecr2*$^{tm1.1Hemc}$ genotyping

Generation of the FVB/N *Cecr2*$^{tm1.1Hemc}$ mutant mouse line was described in detail in Fairbridge et al. (2010). *Cecr2*$^{tm1.1Hemc}$, is a presumptive loss-of-function mutation in which exon 1 and 1 kb upstream of it have been deleted through LoxP-Cre recombination (Fairbridge et al. 2010). Using a multiplex PCR, the *Cecr2* wild-type allele (200 bp amplicon) and the *Cecr2*$^{tm1.1Hemc}$ mutant allele (450 bp amplicon) were simultaneously amplified to determine the *Cecr2* genotype. PCR was done in 20 µl reaction containing 2 µl genomic DNA, 0.3 µl of DreamTaq DNA Polymerase (Life technologies, cat. no. EP0702), 2 µl of 10X DreamTaq Buffer,

0.5 µM of each primer (IngeniousLox1, Ingenious SDL2 and LoxCECR2_DEL3R) and 0.25 mM dNTPs. A Peltier Thermo Cycler PTC-200 (MJ Research) was used for PCR reactions. The cycling began with 94°C for 3 minutes, followed by 37 cycles of 94°C for 15 seconds, 60°C for 20 seconds, and 68°C for 40 seconds, and a final extension at 68°C for 5 minutes. For sequences of the primers see Appendix A.

2.2.3. *Cecr2^{Gt45Bic}* genotyping

The *Cecr2^{Gt45Bic}* mutant mouse line was generated in McDermid' lab (Banting et al. 2005). In *Cecr2^{Gt45Bic}* a βgeo cassette is spliced to exon 7 and leads to a CECR2-β-galactosidase fusion polypeptide containing the first 7 CECR2 exon fused to β-galactosidase. Using a multiplex PCR, the wild-type *Cecr2* allele (376 bp amplicon), the *Cecr2^{Gt45Bic}* (573 bp amplicon) and the male specific SRY gene (266 bp amplicon) were simultaneously amplified to determine the *Cecr2* genotype and the sex. The PCR reactions were performed in 20 µl reaction containing 2 µl genomic DNA, 0.3 µl of DreamTaq DNA Polymerase (Life technologies, cat. no. EP0702), 0.5 µM of each primers (Mmu Intron7 F4, Mmu Intron7 R4, pGT1R4, SRY FOR and SRY REV), 2 µl 10X DreamTaq Buffer and 0.25 mM dNTPs. A Peltier Thermo Cycler PTC-200 (MJ Research) was used for PCR reactions. The cycling began with 94°C for 3 minutes, followed by 37 cycles of 94°C for 15 seconds, 60°C for 20 seconds, and 68°C for 40 seconds, and a final extension at 68°C for 5 minutes. For sequence of the primers see Appendix A.

2.2.4. Agarose gel elctrophoresis

To analyze the PCR products, the PCR reactions were separated via electrophoresis through a 2% agarose gel containing 0.1 µg/ml of ethidium bromide in 1X TAE buffer (40 mM Tris-acetate, 1 mM EDTA) at 130 V for 1 hour along with a nucleic acid marker. DNA bands were visualized by UV fluorescence gel imager (Alpha Innotech). For *Cecr2$^{tm1.1Hemc}$* the wild-type band was 250 base-pairs (bp) and the mutant band was 500 bp. For *Cecr2^{Gt45Bic}* the wild-type band of 376 bp was produced by the Mmu Intron7 F4 and Mmu Intron7 R4 primers. The Mmu Intron7 F4 and pGT1R4 primers produced a 573 bp band From the *Cecr2^{Gt45Bic}* allele. The SRY primers produced a 266 bp band in male mice.

2.3. Affinity purification of chicken polyclonal antibodies to mouse CECR2

Two peptides of 14 amino acids each from mouse CECR2, KGKRTKRPQPELQH and RSRDTEGSSRKQPP, were selected and conjugated to keyhole limpet hemocyanin via addition of a C-terminal cysteine residue by GenScript Corporation. Two chickens received the two peptides under a regimen of a single injection every two weeks for eight weeks and yolk emulsions from immunized chickens after the fourth injection were provided by GenScript. Affinity purification was used to purify anti-CECR2 peptide antibody from yolk emulsion as shown in Figure 2-1. At first, total IgY was purified from yolk emulsions according to a method reported previously (Jensenius et al. 1981). To delipidate immune yolk emulsion, it was diluted 1:10 by adding milliQ water. After bringing the pH to ~7.0 with 10 mM NaOH, the diluted yolk was placed at -80°C overnight and next day it was thawed at room temperature. The mixture was then centrifuged at 2000 g for 30 minutes at 4°C to separate a lipid pellet from a water-soluble protein fraction (WSPF) supernatant. The WSPF was passed through three layers of gauze and then by adding ammonium sulfate on ice with stirring it was brought to 33% saturation. The pH was adjusted to 5.8-6.0 (Isoelectric point of IgY) with 10% ammonium hydroxide solution. Stirring was continued for an additional hour on ice. The saturated WSPF was centrifuged at 10,000 g for 15 minutes at 4°C. The precipitated IgY was dissolved in a maximum 4-5 ml of phosphate-buffered saline (PBS: 10 mM sodium phosphate, 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.76 mM KH_2PO_4 pH 7.4). To remove ammonium sulfate the yield of total IgY was dialyzed using cellulose dialysis tubing (Fisher Scientific, cat. no. 21-152-15) with a molecular weight cut-off of 12,000-14,000 according to the product manual. The dialyzed total IgY was sterilized using a 0.22 μm syringe filter (Millipore, cat. no. SLGS033SS) and stored at 4 °C until affinity purification. The SulfoLink[TM] kit (Pierce cat. no. PI20325) was used to affinity purify peptide-specific antibody. Approximately 1 mg of the mouse CECR2 peptides were incubated with crosslinked iodoacetyl-activated agarose resin for 30 minutes at room temperature followed by washing three times with 2 ml of 1.0 M NaCl. Washing continued with 2 ml of coupling buffer provided in the kit. The resin was blocked with 2 ml of 50 mM L-Cysteine. HCl for 15 minutes with gentle rocking and for 30 minutes without rocking at room temperature. Then the resin was washed with TBS three times. Approximately 3 ml of total IgY fraction (~100 mg) was mixed with the affinity resin coupled to CECR2 peptides for 1 hour at room temperature with gentle rocking. The resin was washed with one resin volume (2 ml) of TBS (TBS: 25 mM Tris,
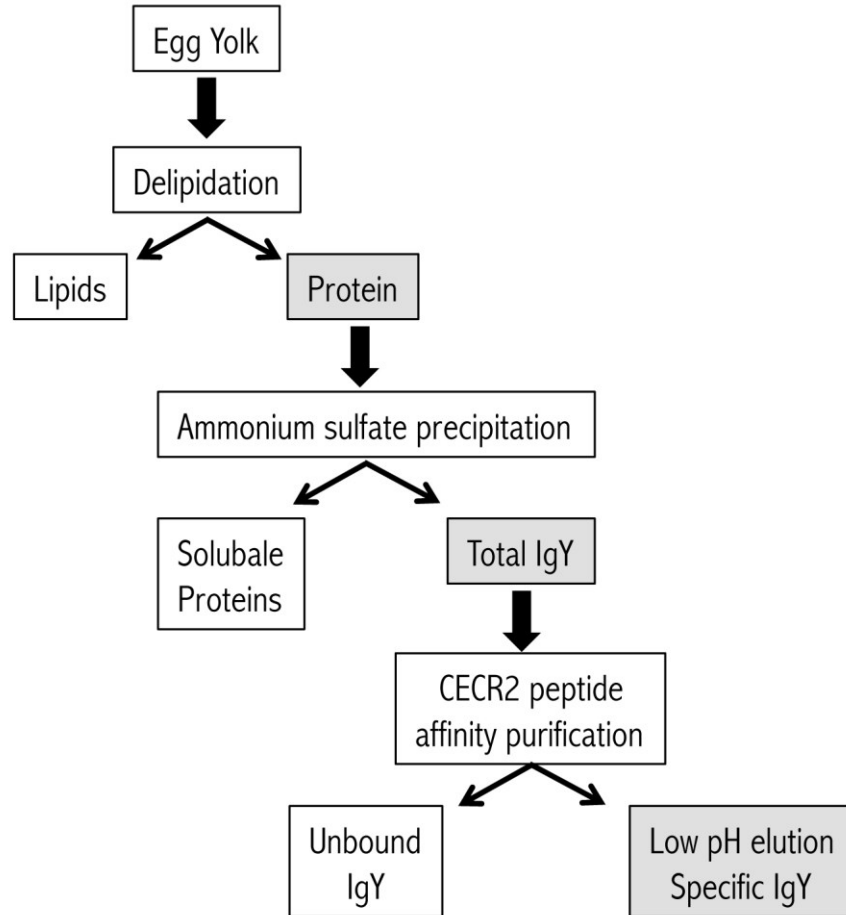
Figure 2-1 Affinity purification of the anti-CECR2 peptide antibody from yolk emulsion. The immune yolk emulsion was delipidated to obtain a water-soluble protein fraction (WSPF). Total IgY was precipitated by adding ammonium sulfate. Precipitated IgY was dissolved in a maximum 4-5 ml of phosphate-buffered saline and was dialyzed using cellulose dialysis tubing with a molecular weight cut-off of 12,000-14,000. Finally, the anti-CECR2 peptide antibody was affinity purified from the total IgY using SulfoLink[TM] (Pierce cat. no. PI20325).

500 mM NaCl pH 7.4) containing 0.5 M NaCl several times until the absorbance of the flow-through at 280 nm (A_{280}) was near zero as measured by a spectrophotometer. Bound IgY was eluted with 0.1 M glycine pH 2.5 and collecting 0.5 ml fractions in 1.5 ml centrifuge tubes containing 50 μl of neutralization buffer (1.5 M Tris-HCl pH 8.8). The A_{280} of the eluted fractions was measured and the fractions with the highest A_{280} were combined and stored at -20°C. Concentration of the affinity purified IgY was calculated by dividing the A_{280} value by 1.4.

2.4. Production of antibodies directed against mouse CECR2 in rabbit

2.4.1. Total RNA extraction

Total RNA was extracted from a whole E12.5 embryo using an RNeasy Lipid Tissue MiniKit (Qiagene cat. no. 74804) according to the instructions provided by the kit. The extracted total RNA was quantified using the NanoDrop Spectrophotometer at the Molecular Biology Service Unit (MBSU).

2.4.2. cDNA synthesis

First strand cDNA was synthesized using the Superscript III Reverse Transcriptase (Invitrogen cat. no. 18080-093). A 20 μl reaction, containing 1 μg total RNA template, 0.5 mM total dNTPs and 50 ng random primers provided in the kit, was heated to 65°C for 5 minutes. 4 μl 5X FirstStrand buffer, 1 μl 0.1M DTT, 40 units of RNaseOUT and 200 units Superscript III Reverse Transcriptase were added to the reactions and incubated at 25°C for 5 minutes followed by a 50 minute incubation at 50°C. To inactivate the reaction, tubes were heated at 70°C for 15 minutes. The RNA template was removed by addition of 2 units E. coli RNase H and incubating at 37°C for 20 minutes.

2.4.3. Amplification of histidine-tagged *Cecr2* fragments

Two fragments of mouse CECR2 protein were selected and used in the production of rabbit polyclonal antibodies (Appendix B). Fragment 1 contained exon 1 to exon 9 and fragment 2 contained part of exon 17 to the end of the coding part of exon 19 of *Cecr2*. Each *Cecr2* gene fragment was amplified using forward and reverse primers containing EcoRI and SalI restriction sites and a sequence coding for an N or C-terminal Hexahisitidine (His6)-tag. Phusion High-Fidelity PCR Kit (Thermo Scientific cat. no. F-530S) was used for the amplification of the selected *Cecr2* fragments. A 20 μl reaction volume was used, containing 4 μl 5xPhusion HF buffer, 200 μM total dNTPs, 0.5 μM of forward and reverse primers, 3% DMSO, 2 μl cDNA and 0.02 U/μl Phusion DNA Polymerase. The cycling condition was: initial heating at 98°C for 30 seconds, followed by 35 cycles of 98°C for 10 seconds, 65°C for 30 seconds, 72°C for 2 minutes and 30 seconds, and a final extension at 72°C for 10 minutes. PCR amplicons were stored at -20°C until cloning.

2.4.4. A-tailing and ligation

PCR fragments generated by Phusion DNA Polymerase in section 2.4.3 were A-tailed. A 10 μl reaction, containing 4 μl PCR fragments, 5 U Taq DNA Polymerase (Invitrogen cat. no. 10342-053), 1 μl 10X PCR Buffer, 1.5 mM $MgCl_2$ and 0.2 mM dATP was incubated at 70°C for 30 minutes. A-tailed DNA was ligated into pGEM®-T Easy Vector (Promega cat. no. A3600) overnight at 4°C in a reaction containing 3 μl A-tailed DNA, 1 μl pGEM-T easy vector, 5 μl 2X Rapid Ligation Buffer and 1 U T4 DNA Ligase. The ligation reaction was transformed into competent *E. coli* (DH5 alpha).

2.4.5. Preparation of competent *E. coli* (DH5 alpha)

E. coli (DH5 alpha) cells were streaked on a Luria-Bertani (LB) media plate (1% Bacto-tryptone, 0.5% Bacto-Yeast, 0.5% Nacl, 0.1% Glucose and 2% Agar, pH 6.7) and incubated at 37°C overnight allowing them to produce colonies. 5 ml liquid LB media (8% Bacto-tryptone, 0.5% Bacto-Yeast, 0.5% NaCl and 0.1% Glucose, pH 6.7) was inoculated with a single colony and incubated at 37°C overnight. 0.5 ml of overnight culture was used to inoculate 50 ml liquid LB media, which then was incubated at 37°C with agitation until reaching OD_{600} of 0.3-0.5 (~2 hours). Cells were centrifuged at 3000 g for 10 minutes at 4°C. Cells then were resuspended in 50 ml ice-cold 0.1 M $MgCl_2$ and incubated on ice for 15-30 minutes. Cells were centrifuged at 3000 g for 10 minutes at 4°C and resuspended in 50 ml ice-cold 30 mM CaCl2 and incubated on ice for 30 minutes. Cells were centrifuged at 3000 g for 10 minutes at 4°C and resuspended in 3 ml 30 mM CaCl2 containing 15% glycerol and stored at -80°C in 100 μl aliquots.

2.4.6. Cloning the ligated-*Cecr2*-fragments

The pGEM-T easy vectors containing *Cecr2* fragments were transformed into competent *E. coli* (DH5 alpha) cells. 100 μl competent cells were added to 10 μl ligation reaction and incubated at 4°C for 30 minutes, 42°C for 90 seconds and 4°C for 2 minutes. After adding 1 ml liquid LB, the reaction was incubated at 37°C for 1 hour and plated on LB/amp (25 mg/ml)/IPTG (0.2 g/ml)/X-gal (50 mg/ml in dimethylformamide) plates at 37°C overnight. Vectors containing cloned *Cecr2* fragments were extracted from the produced colonies and stored at -20°C.

2.4.7. Subcloning into the expression vector

Cecr2 fragments prepared in section 2.4.6 were then subcloned into the expression vector pET-21a (Novagen cat. no. 69740-3) and were transformed into *E. coli* BL21 (DE3). Transformed cells were resuspended in LB media containing 15% glycerol and stored at -80°C.

2.4.8. Expression and purification of recombinant polypeptide (B5P44)

Transformed *E. coli* BL21 (DE3) cells were grown in 5 ml liquid LB media containing 100 μg/ml ampicillin at 37°C with agitation (200 rpm) overnight. Five ml overnight culture was added to 500 ml LB containing 100 μg/ml ampicillin at 37°C with agitation (250 rpm) until reaching OD_{600} of 0.5-0.6. Expression was induced by addition of 0.1 mM IPTG and incubating at 30°C with agitation (200 rpm) overnight. Cells were harvested by centrifugation at 4000 g for 20 minutes at 4°C in a Sorvall RC-35B Refrigerated Superspeed centrifuge and then frozen at -20°C overnight. Cells were thawed on ice for 15 minutes and resuspended in imidazole buffer (50 mM NaH2PO4, 300 mM NaCl, 10 mM imidazole, pH 8, containing 2 mM PMSF) at 4 ml per gram wet weight. Lysozyme was added to a final concentration of 1 mg/ml and incubated on ice for 30 minutes. The cell lysate was then sonicated on ice 6 x 10 second bursts at 200 W with 10 second cooling intervals. The lysate was centrifuged at 10,000 g for 25 minutes at 4°C in a Sorvall RC-35B Refrigerated Superspeed centrifuge and the supernatant was used for Ni+ affinity chromatography purification of the 6xHis-tagged CECR2 fragments using Ni-NTA Agarose (QIAGEN cat. no. 30210) as follows. 1 ml of the 50% Ni-NTA slurry was added to 4 ml of the lysate and rocked for 1 hour at 4°C. The lysate–Ni-NTA mixture was loaded into a column and washed 3 times with wash buffer (50 mM NaH2PO4, 300 mM NaCl, 20 mM imidazole, pH 8) and the recombinant protein was eluted 7 times with 0.5 ml elution buffer (50 mM NaH2PO4, 300 mM NaCl, 250 mM imidazole, pH 8). The purified protein was dialyzed against PBS and stored at -20°C. Purity of the recombinant proteins was determined by Coomassie blue staining of SDS-PAGE gel.

2.4.9. Immunization and production of polyclonal antibodies

2.4.9.1.   Animals

Four Female New Zealand White rabbits (10 - 12 weeks of age) were purchased from Charles River. The rabbits were housed in the SASS facility at the University of Alberta.

2.4.9.2.   Immunization

Two rabbits were immunized with CECR2 fragment 1 (section 2.4.8) and two rabbits with CECR2 fragment 2 (section 2.4.8) according to standard immunization protocol in the SASS facility at the University of Alberta. Briefly, 750 ml of purified CECR2 fragment in PBS (0.5 mg/ml) was mixed 1:1 with Complete Freund's Adjuvant (CFA) and injected into the rabbits at 3 sites subcutaneously. Boosters were given at 3 weeks, 7 weeks, 11 weeks and 15 weeks after the first immunization with 750 ml of purified CECR2 fragment in PBS (0.5 mg/ml) mixed 1:1 with Incomplete Freund's Adjuvant (IFA).

2.4.9.3.   Blood sampling and isolation of serum

Pre-immune blood samples were taken prior to the first immunization. Immune blood samples were taken 5 weeks, 9 weeks, 13 weeks and 17 weeks after the first immunization. Two weeks after the final blood sampling, all rabbits were exsanguinated, conducted by personnel of the SASS facility at the University of Alberta. Blood samples obtained from the rabbits were incubated at room temperature for one hour in glass test tubes to allow the formation of the clot. The clot was loosened from the side of the tubes using a pasteur pipette and incubated at 4°C overnight for the clot to retract. The separated serum was collected from the blood samples and transferred to another tube and centrifuged at 4000 g for 10 minutes at 4°C. The serum samples, which contained anti-CECR2 polyclonal sera, were aliquoted and stored at − 80°C. The immunization and blood sample collections are shown in Table 2-1.

2.4.9.4.   ELISA analysis of sera for anti-CECR2 polyclonal antibodies

To evaluate the specificity of the CECR2 polyclonal antibodies in different sera samples, ELISA with the pre- and post-immune sera on purified recombinant CECR2 fragments was performed. 96-well microplates (Roche, cat. no. 12-03092) were coated with 100 μl of 10 μg/ml purified CECR2 fragments in PBS at 4°C overnight. The coated wells were washed with PBS for 10 minutes at room temperature with agitation. The coated wells then were blocked with 350 μl of 1% bovine serum albumin (BSA) in PBS for one hour at room temperature. After washing 2 times for 10 minutes in PBS, the wells were incubated with 350 μl of different dilutions (1:100, 1:1000, 1:5000, 1:10,000 and 1:25,000) of pre- and post-immune sera for one hour at room temperature. After washing 2 times in PBS for 10 minutes each time, wells were incubated in 350 μl of goat anti-rabbit IgG-HRP (1:5000) (Santa Cruz, cat. no. 12-02585) for one hour at room

temperature. Wells were washed with PBS twice for 10 minutes and then were incubated with 100 μl/well Tetramethylbenzidine (TMB) substrate (Cell Signaling Technology, cat. no. 7004) for 15 minutes at room temperature. The reaction was stopped by addition of 100 μl/well of stop solution. The absorbance at 450 nm was then read using a Plate Reader (ThermoMax). To check for the specificity of antibody binding to polypeptides, control wells coated with only BSA and no antigen were used. The ELISA result was considered as positive when the A_{450} value of post-immune sera was at least twice as high as the A450 value of pre-immune sera.

Table 2-1 Immunization and blood sampling of rabbits. Two rabbits (M10-1 and M10-2) were immunized with purified CECR2 fragment 1 and two rabbits (M10-3 and M10-4) with CECR2 fragment 2. 750 **μl** of purified CECR2 fragment in PBS (0.5 mg/ml) was mixed 1:1 with Complete Freund's Adjuvant (FCA) and injected into the rabbits at 3 sites subcutaneously. Boosters was given at 3 weeks, 7 weeks, 11 weeks and 15 weeks after the first immunization with 750 **μl** of purified CECR2 fragment in PBS (0.5 mg/ml) mixed 1:1 with Incomplete Freund's Adjuvant (FIA). For two of the rabbits an extra injection was performed 2 weeks after the last injection.

Week	Procedure	Rabbit M10-1 (Hocus)[a]	Rabbit M10-2 (Thumper)[b]	Rabbit M10-3 (Spirit)[a]	Rabbit M10-4 (Fang)[b]
0	Pre-immune blood sample	5 ml	5 ml	5 ml	5 ml
1	1st Antigen/FCA injection	1.5 ml	1.5 ml	1.5 ml	1.5 ml
4	2nd Antigen/FIA injection	1.5 ml	1.5 ml	1.5 ml	1.5 ml
6	1st Post-immune blood sample	5 ml	5 ml	5 ml	5 ml
8	3rd Antigen/FIA injection	1.5 ml	1.5 ml	1.5 ml	1.5 ml
10	2nd Post-immune blood sample	5 ml	5 ml	5 ml	5 ml
12	4th Antigen/FIA injection	1.5 ml	1.5 ml	1.5 ml	1.5 ml
14	3rd Post-immune blood sample	5 ml	5 ml	5 ml	5 ml
15	Whole blood collection	-	64 ml	-	47 ml
16	5th Antigen/FIA injection	1.5 ml	-	1.5 ml	-
18	4th Post-immune blood sample	5 ml	-	5 ml	-
20	Whole blood collection	74 ml	-	56 ml	-

2.4.9.5. Affinity purification of rabbit polyclonal antibodies to mouse CECR2

The serum sample from one of the rabbits immunized with the CECR2 C-terminal fragment was chosen to be affinity purified based on the result obtained from ELISA and Western Blot analysis performed using serum samples. The SulfoLink Coupling Resin (Thermo Scientific, cat. no. 20401) was used to affinity purify CECR2-specific antibody. Two ml

(approximately 1 mg) of Ni+ affinity-purified-CECR2-fragment containing exons 17-19 (section 2.4.8) was treated with the reducing reagent DTT to a final concentration of 10 mM and incubated for 30 minutes at room temperature. The protein was dialyzed against Tris-EDTA buffer (50 mM Tris, 10 mM EDTA, pH 8.5) overnight at 4°C. After equilibrating SufoLink Coupling Resin to room temperature, 2 ml of the resin (50% slurry) was loaded into a column and washed with 12 ml (12 x volume of the resin) with Tris-EDTA buffer. The column cap was replaced and reduced CECR2 fragment was then added to the column and the column was rocked for 15 minutes at room temperature. The mixture was incubated for an additional 30 minutes without rocking and then the column was allowed to drain. The resin was washed with 6 ml Tris-EDTA buffer and then blocked with 1 ml of 50 mM L-Cysteine in Tris-EDTA buffer for 15 minutes with rocking and for 30 minutes without rocking at room temperature. Then the resin was washed with 6 ml of 1 M NaCl followed by washing with 6 ml PBS. Six ml of the selected serum was thawed on ice and mixed with 6 ml of 2xPBS containing 0.2% NaN3. The mixture was loaded onto the column and the flow-through was collected with 15 minutes intervals after collecting every 750 μl. All flow-through was then passed through the column for second time without any stopping. The resin was washed with PBS containing 0.5 M NaCl and 0.1% Triton X-100 several times until the absorbance of the flow-through at 280 nm (A_{280}) was near zero as measured by a spectrophotometer. Bound IgG was eluted with 0.1 M glycine pH 2.6 and collecting in 0.5 ml fractions in eppendorf tubes containing 30 μl of neutralization buffer (1.5 M Tris-HCl pH 8.5). The A_{280} of the eluted fractions was measured and the fractions with the highest A_{280} were combined and stored at -20°C. Concentration of the affinity purified anti-CECR2 antibody was calculated by dividing the A_{280} value by 1.4.

2.5. Mouse Embryonic Stem (ES) cell culture

TT2 (a gift from Matt Lorincz's lab from the University of British Columbia) and E14 (a gift from Dr. Laszlo Tora, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), France) mouse ES cells were cultured in Dulbecco's Modified Eagle high glucose (Sigma, cat. no. D6429-1L) supplemented with 15% embryonic stem cell-qualified Fetal Bovine Serum (FBS) (Life technologies, cat. no. 10439-024), 2 mM L-glutamine (Life technoligies, cat. no. 25030-081), 100 mM β-mercaptoethanol, 0.1 mM MEM non-essential amino acids solution

(Life technologies, cat. no. 11140-050), 100 Units/ml Penicillin-Streptomycin (Life technologies, cat. no. 15070-063) and 1000 U/ml recombinant leukemia inhibitory factor (Sigma-Aldrich cat. no. L5158-5UG). The cells were grown on 100 mm Cell Culture Dishes (Thermo Scientific, cat. no. NC0479278) coated with gelatin and incubated at 37°C in a 5% CO2 atmosphere. Cultures were split every two days at a 1:10 ratio and fed on the alternate days. 0.25% trypsin-EDTA solution (Invitrogen, cat. no. 25200-072) was used to split and harvest the ES cells.

2.6. Preparation of nuclear extracts from ES cells

ES cells in 100 mm culture plates were incubated in 2 ml of 0.25% Trypsin-EDTA (Invitrogen, cat. no. 25200-072) for 5 minutes at 37°C. Cells were centrifuged at 1000 g for 5 minutes and washed once in PBS, then centrifuged again at 1000 g for 5 minutes. Cells were then resuspended in 10 pellet volumes of hypotonic cell lysis buffer (10 mM HEPES pH 7.9, 1.5 mM MgCl2, 10 mM KCl, 0.5 mM DTT) supplemented with 1% protease inhibitor cocktail (Sigma-Aldrich, cat. no. I3786). Cells were incubated on ice for 15 minutes and were then lysed by 10 strokes in a glass-glass dounce homogenizer, using the type B pestle. A 10 μl aliquot of cells was stained with 10 μl of 0.4% trypan blue solution (Sigma) to monitor cell lysis. Cell debris and nuclei were collected by centrifugation at 2500 g for 10 minutes at 4°C. The nuclear pellet was resuspended in one volume of nuclear extraction buffer (20 mM HEPES pH 7.9, 420 mM NaCl, 1.5 mM MgCl2, 0.2 mM EDTA, 25% glycerol) supplemented with 1% protease inhibitor cocktail (Sigma-Aldrich, cat. no. I3786). The suspension was sonicated briefly to shear genomic DNA and rocked for 30 minutes at 4°C. The lysate was centrifuged at 21,000 g for 10 minutes at 4°C. The pellet containing nuclear debris was discarded and the supernatant was aliquotted, flash-frozen in liquid nitrogen, and stored at -80°C. The DC Protein Assay™ (Bio-Rad) was used to determine protein concentration.

2.7. Preparation of nuclear extracts from tissues

Tissue was dissected, transferred to ice-cold PBS and minced into small pieces. The minced tissue was centrifuged at 1000 g for 5 minutes. The supernatant was discarded and the minced tissue was then resuspended in 10 pellet volumes of hypotonic cell lysis buffer (50 mM

Tris-HCl pH 7.4, 5 mM MgCl2, 250 mM sucrose, 1 mM DTT, 50 μg/ml spermidine) supplemented with 1% protease inhibitor cocktail (Sigma-Aldrich, cat. no. I3786). The tissue was lysed by 10 strokes in a glass-glass dounce homogenizer, using the type B pestle. A 10 μl aliquot of cells was stained with 10 μl of 0.4% trypan blue solution (Sigma) to monitor cell lysis. Cell debris and nuclei were collected by centrifugation at 21000 g for 10 minutes at 4°C. The nuclear pellet was resuspended in one volume of nuclear extraction buffer (20 mM HEPES, 420 mM NaCl, 1.5 mM MgCl2, 0.2 mM EDTA, 25% glycerol, pH 7.9) supplemented with 1% protease inhibitor cocktail (Sigma-Aldrich, cat. no. I3786). The suspension was vortexed and rocked for 30 minutes at 4°C. The lysate was centrifuged at 21,000 g for 10 minutes at 4°C. The pellet containing nuclear debris was discarded and the supernatant was aliquotted, flash-frozen in liquid nitrogen, and stored at -80°C. The DC Protein Assay™ (Bio-Rad) was used to determine protein concentration.

2.8. Preparation of whole cell lysates from ES cells

ES cells in 100 mm culture plates were washed 3 times with ice-cold PBS. After discarding the PBS, 1 ml of freshly made ice-cold lysis buffer (20 mM Tris pH 8.0, 420 mM NaCl, 10% glycerol, 1% IGEPAL® CA-630) supplemented with 1% protease inhibitor cocktail (Sigma-Aldrich, cat. no. I3786) was added to each 100 mm plate, and cells were scraped off the plate using a plastic cell scraper. The cell suspension was transferred into a pre-cooled microfuge tube. The cells were homogenized by passing the lysate through a 27-gauge needle (5 times) and rocking at 4°C for 30 minutes. Cell debris was collected by centrifugation at 15,000 g for 20 minutes at 4°C. The pellet containing cell debris was discarded and the supernatant was aliquotted, flash-frozen in liquid nitrogen, and stored at -80°C. The DC Protein Assay™ (Bio-Rad) was used to determine protein concentration. To prepare whole cell lysate with low salt concentration, the same procedure was used with the following extraction buffer: 50 mM Tris pH 8.0, 150 mM NaCl, 1% IGEPAL® CA-630 supplemented with 1% protease inhibitor cocktail (Sigma-Aldrich, cat. no. I3786). In some of the experiments I used 150 mM NaCl in the lysis buffer instead of 420 mM.

2.9. Preparation of whole cell lysates from tissues

Tissue was dissected and transferred into an eppendorf tube containing ice-cold lysis buffer (50 mM Tris pH 8.0, 420 mM NaCl, 1% IGEPAL$^®$CA-630) supplemented with 1% protease inhibitor cocktail (Sigma-Aldrich, cat. no. I3786). 1 ml of lysis buffer was used per gram of the tissue. The tissue was then diced into very small pieces using a razor blade. The pieces were then lysed by 20 strokes in a glass-glass dounce homogenizer, using the type B pestle. A 10 μl aliquot of cells was stained with 10 μl of 0.4% trypan blue solution (Sigma) to monitor cell lysis. The cells were further homogenized by passing the lysate through a 27-gauge needle (5 times) and rocking at 4°C for 30 minutes. Cell debris was collected by centrifugation at 15,000g for 20 minutes at 4°C. The pellet containing cell debris was discarded and the supernatant was aliquotted, flash-frozen in liquid nitrogen, and stored at -80°C. The DC Protein Assay™ (Bio-Rad) was used to determine protein concentration. To prepare whole cell lysate with low salt concentration, the same procedure was used using the following extraction buffer: 50 mM Tris pH 8.0, 150 mM NaCl, 1% IGEPAL® CA-630 supplemented with 1% protease inhibitor cocktail (Sigma-Aldrich, cat. no. I3786).

2.10. SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE was performed using the Tris-glycine SDS-PAGE method previously described (Laemmli 1970). Electrophoresis was performed using the Mini Protean III™ electrophoresis system (Bio-Rad) according to the manufacturer's instructions. A 4% acrylamide stacking gel (4% acrylamide, 125 mM Tris-HCl pH 6.8, 0.1% SDS, 0.1% ammonium persulfate, 0.1% TEMED) and a 7.5% resolving gel (7.5% acrylamide, 375 mM Tris-HCl pH 8.8, 0.1% SDS, 0.1% ammonium persulfate, 0.08% TEMED) were used to separate proteins. Protein samples were mixed with 4x NuPAGE loading buffer (Life technologies, cat. no. NP0007) and β-mercaptoethanol to a final concentration of 100 mM and then denatured by heating at 95-100°C for 5 minutes, followed by loading onto the gel. The proteins were separated via electrophoreses in Tris-glycine running buffer (25 mM Tris, 190 mM glycine, 0.1% SDS) at 120 V during the migration of proteins through the stacking gel and then at 175 V during the migration of proteins through the resolving gel. The molecular weight of proteins was estimated by running a pre-stained protein ladder (Thermo Scientific, cat. no. 26616) along side with the protein samples.

2.11. Coomassie Blue staining of polyacrylamide gels

Protein samples were separated by SDS-PAGE as described in previous section. After electrophoresis, polyacrylamide gels were washed 3 times with milliQ water for 15 minutes and then stained with GelCode Blue stain reagent (Fisher Scientific, cat. no. 24590) for one hour on a shaker followed by washing in MQ water for 15 minutes. Gels were photographed using a digital camera.

2.12. Western blot analysis

Proteins were separated by SDS-PAGE as described in section 2.10 and transferred to 0.45 μm pore size PVDF membranes (Millipore, cat. no. IPVH00010) using the Mini Protean III™ submerged tank wet transfer unit (Bio-Rad) for 30 minutes at a constant current of 350 mA in a transfer buffer (25 mM Tris, 190 mM glycine, 10% methanol). Protein transfer was monitored by the transfer of the protein ladder. Membranes were rinsed in TBS containing 0.05% Tween-20 (TBST) and blocked in 5% skim milk in TBST or 3% BSA in TBST for 30 minutes at room temperature with agitation. Membranes were then washed in TBST for 5 minutes and then incubated in the primary antibodies overnight at 4°C. The primary antibodies were diluted in 5% milk/TBST as follows:

Affinity purified rabbit polyclonal anti-CECR2 antibody diluted 1:50,000 in 5% milk/TBST, mouse monoclonal anti-TBP (Abcam, cat. no. ab818) diluted to 0.25 μg/ml (1:5000 dilution) in 5% milk/TBST, mouse monoclonal anti-SNF2H ascite fluid (Active Motif) diluted 1:3000-1:6000 in TBST, rabbit polyclonal anti-SNF2H (Abcam, cat. no. ab3749) diluted 1:4000 in TBST, rabbit polyclonal anti-LUZP1 (Protein tech, cat. no. 17483-1-AP) diluted 1:5000 in 5% milk/TBST, rabbit polyclonal anti-CCAR2 (Protein tech, cat. no. 22638-1-AP) diluted1:1000 in 5% milk/TBST, sheep polyclonal anti-SNF2L (contributed by Dr. David Picketts) diluted 1:1000 in 5% milk/TBST, mouse monoclonal anti-tubulin (Sigma-Aldrich, cat. no. T6199) 1:10000 in 5% milk/TBST and affinity purified chicken polyclonal anti-CECR2 peptide antibody was used at 0.5 μg/ml.

Membranes were washed in TBST three times (15 minutes each) and incubated in

horseradish peroxidase-coupled secondary antibodies for 1 hour at room temperature. Rabbit IgG was detected using goat anti-rabbit IgG (BioRad, cat. no. 170-5046) diluted 1:5000 in 5% milk/TBST. Sheep IgG was detected using rabbit anti-sheep IgG-HRP (Santa CruZ cat. no. sc-2770) diluted 1:5000 in 5% milk/TBST. Mouse IgG was detected using goat anti-mouse IgG-HRP conjugate (Sigma-Aldrich) diluted 1:10,000 in 5% milk/TBST. After 3 washes for 10 minutes each in TBST, antibody-antigen complexes were detected on membranes using SuperSignal West Pico Chemi-Luminescence kit (Thermo Scientific, cat. no. PI34080) according to the manufacturer's instruction. Membranes were then exposed to X-ray film and developed using an automated developer.

2.13. Immunoprecipitation

10 μl of Dynabeads® Protein A (Life technologies, cat. no. 10001D) was used for each IP reaction. Dynabeads were separated from buffer using a Magnetic Separation Stand and resuspended in 200 μl of PBS containing 0.02% Tween-20. After addition of 1 μg antibody, the beads were rocked for 10 minutes at room temperature to allow the antibody to bind to the beads. The beads were separated from the supernatant and were washed with 200 μl of PBS containing 0.02% Tween-20. Whole cell lysate was prepared concurrently as described in section 2.7 and 2.8. Approximately 500-1000 μg of whole cell lysate was diluted with immunoprecipitation dilution buffer (50 mM Tris pH 8.0, 1% IGEPAL® CA-630) supplemented with 1% protease inhibitor cocktail to obtain a final concentration of 150 mM of NaCl in the solution. Diluted whole cell lysate was then added to the beads bound to the antibody. IP reactions were rocked for 2 hrs to overnight at 4°C. The best incubation time was obtained by optimizing each antibody separately. Following the incubation, beads were washed three times for 5 minutes each in 0.5 ml of lysis buffer containing 0.1% Tween-20. To elute immunoprecipitated proteins, beads were transferred to new tubes and 10 μl of 4x NuPAGE™ loading buffer (Life Technologies, cat. no. NP0007) and 10 μl of 1M β-mercaptoethanol were added to the beads, which were then heated at 95-100°C for 5 minutes. The supernatant containing the eluted proteins was separated from the beads and used in SDS-PAGE and Western blot analysis as described in sections 2.9 and 2.11, respectively. 25% of the IP reaction was analyzed in each lane for Western blots. For a negative control, normal IgG from the same species as the experimental antibody was used in an IP

reaction. As a negative control for immunoprecipitation experiments with CECR2, SNF2H (abcam, cat. no. ab3749), LUZP1 (Proteintech, cat. no. 17483-1-AP), CCAR2 (Proteintech, cat. no. 22638-1-AP) antibodies, 1 μg of normal rabbit IgG (New England Biolabs, cat. no. 2729) was used. For SNF2L antibody 1 μg of normal sheep IgG (Santa Cruz Biotechnology, cat. no. sc-2717) was used. For IP from gel filtration fractions, the fraction of interest was used in the IP reaction in the same way whole extract was used as described earlier in this section.

2.14. Gel filtration

A column of Sephacryl S-400 HR™(GE Healthcare, cat. no. 17-0609-10) media was prepared for gel filtration. The column was 50 cm in height and 1 cm in diameter. The column was packed using a peristaltic pump (Gilson, MINIPULS® 3) with the constant flow rate of 1 ml per minute to obtain a bed volume of ~45 cm equilibrated in gel filtration buffer (TBS buffer). The column was calibrated by loading the column with 1 mg of blue dextran (Sigma-Aldrich, cat. no. D5751-1G) with a molecular weight of 2 MDa in 0.3 ml of TBS, immediately followed by loading a mixture of 2 mg thyroglobulin (Sigma-Aldrich, cat. no. T9145-1VL) with a molecular weight of 670 kDa and 2 mg of BSA (Sigma-Aldrich) with a molecular weight of 66 kDa in 0.3 ml TBS. Immediately after applying the standards, fractions of 0.5 ml were collected by applying gel filtration buffer to the column. A steady flow rate of 0.5 ml per minute was kept using a peristaltic pump. Absorbance of the fractions was measured at 280 nm (A_{280}) using a spectrophotometer. The elution volume (V_e) for each standard protein was determined by measuring the volume of fractions collected from the point of loading of the standards to the fractions showing the peak A_{280} value. The elution volume for Blue dextran was taken as the void volume (V_0, $\geq$8 MDa) for the column. The total elution volume (V_t, <20 kDa) was taken as one column bed volume (~40 ml). The molecular weight of each standard protein was plotted against V_e/V_0 for each standard protein to obtain the calibration curve as shown in Figure 2-2.

Whole cell lysate was prepared as described previously using lysis buffer (50 mM Tris pH 8.0, 150 mM NaCl, 1% IGEPAL® CA-630) supplemented with 1% protease inhibitor cocktail and applied to the column. Fresh lysate was **always** used in gel filtration experiments. I found that freezing the lysate leads to dissociation of the CECR2-containing complexes. The column was equilibrated to room temperature and 2 volumes of running buffer were run through the

column. Approximately 1 mg of the lysate in 0.3-0.5 ml was loaded into the column and immediately fractions of 0.5 ml were collected by applying gel filtration buffer to the column. A flow rate of 0.5 ml per minute was maintained during the experiment using a peristaltic pump. Collected fractions were stored at -20°C or kept on ice, either to be prepared for the next assays including immunoprecipitation, or precipitated with trichloroacetic acid (TCA) to be used in western blot analysis as described in section 2.15.
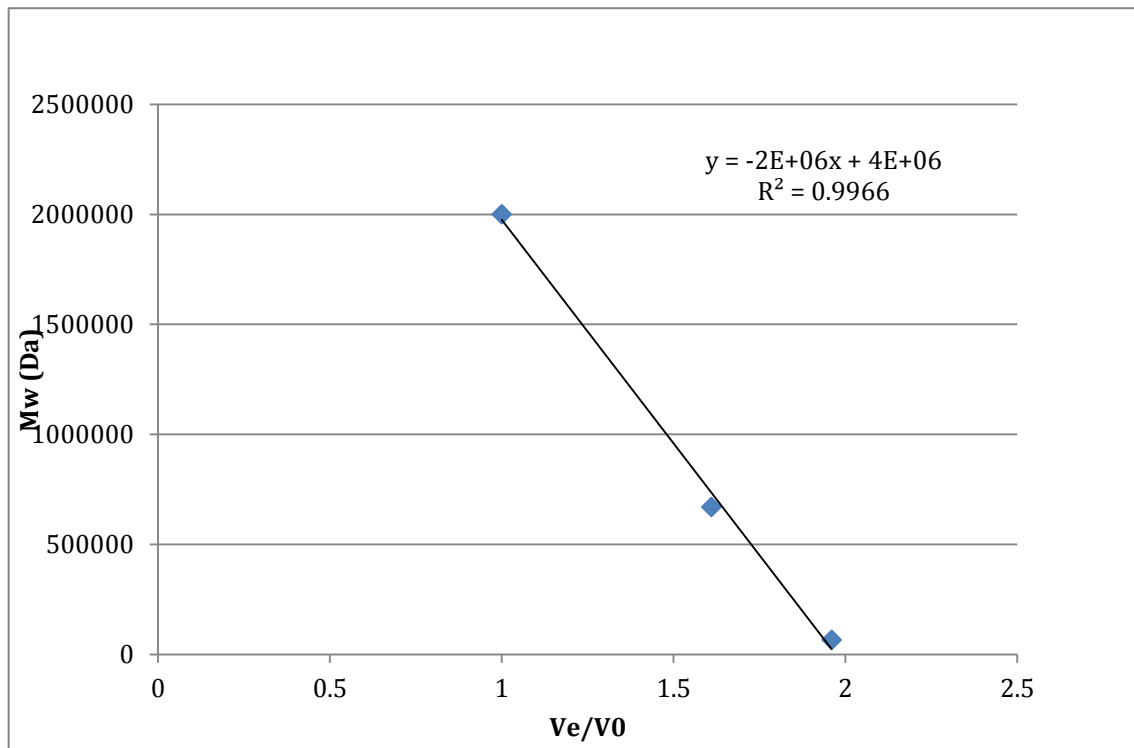
Figure 2-2 Calibration of gel filtration column for determination of protein size. A column of 50 cm in height and 1 cm in diameter was packed with Sephacryl S-400 HR. The column was calibrated by loading three size standards: 1 mg of blue dextran (2 MDa) in 0.3 ml of TBS was loaded first, immediately followed by loading a mixture of 2 mg thyroglobulin (670 kDa) and 2 mg of BSA (66 kDa). Calibration curve was obtained by plotting the molecular weight of each protein against V_e/V_0 for each protein. V_0 (the void volume): the elution volume for Blue dextran was taken as the void volume (V_0, ≥8 MDa) for the column. V_e (the elution volume): the elution volume (V_e) for each standard protein was determined by measuring the volume of fractions collected from the point of loading of the standards to the fractions showing the peak A_{280} value.

2.15. TCA precipitation

A 20% W/V solution of trichloroacetic acid (TCA) was prepared by dissolving TCA in water. An equal volume of 20% TCA was added to the protein solutions to be precipitated and incubated on ice for 30 minutes. After centrifugation at 21,000g for 15 minutes at 4°C, the supernatant was removed as much as possible and protein pellets were washed by adding 0.5 ml of ice-cold 100% acetone to the pellets and centrifuged at 21,000g for 10 min at 4°C. The supernatant was discarded, and pellets were air-dried for 10 minutes under the fumehood at room temperature. The pellets were prepared for SDS-PAGE and western blot by adding 5 μl of 4X NuPAGE™ loading buffer (Life technologies, cat. no. NP0007) and 5 μl of 1 M β-mercaptoethanol followed by boiling for 5 minutes. The samples were immediately used in SDS-PAGE or stored at -20°C.

2.16. Ethidium bromide and RNase A treatment of protein extracts

Treating lysates with ethidium bromide and RNase A before co-immunoprecipitation or gel filtration assays can eliminate the DNA and RNA dependant interactions between proteins respectively (Lai & Herr 1992, Lefebvre et al. 2002). After preparing whole cell lysate, an ethidium bromide solution (Fluka Biochemika, cat. no. 46067) was added to the lysate to a final concentration of 50 μg/ml and incubated for 1 hour on ice. After treatment with ethidium bromide, RNase (Thermo Scientific, cat. no. EN0531) was added to the solution to a final concentration of 0.33 mg/ml and incubated at room temperature for 20 minutes. The treated lysate was then used in immunoprecipitation or gel filtration experiments as described in sections 2.12 and 2.13, respectively.

2.17. Mass spectrometry

2.17.1. Crosslinking antibody to Dynabeads

10 μl of Dynabeads® Protein A (Life technologies, cat. no. 10001D) was used for each IP reaction. Dynabeads were separated from buffer using a Magnetic Separation Stand and resuspended in 200 μl of PBS containing 0.02% Tween-20. After addition of 1 μg antibody or corresponding IgG, beads were rocked for 15 minutes at room temperature to allow the antibody

to bind to the beads. The beads were separated from the supernatant and were washed with 500 μl of PBS containing 0.02% Tween-20 twice. The beads were then resuspended in 500 μl of 0.2 M triethanolamine pH 8.2 and rocked for 10 minutes at room temperature. After discarding the supernatant the beads were resuspended in 500 μl of freshly made 20 mM DMP in 0.2 M triethanolamine pH 8.2 and rocked for 30 minutes at room temperature to cross link the antibody to the dynabeads. Dynabeads were separated from the supernatant and resuspended in 1 ml of 50 mM Tris-HCl pH 7.5 and rocked for 15 minutes at room temperature to stop the cross-linking reaction. Dynabeads were separated from the Tris buffer using a Magnetic Separation Stand and washed in 500 μl of PBS containing 0.02% Tween-20. To remove uncoupled immunoglobulin, dynabeads were resuspended in 1 ml of 0.58% v/v acetic acid containing 150 mM NaCl and the supernatant was discarded immediately. This step was repeated 9 more times. The dynabeads were washed 3 times in 1 ml of PBS containing 0.02% Tween-20 three times and used in the immunoprecipitation assay or stored at 4°C up to 2 weeks.

2.17.2. Large scale IP using ES cells

ES cells were harvested from six 10 cm culture plates and whole cell lysate was prepared as described previously except for 2 experiments, one for ES cells and one for testis experiments, where I used 150 mM NaCl during the lysate preparation instead of 420 mM. 10 tubes containing 10 μl of Dynabeads® Protein-A (Life technologies, cat. no. 10001D) were cross-linked to 1 μg of CECR2 antibody as described in section 2.16.1. For a negative control, 10 tubes each containing 10 μl of Dynabeads® Protein-A were cross-linked to 1 μg of normal rabbit IgG (New England Biolabs, cat. no. 2729). Approximately 1 mg of ES cell lysate was added to each IP reaction and incubated for a range of 2 hrs to overnight (according to the optimized condition for each antibody) at 4°C on a rocker. The IP conditions were exactly the same as described in section 2.12. Beads were washed 5 times in 0.5 ml of lysis buffer. All CECR2 IP beads were pooled together in a tube and bound proteins were eluted by adding 20 μl of 4x NuPAGE™ loading buffer (Life technologies, cat. no. NP0007) and 20 μl of 1 M β-mercaptoethanol followed by heating at 95-100°C for 5 minutes. Similarly, all IgG control beads were pooled and eluted. The supernatant containing the eluted protein was separated from the beads. 2 μl of the elution was used to analyse the efficiency of the immunoprecipitation using western blot analysis. The rest of the samples were prepared for mass spectrometry analysis as described in section 2.17.4.

2.17.3. Large scale IP using testis

Testes were dissected out from the animal and the outer layer of the testes (tunica albuginea) removed in PBS on ice. Whole tissue lysate was prepared as described in section 2.9, except that 1100 µl lysis buffer was used per animal. 10 IP reactions were performed using CECR2 antibody as described for ES cells in section 2.17.2. 10 IP reactions were also performed using normal rabbit IgG as a negative control using the same conditions. 1 mg of testis lysate was used for each IP reaction.

2.17.4. In-Gel tryptic digestion

Keratin is a common contaminant that interferes with mass spectrometry experiments causing signal loss. It is very important to have a keratin-free environment as much as possible during the experiment. Keratin originates from skin, hair, nails and woolen cloths and it adheres to dust. Therefore, to avoid keratin contamination, all the steps hereafter were performed in a biological safety cabinet. Protective clothing (clean lab coat, powder-free nitrile gloves, hair net and a facemask) was used to minimize the exposure of bare skin to the biological safety cabinet. All the pipettors used for the experiment were wiped with ethanol.

To prepare immunoprecipitated CECR2-containing complexes for mass spectrometry, the In-Gel Tryptic Digestion Kit (Thermoscientific, cat. no. 89871) was used with some modifications as follows. After performing large scale IP reactions as described in sections 2.17.2 and 2.17.3, the samples were subjected to SDS-PAGE analysis. To minimize keratin contamination SDS-PAGE was performed using pre- cast gels in a biological safety cabinet. The samples were run approximately 10 mm or all the way down (~5 cm) into a 4-20% precast SDS-PAGE gel (BioRad, cat. no. 456-1094). The gel was stained with Coomassie Blue as described in section 2.11. The gel containing all the proteins was cut into bands with the width of ~ 5 mm and placed in eppendorf tubes. 200 µl of destaining solution (25 mM ammonium bicarbonate in 50% acetonitrile) was added to each tube and the mixture was agitated at 37°C for 30 minutes at room temperature. The destaining solution was removed from the gel pieces and destaining was repeated. The samples were reduced by adding 30 µl of freshly prepared reduction buffer (50 mM TCEP in digestion buffer) to the tube and incubated at 60°C for 10 minutes. After cooling the samples, the reduction buffer was removed and the samples were alkylated by adding 30 µl of freshly prepared alkylation buffer (500 mM iodoacetomide in digestion buffer) to the tube and

incubated for 60 minutes in the dark at room temperature with agitation. After removing the alkylation buffer, the samples were washed by adding 200 µl of destaining buffer and incubated at 37°C for 15 minutes with agitation. The washing step was repeated once more. To shrink the gel pieces 50 µl of acetonitrile was added to each tube and the mixture was incubated for 15 minutes at room temperature. After removing acetonitrile, the gel pieces were air-dried for 5 minutes. The gel pieces were then covered by adding 10 µl of activated trypsin solution (Pierce cat. no. 1862748) (~10 ng/µl in digestion buffer) and incubated for 15 minutes at room temperature. 25 µl of digestion buffer (25 mM ammonium bicarbonate in water) was added to each tube and the mixtures were incubated at 37°C for 4 hours with vigorous shaking. The digestion solution was transferred into new tube. The gel pieces were covered with 30 µl of 50% acetonitrile in 0.2% trifluoroacetic acid (Sigma-Aldrich, cat. no. T6508) and vortexed for 20 minutes at room temperature. The solution was collected from the mixture and transferred to the same new tube containing the digestion solution. The gel pieces were covered again with 30 µl of 50% acetonitrile in 0.2% trifluoroacetic acid, vortexed for 20 minutes at room temperature and then the solution was collected from the mixture and added to the same new tube containing the elutes.

2.17.5. Purification of peptides after In-gel digestion and LC-MS/MS

The peptide samples obtained from in-gel digestion were purified using C18 Spin Columns (Pierce, cat. no. 89873). To clean the peptides using the spin column, the peptide samples were completely dried using a *Savant* SpeedVac *SC 110A* Concentrator at the MBSU. 60 µl of 0.5% trifluoroacetic in 5% acetonitrile was added to each tube. 200 µl of activation solution (50% acetonitrile) were applied to the spin columns and the columns were centrifuged at 1500g for 1 minute. The activation step was repeated once more and the flow-through was discarded. 200 µl of equilibration solution (0.5% trifluoroacetic in 5% acetonitrile) was added to each spin column and the columns were centrifuged at 1500 g for 1 minute. The equilibration step was repeated once more and the flow-through was discarded. Peptide samples were loaded into the spin columns followed by centrifugation at 1500 g for 1 minute. The flow-through was loaded into the spin columns again and after centrifugation at 1500 g for 1 minute, the flow-through volumes were discarded. The spin columns were washed by adding 200 µl of wash solution (0.5% trifluoroacetic in 5% acetonitrile) followed by centrifugation at 1500 g for 1 minute. The

wash was repeated. The cleaned peptides were eluted twice by adding 20 μl of elution buffer (0.1% formic acid in 50% acetonitrile) to each spin column and collecting the flow-through by centrifugation at 1500 g for 1 minute. The eluted peptides were completely dried using a *Savant SpeedVac SC 110A* Concentrator at the MBSU. The peptides were dissolved in 0.1% formic acid and were sent to the Alberta Proteomics and Mass Spectrometry Facility in the Department of Biochemistry or the Mass Spectrometry Facility in the Department of Chemistry at the University of Alberta for liquid chromatography tandem mass spectrometry (LC-MS/MS) identification of the bound proteins. In some of the experiments the Alberta Proteomics and Mass Spectrometry Facility performed the in-gel digestion.

2.17.6. The analysis of the mass spectrometric data

The analysis of the mass spectrometric data was performed using Proteome Discoverer 1.4 (Thermo Scientific). The generated result was searched against Uniprot (uniprot.org) Mus database using SEQUEST (Thermo Scientific) with the parameters including a precursor mass tolerance of 10 ppm and a fragment mass tolerance of 0.8 Da. Carbamidomethyl cysteine as a static modification and oxidized methionine and deamidated glutamine and asparagine as dynamic modifications were set for all searches. Results were also searched against a customized database including the sequences of CECR2, SNF2H, SNF2L, LUZP1 and RUVBL1.

Gene ontology analysis was done using PANTHER database (http://www.pantherdb.org/) for identification of the enrichments in three categories, including protein class, molecular function and biological process.

2.18. Immunofluorescence staining of neural tube sections

2.18.1. Embryo collection

To obtain embryos of a specific embryonic stage, a male and two females of the desired genotype for *Cecr2* were set up for timed mating. To monitor successful mating and to know the beginning of pregnancy, females were tested for the presence of a vaginal copulatory plug each morning. The morning of positive plug testing was considered as day 0.5 of the embryonic stage (E0.5). At the specific embryonic stage, the pregnant females were euthanized by cervical dislocation. The embryos were dissected out of the uterus and the amnion was separated from the embryo to be used for DNA extraction and genotyping purposes. The embryos were stored at -

20°C to be used for or prepared directly for histological experiments as described in section 2.18.2. For studying the embryos during the neural tube closure histological sections were prepared from *Cecr2* wild-type and homozygous *Cecr2$^{Tm1.1Hemc}$* mutant embryos at E9.5 at different developmental stages of 12-17 somites.

2.18.2. Preparing histological sections from embryos

E9.5 embryos were collected as described in section 2.18.1 and immediately transferred to ice-cold PBS. The embryos then were transferred to 4% paraformaldehyde (PFA) in PBS and incubated on ice for 90 minutes. The embryos were then washed 3 times for 5 minutes in PBS, followed by 5 minutes in 25% ethanol, 5 minutes in 50% and finally 5 minutes in 70% ethanol. All ethanol solutions were diluted in 1X PBS. After the last wash, the embryos were transferred into 70% ethanol diluted in water and stored at -20°C until proceeding to paraffin embedding.

To embed the embryos, the samples were submitted to the microscopy unit in Biological Sciences for overnight processing into paraffin using a Fisher Histomatic Tissue Processor (Model 166). After automatic processing, the embryos were placed in embedding moulds containing melted paraffin in desired orientations, and then allowed to cool down at room temperature. After solidification of paraffin, the embryos now in paraffin blocks were sectioned at 5 μm thickness using a using a microtome (Reichert-Jung 2040). Sections were floated in a water bath at 40°C and mounted on Superfrost Plus Microscope Slides (Fisher Scientific cat. no. 12-550-15). The slides were dried at 37°C overnight and then stored at room temperature until used for immunofluorescence or apoptosis experiments.

2.18.3. Immunofluorescence staining of embryos

Sections prepared from wildtype *Cecr2* and homozygous *Cecr2$^{tm1.1Hemc}$* mutant embryos were prepared as described in section 2.18.2. The embryo sections on microscopic slides were deparaffinized by washing 3 times in toluene for 5 minutes for each wash, and then rehydrated by washing 2 times in 100% ethanol for 10 minutes each, 2 times in 95% ethanol for 10 minutes each and finally 2 time in milliQ water for 5 minutes. The slides were heated in antigen retrieval solution (10 mM Tris Base, 1 mM EDTA, 0.05% Tween 20, pH 8.0) until boiling. Antigen retrieval was done three times with 10 minutes cooling period after each boiling step. The sections were blocked in 10% normal goat serum (Sigma, cat. no. G9023) and 0.6% Triton X-100 (Sigma, cat. no. T8787) dissolved in PBS for 60 minutes at room temperature. The sections were

incubated in primary antibody diluted in antibody dilution buffer (1% bovine serum albumin, 0.3% Triton X-100 in PBS) at 4°C overnight inside a humidifying chamber. The primary antibody dilutions were as follows: purified rabbit polyclonal anti-CECR2 diluted to 1:10,000 and rabbit polyclonal anti-SHH (Santa Cruz Biotechnology, cat. no. sc-9024) diluted to 1:1000. The sections were washed 3 times in PBS for 5 minutes and incubated in secondary antibody (AlexaFluor 488 goat anti-rabbit, Life Technologies, cat. no. A11008) diluted to 1:200 in antibody dilution buffer (1% bovine serum albumin, 0.3% Triton X-100 in PBS) for 2 hours at room temperature in the dark. The sections were then stained with 0.1% DAPI in PBS for 5 minutes followed by 3 washes for 5 minutes in PBS. The slides were then mounted with Fluoromount G (SouthernBiotech, Cat. no. 0100-01) and analyzed with a Nikon Eclipse 80i confocal microscope with a CVI Melles Griot Ion Laser and using NIS-Elements v4.0 software.

2.19. Apoptosis assay (TUNEL)

Sections were prepared from wild-type *Cecr2* and homozygous *Cecr2^{tm1.1Hemc}* mutant embryos at E9.5 embryos at the 16-18 somite stage as described in section 2.18.2. The embryo sections on microscopic slides were deparaffinized by washing 3 times in xylene for 5 minutes for each wash, and then rehydrated by washing 2 times in 100% ethanol for 5 minutes, once in 95% ethanol for 3 minutes, once in 70% ethanol for 3 minutes and finally once in PBS for 5 minutes. The ApopTag Plus Peroxidase In Situ Apoptosis Detection Kit (Millipore, cat. no. S7101) was used to detect apoptotic cells. 20 ug/ml of proteinase K (Invitrogen) in water was directly added to the sections and incubated for 15 minutes at room temperature. The sections were washed twice in MQ water for 2 minutes. 3% hydrogen peroxide (Sigma, cat. no. 216763) in PBS was directly added to the sections with 5 minutes incubation at room temperature followed by 2 times washes in PBS for 5 minutes. The sections were incubated at room temperature after adding Equilibration Buffer (Millipore, cat. no. 90416) for 5 minutes. TdT enzyme (Millipore, cat. no. 90418) was applied on the sections and the sections were incubated for 60 minutes at 37°C inside a humidifying chamber. The slides were then incubated in Stop/Wash buffer (Millipore, cat. no. 90419) for 10 minutes at room temperature. After 3 washes in PBS for 1 minute, the Anti-Digoxigenin Peroxidase (Millipore, cat. no. 90420) was applied on the sections and the sections were incubated for 30 minutes at room temperature inside a humidifying chamber. After 4 washes in PBS for 2 minute, the Peroxidase Substrate containing 3

μl of DAB Substrate (Millipore, cat. no. 90423) in 147 μl of DAB Dilution Buffer (Millipore, cat. no. 90424) was applied on each section and the slides were incubated at room temperature and constantly monitored under a light microscope for monitor color. The slides were washed 3 times for 1 minute and once for 5 minutes in MQ water. The sections were counterstained in 0.5% (w:v) methyl green for 10 minutes at room temperature. The slides were dipped 10 times in MQ water. This was done twice followed by another wash in MQ water for 30 seconds. The slides were then washed by 10 times by dipping in 100% N-butanol. This was done twice followed by another 30-second wash in 100% N-butanol. The sections were dehydrated by washing 3 times in xylene for 2 minutes and mounted under a glass coverslip with DPX mounting medium. The slides were studied under a Carl Zeiss Axio Scope.A1 microscope and Optronics camera with PictureFrame software.

2.20. ChIP-sequencing

2.20.1. Preparation of chromatin from ES cells

ES cells from a 100 mm culture plate (70-80% confluence) were used for each IP reaction. Media was discarded and fresh media containing 1% formaldehyde (Thermo Scientific, cat. no. 28908) was added to the cells. The plate was agitated for 10 minutes at room temperature to crosslink proteins to DNA. To stop the crosslinking, freshly made glycine solution was added to the plates to a final concentration of 125 mM and agitation was continued for 5 minutes at room temperature. Cells were washed twice with 10 ml of ice-cold PBS for 5 minutes. PBS was discarded as much as possible. To harvest the cells, 2 ml of ice-cold PBS was added to the plates and cells were scraped using a plastic cell scraper and transferred to a 15 ml tube (Falcon). Plates were rinsed with 4 ml of PBS to collect the remaining cells, which were added to the falcon tube. The cell suspension was centrifuged at 1000 rpm for 5 minutes at room temperature. Cells were resuspended in 1 ml of ice-cold PBS, transferred to a 1.5 ml tube and centrifuged at 3500 g for 5 minutes. To prepare cell lysate, the supernatant was discarded and the cell pellet obtained from each 10 cm plate was suspended in 300 μl of RIPA 500 lysis buffer (50 mM Tris pH 8.0, 500 mM NaCl, 1% IGEPAL, 0.1% SDS, 1 mM EDTA, 0.5%, sodium deoxycholate) supplemented with 1% protease inhibitor cocktail (Sigma-Aldrich, cat. no. I3786) and rocked at 4°C for 30 minutes. The cell lysate, in a crushed ice water bath at 4°C, was sonicated for 40 cycles of 30 sec

ON/ 30 sec OFF at high setting with a Bioruptor UCD-200 ultrasound sonicator to get DNA fragments at 200-500 bp. Tubes were briefly vortexed and centrifuged after each run of 10 cycles. Ice cubes were added to the sonicator water bath every 10 cycles. Samples were centrifuged at 20,000 g for 20 minutes at 4°C and the supernatants were transferred into new tubes to be used in the chromatin immunoprecipitation assay. 10% of the chromatin preparation was transferred into another tube (~30 μl) as the input for the ChIP experiment. The chromatin solution can be stored at -20°C, but it is better to use fresh sample for chromatin immunoprecipitation.

2.20.2. Preparation of chromatin from testis

Testes were dissected out from the animal and the tunica albuginea were removed in PBS on ice. Each testis tissue was placed in separate containers containing ice-cold PBS. The testis tissue was completely minced with a new razor blade in 1 ml of PBS. Then PBS containing the minced testis tissue was transferred to a new 2 ml tube and centrifuged at 6000 rpm for 5 minutes at room temperature. After discarding the supernatant, the pellet was resuspended in 1700 μl of PBS containing 1% formaldehyde (Thermo Scientific, cat. no. 28908) and rocked for 15 minutes at room temperature to crosslink proteins to DNA. Crosslinking was performed as described in section 2-20-1. The suspension was centrifuged at 6000 rpm for 5 minutes at room temperature. After discarding the supernatant, the pellet was washed with PBS for 5 minutes twice. The rest of the procedure, preparation of cell lysate and sonication, was performed as described in section 2-20-1 except for using 300 μl of RIPA 500 lysis buffer for each testis.

2.20.3. Checking the efficiency of sonication

To reverse the crosslinking, 5 μl of the sonicated chromatin was mixed with 20 μl of water and 2 μl of 5 M NaCl and boiled for 5 minutes. 10 μl of the sample was electrophoresed in a 2% agarose gel containing 0.1 μg/ml of ethidium bromide in 1X TAE buffer (40 mM Tris-acetate, 1 mM EDTA) at 130 V for 30 minutes along with a nucleic acid marker. The DNA smear was visualized by a UV fluorescence gel imager (Alpha Innotech).

2.20.4. Chromatin immunoprecipitation

Immunoprecipitation was performed as described in section 2.13 except that here 20 μl of Dynabeads® Protein A and 250 μl (500-1000 μg) of sonicated chromatin (prepared according to the procedure in section 2.20.1 and 2.20.2) was used for each IP reaction. IP reactions were

rocked overnight at 4 °C. Following the overnight incubation, beads were washed 5 times for 5 minutes in 0.5 ml of RIPA 500 lysis buffer. To elute immunoprecipitated chromatin, beads were transferred to a clean tube and resuspended in 120 µl of elution buffer (100 mM NaHCO3, 1% SDS, freshly added 20 mM DTT) and vortexed for 15 minutes at setting 1 at room temperature. The supernatant containing the eluted chromatin was separated from the beads. The eluted chromatin was stored at -20°C or used directly in the next step.

2.20.5. Reverse crosslinking of DNA-protein complexes

5 µl of 5 M NaCl was added to each elution obtained from each chromatin immunoprecipitation reaction and boiled for 5 minutes to reverse DNA-protein crosslinking. As for the input sample, 70 µl of water and 5 µl of 5 M NaCl were added to the 50 µl input samples set aside in section 2.20.1 and 2.17.2 and boiled for 5 minutes. 4 µl of 25 mg/ml Proteinase K (~0.8 mg/ml) was added to each ChIP and input reaction and the tubes were incubated at 65 °C overnight for protein digestion. 1 µl of 10 mg/ml RNase A (Thermo Scientific, cat. no. EN0531) was added to each reaction and incubated at 37°C for 10 minutes. The ChIP and input samples were boiled for 5 minutes to denature the Proteinase K and RNAse A.

2.20.6. Purifying the immunoprecipitated DNA (Phenol:Chloroform:Isoamyl Alcohol extraction)

For each antibody, 2 reactions were prepared and after reverse crosslinking the replicates were mixed to give a higher concentration. Each reaction volume was brought up to 400 µl by adding elution buffer (100 mM NaHCO3, 1% SDS, freshly added 20 mM DTT). Each sample was mixed with equal volume (400 µl) of Phenol:Chloroform:Isoamyl alcohol (25:24:1) (Life technologies, cat. no. 15593-031) until an emulsion was formed. After centrifugation at 12,000 g for 3 minutes at room temperature the aqueous phase (upper layer) was transferred into a clean tube. The aqueous phase was mixed with an equal volume of chloroform and centrifuged at 12,000 g for 3 minutes at room temperature. The separated upper phase was transferred into a clean tube and the DNA fragments were precipitated as follows.

The DNA solution was mixed with 1/10 volume of 3 M sodium acetate. Glycogen (Thermo Scientific cat. no. R0561) was added to a final concentration of 1 ug/ µl and then gently mixed with 2.5 volumes of 100% ethanol. The mixture was incubated at -20°C for up to 60 minutes. After centrifugation at 10,000 rpm for 15 minutes at room temperature and discarding

the supernatant, the pellet was washed with 500 µl of 70% ethanol followed by centrifugation. The pellet was air-dried, dissolved in 30 µl of nuclease free water and stored at -20°C

2.20.7. Quality control by Bioanalyzer

Before building the library, the immunoprecipitated DNA and input DNA were quantified and quality-assessed using the Agilent High Sensitivity DNA kit (Agilent Technologies, cat. no. 5067-4626) on an Agilent 2100 Bioanalyzer and Qubit following the manufacturer's protocol. Briefly, after allowing reagents to equilibrate to room temperate for 30 minutes, 15 µl of the High Sensitivity DNA Dye Concentrate was added to the High Sensitivity DNA Gel Matrix vial and the mix was vortexed for 10 seconds to properly mix the gel and dye. The gel-dye mix was passed through a spin filter and centrifuged at room temperature at 2240 g for 10 minutes. 9 µl of gel-dye mix, 5 µl of the High Sensitivity DNA marker, 1 µl of the High Sensitivity DNA ladder and 1 µl of each sample (the immunoprecipitated and Input DNA) were loaded into marked wells of a High Sensitivity DNA Chip. The loaded chip was agitated on an IKA vortex mixer at 2400 rpm for 1 minute. The chip was inserted in the Agilent 2100 Bioanalyzer and the 2100 Expert software was started at the computer connected to the Bioanalyzer. After finishing, the data were saved as PDF.

2.20.8. Preparing libraries

Thirteen libraries (ChIP samples and an input DNA) were made using the NEXTflex ChIP-Seq kit (Bioo Scientific, Cat. no. 5143-01) by Delta Genomics in Edmonton as follows:

To prepare each library, 1-10 ng of immunoprecipitated DNA or input DNA were used. The immunoprecipitated and input DNA fragments-with-overhangs were repaired to obtain blunt-ended DNA. To do this, for each sample, 1-10 ng of DNA fragments were mixed with blunting mix and brought up to 50 µl by adding nuclease free H2O in a PCR tube. The mixtures were incubated on a thermocycler at 22°C for 30 minutes. In order to select DNA fragments between 200-400 bp, a gel-free size selection clean-up kit was used. Then, the 3' end of the DNA fragments were adenylated by incubating a mixture of 16 µl of DNA fragments and 4.5 µl of ChIP Adenylation Mix buffer at 37°C for 30 minutes in a thermocycler for each sample. After adenylation of the 3' ends, DNA adapters were ligated to adenylated DNA fragments. For each sample, 20.5 µl of adenylated DNA fragments, 27.5 µl of ChIP ligation mix and 2 µl of ChIP Adapter were mixed and incubated on a thermocycler at 22°C for 15 minutes. Adapter sequences

are available in Appendix A. The DNA fragments were cleaned again as mentioned previously by a gel-free size selection clean-up kit. DNA fragments were eluted in 38 µl of Resuspension Buffer and 36 µl of the supernatant was transferred into a clean tube and used in the PCR Amplification step.

Each PCR reaction contained 36 µl of the adapter ligated DNA fragments, 12 µl of ChIP PCR Master Mix and 2 µl of ChIP Primer Mix (see Appendix A for sequences). The cycling conditions were: initial heating at 98°C for 2 minutes, followed by 11 cycles of 98°C for 30 seconds, 65°C for 30 seconds, 72°C 60 seconds, and a final extension at 72 °C for 4 minutes. The PCR reactions were cleaned using the same gel-free size selection clean-up kit and DNA fragments (libraries) were eluted in 30 µl of Resuspension Buffer. Libraries were qualified and quantified by Agilent 2100 Bioanalyzer and High Sensitivity D1000 ScreenTape before sequencing as described in section 2.20.5.

For optimal cluster density, qPCR was used to quantitate DNA library templates utilizing the KAPA Library Quantification Kit (Kapabiosystems cat. no. KK4827) according to the manufacturer's instruction by Delta Genomics in Edmonton. Briefly, the purified libraries were diluted in library dilution buffer (10 mM Tris-HCl, pH 8.0 + 0.05% Tween 20) with dilution factors of 1:1000, 1:2000, 1:4000 and 1:8000 and vortexed for 10 seconds. The qPCR was done in 20 µl reaction containing 12 µl of KAPA SYBR® FAST qPCR Master Mix with Primer Premix, 4 µl of diluted library DNA or DNA standard (1-6) and 4 µl of PCR-grade water. Each reaction including the six standards (supplied in the kit) and each diluted library DNA was performed in triplicate. A real-time thermocycler was used. The cycling began with an initial heating at 95°C for 5 minutes, followed by 35 cycles of 95°C for 30 seconds and 60°C for 45 seconds. The concentration of each library was calculated as indicated in the following table:

Library name	Conc. In pM calculated by qPCR instrument (triplicate date)			Avg. conc. (pM)	Size adjusted concentration (pM)	Concentration of undiluted library stock (pM)
Library 1:1000	A1	A2	A3	A	$A \times \dfrac{452}{\text{Avg. fragment length}} = W$	W x 1000

The calculated concentrations of the library dilutions were obtained from the instrument and a size adjustment calculation was performed for the size difference between the average

length of the library and the DNA standard (452 bp). The concentration of the undiluted libraries was calculated by applying the dilution factors.

2.20.9. High-throughput sequencing

The sequencing of the libraries was performed using the Illumina NextSeq® 500 System at the Molecular Biology Service Unit at the University of Alberta in collaboration with Delta Genomics in Edmonton.

2.20.9.1.   Denaturing and Diluting Libraries for the NextSeq System

The 13 individually bar-coded libraries were combined in equal concentrations and the final concentration was adjusted to 4 nM by adding an appropriate volume of water. 5 µl of library mix was combined with 5 µl of freshly made 0.2 N NaOH, briefly vortexed and then centrifuged at 280 g for 1 minute. The mixture was incubated at room temperature for 5 minutes to denature the libraries into single strands. 5 µl of 200 mM Tris-HCl, pH 7 was added to the mixture and after brief vortexing it was centrifuged at 280 g for 1 minute. To obtain the concentration of 20 pM, 985 µl of Hybridization Buffer (HT1) was added to the tube of denatured libraries and after brief vortexing it was centrifuged at 280 g for 1 minute. To obtain the final volume of 1300 µl of the loading concentration of 1.5 pM, 97.5 µl of denatured library solution was combined with 1202.5 µl of HT1. The solution was inverted several times and then centrifuged and stored on ice.

2.20.9.2.   Denaturing and Diluting PhiX Control

A PhiX library as a 10% spike-in was used as a sequencing control. In a 2 ml tube, 10 µl of 10 nM PhiX (Illumina, cat. no. FC-110-3002) was mixed with 15 µl of Resuspension Buffer (RSB) (Illumina, cat. no. FC-110-3002) to obtain a total volume of 25 µl at a final concentration of 4 nM. The mixture was briefly vortexed and then centrifuged. To denature the PhiX library, in a 2 ml tube, 5 µl of 4 nM PhiX library was mixed with freshly prepared 0.2 N NaOH and after vortexing the mixture was incubated at room temperature for 5 minutes. The mixture was then briefly vortexed and centrifuged at 280 g for 1 minute. 5 µl of 200 mM Tris-HCl, pH 7 was added to the mixture and the mixture was vortexed followed by centrifugation at 280 g for 1 minute. To obtain 1 ml of 20 pM of PhiX library, 985 µl of prechilled HT1 was added to the tube.

2.20.9.3. Combining the Combined Libraries and PhiX Control

A PhiX library as a 10% spike-in was used as a sequencing control. 1.2 μl of denatured and diluted PhiX control (20 pM) was combined with 1299 μl of denatured and diluted library mix (1.5 pM) and the mixture was loaded onto the reagent cartridge of the sequencer.

2.20.9.4. Sequencing

The sequencing of the libraries was performed using the Illumina NextSeq® 500 System at Molecular Biology Service Unit at the University of Alberta using NCS v2.02. After the sequencing was finished, the raw data was available as FastQ files.

2.21. ChIP-sequencing data analysis

2.21.1. Peak calling using a model based analysis of ChIP-seq (MACS) and visualization

The FASTQ data obtained from the sequencing step were aligned to NCBI m37 mouse assembly (mm9) using Burrows-Wheeler Aligner (BWA) program in collaboration with Dr. Paul Stothard. Since the majority of the published data were analyzed by mm9, I also used mm9 instead of the more recent mm10 to be able to compare my results with the published data. BWA was used for its ability to map paired-end reads and its ultrafast processing feature (Li & Durbin 2009). All the subsequent analyses were done after integrating the two technical replicates. To predict the genomic regions where the chromatin immunoprecipitated proteins bind, a peak caller program, Model-based Analysis of ChIP-Seq (MACS v1.4.2), was used using the following parameters, band width=300 and P-value cut off of 1.00e-05 for the ES cells dataset and band width=300 and P-value cut off of 1.00e-03 for the testis dataset (http://liulab.dfci.harvard.edu/MACS/). MACS finds regions on the genome with significant numbers of aligned reads compared to the control samples (Zhang et al. 2008). For the ES cells dataset, INPUT control and for the testis IgG controls were used as background.

The BEDTools suite (V2.23.0) was used to merge the biological replicates and to generate overlapping peaks for CECR2, SNF2H and LUZP1 (http://bedtools.readthedocs.io/en/latest/index.html) (Quinlan & Hall 2010). The BED format files produced by MACS were used in the Integrative Genomics Viewer (IGV, version 2.3) for

visualizing peak position on the mm9 genome assembly (Robinson et al. 2011). The Integrative Genomics Viewer was also used for peak graph production.

2.21.2. Genomic distribution and Gene Ontology (GO) analysis

The Genomic Regions Enrichment of Annotation Tool (GREAT) was used to carry out gene ontology enrichment analysis and to find the distance to nearest TSS (http://bejerano.stanford.edu/great/public/html/index.php). The genomic location of the peaks (from MACS) was submitted to GREAT (McLean et al. 2010). To evaluate the enrichments I used the default association rule [-5 kb to +1 kb of TSS plus an extension (up to 1 Mb) and curated regulatory domains for each gene] and default significance thresholds (region-based binomial fold=2, region-based binomial FDR= 0.05, gene-based hypergeometric FDR=0.05).

To find the genes bearing the binding sites of each protein or the overlapping binding sites in the promoter region, the DNA sequence of promoter regions (TSS to 5 Kb upstream) of the Mus musculus (mm9) assembly was obtained from the UCSC Table browser as a BED file format (http://genome.ucsc.edu/index.html). Overlap between the sequences of the promoters and the binding sites obtained from ChIP-seq data were generated using bedtools intersect.

Genomic features (TSS, intron, exon, promoter) associated with peaks were determined by the Cis-regulatory Element Annotation System (CEAS) (Shin et al. 2009) in the Galaxy/Cistrome platform (http://cistrome.org/ap/root) (Liu et al. 2011).

2.21.3. Motif analysis

Analysis of DNA sequence to identify enriched motifs within the binding sites were performed on the ChIP-seq data using MEME-ChIP, which is designed for motif analysis in large ChIP-seq datasets (http://meme-suite.org/) (Machanick & Bailey 2011). In order to use MEME-ChIP the 250 bp sequences flanking each side of the peak summit were extracted as FASTA data using the peak coordinates in the Galaxy platform using a web browser (https://usegalaxy.org/). The repeat sequences were removed manually with Microsoft excel and the remaining unique sequences were used as inputs for MEME-ChIP for motif discovery. MEME-ChIP was run with default settings and to identify known motifs the JASPAR CORE vertebrates motif database and the UniPROBE mouse motif database were used. Motifs with a P-value $< 10^{-5}$ were selected for further analysis. PANTHER, which accepts gene symbols as input, was used for the Gene Ontology analysis of the identified transcription factors (http://pantherdb.org/) (Mi et al. 2016).

For the unknown novel motifs the genome coordinates from corresponding binding sites were used in GREAT for GO analyses.

2.21.4. Conservation of the ChIP-seq peaks

Since functional genomic regions are conserved, the evolutionary conservation of DNA sequences of the ChIP-seq peaks were compared with flanking regions. The higher evaluation conservation of the peak regions is an indicator of good quality of the ChIP-seq data. The 250 bp sequences flanking each side of the peak summit in BED files were used in Conservation Plot tool in the Galaxy/Cistrome platform (http://cistrome.org/ap/root) (Liu et al. 2011). The PhastCons conservation scores were used to obtain the average conservation score profiles.

3. Results

Aim 1: Composition of CECR2-containing complexes

3.1. Previous attempts to produce Anti-CECR2 antibodies

To study the CECR2 protein the McDermid lab tried to raise anti-CECR2 antibodies from peptides numerous times in the past with no success. Five peptide antibodies were generated previously in rabbits (Figure 3-1). Four peptides (Pep A, Pep B, Pep C and Pep D) predicted by ProtScale (http://web.expasy.org/protscale/) to be on the surface of the native mouse CECR2 were selected as immunogens and injected into rabbits in the Science Animal Support Services (SASS) facility at the University of Alberta. None of the rabbit anti-peptide antibodies were able to detect a CECR2-specific protein in Western blot analysis, as determined by comparing normal tissue to Cecr2-mutant tissue (Tanya Ames, unpublished data).

In another attempt, a rabbit antibody was generated against a peptide (Pep WB) near the C-terminal end of mouse CECR2 by Washington Biotechnologies (Figure 3-1). This antibody preparation was also not successful in detecting CECR2 protein in Western blot analysis (Tanya Ames, unpublished data).

3.2. Characterization of chicken polyclonal antibody to mouse CECR2 peptide

After the failure of the rabbit anti-peptide antibodies, two unique CECR2 peptide antibodies were generated in chicken by GenScript Corporation. After purification of the total IgY from egg yolks, I affinity-purified the antibodies against the peptides. The purified antibodies were used in Western blot analysis on nuclear extracts prepared from wild-type, $Cecr2^{Gt45Bic}$ homozygous mutant (Gt/Gt) and $Cecr2^{Tm1.1Hemc}$ homozyogous mutant (Del/Del) embryos to test the specificity of the chicken antibodies for mouse CECR2. $Cecr2^{Gt45Bic}$ is a genetrap with ~14-fold reduction in $Cecr2$ transcripts in Gt/Gt neurulating embryos and $Cecr2^{Tm1.1Hemc}$ is an engineered deletion of the first exon which shows a ~200 fold reduction in

Cecr2 expression in Del/Del embryos heads at the time of neurulation. Since E14.5 mouse embryos had been previously shown to express *Cecr2* (Banting et al. 2005), the nuclear extracts were prepared from embryos to test the antibody preparations. The extract prepared from *Cecr2^{tm1.1Hemc}* Del/Del mutants was used as the negative control for the presence of specific CECR2 bands in Western blot analysis. The affinity-purified chicken antibodies did not detect any specific band for CECR2 in wild-type samples compared to mutant samples (Figure 3.2).
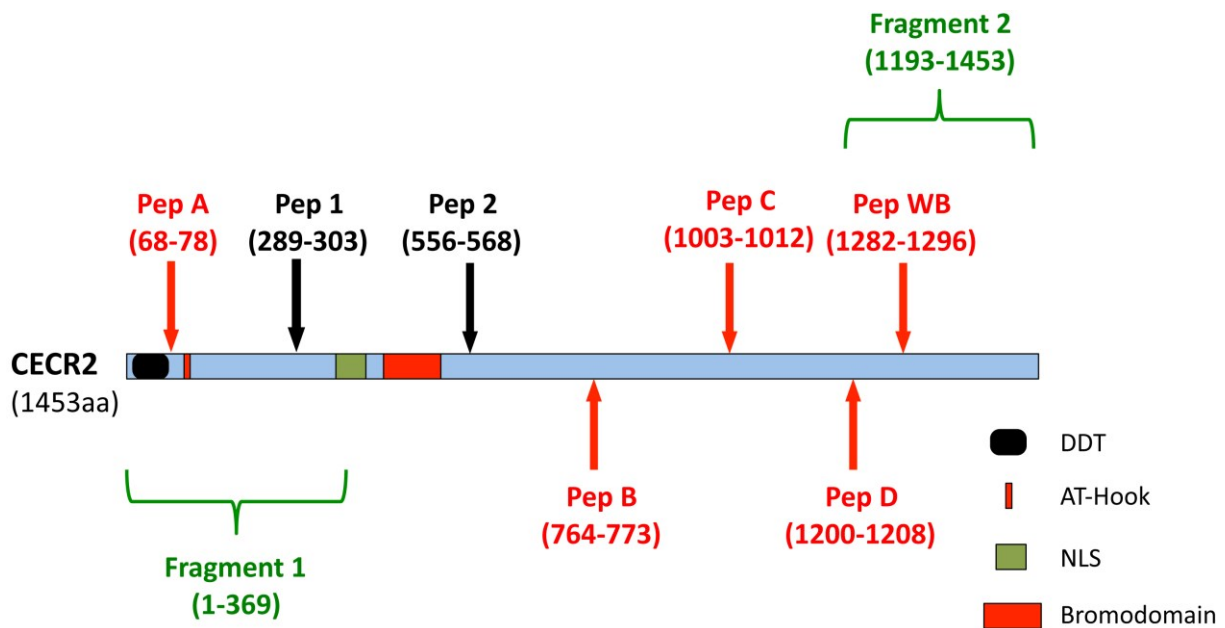
Figure 3-1 Location of peptides and fragments used for raising CECR2 antibodies. The positions of the functional domains – DNA binding homeobox and different transcription factors (DDT), AT-Hook, bromodomain and nuclear localization signal (NLS) - are shown on the CECR2 protein. The positions of the epitopes for the previous unsuccessful rabbit antibodies for CECR2 are shown in red. The positions of the epitopes for chicken antibodies for CECR2 are shown in black. Fragment 1 and Fragment 2 (green) are the fragments of CECR2 used in this study in raising CECR2 antibodies in rabbit. The numbers show the amino acid sequence positions on mouse CECR2 (MGI ID: OTTMUSP00000028044).
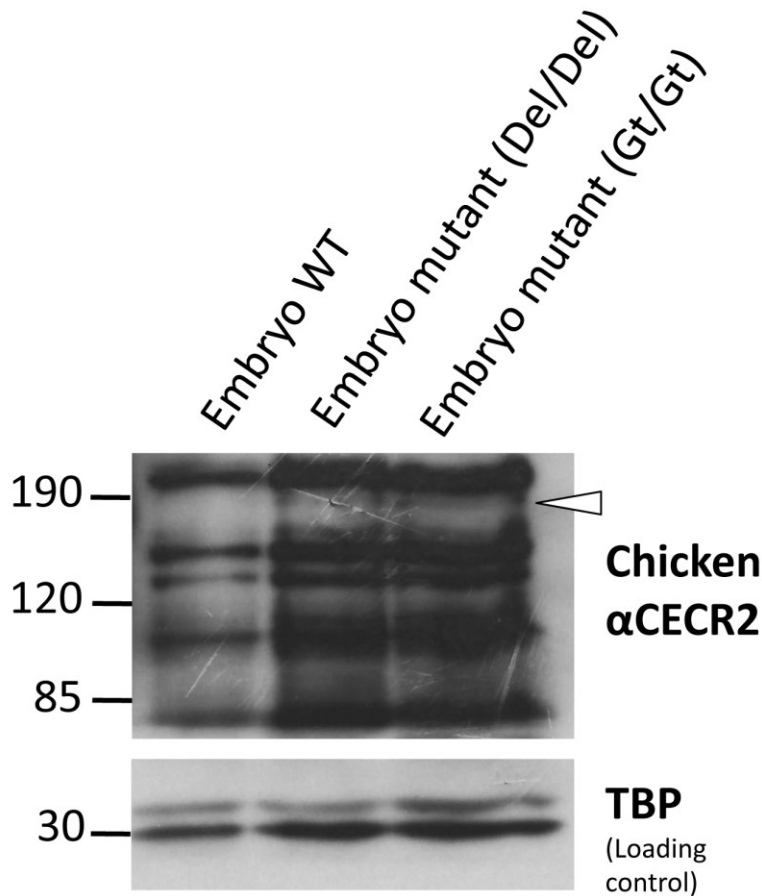
Figure 3-2 The chicken anti-CECR2 did not detect any specific band for CECR2. Nuclear extracts were prepared from 12.5 dpc whole mouse embryos and were used in Western blot analysis using the chicken anti-CECR2. No specific CECR2 band is detected in wild type embryos (WT) compared to $Cecr2^{Gt45Bic}$ homozygous mutant (Gt/Gt) and $Cecr2^{Tm1.1Hemc}$ (Del/Del) embryos. TBP was detected as a loading control. The molecular weight standards are shown as kDa on the left side of the figure. Arrowhead shows the predicted region where CECR2 would be detected.

3.3. Analysis of rabbit polyclonal antibodies to mouse CECR2 fragments

The results from testing different antibodies generated against peptides in various parts of CECR2 indicated that the peptide antibodies may not be a good approach to raising antibodies against CECR2. None of the peptide antibodies were able to detect any specific band for CECR2 in Western blot analysis under the conditions tested.

In my project I was addressing questions about the CECR2 protein and the CECR2-containing complexes that had been previously identified using the CECR2 fusion protein (CECR2^{Gt45Bic}) (Banting et al. 2004, Thompson et al. 2012). Therefore, I needed an antibody that

would specifically bind to CECR2 protein in its denatured form for Western blot analysis and to its native form for biochemical assays including immunoprecipitation (IP) and chromatin immunoprecipitation (ChIP). Because of the failure of all peptide antibodies, I decided to generate an antibody to mouse CECR2 using big fragments of CECR2 as the immunogens to raise antibody in rabbits.

To generate a new CECR2 antibody, two fragments of CECR2 were selected (Figure 3-1). To avoid cross-reactions, the bromodomain was avoided when selecting the protein fragments. Fragment #1 spanned exon 1 to the end of exon 9 and contained an N-terminal His-tag and Fragment #2 spanned the middle of exon 17 to the end of the CECR2 protein and contained a C-terminal His-tag. Both were cloned in pET-21a (Novagen cat. no. 69740-3) and were transformed into *E. coli* BL21 (DE3) (Appendix B). Sequencing confirmed that the vectors contained the *Cecr2* fragments. Western blot analysis using an antibody for histidine (Genscript Corporation, cat. no. A00174) showed that the expression of recombinant CECR2 fragments at the appropriate sizes was successful (Figure 3-3). Purified recombinant CECR2 fragments were used to immunize rabbits. Post-immune blood samples were tested for the presence of CECR2 specific antibodies using ELISA (Figure 3-4) and Western blot analysis using the purified CECR2 fragments.

To select the best antiserum, the four post-immune sera were used in Western blot analysis to detect the CECR2 protein. The Western blot was performed on nuclear extracts prepared from E12.5 wild-type and *Cecr2*$^{Tm1.1Hemc}$ mutant embryos. qRT-PCR on heads of E9.5 embryos previously showed that the there was 200-fold decrease in *Cecr2* transcript in *Cecr2*$^{Tm1.1Hemc}$ mutants (Fairbridge et al. 2010), making it a suitable candidate to be used in judging the specificity of the detected bands by the CECR2 antibody. Only post-serum from M10-3 detected a band at a size of ~170 kDa in the wild-type (Figure 3-5). This band was absent in the mutant. Because this band was only present in wild-type and because it's size was very close to the calculated molecular weight of mouse CECR2 (158 kDa), it was concluded that this represents mouse CECR2. The detection of the CECR2 band showed the specificity of the antibodies present in the post-immune serum obtained from M10-3. To decrease the background and remove the non-specific bands seen in Western blot analysis (Figure 3-5), the antibody was affinity purified according to the protocol explained in the Materials and Methods section. All further CECR2 experiments in this thesis were performed using this affinity purified antibody.
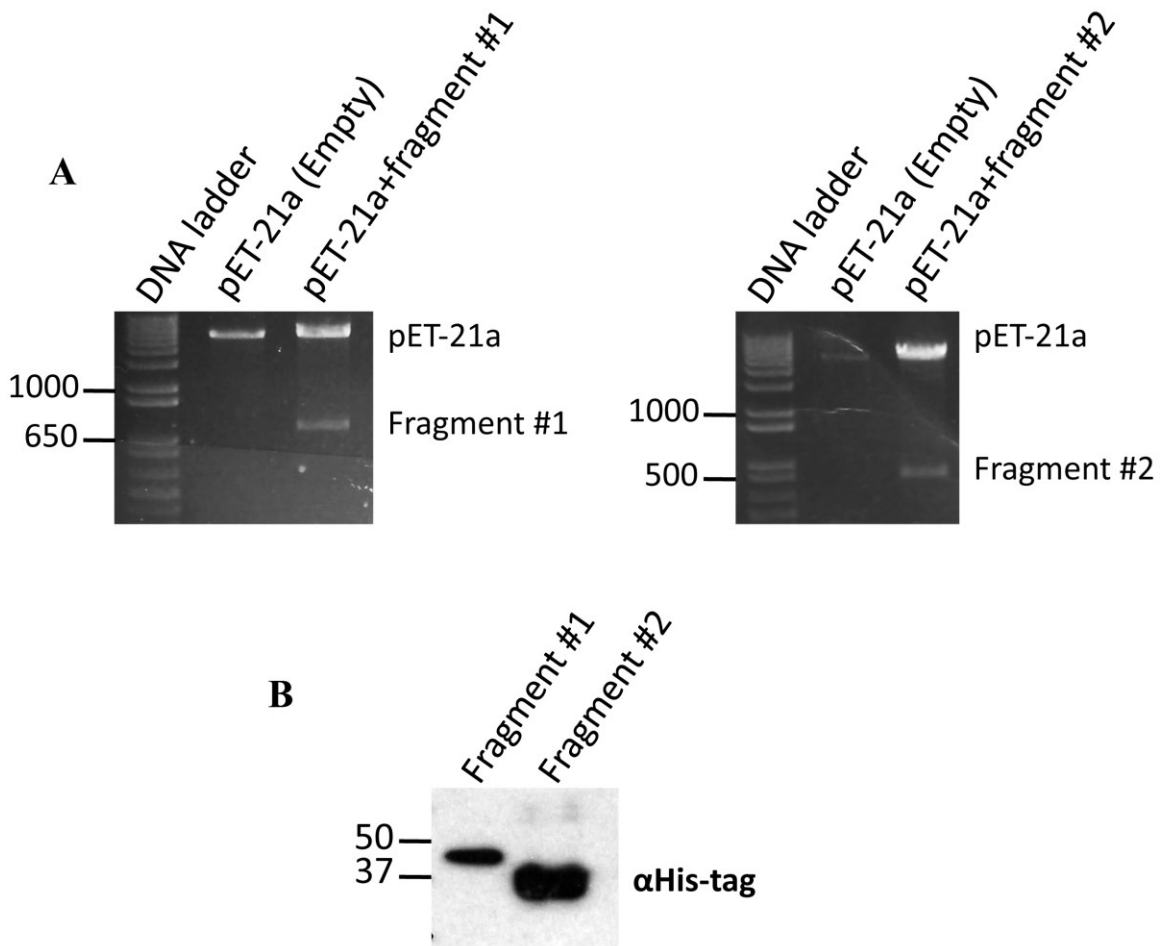
Figure 3-3 *Cecr2* fragments were successfully cloned and expressed in pET-21a. A) Each fragment was amplified with Sall and EcoRI restriction sites and an N or C terminal His[6] tag incorporated into the fragments. Amplicons were then cloned into pET-21a. Cutting the vector with Sall and EcoRI shows the presence of *Cecr2* fragments in the expression vector. Sequencing the vectors confirmed the presence *Cecr2* fragments. B) Expression of recombinant CECR2 fragments in *E. coli*. After inducing the expression of CECR2 fragments in E. coli, bacteria lysates were prepared and subjected to western blot analysis using an antibody for the His-tag. Fragment #1 includes exon 1-9 (predicted size of 45 kDa with His-tag) and fragment #2 includes exon 17-19 (predicted size of 30 kDa with the His-tag). The molecular weight standard are shown as kDa on the left side of the figure.
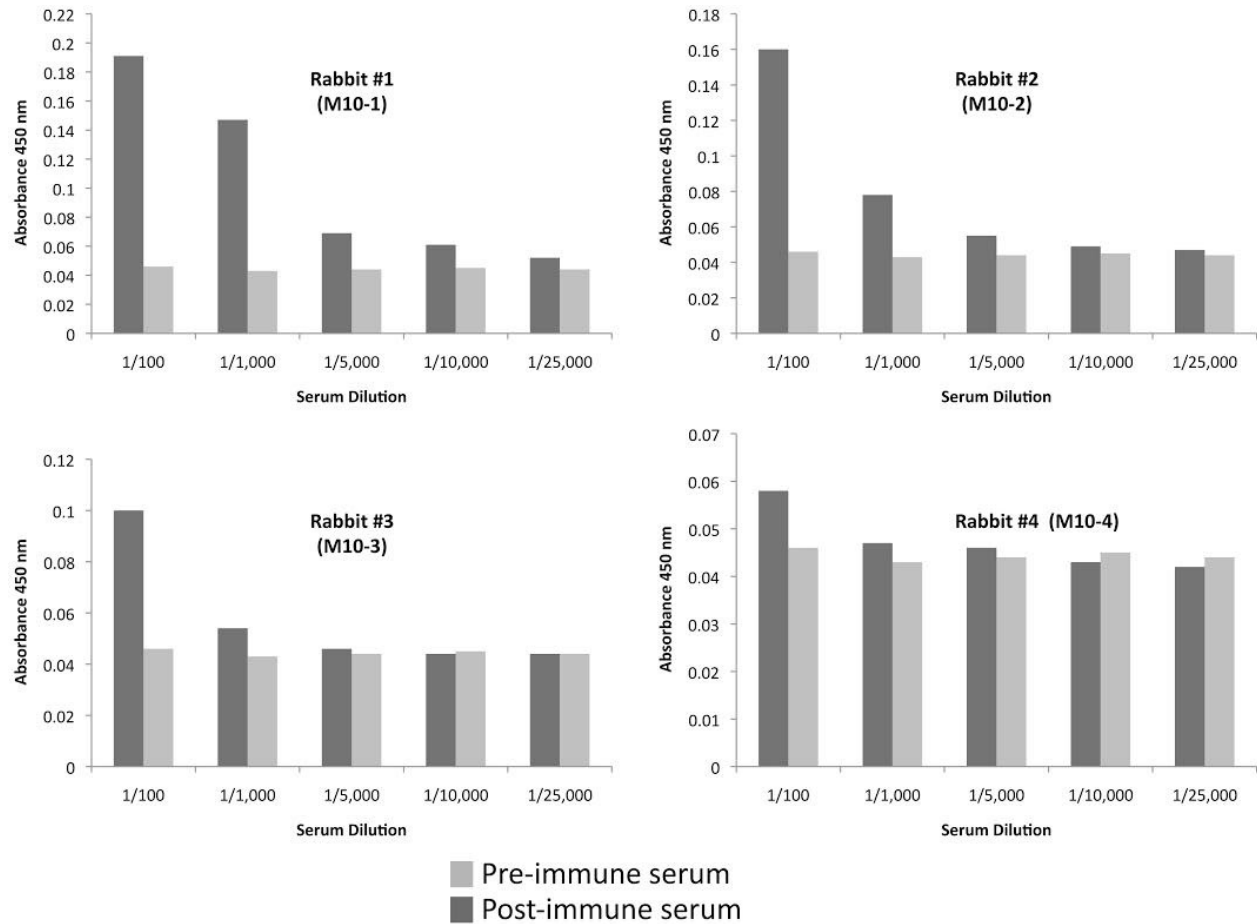
Figure 3-4 Absorbance values of post-immune sera compared to pre-immune sera at different dilutions. ELISA plates were coated with recombinant CECR2 protein fragment #1 or #2 and incubated in serial dilutions of serum prepared from first blood samples of rabbits (M10-1, M10-2, M10-3 and M10-4) followed by incubation in anti-rabbit IgG-HRP. The secondary antibody was then detected by treating the plates with tetramethylbenzidine (TMB) substrate and recording the absorbance value at 450 nm. Each graph represents the OD 450 of pre-immune serum and post-immune serum of each rabbit. The absorbance values of post-immune sera are higher than that of pre-immune sera indicating the presence of functional antibodies against each of the recombinant fragments in the post-immune sera.

Western blot analysis using the affinity purified CECR2 antibody resulted in decreased background (Figure 3-6A). This also confirmed that there was not any detectable CECR2 in the mutant embryos. Exposing the same blot to film for a longer period of time detected an extra band approximately 10 kDa smaller than the CECR2 band in homozygous $Cecr2^{Tm1.1Hemc}$ mutant embryos (Figure 3-6B). The same band can be seen in wild-type tissues when electrophoresed for a longer period of time. This band could be a non-specific band or could be a low-abundance

isoform of CECR2 present in both wild-type and mutant embryos, not affected by the exon 1 deletion. This band will be discussed in more detail in section 3.5.

To find the optimal concentration of the affinity purified antibody, whole cell extracts were prepared from TT2 ES cells and used in a Western blot analysis using different concentrations (1:5000, 1:10000, 1:20000, 1:50000) of the CECR2 antibody. As shown in figure 3-7, a very strong band was detected at approximately 170 kDa in all of the antibody concentrations, which shows the purity and good quality of the antibody in detecting CECR2 protein in its denatured form.
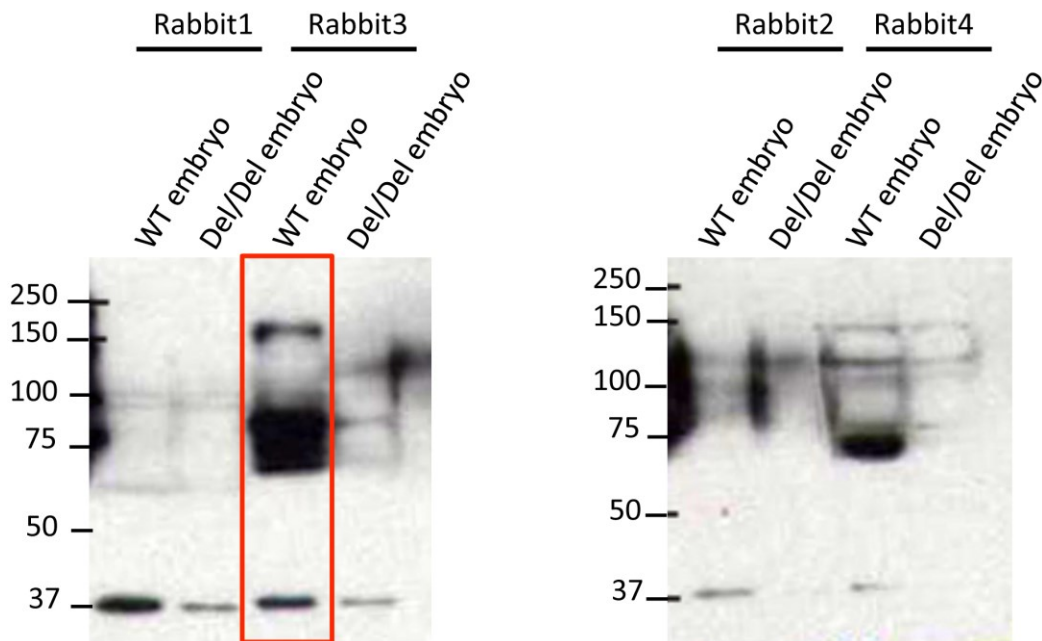
Figure 3-5 Post-immune serum (fragment 17-19) from rabbit #3 was able to detect CECR2 protein at the molecular weight of ~170kDa in western blot analysis (shown inside the red rectangle). Nuclear extracts prepared from E12.5 whole wild-type and *Cecr2*^{Tm1.1Hemc} (Del/Del) embryos were visualized by western blot analysis using post-immune sera from four rabbits. TATA-binding protein (TBP) was used as a loading control. The molecular weight standards are shown as kDa on the left side of the figure. The smaller bands detected in the lane inside the red rectangle are nonspecific bands that were not detected using affinity-purified CECR2 antibody.
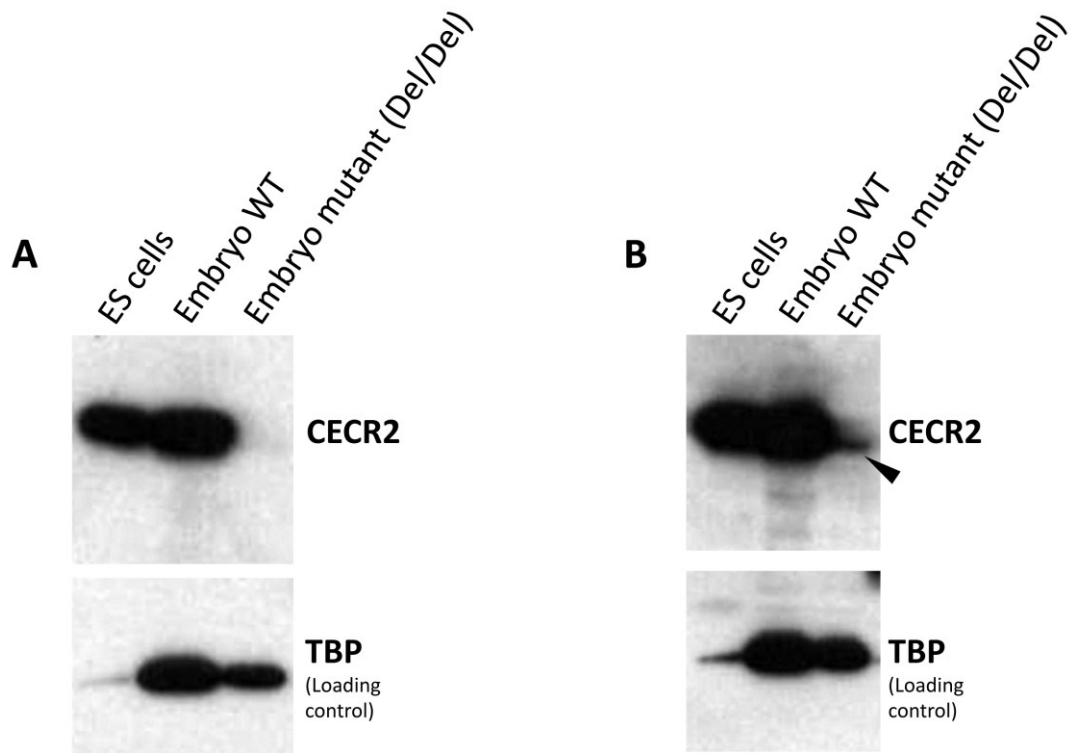
Figure 3-6 Western blot analysis showing the specificity of the rabbit polyclonal CECR2 antibody. A) Protein extracts were prepared from ES cells and 12.5 dpc embryos and subjected to Western blot analysis using the affinity purified antibody. A specific band was detected very close to the calculated molecular weight (158 kDa) in ES cells and wild-type embryos, which was not detected in homozygous $Cecr2^{Tm1.1Hemc}$ mutant (del/del) embryo. B) The same blot was exposed to film for a longer period of time. Arrowhead shows an extra band detected after longer exposure both in wild-type and mutant embryos.
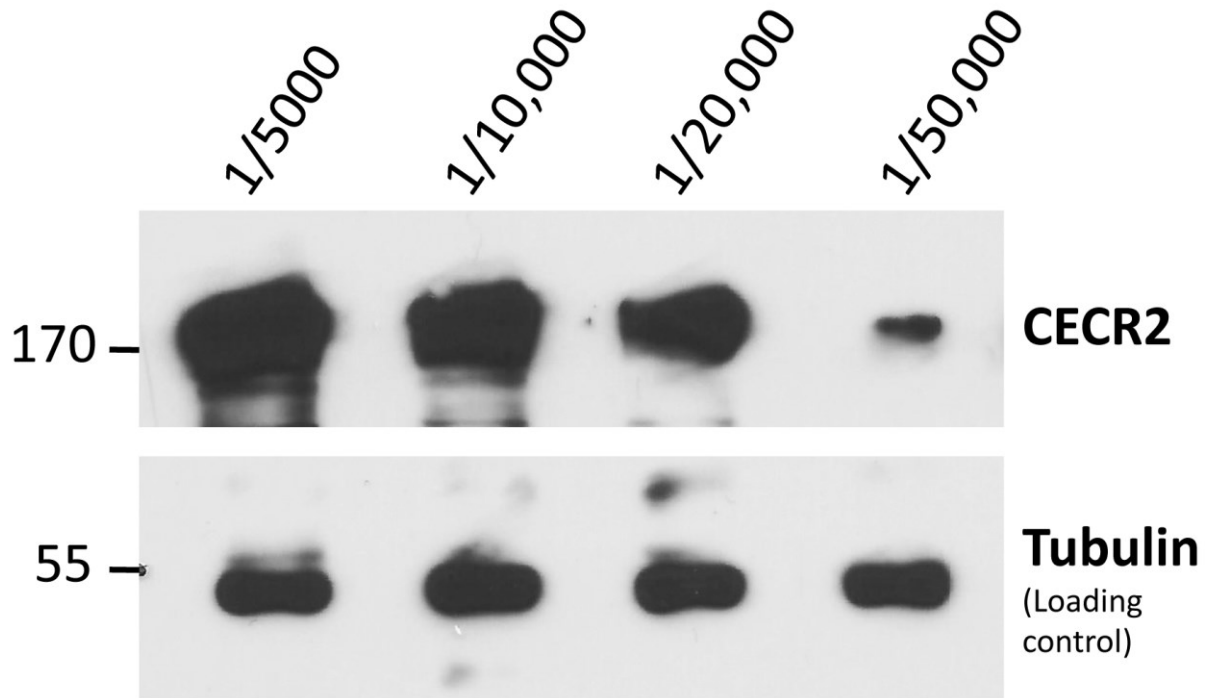
Figure 3-7 Affinity purified antibodies from rabbit M10-3 detects a very strong band for CECR2 in ES cell extracts. To obtain the best concentration of the CECR2 antibody, Western blot was performed on whole cell extract from ES cell using different concentrations of the antibody. The concentrations are indicated at the top of the blot. Tubulin was used as a loading control. The molecular weight standard are shown as kDa on the left side of the figure.

3.4. Tissue distribution of CECR2

The purified rabbit antibody was used in Western blot analysis to determine the presence of CECR2 in various tissues. Whole cell lysates were prepared from ES cells as well as wild-type and *Cecr2*$^{Tm1.1Hemc}$ homozygous mutant tissues, including E12.5-E14.5 embryos, adult testis, adult ovary, adult liver and adult kidney (Figure 3-8). A band at ~170 kDa was detected in ES cells and embryos as expected. From the adult tissues, the same band was only detected in testis and ovary. The absence of this band in *Cecr2*$^{Tm1.1Hemc}$ homozygous mutant tissues demonstrated that the detected band in wild-type embryo and testis is CECR2. *Cecr2* expression in adult liver and adult kidney was below the detection sensitivity of Western blot analysis (Figure 3-8).

Western blot analysis also detected a CECR2 band in our insect SF9 MIB-CECR2 cell line, but not in the control SF9 cells. SF9 MIB-CECR2 carries a human *Cecr2* gene construct, indicating that our antibody cross-reacts with human CECR2 and can be used in human related research projects (Figure 3-8).
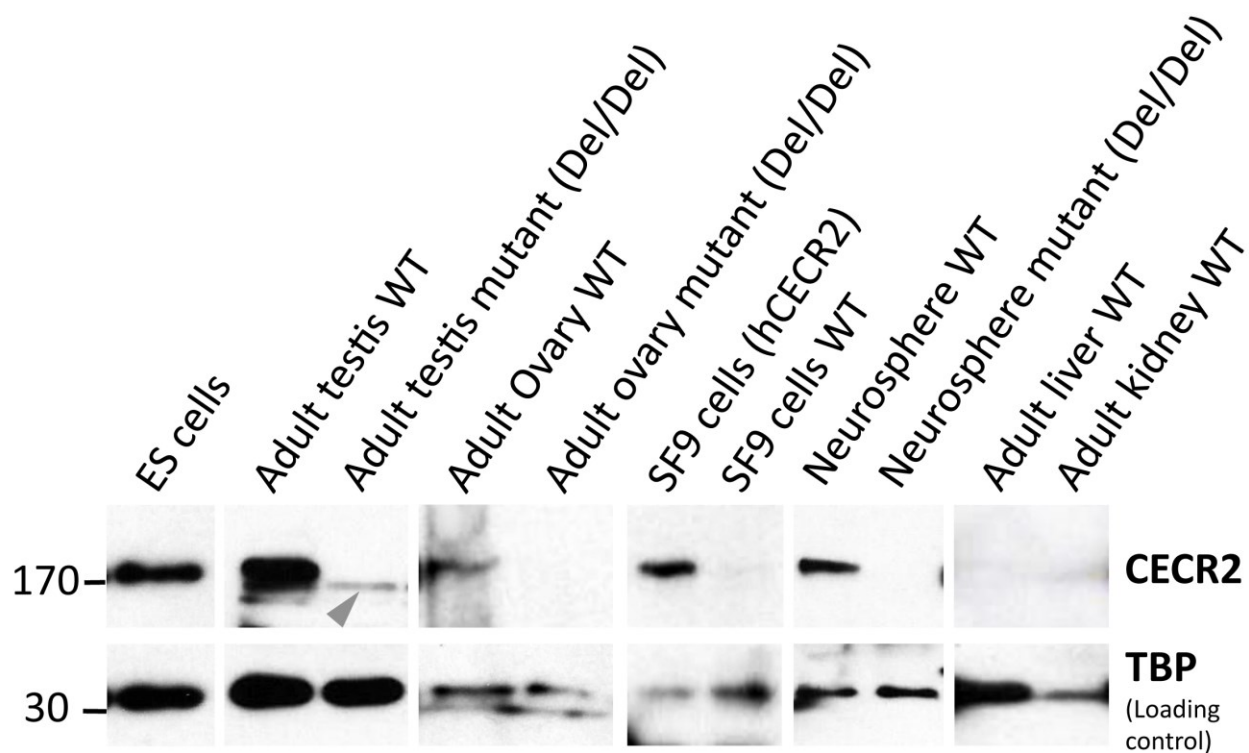
Figure 3-8 CECR2 expression in different tissues. Protein extracts from different tissues were used in a western blot analysis to detect the presence of CECR2 protein using the purified CECR2 antibody. The ES cell line is TT2. Neurospheres originated from E13.5 mouse embryos. The SF9 insect cell line (hCECR2) was transfected with human *Cecr2* (Graham Banting, unpublished). The mutants (Del/Del) are all homozygous *Cecr2^{Tm1.1Hemc}*. TBP was used as a loading control. The molecular weight standards are shown as kDa on the left side of the figure. Arrowhead shows the extra band (possibly a CECR2 isoform) detected in both wild-type and mutant testis.

Western blot analysis using protein extracts from different parts of the adult brain showed the expression of CECR2 in cerebellum (Figure 3-9), which is consistent with the expression of CECR2^{Gt45Bic} fusion protein as shown by X-gal staining in the brain of a *Cecr2^{Gt45Bic}* homozygous mutant (Figure 3-10). CECR2 was also detected in wild-type neurospheres isolated from E13.5 embryos (Figure 3-8). X-gal staining of the adult brain also shows the expression of *Cecr2^{Gt45Bic}* in the dentate gyrus of the hippocampus (Figure 3-10).
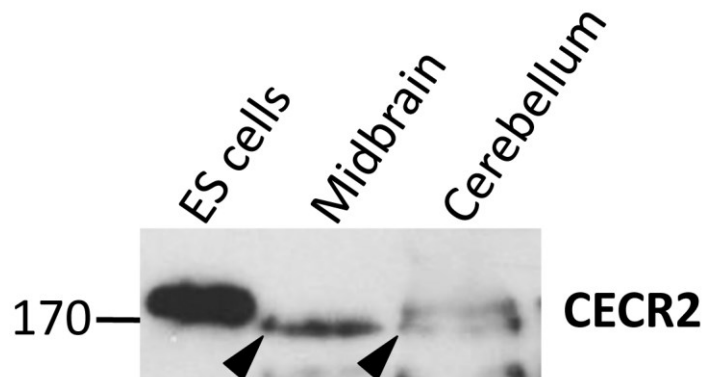
Figure 3-9 CECR2 is detected in cerebellum of adult mouse. Protein extracts from cerebellum and midbrain of adult mouse were used in a western blot analysis to detect the presence of CECR2 protein. ES cells were used as a control for the comparison of the sizes of the bands. CECR2 band was detected in ES cells as expected and in cerebelleum. Arrowheads show an extra band on cerebellum and midbrain, which is smaller than CECR2. The molecular weight standard is shown as kDa on the left side of the figure.
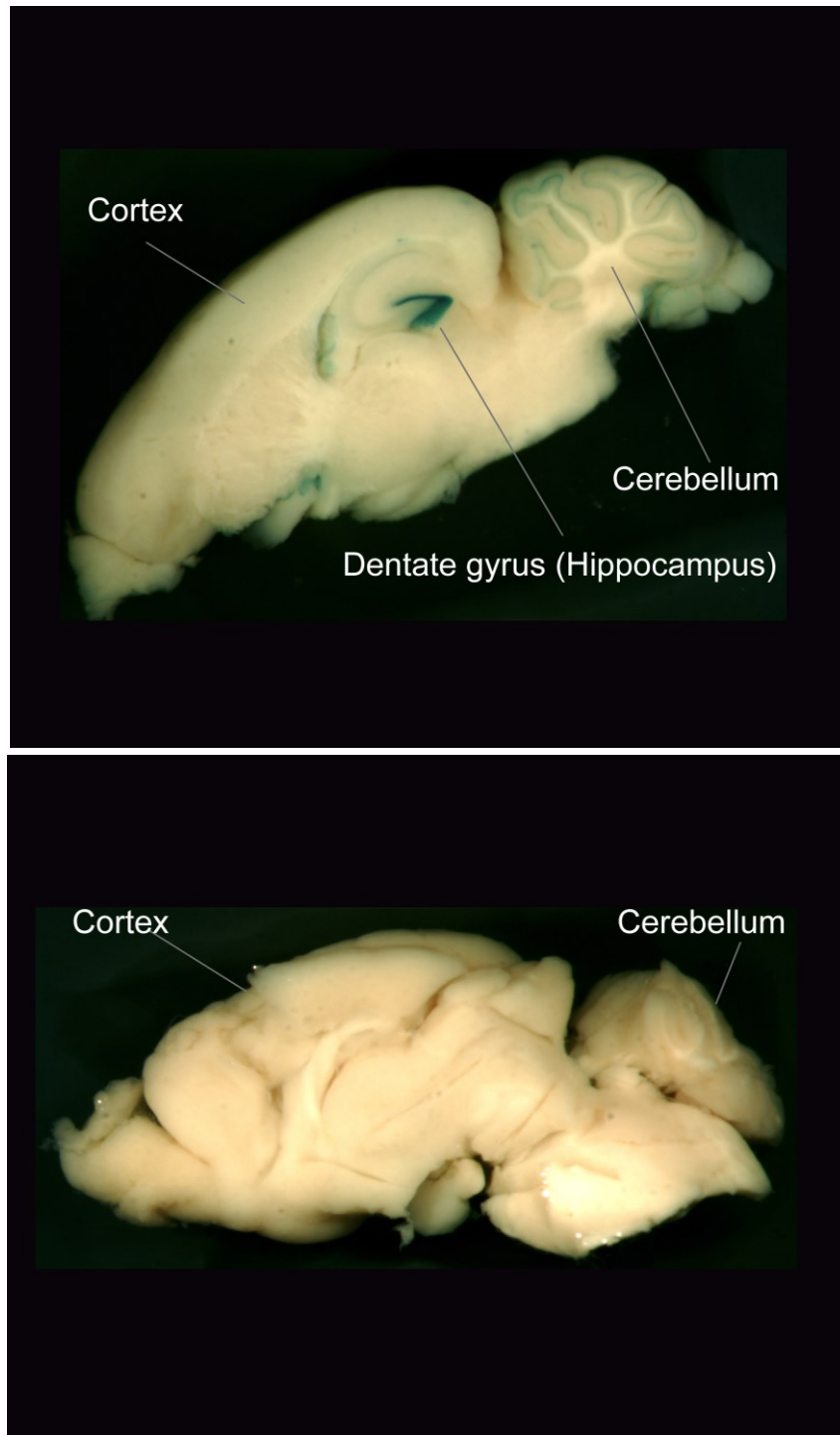
Figure 3-10 *Cecr2*^{Gt45Bic} **is expressed in dentate gyrus (hippocampus) and cerebellum of adult mouse brain.** Brains of *Cecr2*^{Gt45Bic} heterozygous (top) and wild-type (bottom) adult mice were sectioned in half and stained with X-gal to reveal the presence of CECR2^{Gt45Bic} in the brain.

All the subsequent experiments in this project were performed using ES cells and adult testes. Since collecting embryos during the neural tube closure stage is very difficult and is almost impossible to collect enough tissues to do biochemical experiments, I picked ES cells to look for the composition of the CECR2 complex and its genomic targets. The testis is one of the few adult tissues with strong expression of *Cecr2*, and *Cecr2*[tm1.1Hemc] mutation leads to subfertility in both sexes on the FVB/N strain indicating the involvement of *Cecr2* in reproduction. To study the CECR2 complex in adult testes, I used the FVB/N strain, as all the BALB/c *Cecr2*[tm1.1Hemc] mutants die perinatally.

3.5. Possible low abundance CECR2 isoform

Western blot analysis regularly showed a fainter band slightly smaller than the "canonical" CECR2 in embryos and testis (Figure 3-6B and 3-8). This band was present in protein extracts prepared from all genotypes including the *Cecr2*[Gt45Bi] homozygous and *Cecr2*[Tm1.1Hemc] homozygous mutants. This band was also detected in midbrain and cerebellum of adult mouse (Figure 3-9) and neurospheres (data not shown). The same band is the only band detected in the immortalized GC1 spermatogonia B cell line (ATCC, cat. no. CRL-2053) (Figure 3-11). This cell line was originally used to test the expression of CECR2 in spermatogonia. Immunofluoroscence using the CECR2 antibody showed that in adult testis CECR2 is limited to spermatogonia (Figure 3-12) (Norton, unpublished data). Morphological observation of cells expressing *Cecr2* in adult testis indicated that CECR2 is only in spermatogonia type A cells located in the margins of the seminiferous tubules. This was supported by a lack of a canonical CECR2 band in GC1 cells.

This smaller band could be a product of the cross-reactivity of the CECR2 antibody with a non-specific polypeptide or it could be a low-abundance isoform of CECR2 that is present in the mutants. To determine the presence of any possible transcript of *Cecr2* in the mutant mouse, cDNA was synthesized using the total RNA obtained from homozygous *Cecr2*[tm1.1Hemc] embryos (E9.5 and E14.5) and was used in RT-PCR reactions using the primers in various combinations. Since exon 1 has been deleted in the *Cecr2*[tm1.1Hemc] mutant allele, exon 1 was not expected to be detected. Interestingly, the data from the RT-PCR assay indicated that there is a transcript of *Cecr2* in mutants containing all of the *Cecr2* exons except for the first exon (Figure 3-13). This

transcript might be transcribed from an unidentified alternate upstream exon and translated to a polypeptide that is detected in Western blot analysis of mutant tissues. Our rabbit
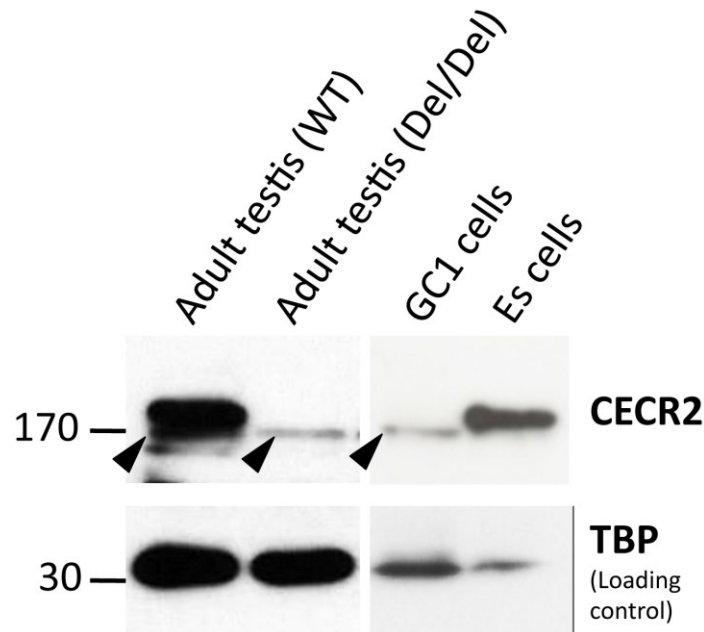
Figure 3-11 A possible low abundance CECR2 isoform or non-specific cross-reacting protein in adult testis. Protein extracts were prepared from ES cells, wild-type and homozygous *Cecr2^{Tm1.1Hemc}* adult testis (Del/Del) and mouse GC1 cell line (Type B spermatogonia immortalized by transfection with pSV3-neo). Extracts were used in a Western blot analysis to detect the presence of CECR2 protein. ES cell line is TT2 cell line. Arrowheads show the non-canonical CECR2 band. TBP was used as a loading control. The molecular weight standards are shown as kDa on the left side of the figure.

polyclonal antibody was raised against the C-terminal section of CECR2; therefore the presumptive CECR2 isoform could be detected in Western blot analysis using the same antibody. CECR2 contains a DDT domain that is located in exons 1 and 2. It has been shown that the DDT domain is found in all the large non-catalytic components of ISWI complexes and it binds to SNF2H (Dong et al. 2013, Eberharter et al. 2004). The lack of exon 1 in the CECR2 isoform means that the DDT domain is disrupted and the CECR2 isoform probably could not form a complex with SNF2H. Interestingly, the smaller isoform was never detected in the big CECR2-containing complexes as seen when calculating the CECR2 complex size using gel filtration (Figure 3-19).
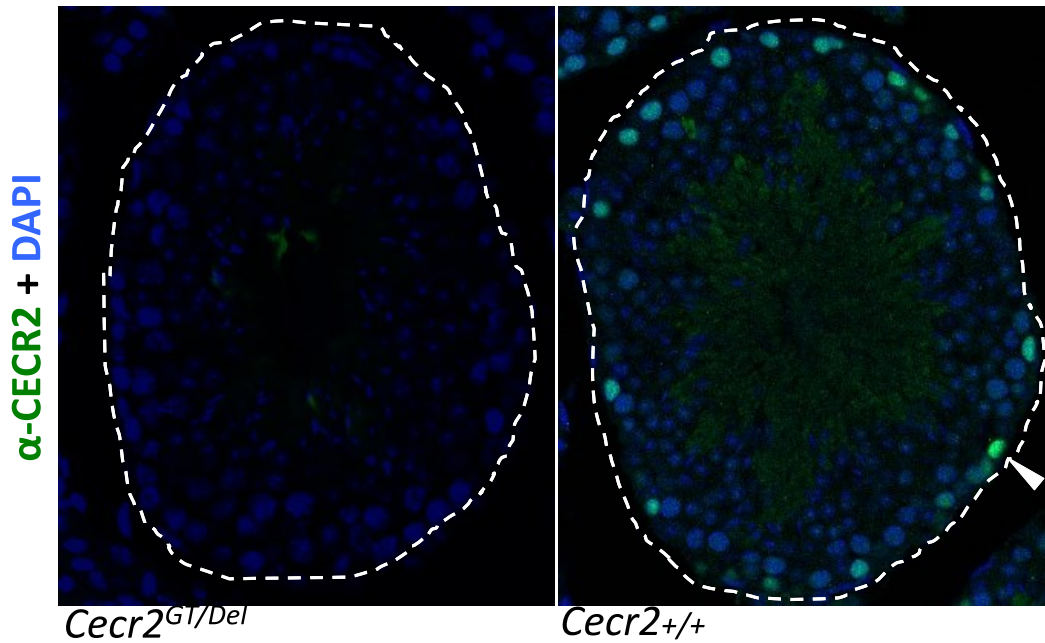
Figure 3-12 CECR2 is only detected in spermatogonia type A in adult testis. Testis from wild-type (*Cecr2 +/+*) and compound heterozygous *Cecr2*$^{Tm1.1Hemc/Gt45Bic}$ (*Cecr2*$^{GT/Del}$) adult mice were sectioned and the presence of CECR2 was detected by immunoflourescence staining using the CECR2 antibody and counterstained with DAPI to visualize DNA. Arrowhead indicates staining of spermatogonia located at the marginal regions of seminiferous tubule in wild-type testis. Dashed lines mark the boundary of the seminiferous tubule. (Kacie Norton, unpublished data).

Cecr2^{tm1.1Hemc} cDNA

Ex1	Ex2	Ex 3	Ex4	Ex5	Ex6	Ex7	Ex8	Ex9	Ex10	Ex11	Ex12	Ex13	Ex14	Ex15	Ex16	Ex17	Ex18	Ex19
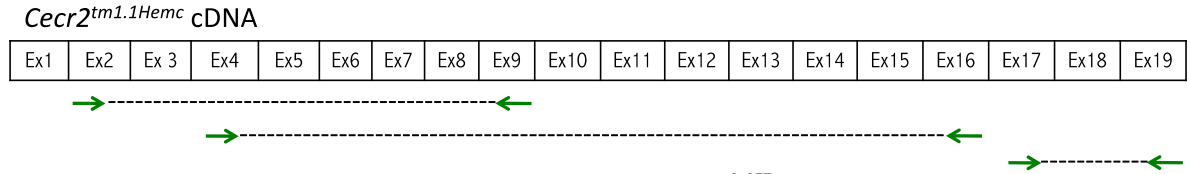
Figure 3-13 A summary of RT-PCR results for *Cecr2^{tm1.1Hemc}*. cDNA obtained from mutant embryos (Del/Del) was used to investigate the presence of any possible *Cecr2* transcript. Dashed lines showed the DNA fragments produced using the primers shown with green arrows. Ex: exon.

3.6. CECR2 is a nuclear protein

It has been previously demonstrated that the CECR2^{Gt45Bic} fusion protein localizes to the nucleus in CT45 cells (Thompson et al. 2012). CECR2 also contains a putative nuclear localization signal (NLS) (Figure 1-5) (Banting et al. 2005). To confirm that the native protein is a nuclear protein, a cell fractionation technique was used and cytoplasmic and nuclear fractions were prepared from ES cells and E12.5 embryos. Western blot analysis of cytoplasmic and nuclear extracts using the rabbit CECR2 antibody confirmed the nuclear localization of CECR2 (Figure 3-14). Western blot analysis of tubulin (cytoplasmic marker protein) and TBP (nuclear marker protein) showed that the cell fractionation was successful.

Immunofluorescence was also used to monitor the localization of CECR2 in P19 cells (Figure 3-15). Immunostaining of P19 cells showed that CECR2 localized to nuclei, colocalizing with DAPI. No-primary antibody control did not show any background staining (Figure 3-15E).

Since preparing whole cell lysate increased the yield of CECR2 and was less time-consuming than nuclear preparations, all of the following experiments were performed using whole cell lysates.

3.7. CECR2 antibody specifically binds to native CECR2 protein

It has been previously reported that human CECR2 forms a complex with SNF2L in HEK293 cells (Banting et al. 2005) and biochemical analysis of CECR2^{Gt45Bi} fusion protein indicated that it is part of a complex with SNF2H in mouse ES cells and adult testis (Thompson et al. 2012). To confirm that native CECR2 can also interact with other proteins in a complex, CECR2 should be isolated in its native form. The results in the previous sections showed that CECR2 antibody successfully detects CECR2 in its denatured form as judged by Western blot analyses. Therefore the CECR2 antibody was characterized with respect to its reactivity to native CECR2 by IP followed by Western blot analysis. The CECR2 antibody successfully immunoprecipitated CECR2 from ES cells (Figure3-16A). As a negative control, the same

experiment was performed with normal rabbit IgG instead of CECR2 antibody, and there was no band for CECR2 in the control lane, indicating specificity.
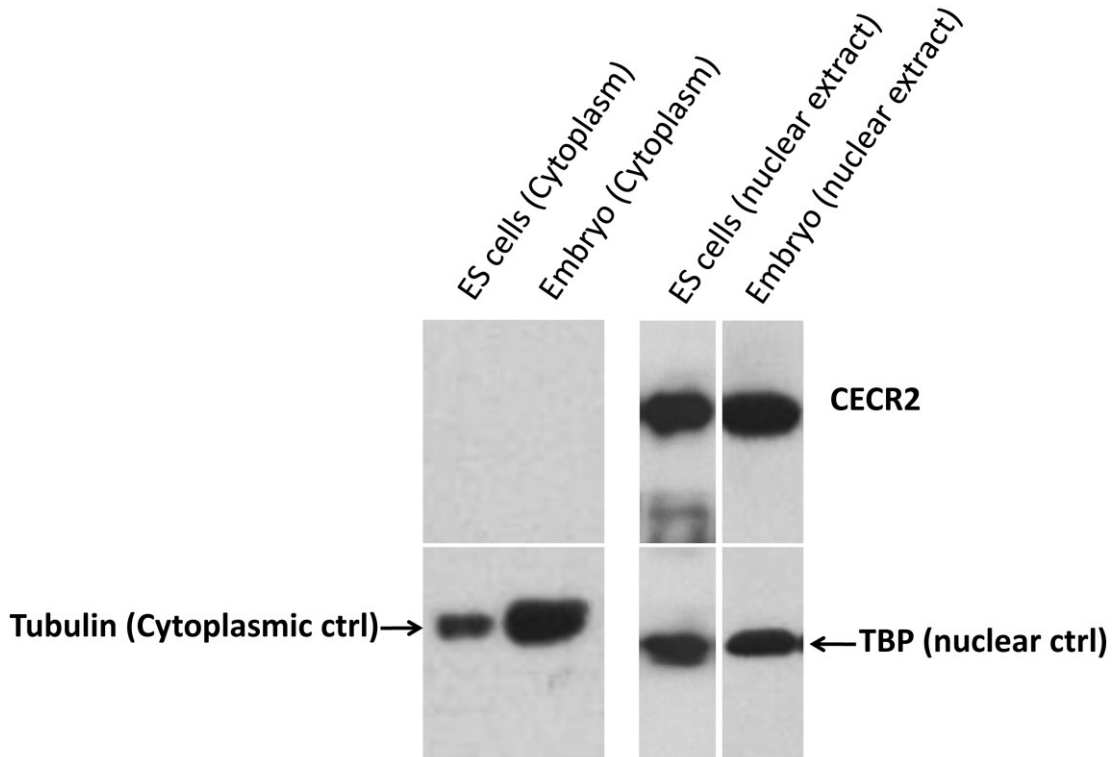
Figure 3-14 Cell fractionation reveals that CECR2 is a nuclear protein. Nuclear and cytoplasmic extracts prepared from 12.5 dpc whole embryos and ES cells were used in Western blot analysis using the rabbit anti-CECR2. No CECR2 band was detected in cytoplasmic extracts. CECR2 band was detected in nuclear extracts at the expected size. TBP was used as a nuclear marker protein and tubulin was used as a cytoplasmic marker protein.
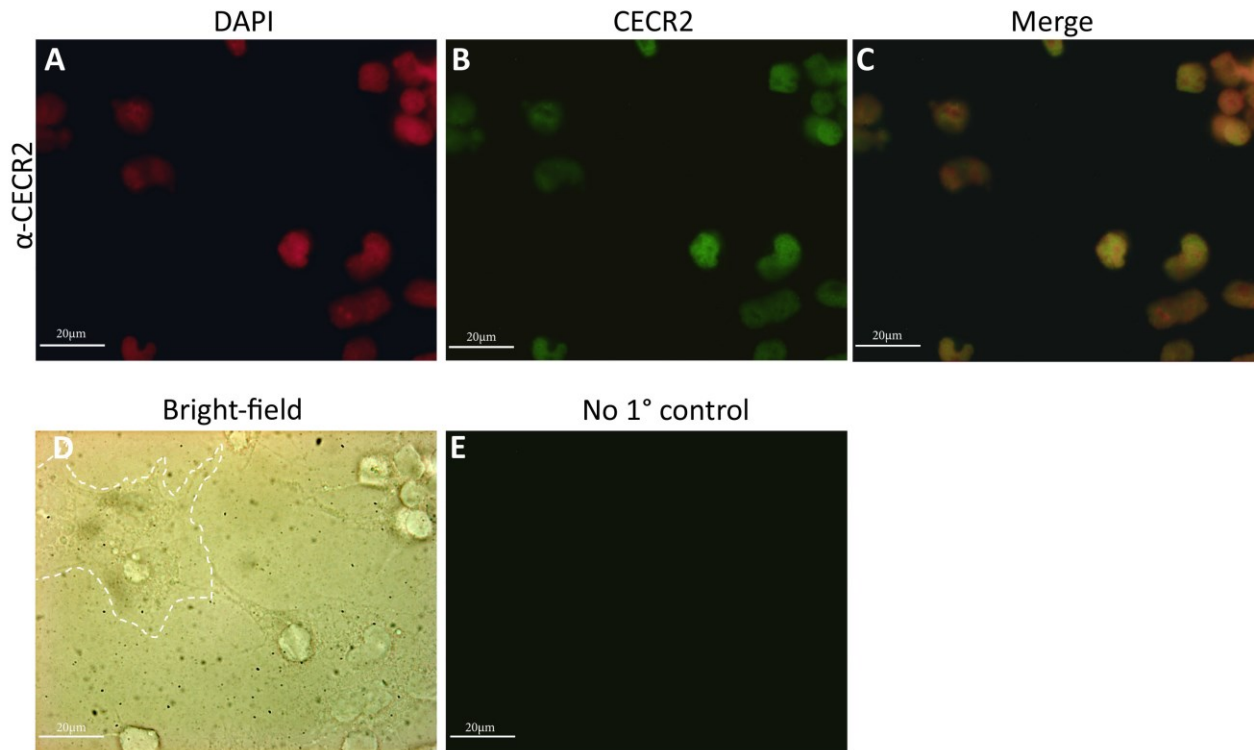
Figure 3-15 Immunoflourescence staining of P19 cells indicates that CECR2 is a nuclear protein. P19 cells were grown on coverslips, fixed with formaldehyde, stained with CECR2 antibody (green) (B), and counterstained with DAPI (red) (A). The merged image of CECR2 and DAPI showed the nuclear localization of CECR2 (C). The bright-field image of the same cells was taken to show the nucleus and cytoplasm of the cells together (D). No background staining was detected in the no 1° antibody control (E). Dashed lines mark the boundary of a cluster of cells.

The same IP assay was performed using adult testis. Adult testis is one of the few adult tissues with strong expression of *Cecr2*. To show the specificity of the CECR2 antibody, whole cell lysate was prepared from testis of a homozygous *Cecr2*^*tm1.1Hemc*^ (mutant) adult mouse. As expected, there was no band detected in this control, indicating that the CECR2 band detected after IP is indeed CECR2 (Figure 3-16B). The reactivity of the polyclonal CECR2 antibody to native CECR2 allowed us to use this antibody in biochemical analyses, including investigating the CECR2-containing complexes
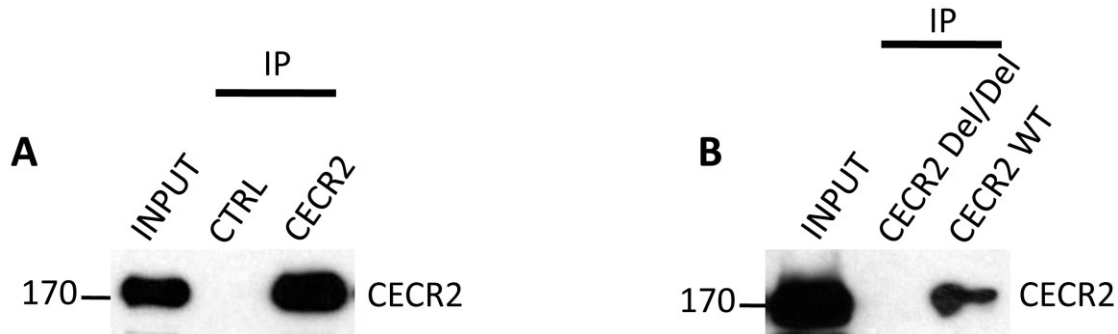
Figure 3-16 CECR2 antibody specifically binds to native CECR2 protein. A) The CECR2 antibody was bound to Dynabeads® proteinA and then added to whole cell lysate prepared from ES cells to immunoprecipitate native CECR2 protein. Normal rabbit IgG was used as a negative control. Western blot analysis after IP showed that the antibody successfully immunoprecipitated CECR2 protein. B) IP followed by Western blot analysis was performed as described in section A except that whole cell lysate was obtained from adult FVB/N testis. For a negative control in addition to the IgG control, CECR2 IP was performed using whole cell lysate prepared from testes of homozygoust $Cecr2^{tm1.1Hemc}$ (*Cecr2* Del/Del) adult mouse.

3.8. CECR2-containing complexes

The CECR2-containing Remodeling Factor complex (CERF) was first reported in HEK293 cells (Banting et al. 2005), where human CECR2 interacts with SNF2L in a complex with a size of approximately 600 kDa. SNF2L is one of two mammalian homologues of the ISWI family of ATP-dependent chromatin remodellers. The other mammalian ISWI homologue is SNF2H. In a previous study it has been shown that CECR2^{Gt45Bic} fusion protein interacts with SNF2H in mouse ES cells and adult mouse testis (Thompson et al. 2012). The CECR2^{Gt45Bic}/SNF2H-containing complex isolated from ES cells was different in size from the complex isolated from adult testis. In ES cells the complex size was estimated to be ~300-400 kDa and in adult testis the size was ~0.9-1 MDa. These measurements were done using the CECR2^{Gt45Bic} fusion protein and an antibody against β-galactosidase (Thompson et al. 2012). In *Cecr2^{Gt45Bic}* a βgeo cassette is spliced to exon 7 and leads to a CECR2-βgalactosidase-neomycin phosphotransferase fusion polypeptide, which lacks the bromodomain (Figure 1-5). It is possible that CECR2^{Gt45Bic} acts differently from the native CECR2 protein. To exclude potential artifacts

caused by the reporter fusion protein and to confirm the identity of CECR2-containing complexes in mouse, the rabbit polyclonal CECR2 antibody was used to analyse CECR2 complexes.

3.8.1. The CECR2 complex size is affected by NaCl concentration during extraction and gel filtration

To characterize the native CECR2-containing complexes in ES cells, Sephacryl S-400 High Resolution medium was used to perform gel filtration on protein extracts prepared from ES cells. After gel filtration, column fractions were subjected to Western blot analysis to detect CECR2. The size of the CECR2-containing complex was affected by the salt concentration used in protein extraction and gel filtration steps.
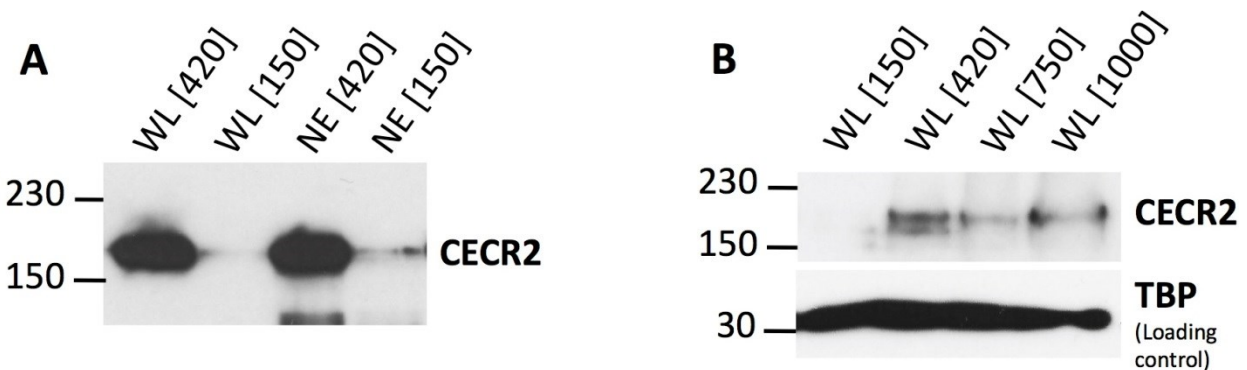
Figure 3-17 Release of CECR2 from chromatin using different concentrations of NaCl. A) Western blot analysis of whole cell lysates (WL) and nuclear extracts (NE) prepared from ES cells using two different NaCl concentrations. Whole cell lysates and nuclear extracts were prepared using the indicated concentrations of NaCl and the release of CECR2 was monitored by Western blot analysis using the CECR2 antibody. B) Western blot analysis of whole cell lysates (WL) prepared from adult testis using four different NaCl concentrations. Whole cell lysates were prepared using the indicated concentrations of NaCl and the release of CECR2 was monitored by Western blot analysis using the rabbit anti-CECR2 antibody. NaCl concentrations used in this experiment are shown on top of the blots. TBP was used as a loading control for adult testis. The molecular weight standards are shown as kDa on the left side of the figure.

During the optimization of protein extraction for biochemical analyses, the best salt concentration for preparing whole cell lysate and nuclear extracts from cell lines and tissues including testis was 420 mM of NaCl (Figure 3-17). Physiological NaCl concentration (150 mM) lead to a very low yield of CECR2 as judged by Western blot analysis. Salt is added to the lysis

buffer to increase protein solubility by increasing the ionic strength (Wit & Kessel 1996). Changes in solubility caused by altering the concentration of salt during the protein extraction has been used to monitor the binding strength of nuclear proteins to chromatin (Lichota & Grasser 2001). The higher the concentration of salt needed for the extraction of nuclear protein, the stronger the protein binds to chromatin. As CECR2 was most efficiently extracted at a high NaCl concentration, this indicates strong binding to chromatin. Therefore for most of the experiments the protein extracts were prepared using the lysis buffer containing 420 mM of NaCl. The problem that I had to consider was that an increase in salt concentration usually affects protein-protein interactions (Zhang et al. 2011).

Using 420 mM NaCl, the elution profile of CECR2-containing complexes from adult testis obtained from gel filtration indicated the presence of a small complex (Figure 3-18B). CECR2 was eluted from fraction 36 to fraction 46 corresponding to molecular weight of 1 MDa to 180 kDa (the latter being the estimated size of CECR2 alone) with the peak in fraction 42 corresponding to a molecular weight of ~500 kDa. This result was different than the previous report using buffers containing 420 mM of NaCl, which indicated that the CECR2-containing complex size was ~0.9-1 MDa (Thompson et al. 2012). Changing the concentration of NaCl in the lysis buffer and the gel filtration running buffer (the same as lysis buffer) showed that NaCl affects the elution profile of the complex. On the other hand, using the normal (physiological) salt concentration of 150 mM in the lysis buffer and gel filtration running buffer resulted in detection of a very large CECR2-containing complex in adult testis (Figure 3-18A). CECR2 was detected from fraction 24-34 corresponding to molecular weight of 1.2 MDa to 2 MDa with the peak at approximately 1.8 MDa (Figure 3-18A). This result suggests that the CECR2-containing complexes do not maintain stability in the gel filtration assay with high salt condition.

To investigate the effect of NaCl concentration on CECR2-containing complexes in ES cells, both protein extraction and gel filtration assays were performed in low (the physiological NaCl concentration of 150 mM) and high salt (NaCl concentration of 420 mM) conditions. The NaCl concentration of 150 mM resulted in detection of a very big complex in ES cells (figure 3-19A) different from the small size reported previously by Thompson *et al* using the NaCl concentration of 420 mM. CECR2 eluted over a range of 1.6 MDa to 2 MDa with the peak corresponding to a relative mobility of ~2 MDa in the physiological NaCl concentration. In the high NaCl condition of 420 mM, the CECR2-contaning complexes eluted over a mobility range

from 180 kDa to 1 MDa and the peak was in fraction 40 corresponding to a molecular weight of ~670 kDa (Figure 3-19B). The data showed that the CECR2-containing complexes, unlike the reported sizes for the CECR2[Gt45Bic]-containing complexes (Thompson et al. 2012) are big in both ES cells and testis. Although there was still a little difference between the size of complexes in ES cells and testis, the calculation of the exact difference in the size was not possible because of the limitations of gel filtration assay in distinguishing small differences in big sizes.

The effect of salt on the protein-protein interaction is very complicated and depends on the biophysical properties of proteins including the net charge, surface charge density and hydrophobicity of a protein (Zhang 2012). Nevertheless, the increase in salt concentration weakens the protein-protein association specially for hetero-complexes (Zhang et al. 2011). My data showed that using the high concentration of NaCl for protein extraction and during the gel filtration assay leads to disassociation of CECR2-containing complexes both in ES cells and testis.

3.8.2. CECR2 interacts with SNF2H in ES cells

Co-immunoprecipitation (co-IP) assay followed by Western blot analysis was used to assess if there is an interaction between native CECR2 and SNF2H in ES cells. Whole cell lysate was prepared from mouse ES cells and used in the reciprocal co-IP assay using CECR2 and SNF2H antibodies. CECR2 antibody immunoprecipitated SNF2H (Figure 3-20A). SNF2H was detected at ~130 kDa which is consistent with the reported size for SNF2H in a previous study in mouse ES cells (Stopka & Skoultchi 2003). The reciprocal assay showed that CECR2 was co-IP with SNF2H (Figure 3-20B). The control IP using the normal rabbit IgG instead of CECR2 or SNF2H antibodies indicated that protein binding in this reciprocal IP experiment was specific. The interaction of CECR2 and SNF2H was stable in high concentration of NaCl.
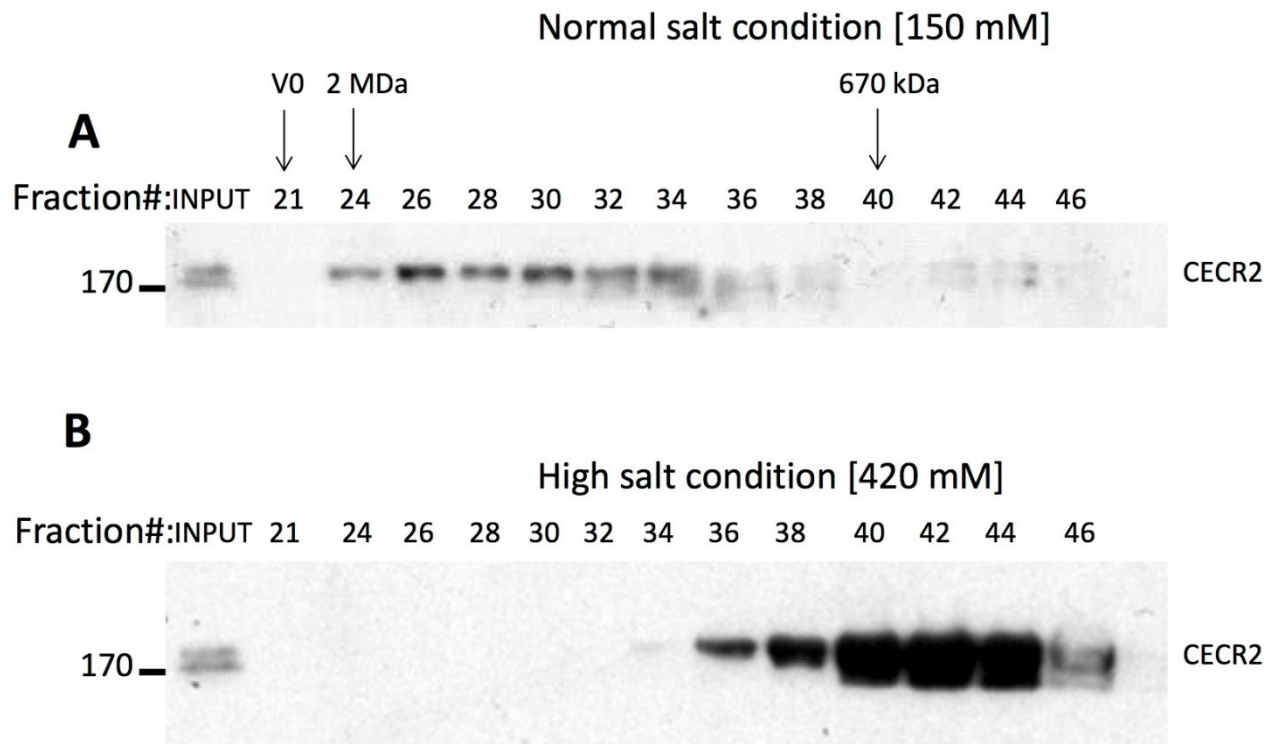
Figure 3-18 High salt concentration during protein extraction leads to dissociation of CECR2-containing complex members in testis. A) Elution profile of CECR2-containing complexes of adult mouse testis on size-exclusion chromatography. Whole cell extract was prepared from testis using a physiological salt concentration [150 mM] and separated on a Sephacryl S-400 HR gel filtration column. Fractions were TCA precipitated and separated by SDS-PAGE and subjected to Western blot analysis using the antibody for CECR2. CECR2 was detected in complexes ranging from 1.2 MDa To 2 MDa with a peak at approximately 1.8 MDa (Fraction 26). B) Gel filtration followed by western blot analysis as described in part A, except that a high concentration of salt [420 mM] was used during protein extraction. CECR2 was detected in complexes ranging from 180 kDa to 1 MDa and with a peak at approximately 500 kDa (Fraction 42). INPUT represents ~1% of the extract loaded on the column. V0 is the void volume of the column. Fraction numbers are shown above the blots, with bigger numbers indicating later elution fractions and therefore smaller molecular weights. Elution of protein markers is given above the blot. Markers are: 2 Mda: Blue Dextran; 670 kDa: Thyroglobulin. The molecular weight standard is shown as kDa on the left side of the figure.
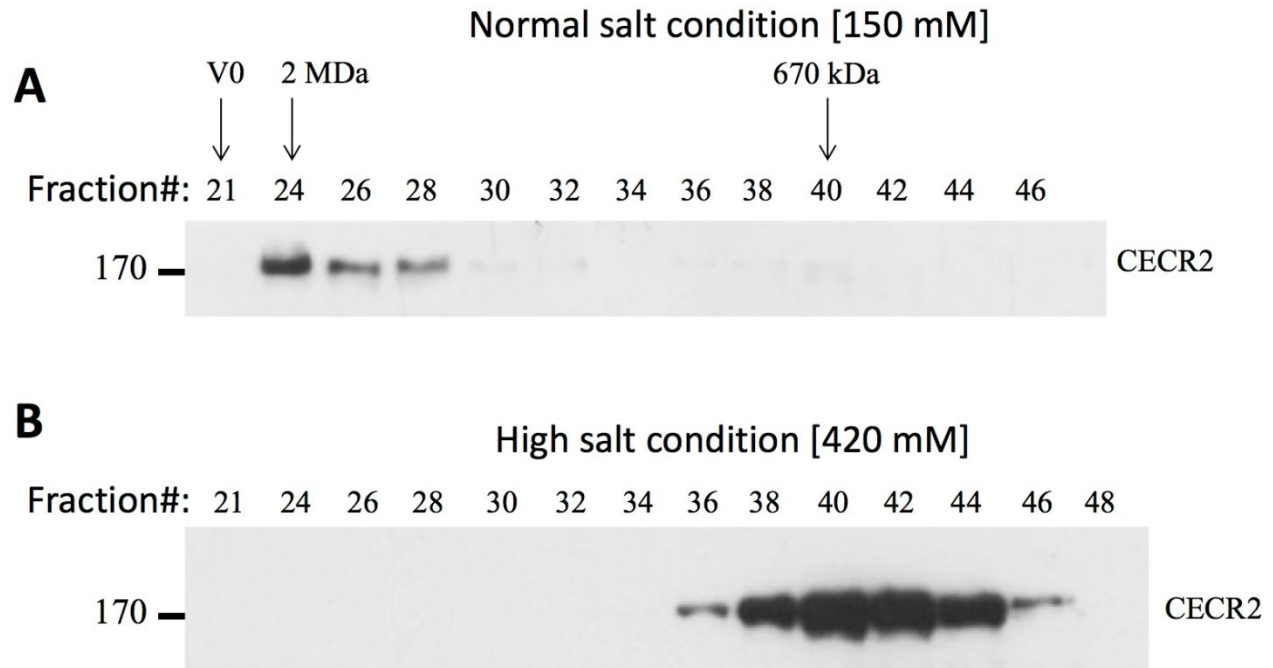
Figure 3-19 High salt concentration during protein extraction leads to dissociation of CECR2-containing complex members in ES cells. Gel filtration followed by Western blot were performed as described in Figure 3-15, except that whole cell lysate from ES cells extract was used. A) In NaCl concentration of 150 mM, CECR2 was detected in complexes ranging from 1.6 to 2 MDa, with the peak at approximately 2 Mda (Fraction 24). B) In NaCl concentration of 420 mM, CECR2 was detected in complexes ranging from 1 MDa to 180 kDa with a peak at approximately 670 kDa (Fraction 40).
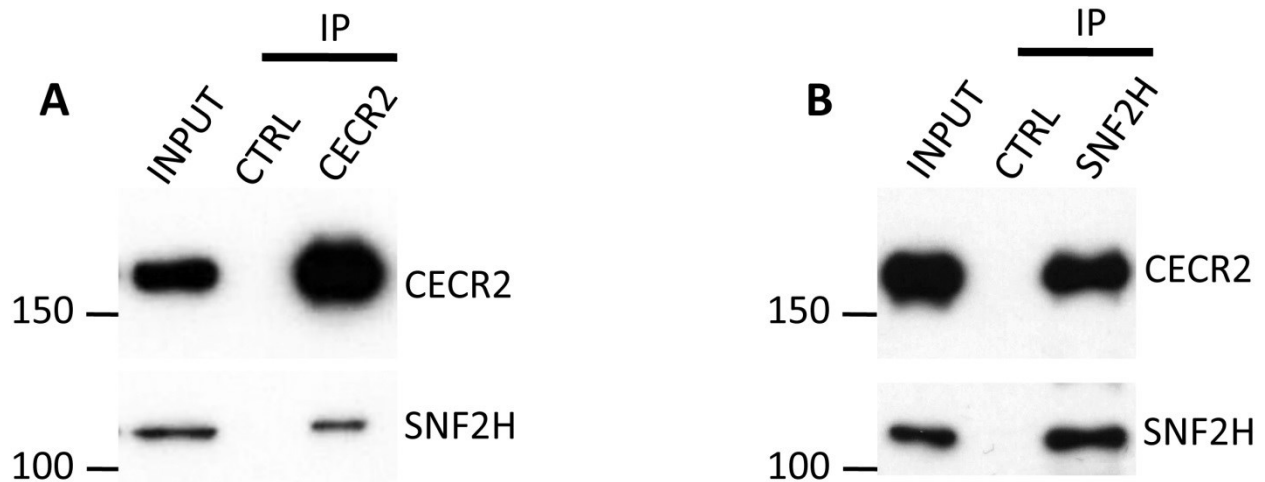
Figure 3-20 Reciprocal co-IP of CECR2 and SNF2H from TT2 ES cell extracts. A) Western blot for CECR2 and SNF2H following IP with CECR2 antibody or rabbit IgG (CTRL). B) Western blot for CECR2 and SNF2H following IP with a rabbit SNF2H antibody (SNF2H) or rabbit IgG (CTRL). The input represents 5% of the extract used in the IP reaction. Approximately 20% of each IP reaction was used in each Western blot analysis. The molecular weight standards are shown as kDa on the left side of the figure.

3.8.3. CECR2 interacts with SNF2H in testes

Cecr2 is highly expressed in testes, as shown by Western blot analysis (Figure 3-8). Also, it has been previously shown that CECR2[Gt45Bic] fusion protein interacts with SNF2H in testis (Thompson et al. 2012). SNF2H has been detected in spermatogonia (Dowdle et al. 2013). Reciprocal co-IP assay using the anti-CECR2 and anti-SNF2H antibodies confirmed the interaction between CECR2 and SNF2H in adult testis (Figure 3-21). As an additional negative control, a co-IP assay was repeated using whole cell lysate prepared from homozygous *Cecr2[tm1.1Hemc]* mutant adult FVB/N mouse and no non-specific band was detected (Figure 3-22). The result of the reciprocal co-IP of CECR2 and SNF2H indicated that both of them are part of a complex, which is consistent with the results using CECR2[Gt45Bic] fusion protein (Thompson et al. 2012).
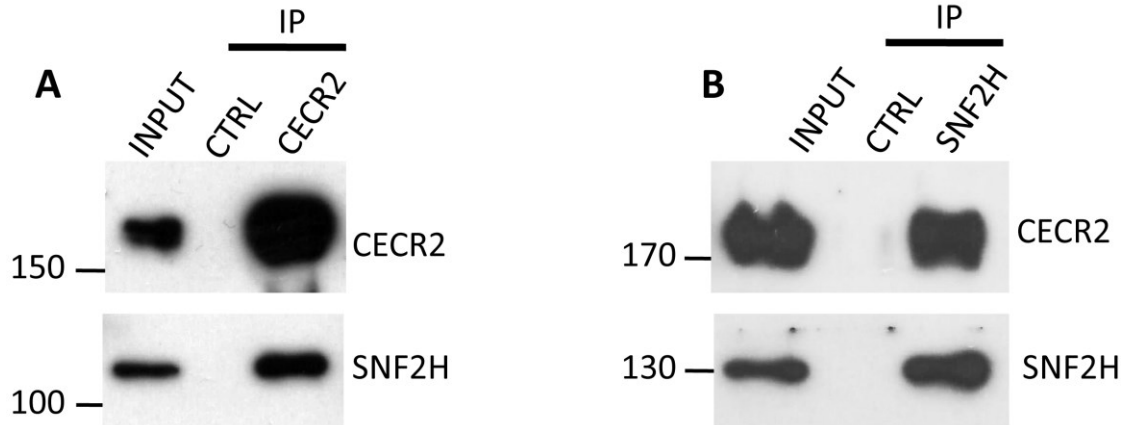
Figure 3-21 Reciprocal co-IP of CECR2 and SNF2H using whole cell lysate prepared from adult mouse testis. Co-IP followed by Western blot were performed as described in Figure 3-20, except that whole cell lysate from testis was used. The input represents 10% of extract used in the IP reaction. Approximately 20% of each IP reaction was used in each Western blot analysis. The molecular weight standards are shown as kDa on the left side of the figure.
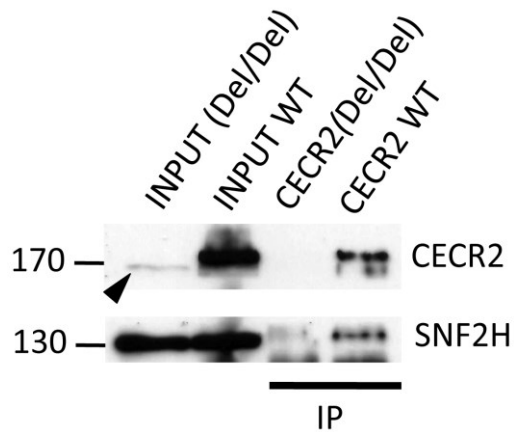
Figure 3-22 CECR2 interacts with SNF2H in adult testis. Co-IP followed by Western blot were performed as described in Figure 3-21, except that whole cell lysate from a homozygous FVB/N *Cecr2^{tm1.1Hemc}* mutant (CECR2 Del/Del) was used as a negative control. The input represents 10% of the extract used in the IP reaction. INPUT (Del/Del) is the original lysate prepared from homozygous *Cecr2^{tm1.1Hemc}* testis and INPUT WT is the original lysate prepared from wild-type testis. Approximately 20% of each IP reaction was used in each Western blot analysis. Arrowhead shows the non-canonical CECR2 band detected in homozygous *Cecr2^{tm1.1Hemc}* mutant testis.

3.8.4. The interaction between CECR2 and SNF2H is not mediated by DNA or RNA

A potential problem with showing protein interactions by IP assay is the possibility of a false positive result caused by the contamination of the lysate with DNA and/or RNA during preparation. A negatively charged nucleic acid can bind to basic surfaces on proteins, and this binding could mediate interactions between two proteins (Nguyen & Goodrich 2006). DNA and RNA can mediate interaction between proteins, especially proteins containing a DNA-binding motif. The CECR2 protein has an AT-hook domain (Banting et al. 2005) that could bind to DNA. Thus a concern is that proteins bound to adjacent DNA sequence might be immunoprecipitated with CECR2. To eliminate possible false interactions, one approach is to apply ethidium bromide to the lysate (Lai & Herr 1992). Ethidium bromide intercalates between DNA strands and deforms it causing a decreased affinity for proteins. Treating lysates with RNase before a co-IP assay can eliminate the RNA-dependent interactions between proteins (Lefebvre et al. 2002). To show that the interaction between CECR2 and SNF2H is independent of DNA and RNA, the ES cell and testes lysates were treated with 50 µg/ml ethidium bromide followed by treating with 0.33 mg/ml RNase. Co-IP with the cell lysates treated with ethidium bromide and RNase showed that neither DNA nor RNA mediates the interaction between CECR2 and SNF2H and these proteins are interacting in a protein complex (Figure 3-23).

3.8.5. CECR2 interacts with SNF2H in an ~ 2 MDa complex in ES cells and adult testis

The CECR2-containing complex is ~2 MDa in ES cells (Figure 3-19A). To show that SNF2H is part of this big complex, the fractions from the gel filtration analysis shown in Figure 3-19A, performed in physiological salt concentrations (150 mM), were separated by SDS-PAGE and subjected to Western blot analysis using antibodies for SNF2H, which showed the co-elution of SNF2H with CECR2 (Figure 3-24A). SNF2H eluted in fractions 24-32 corresponding to a calculated molecular weight of 1.3 MDa to 2 MDa. The larger range of elution for SNF2H is the result of its presence in other complexes with various molecular weights. Co-elution of SNF2H and CECR2 supported their interaction in the big complex.

To obtain further evidence for an association between CECR2 and SNF2H in the big complex, in another similar experiment fractions 23-26 containing the CECR2 complex were combined and used in a co-IP with CECR2 antibody followed by Western blot analysis (Figure

3-24B). Co-IP of SNF2H and CECR2 from the combined fractions could be seen, which was not detected in the IgG control.
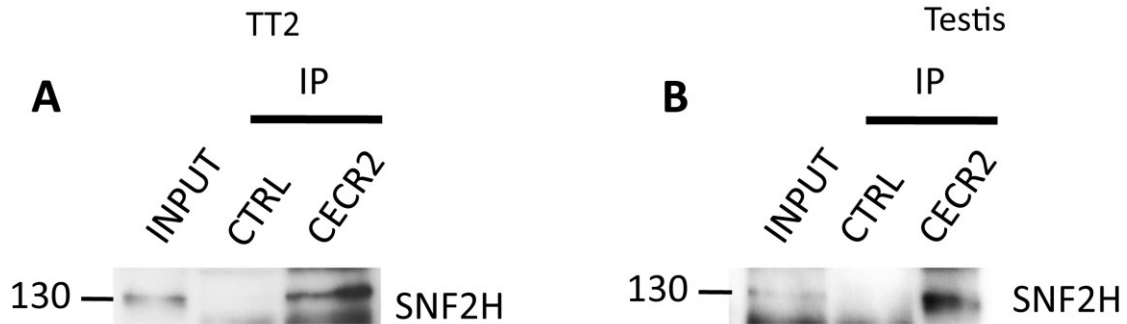
Figure 3-23 The CECR2 and SNF2H interaction is independent of DNA or RNA. Western blot detected SNF2H following IP with the CECR antibody or rabbit IgG (CTRL) using whole cell lysates prepared from ES cells (A) and adult mouse testis (B). The ES cell and testis lysates were treated with 50 μg/ml ethidium bromide followed by treating with 0.33 mg/ml RNase A prior to co-IP. The input represents 5% of extract used in the IP reaction. Approximately 20% of each IP reaction was used in each Western blot analysis. The molecular weight standards are shown as kDa on the left side of the figure.

These data indicate that CECR2 and SNF2H form a complex with a molecular weight of ~2 MDa in mouse ES cells. LUZP1 seen in the figure 3-24B is a novel binding partner of CECR2 and will be discussed in section 3.10.2.4.

The same method was used to show the interaction of CECR2 and SNF2H in the big complex in adult testis. Gel filtration fractions prepared from adult testis in figure 3-18A [performed in physiological salt concentrations (150 mM)] were used. Western blot analysis of SNF2H showed the presence of SNF2H in fractions 24-30 (Figure 3-25A). In another gel filtration experiment under the same conditions, fractions 24-29 were combined for co-IP with CECR2 antibody (Figure 3-25B). SNF2H co-immunoprecipitated with CECR2, which was not seen in the IgG control. The result indicates that CECR2 and SNF2H form a complex with a molecular weight of ~2 MDa in adult mouse testis.
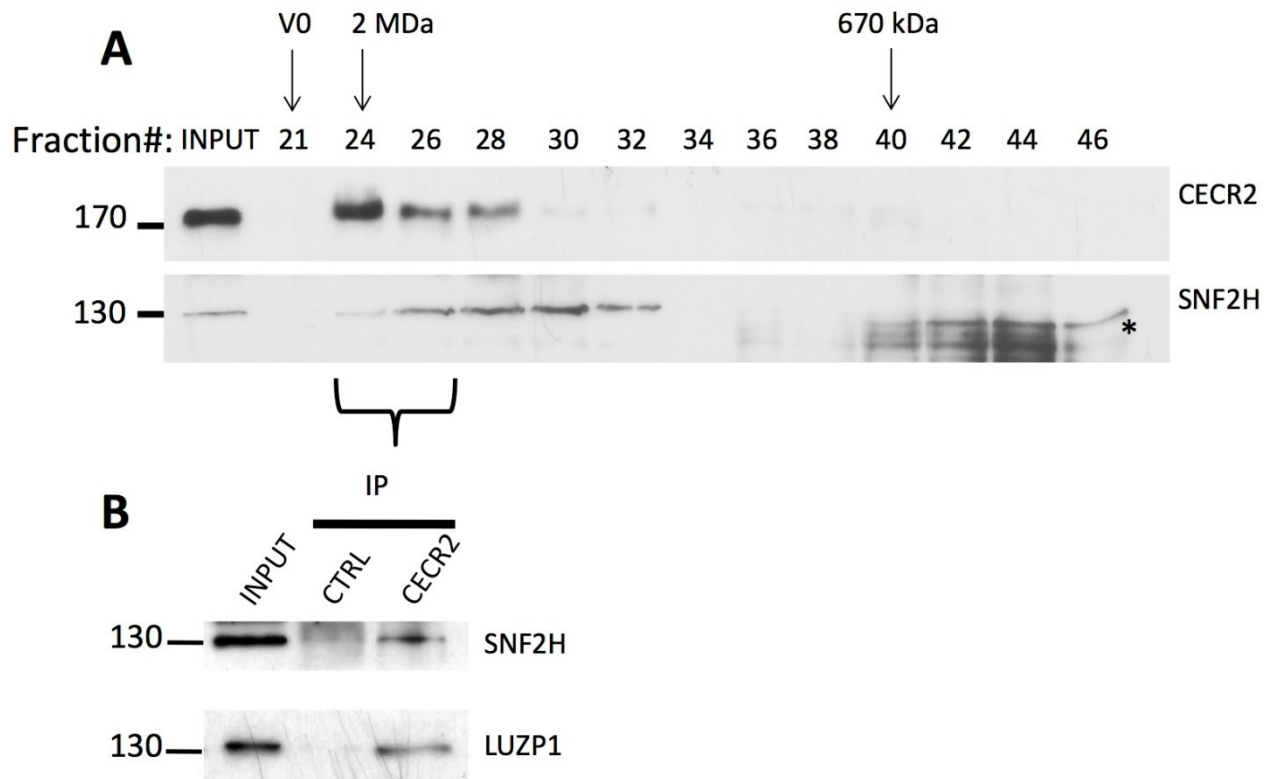
Figure 3-24 CECR2, SNF2H and LUZP1 form a complex with a molecular weight of approximately 2 MDa in mouse ES cells. A) The elution profile of SNF2H in ES cells next to the same CECR2 elution profile seen in Figure 3-19A. B) SNF2H and LUZP1 co-immunoprecipitated with CECR2 from gel filtration fractions 23-26 (~2 MDa). Fractions 23-26 were pooled and subjected to CECR2 or rabbit IgG (CTRL) IP assays and analyzed by western blot with antibodies for SNF2H and LUZP1. V0 is the void volume of the column. (*) non-specific bands detected by SNF2H antibody.
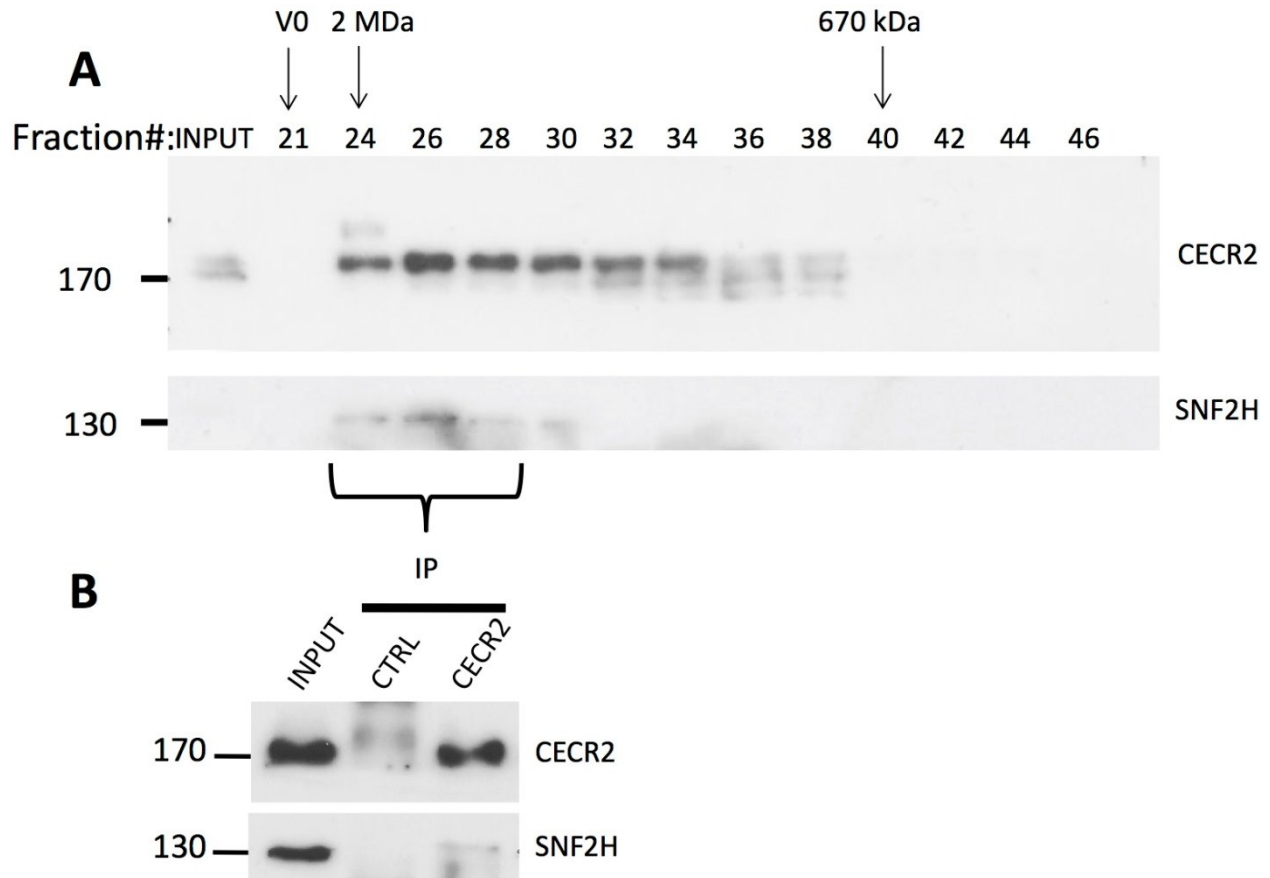
Figure 3-25 CECR2 and SNF2H form a complex with a molecular weight of approximately 2 MDa in adult mouse testis. A) Fractions obtained from the previous gel filtartion (Figure 3-18A) were subjected to Western blot using using antibodies for CECR2 and SNF2H. B) SNF2H co-immunoprecipitates with CECR2 from gel filtration fractions 24-29 (~2 MDa). Fractions 24-29 were pooled and subjected to CECR2 or rabbit IgG (CTRL) IP assays and analyzed by western blot with antibodies for SNF2H and CECR2.

3.9. CECR2-containing complex composition

3.9.1. CECR2-containing complex composition

Only two components of CECR2-containing complexes in mice are known, CECR2 and SNF2H. These two proteins are part of an ~ 2 MDa complex in ES cells and adult testis. The calculated molecular weight of a complex containing only CECR2 and SNF2H is ~ 300 kDa.

This suggests that these 2 MDa complexes must include other interacting proteins. Other ISWI complexes such as CHRAC and NuRD/cohesin have been shown to harbor 4 and 10+ interacting protein components, respectively (Dirscherl & Krebs 2004). Identifying the other interacting protein members of CECR2-containing complexes isolated from ES cells and testis would help to understand their function. Different interacting members in a complex can lead to differences in its function as shown previously for other chromatin remodellers (Barnett & Krebs 2011, Chen et al. 2013b).

To investigate the composition of CECR2-containing complexes, an IP approach followed by mass spectrometry was used. CECR2-containing complexes were isolated from whole cell lysate prepared from mouse ES cells and adult FVB/N testis using the CECR2 antibody, which should also precipitate interacting proteins. To eliminate non-specific binding proteins, normal rabbit IgG was used in parallel IP reactions. As an additional negative control for testis, the CECR2 IP reaction was performed using whole cell lysate from homozygous FVB/N $Cecr2^{Tm1.1Hemc}$ mutant testes. Interacting proteins were separated by SDS-PAGE gel. Gel slices were in-gel digested with trypsin and extracted, then submitted to liquid chromatography tandem mass spectrometry analysis (LC-MS/MS). In the first experiment I only sent the gel slices containing the molecular sizes of ~130 and 170 kDa to detect the presence of CECR2 and SNF2H. The MS analysis showed the presence of both these two proteins indicating that the technique was working (Table 3-1). In the subsequent experiments various conditions were used and specific protein bands or whole gel were analyzed by mass spectrometry to identify interacting proteins (Appendix C). Any protein identified in the IgG control were considered as non-specific and removed from the list. In addition, all the background contaminants that are seen in mass spectrometry analyses were deleted according to the list of contaminants in Contaminant Repository for Affinity Purification (the CRAPome) (Mellacheruvu et al. 2013). For the testis negative control, whole cell lysate from homozygous FVB/N $Cecr2^{Tm1.1Hemc}$ mutant, there was detection of CECR2 (appendix L and M). Because of the evidence showing the presence of a $Cecr2$ gene isoform in the $Cecr2^{Tm1.1Hemc}$ mutant (section 3.5) I ignored this control and I only used the IgG control to remove the non-specific proteins. In fact, the presence of CECR2 protein in the homozygous $Cecr2^{Tm1.1Hemc}$ mutant testis supports the idea of the existence of a CECR2 protein isoform in the mutants (Appendix M).

Table 3-1 CECR2 and SNF2H were detected by mass spectrometry analysis from the samples immunoprecipitated by CECR2 antibody (highlighted). CECR2-containing complexes were isolated from ES cell lysate. Immunoprecipitated proteins were separated by SDS-PAGE gel and stained with coomassie blue. Bands detected in ~130 kDa (SNF2H size) and ~170 kDa were excised from the gel and submitted to liquid chromatography (LC-MS/MS) mass spectrometry analysis. In this experiment no IgG control was used and the list contains specific and non-specific proteins.

Accession (UniProt)	Gene symbol	Score[1]	Coverage%[2]	# Peptides[3]
Q6PR54	*Rif1*	237.49	27.90	43
Q01320	*Top2a*	79.86	16.75	18
Q91ZW3	*Smarca5 (SNF2H)*	68.51	18.08	17
E9QA25	*Cecr2*	58.98	14.11	13
A8DUK4	*Hbb-b1*	16.00	38.78	4
G3X956	*Supt16h*	12.10	6.30	5
A8DUV1	*Hba-a1*	6.79	26.06	3

[1]Score: (Sequest Score) The sum of the ion scores of all peptides that were identified.
[2]Coverage: The percentage of the protein sequence covered by identified peptides.
[3]# Peptides: The total number of distinct peptide sequences identified in the protein group

Overall, I performed 9 mass spectroscopy experiments using ES cells and 5 experiments using testis samples to establish the best conditions (Appendices C). To prepare the interacting protein list for global GO analysis, I used an experiment for each cell type that analysed the entire IP reactions run ~ 1 cm into the gel and cut into 3 gel slices (Table 3-2 and 3-3). This represented the best single run in each case and was used in gene ontology (GO) analysis. Pooling was not done since many experiments looked at only specific gel slices.

Protein extractions for most experiments were done at a NaCl concentration of 420 mM to allow collection of sufficient protein for mass spectroscopy analysis. Since I have shown that using a NaCl concentration of 150 mM instead of 420 mM during gel filtration leads to the detection of bigger CECR2-containing complexes, in one of the experiments I used 150 mM NaCl to immunoprecipitate CECR2-containing complexes from testis (Appendix L). CECR2 was under the detection sensitivity of the mass spectrometry due to the small amount of protein isolated. SNF2H was detected. A similar result was seen for neurospheres (Appendix O). Using 150 mM of NaCl instead of 420 mM to perform IP was not efficient enough for mass spectrometry analysis.

Table 3-2 List of proteins co-immunoprecipitated with CECR2 in ES cells identified using LC-MS/MS. Proteins co-immunoprecipitated with CECR2 were analyzed by liquid chromatography (LC-MS/MS) mass spectrometry analysis. Proteins also identified in the IgG control were considered as non-specific interactions and removed from the list, as were known background contaminants. The confirmed components of the CECR2 complex are highlighted.

Accession (UniProt)	Gene name	Score[1]	# Peptides[2]
P68033	*Actc1*	514.8	11
K3W4R2	*Myh14*	300.58	10
Q80T68	*Myh10*	199.81	5
Q9DCR1	*Tubb4b*	171.14	10
Q3U6S1	*Vim*	126.7	20
P63087	*Ppp1cc*	82.23	10
Q9WTI7	*Myo1c*	64.53	19
P09405	*Ncl*	60.58	14
Q8CB58	*Ptbp1*	56.76	5
P54775	*Psmc4*	54.39	3
P58771	*Tpm1*	53.42	10
Q05DE7	*Smarca1 (Snf2l)*	49.64	6
P27661	*H2afx*	48.59	5
Q7TPV4	*Mybbp1a*	43.17	11
E9QA25	*Cecr2*	40.31	8
Q62261	*Sptbn1*	37.18	11
Q2KN98	*Specc1l*	37.16	8
D3YUW7	*Cgn*	33.32	11
E9QN52	*Lrrfip2*	32.72	5
P21956	*Mfge8*	30.21	6
Q91YR1	*Twf1*	29.63	5
Q3TTX0	*Matr3*	28.94	6
Q9JHJ0	*Tmod3*	28.56	7
Q8VIJ6	*Sfpq*	27.96	8
Q6ZWV3	*Rpl10*	27.48	4
Q9JIK5	*Ddx21*	27.03	12
Q3U0F8	*Smarca5 (Snf2h)*	26.52	7
P29341	*Pabpc1*	26.39	7
Q9QZB7	*Actr10*	26.22	2
P62918	*Rpl8*	25.81	6
Q3TS88	*Ppp1r12a*	25.56	5
Q08288	*Lyar*	24.76	3
Q6ZWV7	*Rpl35*	23.57	2
P16546	*Sptan1*	23.42	10
Q9QXS1	*Plec*	23.21	7
P11103	*Parp1*	22.94	7

Q6DFW4	*Nop58*	22.16	10
Q9DC51	*Gnai3*	20.08	5
Q8BG81	*Poldip3*	20.08	4
Q6ZWX6	*Eif2s1*	20.06	3
Q9DBC7	*Prkar1a*	19.75	3
Q6P1F6	*Ppp2r2a*	19.11	3
O70133	*Dhx9*	19.05	6
Q9QYJ0	*Dnaja2*	18.57	4
Q8BTM8	*Flna*	18.24	7
P57780	*Actn4*	18.07	6
Q9JJ28	*Flii*	18.06	5
Q8BV76	*Thrap3*	17.9	6
P43275	*Hist1h1a*	17.48	4
Q3TBM4	*Nol6*	17.14	3
Q8QZY9	*Sf3b4*	17.09	3
P62315	*Snrpd1*	17.07	2
B2RRX1	*Actb*	16	4
P25976	*Ubtf*	15.7	6
P61161	*Actr2*	15.62	4
Q8VDD5	*Myh9*	15.6	7
P46471	*Psmc2*	15.42	5
Q8BL97	*Srsf7*	14.92	3
Q60865	*Caprin1*	14.75	3
Q99KK2	*Cmas*	14.62	3
Q9DCE5	*Pak1ip1*	14	3
P62830	*Rpl23*	13.9	3
P68254	*Ywhaq*	13.71	2
B2RQ68	*Luzp1*	13.46	5
Q3UEI6	*Serbp1*	13.41	2
Q9CYL5	*Glipr2*	13.21	2
Q8C1B7	*Sept11*	13.12	3
O54946	*Dnajb6*	12.9	3
Q8BW10	*Nob1*	12.6	4
Q9WTM5	*Ruvbl2*	12.56	4
Q9D8Y0	*Efhd2*	12.13	4
P84099	*Rpl19*	11.89	2
Q9R0Q6	*Arpc1a*	11.65	3
Q9Z0U1	*Tjp2*	11.5	4
Q9WV32	*Arpc1b*	11.07	3
Q9D6Z1	*Nop56*	10.91	4
P57776	*Eef1d*	10.7	3
P61205	*Arf3*	10.58	2
Q3TRK3	*Dbn1*	10.26	2
Q60973	*Rbbp7*	10.22	3

Q01320	*Top2a*	10.06	4
Q4FZF3	*Ddx49*	9.77	3
Q61539	*Esrrb*	9.39	3
P62911	*Rpl32*	9.25	2
O35286	*Dhx15*	9.18	2
Q9D883	*U2af1*	9.09	2
Q60596	*Xrcc1*	9.08	2
O35326	*Srsf5*	9.06	3
Q6PDM2	*Srsf1*	8.8	2
Q9CXY6	*Ilf2*	8.64	2
K3W4L0	*Myo18a*	8.62	3
P62204	*Calm1*	8.61	2
P62996	*Tra2b*	8.34	2
P59999	*Arpc4*	8.3	3
Q3U1C2	*Ruvbl1*	8.22	3
D3YZ57	*Fyn*	8.02	3
G5E839	*Cct4*	7.85	3
Q9CSD9	*Ddx5*	7.85	2
Q99MN1	*Kars*	7.65	3
O70251	*Eef1b*	7.63	2
Q3UMG4	*Ina*	7.29	3
P62196	*Psmc5*	7.25	2
Q922K7	*Nop2*	7.23	3
Q3TJ01	*Rtcb*	7.04	3
Q3TGI9	*Rsl1d1*	6.98	2
E0CZA1	*Cct5*	6.82	2
Q8VH51	*Rbm39*	6.79	3
Q9Z2N8	*Actl6a*	6.55	2
P16381	*D1Pas1*	6.48	3
Q99J62	*Rfc4*	6.47	2
Q91VE6	*Nifk*	6.4	2
Q8C2Q3	*Rbm14*	6.39	3
Q99M28	*Rnps1*	6.37	2
Q64012	*Raly*	6.36	2
Q08943	*Ssrp1*	6.3	2
Q059T9	*Prpf4*	6.17	2
Q3TF87	*Dars*	5.88	2
Q9JJ80	*Rpf2*	5.85	2
Q1PSW8	*Trim71*	5.84	2
Q542I9	*Psmc1*	5.8	2
P70168	*Kpnb1*	5.75	2
Q3UDB1	*Cct7*	5.36	2
Q8BU35	*Rbm25*	5.34	2
Q8C3W4	*Ppp1r9b*	5.32	2

O54825	*Bysl*	5.28	2
D3Z2J3	*Dhx30*	5.27	2
Q9CS06	*Cct8*	5.23	2
Q9EP71	*Rai14*	5.19	2
Q61584	*Fxr1*	5.03	2
P26231	*Ctnna1*	4.99	2
P49718	*Mcm5*	4.76	2
P61255	*Rpl26*	4.65	2
P47964	*Rpl36*	4.65	2
Q587J6	*L1td1*	4.39	2
Q9ERG0	*Lima1*	4.36	2
O35130	*Emg1*	4.34	2
Q501J6	*Ddx17*	4.21	2
P19096	*Fasn*	4.04	2
Q5F2E7	*Nufip2*	3.77	2

[1] Score: (Sequest Score) The sum of the ion scores of all peptides that were identified.
[2] Coverage: The percentage of the protein sequence covered by identified peptides.

Table 3-3 List of proteins co-immunoprecipitated with CECR2 in adult testis identified using LC-MS/MS. Proteins co-immunoprecipitated with CECR2 were analyzed by liquid chromatography (LC-MS/MS) mass spectrometry analysis. Proteins also identified in the IgG control were considered as non-specific interactions and removed from the list, as were known background contaminants. The confirmed components of the CECR2 complex are highlighted.

Accession (UniProt)	Gene name	Score[1]	# Peptides[2]
Q91ZW3	*Smarca5 (Snf2h)*	61.31	12
Q9QXS1	*Plec*	40.88	14
Q3UJB9	*Edc4*	38.52	4
Q99LF4	*Rtcb*	38	8
O35691	*Pnn*	34.45	6
Q9JJ28	*Flii*	33.5	5
Q3UMU9	*Hdgfrp2*	28.91	2
Q3UKC1	*Tax1bp1*	26.92	9
Q6PR54	*Rif1*	23.34	7
Q6A4J8	*Usp7*	22.51	5
Q5SYD0	*Myo1d*	22.11	7
Q8BGC0	*Htatsf1*	21.59	4
O35326	*Srsf5*	20.92	2
O08788	*Dctn1*	18.5	4
Q5SW75	*Ssh2*	16.82	4
Q8VDP4	*Ccar2*	16.63	4
Q569Z6	*Thrap3*	16.56	2
Q5SRX1	*Tom1l2*	14.78	3

Q5XJE5	*Leo1*	14.66	2
Q60749	*Khdrbs1*	14.48	3
Q91Z49	*Fyttd1*	13.53	5
Q5XG73	*Acbd5*	13.16	4
O35343	*Kpna4*	12.79	2
Q6PDQ2	*Chd4*	12.23	4
Q80YT5	*Spata20*	11.62	4
Q9ERG0	*Lima1*	11.32	5
P29788	*Vtn*	10.99	2
Q3TTP0	*Shcbp1l*	10.6	2
Q0P678	*Zc3h18*	10.4	3
Q99KP6	*Prpf19*	9.98	4
Q9QYL0	*Hils1*	9.44	2
Q9WUM3	*Coro1b*	9.11	2
Q9JKK7	*Tmod2*	8.84	2
Q2KN98	*Specc1l*	8.64	2
Q9R190	*Mta2*	8.56	2
Q9JMB7	*Piwil1*	8.48	3
Q03265	*Atp5a1*	8.31	2
P16858	*Gapdh*	8.22	3
Q6X6Z7	*Tekt3*	8.15	3
O09106	*Hdac1*	8.14	2
Q810V0	*Mphosph10*	7.45	3
O35286	*Dhx15*	6.94	2
O35345	*Kpna6*	6.92	2
Q922R8	*Pdia6*	6.87	3
Q60596	*Xrcc1*	6.81	3
Q8VHR5	*Gatad2b*	6.78	2
Q9D8Y0	*Efhd2*	6.49	2
Q149S1	*Tekt4*	6.48	2
P60335	*Pcbp1*	6.38	3
Q99JX7	*Nxf1*	6.3	2
Q91YE5	*Baz2a*	6.25	3
Q8CGC7	*Eprs*	6.05	2
Q99KJ8	*Dctn2*	5.59	2
O54826	*Mllt10*	5.36	2
Q6PFR5	*Tra2a*	5.26	2
P61407	*Tdrd6*	5.23	2
O08810	*Eftud2*	5.21	2
Q6A068	*Cdc5l*	5.17	2
Q3U1J4	*Ddb1*	5.16	2
P35601	*Rfc1*	5.07	2
P61164	*Actr1a*	5.03	2
Q8BKX1	*Baiap2*	4.93	2

P21107	*Tpm3*	4.9	2
Q9QXL2	*Kif21a*	4.84	2
P69566	*Ranbp9*	4.69	2
Q8K019	*Bclaf1*	4.31	2

[1] Score: (Sequest Score) The sum of the ion scores of all peptides that were identified.
[2] # Peptides: The total number of distinct peptide sequences identified in the protein group.
Since CECR2 was not in the mass spectrometry facility database, it was not detected in this experiment. The CECR2 sequence was added manually into the database.

3.9.2. Gene Ontology analysis

A list of the interacting proteins co-immunoprecipitated with CECR2 was subjected to GO analysis to identify possible common pathways (Table 3-2 and 3-3). A total of 141 for ES cells and 68 proteins for adult testis were analyzed by the PANTHER classification system in terms of their protein class, molecular function and biological process. For protein classes, 40.4% of the identified CECR2-interacting proteins from ES cell classified as nucleic acid binding proteins (Figure 3-26). The other protein groups consist of cytoskeletal protein (20.6%), enzyme modulator (9.2%), hydrolase (7%) and transferase (5%). For testis, 35% of proteins were classified as nucleic acid binding proteins, while cytoskeletal protein (19.1%), transcription factor (8.8%), enzyme modulator (7.4%) and transporter (7.4%) were the other protein groups (Figure 3-26). Overall, binding was the most common class for GO molecular function of CECR2 interacting candidate proteins in ES cells (47.5%) and testis (50%) (Figure 3-27). The other molecular function overrepresented for CECR2-interacting candidate proteins was catalytic activity (29.8% for ES cells and 26.5% for testis) (Figure 3-27). For the category of biological process, the proteins were classified into groups consisting of cellular process (ES: 58.9%, testis: 54.40%), metabolic process (ES: 35.3%, testis:44%), cellular component organization or biogenesis (ES: 34.8%, testis 23.5%), localization (ES: 13.50%, testis 17.6%), developmental process (ES: 9.9%, testis 8.8%), response to stimulus (ES: 5%, testis: 8.8%), multicellular organismal process (ES: 5%, testis:7.4%), biological regulation (ES: 5%, testis:5.9%), reproduction (ES: 2.1%, testis:2.9%), immune system process (ES: 3.5%) and biological adhesion (ES: 0.7%) (Figure 3-28). Altogether, these data indicats that CECR2 is interacting with different proteins that are involved in various biological processes.
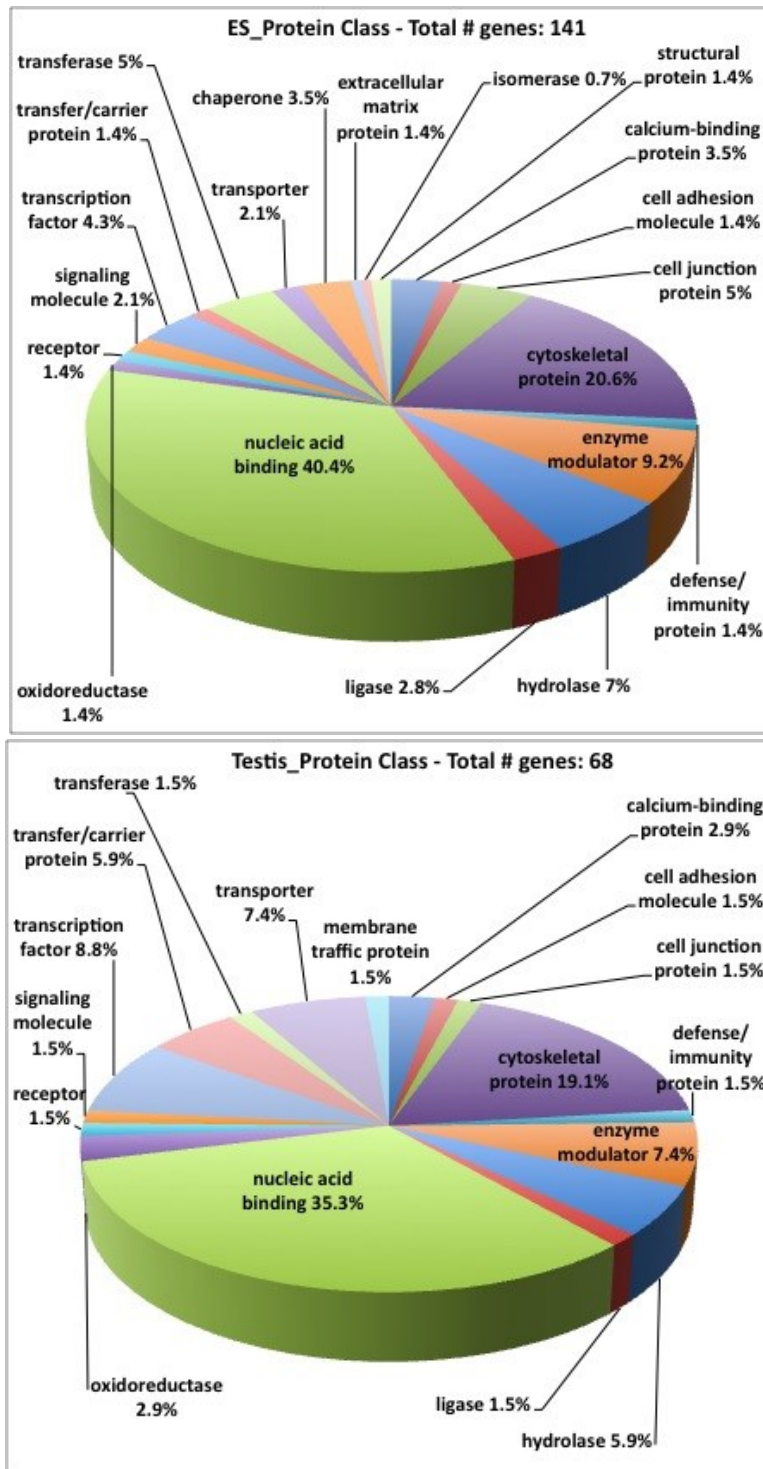
Figure 3-26 Classification of CECR2-interacting proteins detected by mass spectrometry in ES cells and testis according to their protein class. CECR2 and its interacting partners were immunoprecipitated with CECR2 antibody and analyzed by LC-MS/MS. The PANTHER database was used for classification.

ES_Molecular Function
Total # genes: 141

translation regulator activity 2.1%
transporter activity 1.4%
signal transducer activity 0.7%
structural molecule activity 24.1%
receptor activity 1.4%
binding 47.5%
catalytic activity 29.8%

Testis_Molecular Function
Total # genes: 68

translation regulator activity 1.5%
transporter activity 10.3%
structural molecule activity 13.2%
receptor activity 1.5%
binding 50%
catalytic activity 26.5%

Figure 3-27 Classification of CECR2-interacting proteins detected by mass spectrometry in ES cells and testis according to their molecular function. CECR2 and its interacting partners were immunoprecipitated with CECR2 antibody and analyzed by LC-MS/MS. The PANTHER database was used for Gene Ontology analysis.
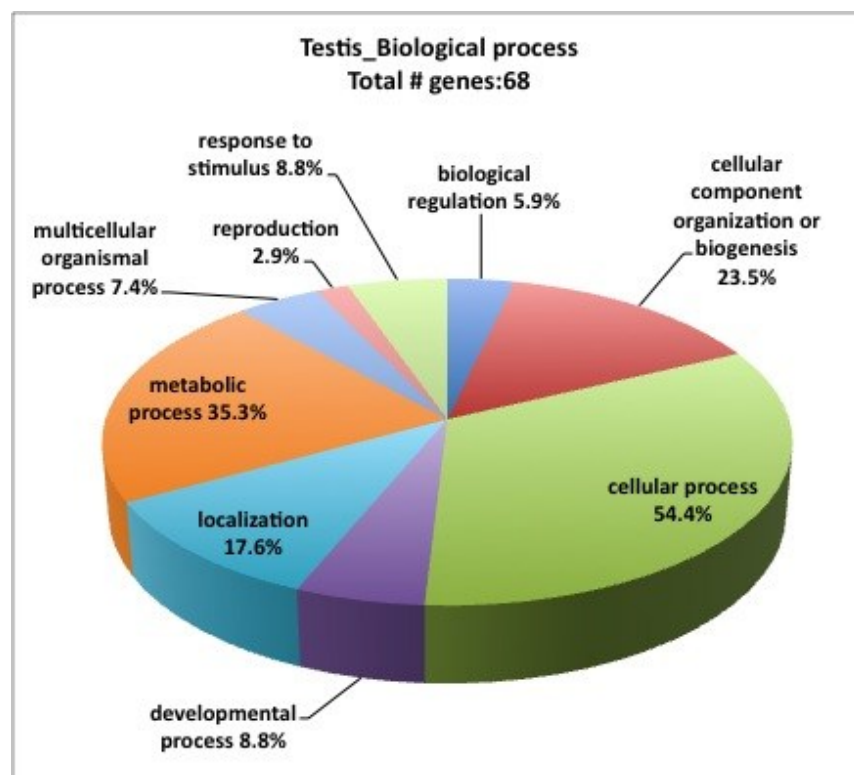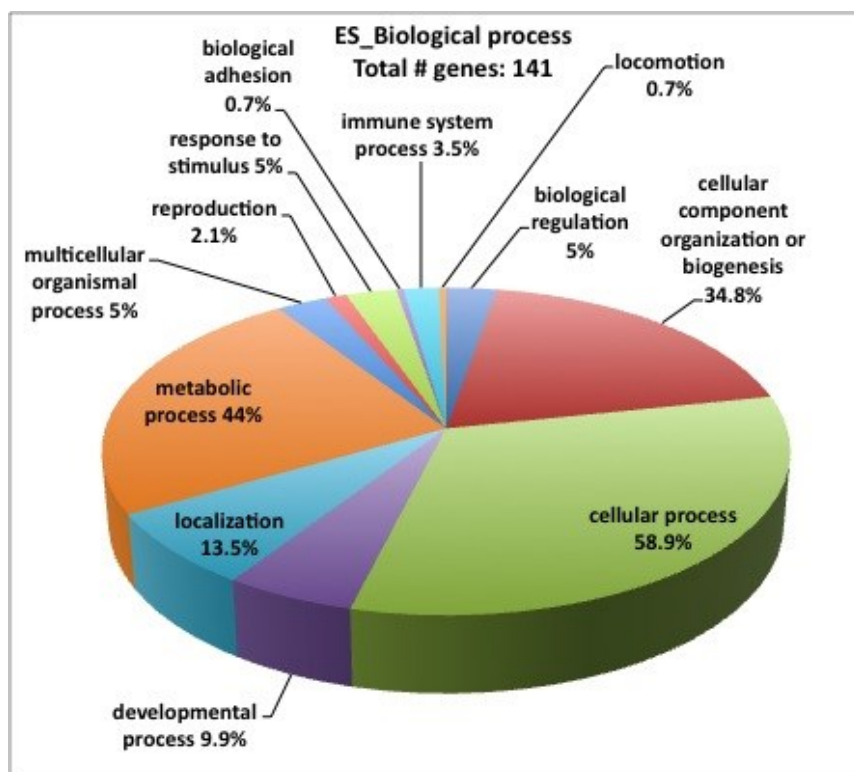
Figure 3-28 Classification of CECR2-interacting proteins detected by mass spectrometry in ES cells and testis according to biological process.

3.10. CECR2 interacting protein candidates

To determine a set of candidate proteins as novel CECR2 interacting partners to study further, two criteria were considered: (1) identification of the protein in more than one experiments and (2) whether they had any similarity to CECR2 according to their known functions, such as processing chromatin binding, DNA binding properties or involvement in neural tube defects (Table 3-4 and 3-5). The data confirmed that SNF2H (also known as SMARCA5) is a component of the CECR2-containing complex in mouse ES cells and adult testis as expected. Three genes were studied further: SNF2L, LUZP1 and CCAR2.

Table 3-4 List of candidate interacting partners of CECR2 in ES cells identified using LC-MS/MS. CECR2 containing complexes were isolated from ES cells whole cell lysate and analyzed using mass spectrometry. Nine independent experiments were performed on ES cells. Proteins detected in more than one independent experiment were considered as the candidate partner of CECR2. Genes order is based on the number of experiments that they were identified (second column from the left). Components of the CECR2 complexes that were confirmed are highlighted.

ES cells			
Gene symbol	# of Experiments	Gene name	Known function
Smarca5 (Snf2h)	8 (Table 3-1 and 3-2, Appendix D,E,G,H,I and J)	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 5	helicase and ATPase activities
Cecr2	5 (Table 3-1 and 3-2, Appendix H,I and J)	Cat Eye Syndrome Chromosome Region, Candidate 2	ATP-dependent chromatin remodeller- Involved in neural tube closure
Smarca1 (Snf2l)	5 (Table 3-2, Appendix G,H,I and J)	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 1	helicase and ATPase activities
Luzp1	4 (Table 3-2, Appendix G,H and J)	Leucine Zipper Protein 1	Involved in neural tube closure
CCAR2	3 (Appendix E,F and G)	Cell Cycle And Apoptosis Regulator 2	Core component of the DBIRD complex
Top2a	4 (Table 3-1, Appendix D,I and J)	topoisomerase (DNA) II alpha	DNA topoisomerase
Matr3	4 (Table 3-2, Appendix E,H and I)	Matrin 3	A nuclear matrix protein
Ruvbl1	3 (Table 3-2, Appendix E and H)	RuvB-like protein 1	ATP-dependent DNA helicase, core component of the chromatin remodeling INO80 complex
Rif1	3 (Table 3-1, Appendix D and I)	Rap1 interacting factor 1 homolog	Involved in fertility and DNA repair
Thrap3	2 (Table 3-2, Appendix E)	Thyroid Hormone Receptor Associated Protein 3	Involved in response to DNA damage.

	2 (Table 3-2, Appendix E)	Tropomodulin 3 (Ubiquitous)	Blocks the elongation and depolymerization of the actin filaments at the pointed end
Tmod3			
Dnttip2	2 (Appendix E and G)	Estrogen Receptor Binding Protein	Involved in chromatin remodeling and gene transcription

Table 3-5 List of candidate interacting partners of CECR2 in testis identified using LC-MS/MS. Five independent experiments were performed independently on testis. Proteins detected in more than one independent experiment were considered as the candidate partner of CECR2. Genes order is based on the number of experiments that they were identified (second column from the left). Components of the CECR2 complexes that were confirmed are highlighted.

Testis			
Gene symbol	**# of Experiments**	**Gene name**	**Known function**
Smarca5 (Snf2h)	5 (Table 3-3, Appendix K,L,M and N)	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 5	helicase and ATPase activities
Cecr2	3 (Appendix K,M,N)	Cat Eye Syndrome Chromosome Region, Candidate 2	ATP-dependent chromatin remodeller- Involved in neural tube closure
Ccar2	3 (Table 3-3, Appendix K and L)	Cell Cycle And Apoptosis Regulator 2	Core component of the DBIRD complex
Tdrd6	3 (Table 3-3, appendix L and M)	Tudor Domain Containing 6	Male infertility associated with arrested spermatogenesis.
Matr3	3 (Appendix K, L and M)	Matrin 3	A nuclear matrix protein
Rif1	2 (Table 3-3, Appendix M)	Replication Timing Regulatory Factor 1	Involved in fertility and DNA repair
Shcbp1l	2 (Table 3-3, Appendix L)	SHC Binding And Spindle Associated 1 Like	Spermatogenesis
Pabpc1	2 (Appendix L and M)	Poly(A) Binding Protein Cytoplasmic 1	Poly(A) binding protein
Ilf2	2 (Table 3-3, Appendix M)	Interleukin Enhancer Binding Factor 2	Embryonic development
Pnn	2 (Table 3-3, Appendix K)	Pinin, Desmosome Associated Protein	Transcriptional activator

3.10.1. SNF2L

Human CECR2 was identified for the first time in embryonic kidney cells (HEK293) and was shown to be part of the CERF chromatin remodeling complex (Banting 2004). In CERF, CECR2 interacts with SNF2L and shows ATP-dependant chromatin remodeling activity. SNF2L and SNF2H are the two mammalian homologues of the ISWI family of ATPases and are approximately 84% identical at the protein level (Lazzaro & Picketts 2001). There are many ISWI complexes isolated from mammalian cells with SNF2H or SNF2L as their catalytic protein member (Erdel & Rippe 2011). SNF2L has been detected in mouse ES cells previously (Bozhenok et al. 2002b) and since SNF2L was in our list of candidate proteins interacting with CECR2 (Table 3-4) in mouse ES cells, I hypothesized the presence of CECR2/SNF2L-containing complex in mouse ES cells.

3.10.1.1.   CECR2 interacts with SNF2L in mouse ES cells

To investigate the interaction between CECR2 and SNF2L, co-IP assays were performed using whole lyates prepared from mouse ES cells. A band on a Western blot was detected at ~ 130 kDa for SNF2L in the samples immunoprecipitated with CECR2 antibody (Figure 3-29). The reverse co-IP using the SNF2L antibody was not successful because of the antibody's inability to immunoprecipitate SNF2L protein under the conditions tested.

3.10.1.2.   CECR2 interacts with SNF2L in the adult testis

Although SNF2L was not detected in mass spectrometry analysis from adult testis samples, *Snf2l* has been shown to be expressed in adult mouse testis (Lazzaro & Picketts 2001, Ye et al. 2009). To determine whether CECR2 interacts with SNF2L in adult testis, different IP conditions were tested. Only in one of the IP conditions SNF2L co-immunoprecipitated with CECR2 (Figure 3-30A). In all the co-IP reactions in this project NaCl concentration of 150 mM was used during the antibody-Ag incubation step. SNF2L only co-immunoprecipitated with CECR2 when the NaCl concentration was reduced to 100 mM. The control IP using the rabbit IgG did not show any nonspecific binding of CECR2 or SNF2L, indicating the specificity of the detected bands. All the IP reaction that I had used for mass spectrometry using testis contained 150 mM of NaCl and didn't result in co-IP of SNF2L with CECR2. Overall, the data collected suggest that CECR2 interacts with SNF2L in adult testis, but this experiment needs to be repeated.
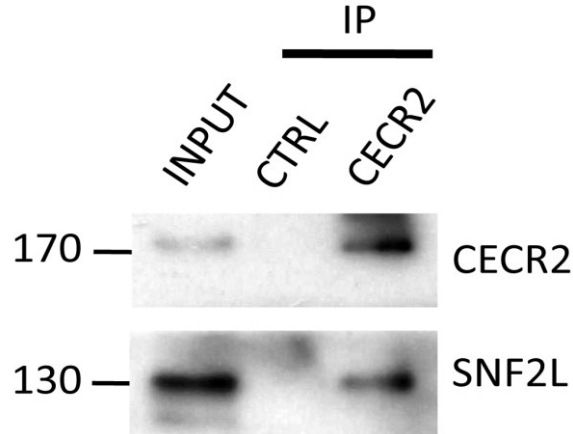
Figure 3-29 CECR2 interacts with SNF2L in mouse ES cells. Western blot for CECR2 and SNF2L following IP with CECR antibody or rabbit IgG (CTRL) using whole cell lysate prepared from mouse ES cells. The input represents 5% of extract used in the IP reaction. Approximately 20% of each IP reaction was used in each Western blot analysis. The molecular weight standards are shown as kDa on the left side of the figure.

3.10.1.3.   CECR2 interacts with SNF2L in the adult ovary

Ovary is another adult tissue with *Cecr2* expression (Figure 3-8). *Snf2h* and *Snf2l* expression has been also detected in ovary (Lazzaro et al. 2006, Lazzaro & Picketts 2001). There is a subfertility phenotype in homozygous *Cecr2*^{Gt45Bic} mutant males that are non-penetrant for neural tube defects and survive to adulthood. The female mutants also produce significantly smaller litters than wild-type females (Norton, unpublished). These data suggest a function for *Cecr2* in ovary, which could involve CECR2 complexes. Western blot analysis of proteins immunoprecipitated with CECR2 showed that SNF2L interacts with CECR2 (Figure 3-30B, Nhu Trieu, unpublished). Western blot analysis showed very low expression of SNF2H in ovary, and using co-IP assays we were not able to show its interaction with CECR2.
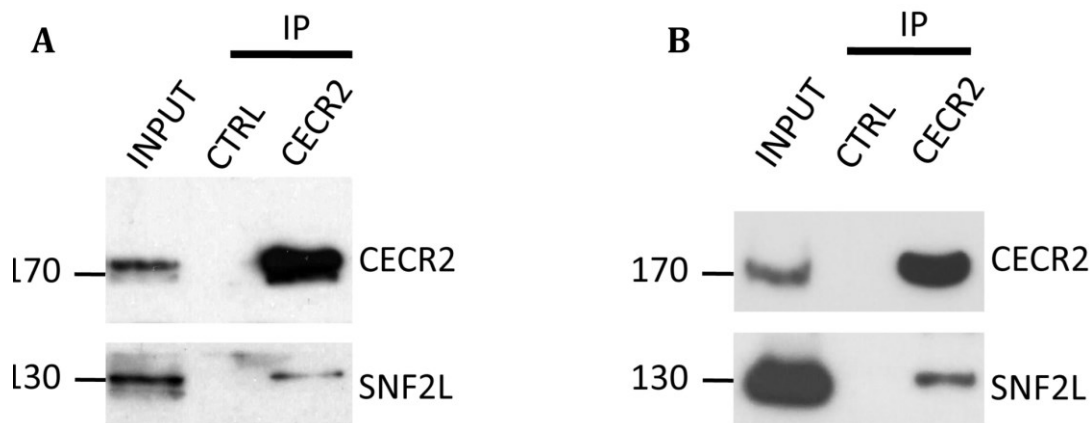
Figure 3-30 CECR2 interacts with SNF2L in adult mouse testis (A) and ovary (B). IP followed by Western blot was performed as described in Figure 3-29 except using whole cell lysates prepared from testis (A) and ovary (B).

3.10.2. LUZP1: a strong candidate for a novel interacting partner of CECR2

Leucine zipper protein 1 (LUZP1) is another candidate protein interacting with CECR2 according to our mass spectrometry analysis. LUZP1 was found in three separate LC-MS/MS runs using independent IP samples and identified by ≥ 5 distinct peptides in mouse ES cells (Table 3-4). LUZP1 was not present in control samples. Interestingly, LUZP1 was not identified in adult mouse testis samples suggesting that the complexes are different.

LUZP1 is a protein with three leucine zipper motifs at its N-terminus (Sun et al. 1996). LUZP1 is present in mouse ES cells (Lee et al. 2001). In mouse development, it has been detected in the heart from E14 to two weeks after birth by Western blot analysis (Hsu et al. 2008). In adult mouse, it is expressed in the brain, with strong detection in the cerebral cortex and hippocampus (Lee et al. 2001). Strikingly, this gene is involved in neural tube closure, producing an exencephaly phenotype similar to *Cecr2*. Other than exencephaly there are additional phenotypes, including defects in ventral body wall closure (omphalocele), cleft palate, cardiovascular malformations. Homozygous *Luzp1* mutant mice (C57BL/6J strain) exhibit exencephaly at 42% penetrance (Hsu et al. 2008). The same NTD phenotype seen in the mutant mice of both genes makes LUZP1 a strong candidate for being a member of CECR2-containing complex.

114

3.10.2.1.　　　LUZP1 interacts with CECR2 in mouse ES cells

To investigate the interactions between LUZP1 and CECR2 in mouse ES cells, reciprocal co-IP assays were performed using whole cell lysates prepared from ES cells. LUZP1 was detected at the molecular size of ~130 kDa (Figure 3-31, Input), which is slightly smaller than a previous report of 140 kDa (Sun et al. 1996). The difference could be caused by the SDS-PAGE gel conditions. The presence of many putative serine/threonine phosphorylation sites (Sun et al. 1996) implies that it might be also caused by post-translational modifications.

Western blot analysis of co-IP samples showed that LUZP1 co-immunoprecipitated with CECR2 (Figure 3-31A). Additionally, CECR2 was co-immunoprecipitated with LUZP1 (3-32B). Comparing the intensity of bands in INPUT and IP lanes shows that a smaller amount of LUZP1 is co-immunoprecipitated with CECR2. This could indicate that not all of LUZP1 in ES cells is interacting with CECR2 and that LUZP1 is member of other complexes, as it has been reported previously (Krebs et al. 2010) or that some LUZP1 is probably not incorporated into a complex. The result of multiple (5 times) reciprocal co-IP of CECR2 and LUZP1 indicated that these proteins were present together in a complex in ES cells.

3.10.2.2.　　　LUZP1 interacts with SNF2H in ES cells

CECR2 forms a complex in ES cells with SNF2H as confirmed in section 3.8.2. There are two possibilities in terms of CECR2-containing complexes. Firstly, CECR2 could form distinct complexes with LUZP1 and SNF2H. The second scenario is that CECR2, SNF2H and LUZP1 are members of the same complex and interact with each other inside a single complex. The fact that the CECR2/SNF2H-containing complex is ~2 MDa in ES cells suggests that this complex contains other protein members and LUZP1 could be one of those binding partners.

To investigate whether LUZP1 interacts with CECR2 in the same complex that also contains SNF2H, a reciprocal co-IP assay was performed using a whole cell lysate prepared from mouse ES cells. As confirmed in section 3.8.2, the immunoprecipitated SNF2H-containing complexes contain CECR2 (Figure 3-20). The same sample was subjected to Western blot analysis using antibodies for LUZP1. SNF2H co-immunoprecipitates LUZP1 (Figure 3-32). Conversely, co-IP of LUZP1 from ES cells confirmed the binding between LUZP1 and SNF2H. The reciprocal experiment showed that LUZP1 co-immunoprecipitated SNF2H (Figure 3-32B). Together, the data showed that CECR2, SNF2H and LUZP1 interacted with each other and this

interaction could be in the same complex or could be in multiple complexes including SNF2H-LUZP1, SNF2H-CECR2 and LUZP1-CECR2.
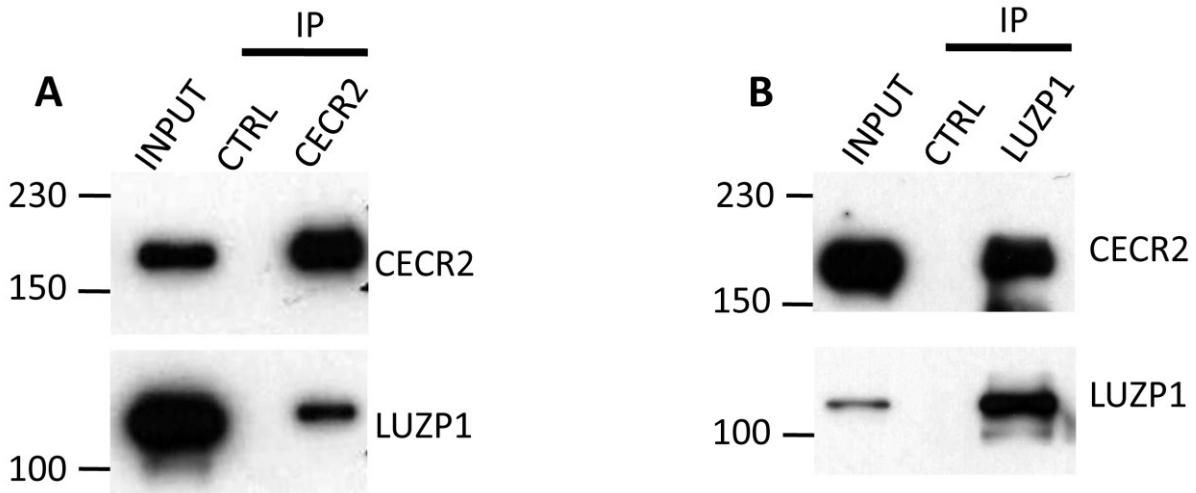
Figure 3-31 Reciprocal co-IP of CECR2 and LUZP1 indicates that the two proteins interact in mouse ES cells. A) Western blot for CECR2 and LUZP1 following IP with CECR antibody or rabbit IgG (CTRL). B) Western blot for CECR2 and LUZP1 following IP with a rabbit LUZP1 antibody (LUZP1) or rabbit IgG (CTRL). The input represents 5% of extract used in the IP reaction. Approximately 20% of each IP reaction was used in each Western blot analysis. The molecular weight standards are shown as kDa on the left side of the figure. This result is a representative of many repeats with similar results.
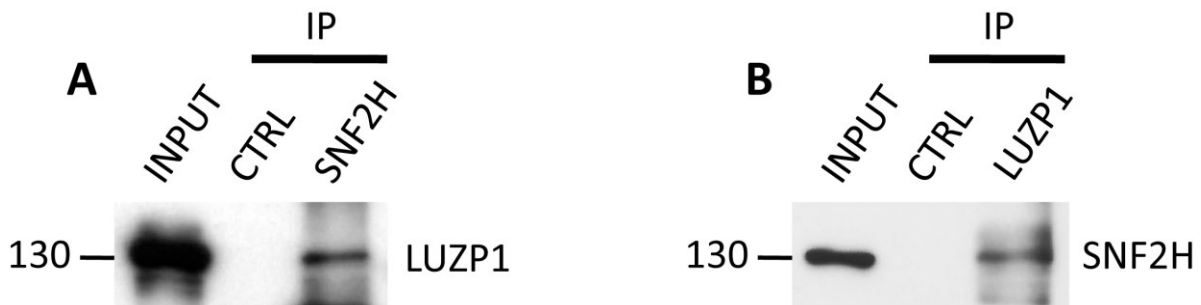
Figure 3-32 SNF2H and LUZP1 are components of the same complex in mouse ES cells. A) Western blot for LUZP1 following IP with SNF2H antibody or rabbit IgG (CTRL). B) Western blot for SNF2H following IP with LUZP1 antibody (LUZP1) or rabbit IgG (CTRL). The input represents 5% of extract used in the IP reaction. Approximately 25% of each IP reaction was used in each Western blot analysis. The molecular weight standards are shown as kDa on the left side of the figure.

3.10.2.3. Interaction between CECR2 and SNF2H is not mediated by LUZP1

LUZP1 has been shown to be part of a molecular bridge that connects the Ada-Two-A-containing (ATAC) histone Acetyltransferase and Mediator co activator complexes, forming a big complex (>1.5 MDa) called meta-coactivator complex (MECO) (Guelman et al. 2009, Krebs et al. 2010). To determine whether LUZP1 is mediating the interaction between CECR2 and SNF2H, co-IP assays were performed using a whole cell lysate prepared from homozygous *Luzp*-KO/lacZ-KI ES cells (Luzp1 -/-) (Lee et al. 2001). In homozygous *Luzp*-KO/*lacZ*-KI ES cells, the expression of *Luzp1* was not detected at either the transcriptional or translational level (Lee et al. 2001). Co-IP from homozygous *Luzp*-KO/*lacZ*-KI ES cells using the CECR2 antibody showed that CECR2 and SNF2H interact with each other in the absence of LUZP1 (Figure 3-33) and therefore the presence of LUZP1 is not necessary for this interaction.
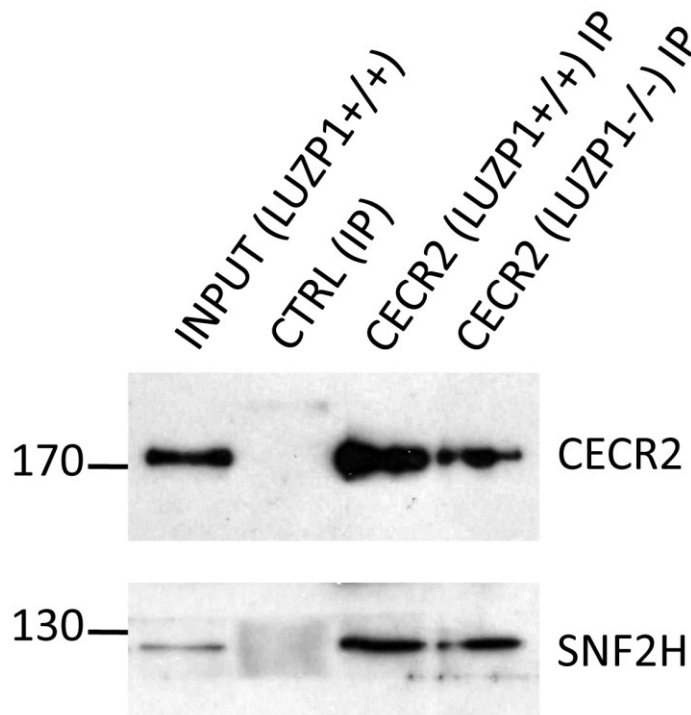
Figure 3-33 CECR2 interacts with SNF2H in homozygous *Luzp*-KO/lacZ-KI ES (LUZP1-/-) cells in the absence of LUZP1. Western blot for CECR2 and SNF2H following immunoprecipitation (IP) with CECR antibody or rabbit IgG (CTRL) using whole cell extracts prepared from wild type ES cells (LUZP1+/+) and homozygous *Luzp*-KO/lacZ-KI ES (LUZP1-/-) cells. The input represents 5% of extract used in the IP reaction. Approximately 20% of each IP reaction was used in each Western blot analysis. The molecular weight standards are shown as kDa on the left side of the figure. Luzp1 mutant and its corresponding wild type ES cells are a gift from Dr. Laszlo Tora, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), France.

3.10.2.4. LUZP1 interacts with CECR2 in a 2 MDa complex in CT45 ES cells

Multiple reciprocal co-IP assays suggested that LUZP1 is a member of CECR2/SNF2H-containing complexes in mouse ES cells (Figure 3-31 and 3-32). To investigate the presence of LUZP1 in the 2 MDa CECR2-containing complex detected in ES cells (E14 cell line), gel filtration was performed using NaCl concentration of 150 mM to prevent the disassociation of the 2 MDa CECR2 complexes. The co-elution of LUZP1 with CECR2 was assessed by Western blot analysis (Figure 34A). LUZP1 eluted over a broad range of molecular weights from 66 kDa to 2 MDa, which suggests that LUZP1 is part of other protein complexes with different sizes. LUZP1 is also eluted at the low molecular weight range of < 200 KDa, possibly suggesting that not all LUZP1 molecules are incorporated into complexes and may be present as single molecules. This also may indicate that some LUZP1-containing complexes were not stable under the tested gel filtration conditions and as a result they were dissociated leading to detection of LUZP1 at the low molecular sizes (Figure 3-34A).
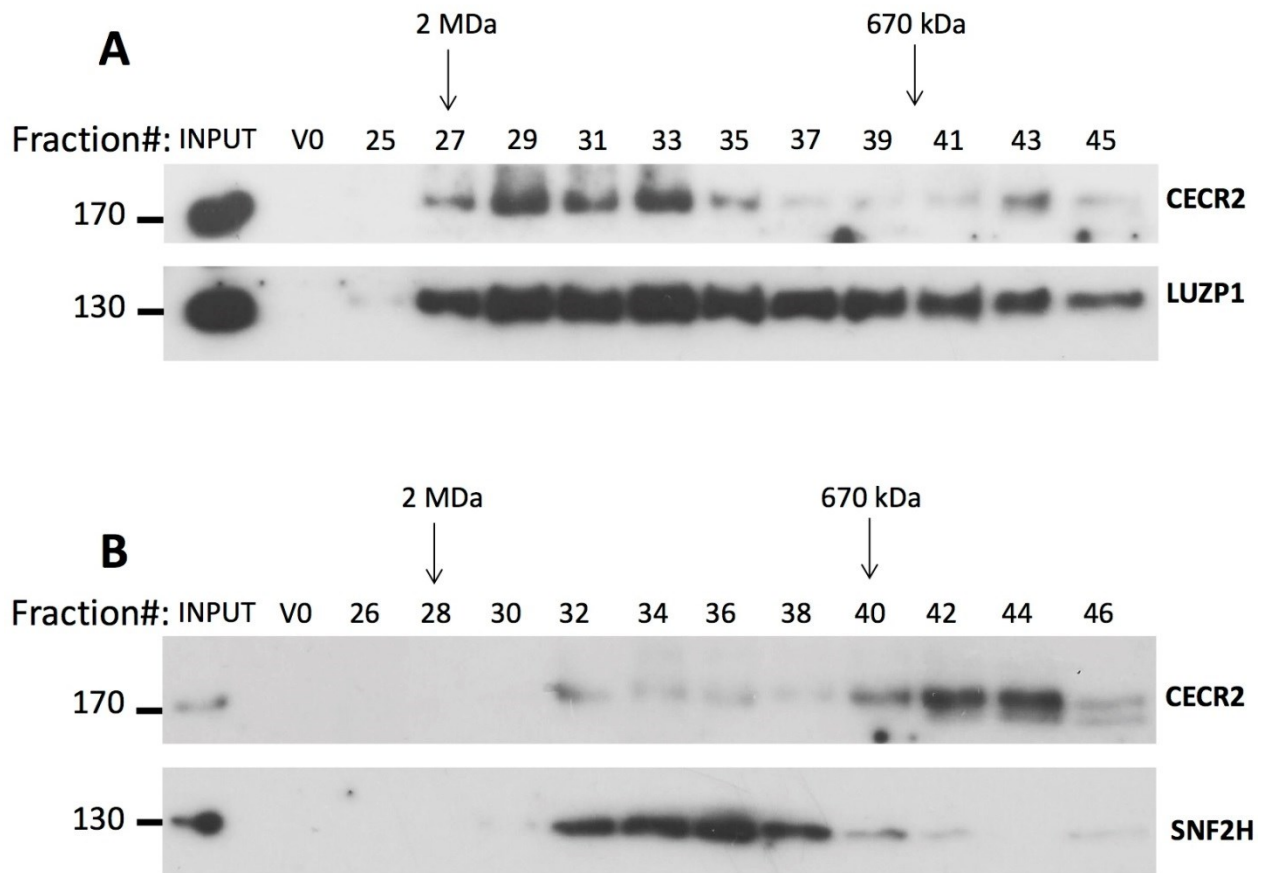
Figure 3-34 Elution profile of the CECR2-containing complex of wild type and homozygous *Luzp1* **mutant ES cell lines on size-exclusion chromatography show that the size of the CECR2 complex changes in the absence of LUZP1.** A) Whole cell extract was prepared from wild-type LUZP1 ES cells and separated on a Sephacryl S-400 HR gel filtration column ([NaCl]=150mM). Fractions indicated above the blot were separated by SDS-PAGE and subjected to Western blot analysis using antibodies for CECR2 and LUZP1. B) Gel filtration followed by western blot analysis as described in part A, except that whole cell extract from homozygous *Luzp1* mutant ES cells was used and CECR2 and SNF2H were examined. INPUT represent ~1% of the extract loaded on the column. V0 is the void volume of the column.

To confirm the incorporation of LUZP1 in the 2 MDa CECR2-complex, co-IP from a previous gel filtration assay, using the pooled fractions 23-26 containing the 2 MDa complex, was performed with the CECR2 antibody (Figure 3-24B). LUZP1 was co-immunoprecipitated with CECR2 from the combined fractions, supporting the presence of LUZP1 in the 2 MDa CECR2-containing complexes in mouse ES cells.

3.10.2.5. Differences in CECR-containing complex sizes in the absence of LUZP1 suggests that LUZP1 acts as a bridging protein in the CECR2-containing complex

The LUZP1-containing MECO complex consists of 2 separate complexes: the Ada-Two-A-containing (ATAC) histone acetyltransferase and the Mediator coactivator (MED) complexes. LUZP1 is part of the molecular bridge connecting ATAC to MED in mouse ES cells (Krebs et al. 2010). In the absence of LUZP1, the MED complex in an ATAC immunoprecipitation is decreased by 50%. This indicates that while LUZP1 is part of the molecular bridge and facilitates binding, the absence of LUZP1 does not completely destroy the complex. Therefore, I tested the possibility that the 2 MDa CECR2-complex is also an assembly of more than one complex connected by LUZP1. The size of the ES cell CECR2-containing complex was estimated by performing gel filtration on whole cell lysate prepared from homozygous *Luzp*-KO/lacZ-KI ES cells (E14 cell line) compare to the wild-type ES cells (E14 cell line) ceck Pickets (Figure 3-34B). Western blot analysis of the eluted fractions using CECR2 and SNF2H antibodies showed the presence of the 2 MDa CECR2-containing complex in wild-type ES cells as expected although there are smaller complexes. However, in the *Luzp*-KO/lacZ-KI ES cells there was a drastic increase in smaller sizes and a decrease in the larger size complexes. CECR2 and SNF2H co-eluted in a maximum molecular weight of ~1.6 MDa (Figure 3-34B, Fractions 31-38) although due to evidence of some disassociation in the control (Figure 3-34A) the 2 MDa complex may not be eliminated by the loss of LUZP1. Overall, the data indicated an increase in smaller CECR2-containing complexes in *Luzp1* mutant ES cells compared to the complex in wild-type ES cells. Elution of CECR2 in a broad range of mobilities from 1.6 MDa to > 200 kDa (fractions 31-46) suggests that LUZP1 has a role in the stability of CECR2-containing complex (Figure 3-34B). Figure 3-34B is a representative of 3 independent experiments with similar results.

3.10.2.6. LUZP1 does not interact with CECR2 in adult mouse testis

LUZP1 is a strong candidate for a novel member of the CECR2-containing complex in ES cells based on co-IP assays showing that LUZP1 interacts with both CECR2 and SNF2H in ES cells. CECR2-containing complexes of ~2 MDa are present in adult testis as well. However, LUZP1 was never identified in our multiple mass spectrometry analyses as a candidate interacting partner of CECR2 in testis (Table 3-5). To further assess whether LUZP1 interacts

with CECR2 in adult testis, co-IP assays were performed using a whole cell lysate prepared from adult testis. Western blot analysis using the LUZP1 antibody detected a band at the same size that was detected in ES cells (~130 kDa) (Figure 3-35A, INPUT). The CECR2 antibody immunoprecipitated CECR2 as expected, but no LUZP1 band was detected in IP sample. The same co-IP assay using the SNF2H antibody showed the successful IP of the CECR2/SNF2H-containing complex. Together the results indicated that despite the presence of LUZP1 in adult testis, it was probably not incorporated into a complex with CECR2.
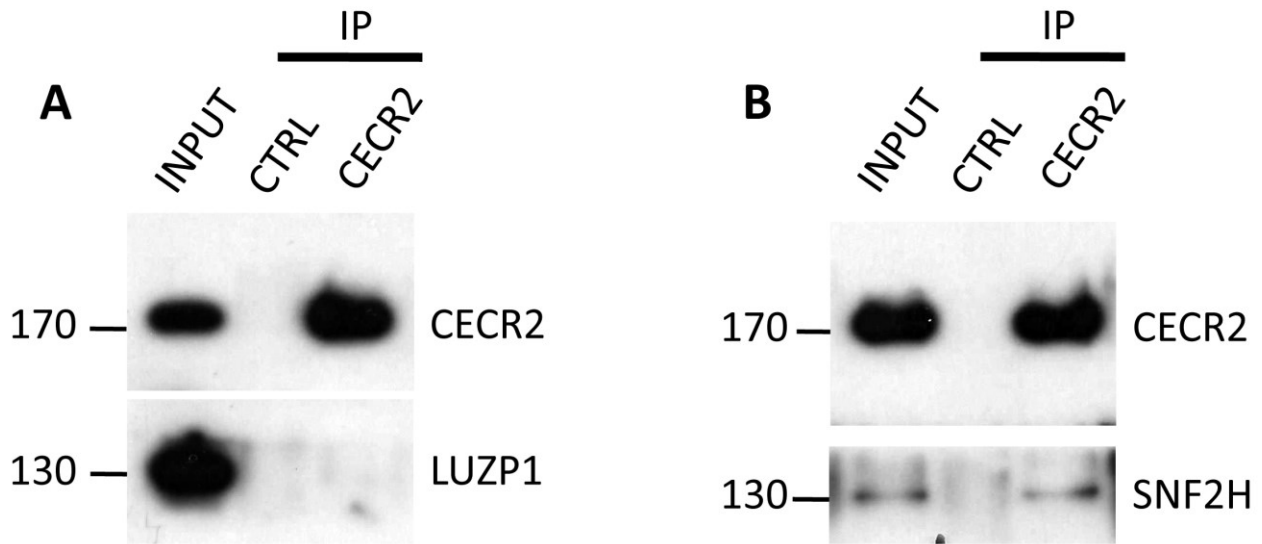
Figure 3-35 LUZP1 does not interact with CECR2 in adult testis. Co-IP of CECR2 and LUZP1 from whole cell lysate prepared from adult mouse testis. A) Western blot for CECR2 and LUZP1 following immunoprecipitation (IP) with CECR antibody or rabbit IgG (CTRL). B) Western blot for CECR2 and SNF2H using the IP reactions from the part A, showing successful immunopreciptiation of SNF2H. The input represents 5% of extract used in the IP reaction. Approximately 20% of each IP reaction was used in each Western blot analysis. The molecular weight standards are shown as kDa on the left side of the figure.

3.10.2.7.   LUZP1 physically interacts with CECR2 in neurospheres

Neurospheres also express CECR2 (Figure 3-8). Because neurospheres contain neural stem cells they may act as a model for some aspects of the developing neural tube. To investigate the interaction of CECR2 with LUZP1 in neurospheres a co-IP assay was done, showing interaction with LUZP1 (Figure 3-36). The reverse co-IP using the LUZP1 antibody was not

performed due to time limits. Further experiments are required to characterize CECR2/LUZP1-containing complexes in the embryonic neurospheres.
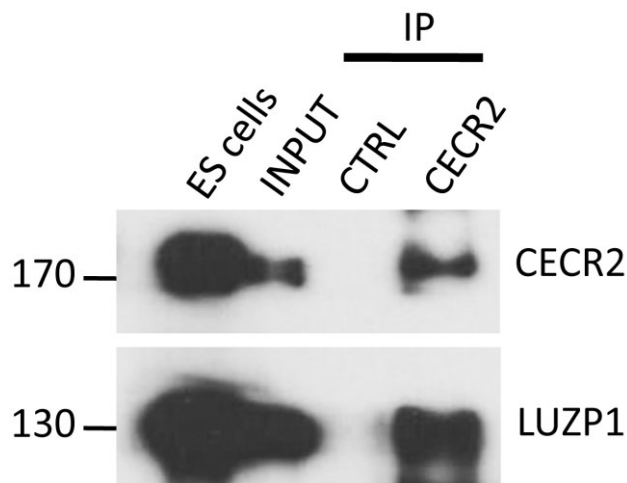
Figure 3-36 CECR2 interacts with LUZP1 in mouse neurospheres. Co-IP followed by Western blot were performed as described in Figure 3-35A, except using wild-type neurospheres. The input represents 10% of extract used in the IP reaction. Approximately 25% of each IP reaction was used in each Western blot analysis. The molecular weight standards are shown as kDa on the left side of the figure.

3.10.3. Sonic hedgehog pathway

Approximately 42% of *Luzp1* mutant mice exhibit exencephaly (Hsu et al. 2008), the same neural tube closure defect seen in *Cecr2* mutants (Banting 2004). An increase in SHH signal transduction [ectopic expression of *Shh* (Ybot-Gonzalez et al. 2002), *Ptch-1* and *Gli-1* (Stottmann et al. 2006)] has been shown to prevent the formation of dorsolateral hinge points of the neural tube, resulting in exencephaly. Ectopic expression of *Shh* in the dorsolateral neuroepithelium and lateral mesenchyme during the neural tube closure (12–17 somite stage) has been seen in LUZP1 mutant mice (Hsu et al. 2008). Since LUZP1 interacts with CECR2 in ES cells and neurospheres, I hypothesized *Cecr2* mutants will also show enhanced SHH signaling.

3.10.3.1. Normal expression of Shh in the neural folds of Cecr2 mutants

Immunofluorescence staining was performed using a polyclonal SHH antibody on *Cecr2* mutant embryos sections from *Cecr2* wild-type and homozygous $Cecr2^{Tm1.1Hemc}$ mutant embryos during the time of neural tube closure (E9.5, 12-17 somites). Immunostaining of wild-type and mutant embryo sections did not revealed any difference in the expression of *Shh* in the neural tube folds at 12-somite (data not shown), 14-somite (Figure 3-37) and 17-somite (Figure 3-38) embryonic stages. This suggests that the exencephaly seen in the *Cecr2* mutants might not be caused by the same process seen in *Luzp1* mutants.

3.10.3.2. Normal cell death rate in neural tube and head mesenchyme of *Cecr2* mutants

One of the events causing neural tube defects during the dorsolateral bending stage is an increased level of apoptosis in the hindbrain (Massa et al. 2009, Migliorini et al. 2002) and the surrounding mesenchyme (Zohn et al. 2007). *Luzp1* exencephalic mutant mice display an increased level of apoptosis in the hindbrain and the underlying mesenchyme at the stage of dorsolateral bending of neural tube (Hsu et al. 2008).

To determine whether elevated apoptosis in the hindbrain occurs in *Cecr2* exencephalic mutant mice, apoptosis was evaluated by the TUNEL assay at E9.5 embryos at the 16-18 somite stage. No apparent difference in apoptosis levels were detected between wild-type and homozygous $Cecr2^{Tm1.1Hemc}$ mutant embryos in whole mount preparations (Figure 3-39) or sections (Figure 3-40). This again suggests a difference in mechanism between *Cecr2* and *Luzp1* NTDs.
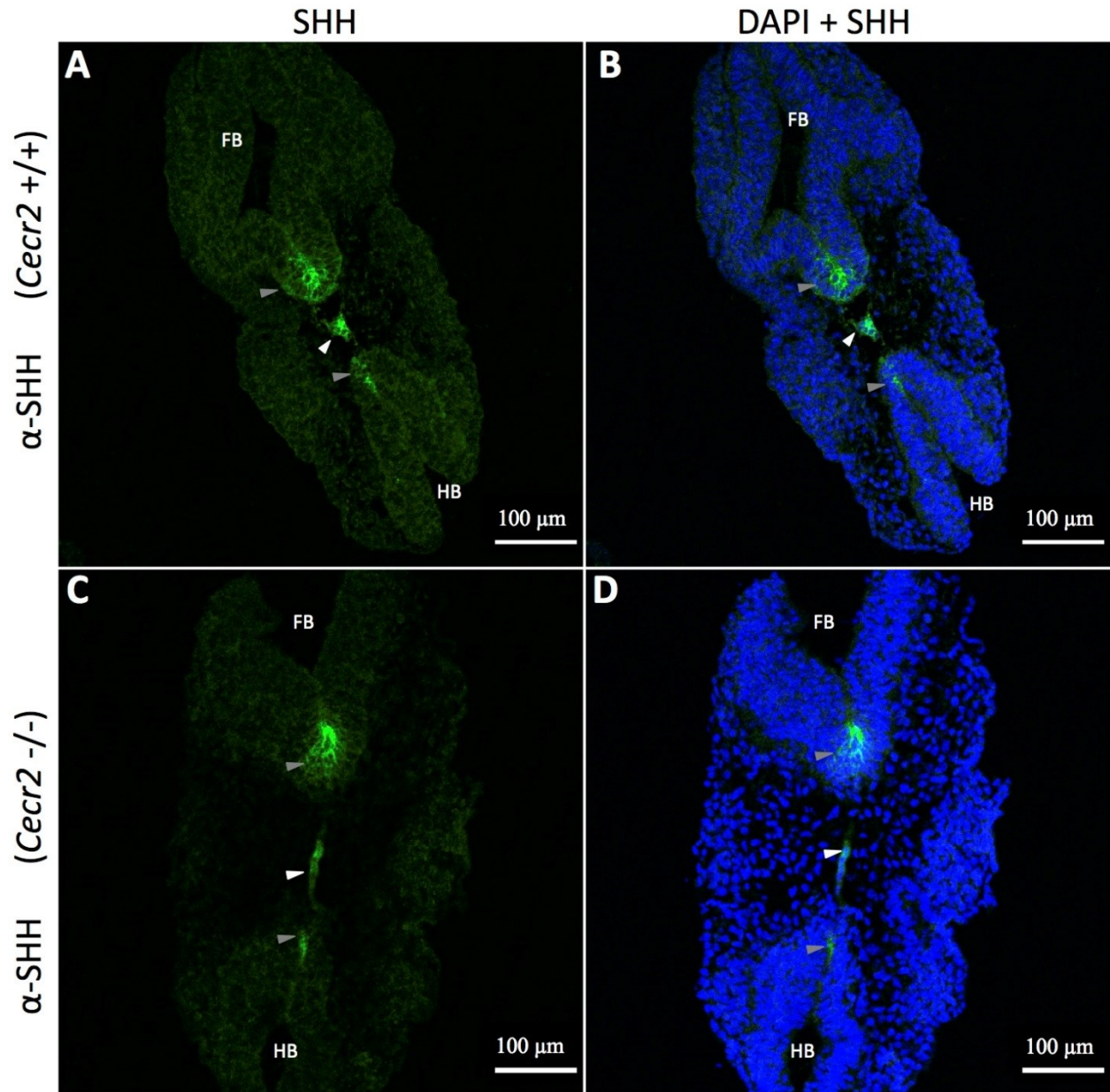
Figure 3-37 SHH expression during the closure of the neural tube of E9.5 embryos (14 somites). E9.5 embryos at 14 somite stage were collected and sectioned and the expression of SHH was detected by immunoflouroscence staining. DAPI was used to stain the nuclei and facilitate microscopic examinations. Expression of SHH is detected in the floor plate (white arrowhead) of both forebrain (FB) and hindbrain (HB) and notochord (grey arrowhead). There was not any detectable difference between wild-type embryos (A,B) and homozygous *Cecr2^Tm1.1Hemc* mutant embryos (C,D). Panels B and D are the merged image of SHH and DAPI. Scale bar: 100 μm.
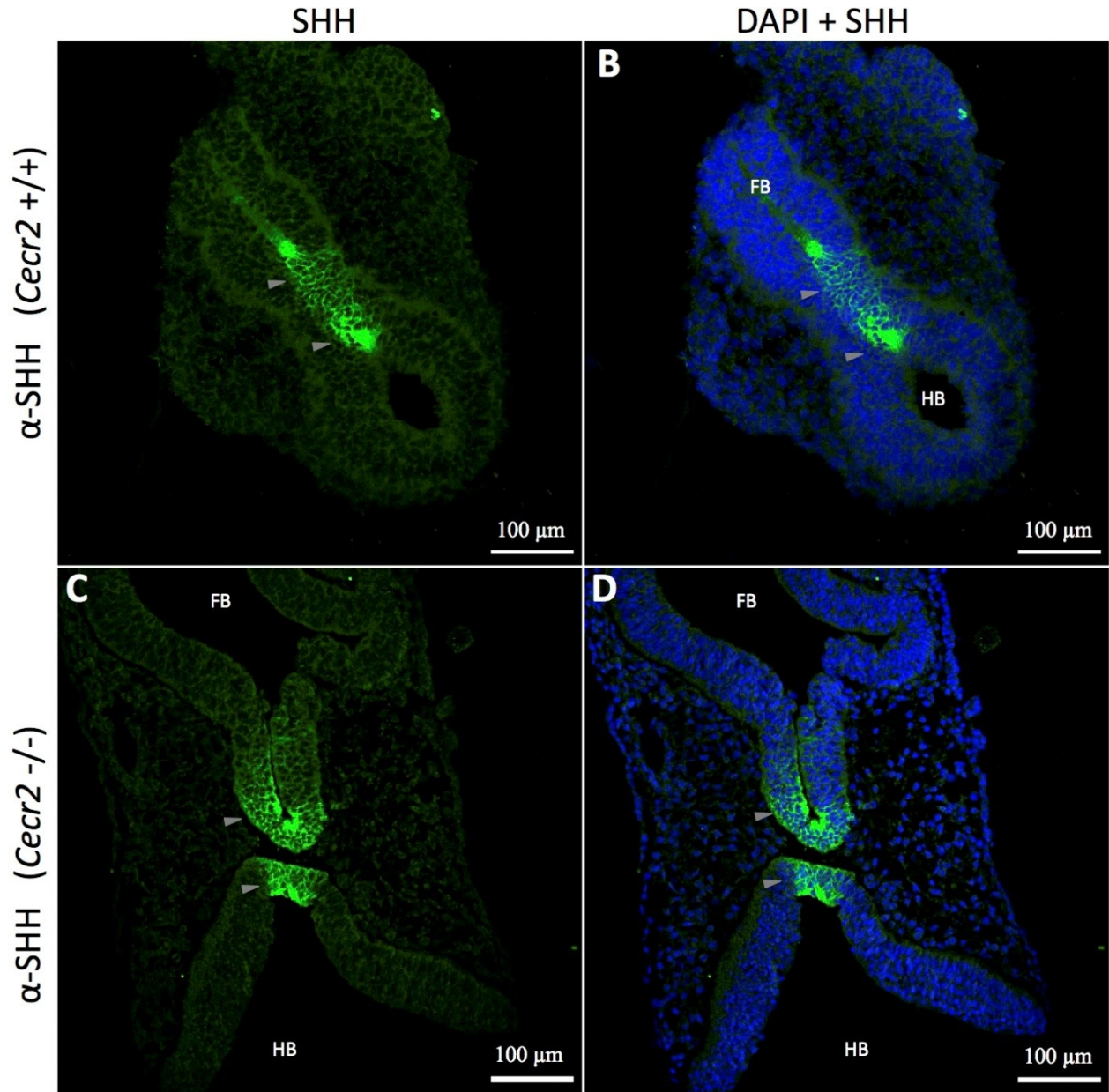
Figure 3-38 SHH expression during the closure of the neural tube at 17 somite embryos. Sectioning and staining were performed as described in Figure 3-37, except using 17 somite embryos. Expression of SHH is detected in the floor plate (white arrowheads) of both forebrain (FB) and hindbrain (HB) and notochord (grey arrowheads). There was not any detectable difference between wild-type embryos (A,B) and homozygoust $Cecr2^{Tm1.1Hemc}$ mutant embryos (C,D). Panels B and D are the merged image of SHH and DAPI. Scale bar: 100 μm.
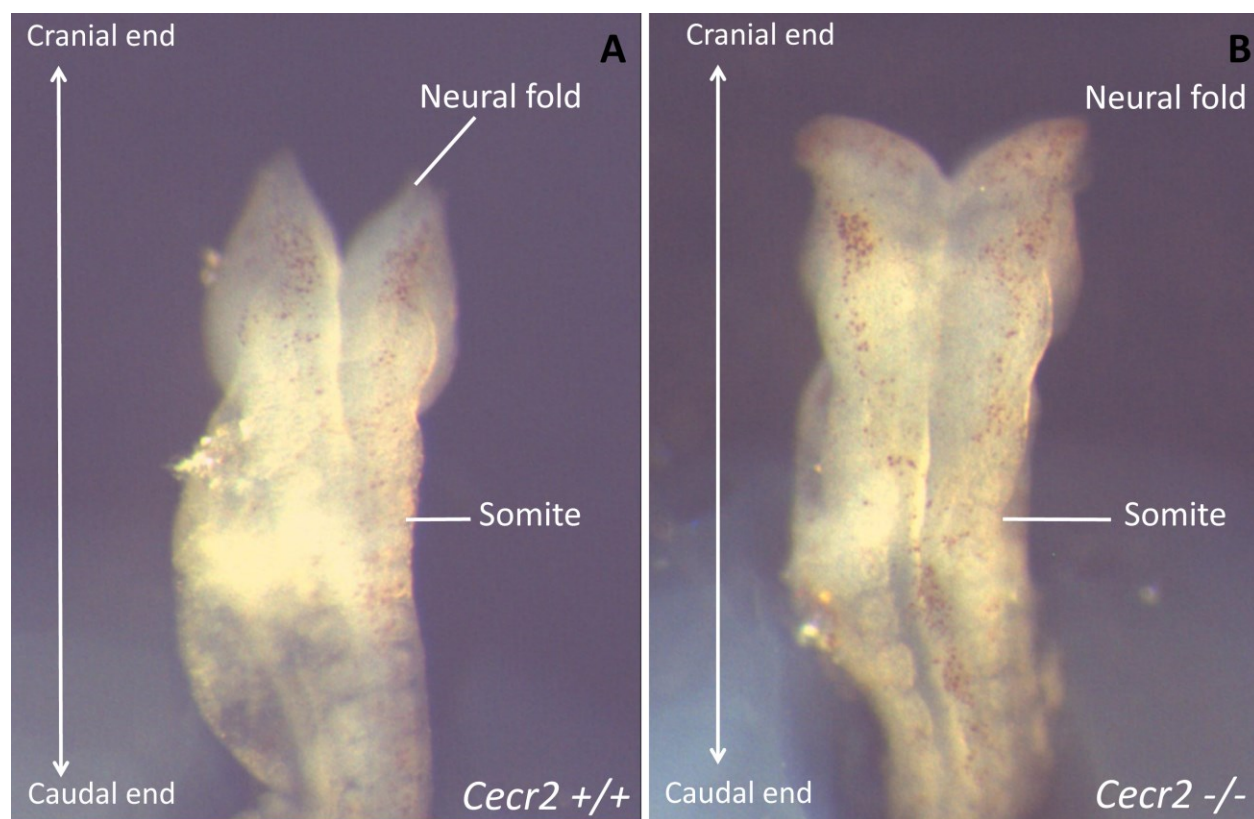
Figure 3-39 Apoptosis is not increased in *Cecr2* mutant embryos during the closure of the neural tube, as seen in whole mounts. E9.5 embryos were collected and the pattern of apoptosis was assessed by TUNEL staining in whole mount embryos. Apoptotic cells were stained in brown colour. There does not appear to be a difference between the numbers of apoptotic cells in wild-type embryos (*Cecr2$^{+/+}$*) compared to homozygoust *Cecr2$^{Tm1.1Hemc}$* mutant embryos (*Cecr2$^{-/-}$*). (Justin Elliott).
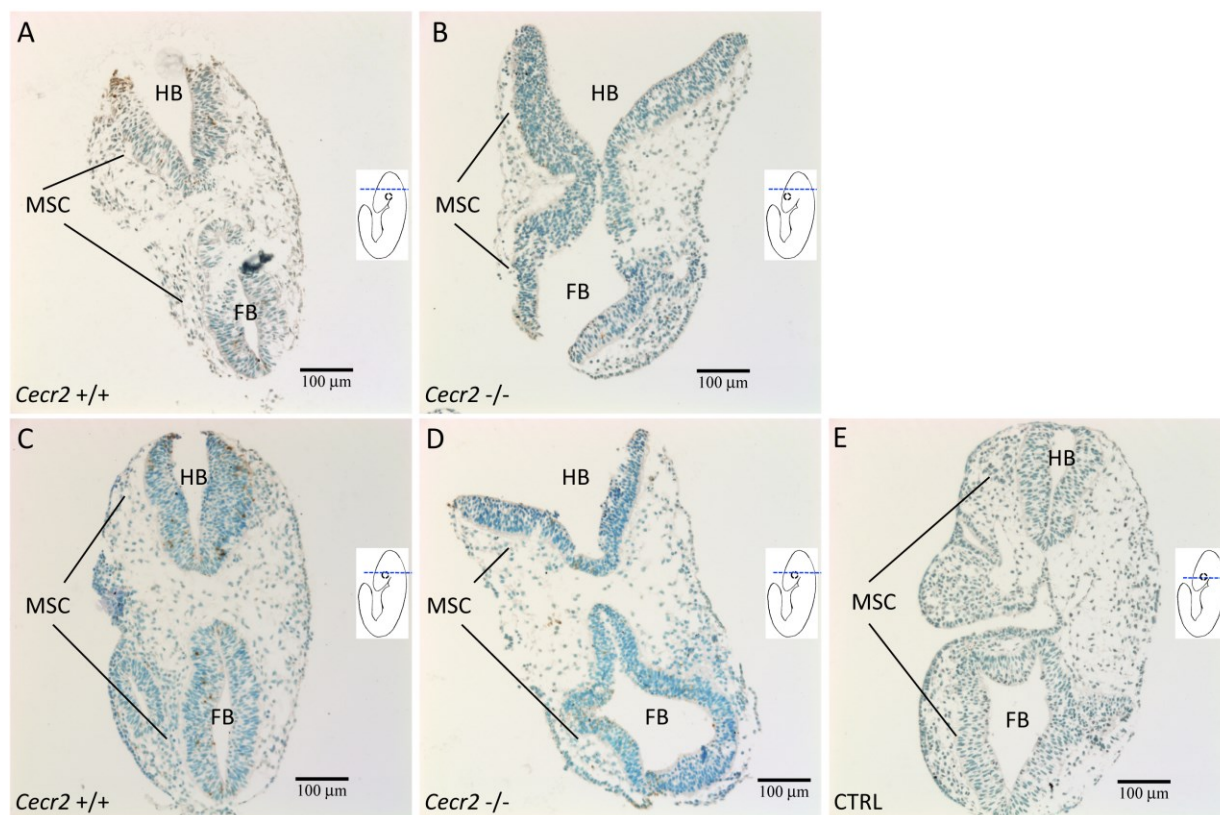
Figure 3-40 Apoptosis is not increased in CECR2 mutant embryos during the closure of the neural tube, as seen in sections. E9.5 embryos were collected and sectioned and the pattern of apoptosis was assessed by TUNEL staining. Apoptotic cells were stained in brown colour. There is not a detectable difference between the numbers of apoptotic cells in wild-type embryos (A,C) compared to homozygout $Cecr2^{Tm1.1Hemc}$ mutant embryos (B,D). Control section (CTRL) is stained using the same procedure except for not applying TdT enzyme to be used as a negative control. Forebrain (FB), Hindbrain (HB), Head mesenchyme (MSC), Dorsolateral hinge point (DLH). Drawings on the right side of the panels show the sectioning planes.

3.10.4. CCAR2: a strong candidate for a novel interacting partner of CECR2

3.10.4.1. CCAR2 is a new member of CECR2-containing complexes in mouse ES cells

CCAR2 (Cell Cycle And Apoptosis Regulator 2) was detected in three separate LC-MS/MS runs using separate IP samples in ES cells (2, 4 and 16 distinct peptides) and adult testis (3, 4 and 6 distinct peptides). CCAR2, also known as deleted in breast cancer 1 (DBC1) and KIAA1967, is a protein with a very wide range of functions, including cellular proliferation, transcription regulation, apoptosis and double-stranded DNA damage repair (Sakurabashi et al. 2015). CCAR2 is also involved in chromatin remodeling by regulating selected deacetylases and histone methyltransferases (Joshi et al. 2013). CCAR2 uses an N-terminal putative Leucine zipper to bind other proteins (Qin et al. 2015)

To investigate the interactions between CCAR2 and CECR2 in mouse ES cells, reciprocal co-IP assays were performed. In the input, CCAR2 was detected at the molecular size of ~130 kDa (Figure 3-41A and B, INPUT), which is consistent with a previous study (Park et al. 2014). CCAR2 co-immunoprecipitated with an antibody against CECR2 (Figure 3-41A). A reverse co-IP using an antibody for CCAR2 produced an interacting band for CECR2 (3-42B). Multiple reciprocal co-IPs of CECR2 and CCAR2 indicated that these proteins are present together in a complex in ES cells.

3.10.4.2. CCAR2 is a new member of CECR2-containing complex in adult mouse testis

Mass spectrometry analysis of CECR2-containing complexes isolated from adult mouse testis in three independent experiments suggested that CCAR2 is interacting with CECR2 in adult mice testis as well (Table 3-4). To confirm, reciprocal co-IP assays were performed on adult mouse testis lysates (Figure 3-42). CCAR2 co-immunoprecipitated using a CECR2 antibody (Figure 3-42A). A reverse co-IP with an anti-CECR2 antibody showed that CECR2 immunoprecipitated with CCAR2 (3-42B).

CCAR2 is expressed very strongly in testis and ES cells as judged by Western blot analysis (Figures 3-41 and 3-42, A and B, INPUT), but a very small amount of CCAR2 was co-immunoprecipitated by CECR2 (Figures 41 and 3-42, compare INPUT and IP lanes). This is consistent with the fact that CCAR2 is involved in other functions and complexes such as the DBIRD complex (Close et al. 2012). While IP of CCAR2 was consistent in ES cells, it often failed to immunoprecipitate in testes. Additionally, immunofluorescence staining of testis showed

that CECR2 and CCAR2 do not co-localize (Alaina Terpstra, unpublished). Thus CCAR2 may interact with CECR2 in mouse adult testis but more experiments are needed to confirm this interaction.
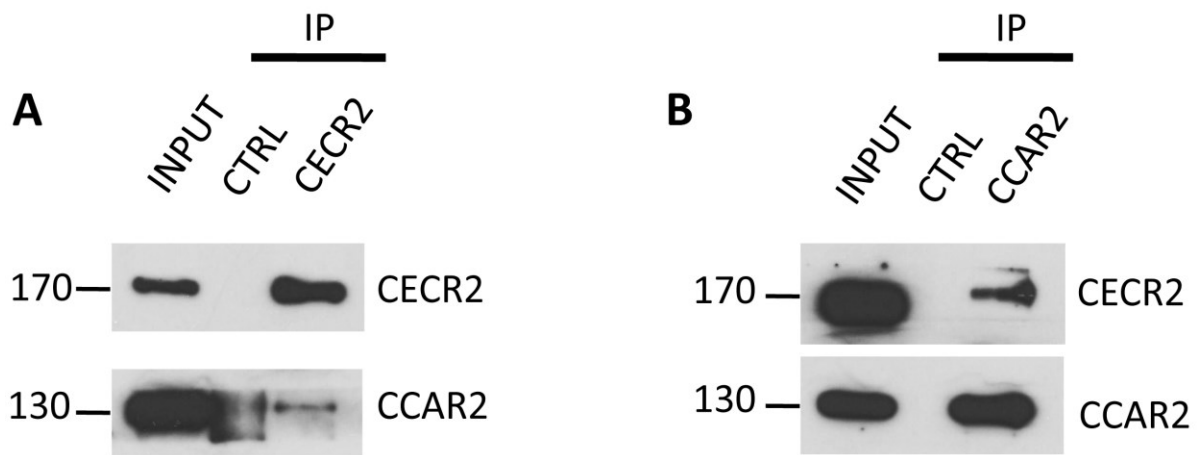
Figure 3-41 Reciprocal Co-IP of CECR2 and CCAR2 from whole cell extract prepared from mouse ES cells. A) Western blot for CECR2 and CCAR2 following IP with CECR2 antibody (CECR2) or rabbit IgG (CTRL). B) Western blot for CECR2 and CCAR2 following the reverse IP with CCAR2 antibody (CCAR2) or rabbit IgG (CTRL). The input represents 5% of extract used in the IP reaction. Approximately 25% of each IP reaction was used in each Western blot analysis. The molecular weight standards are shown as kDa on the left side of the figure.
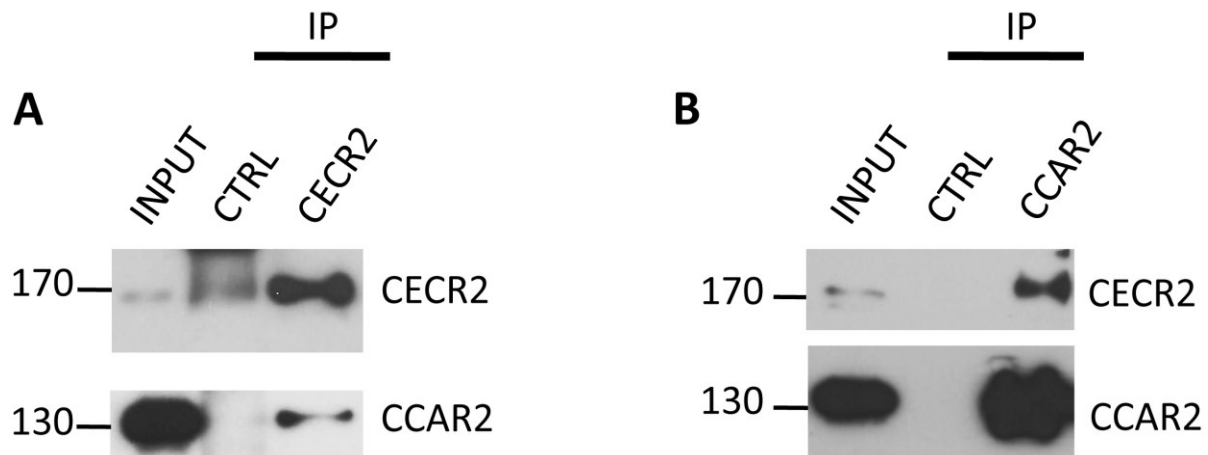
Figure 3-42 Reciprocal Co-IP of CECR2 and CCAR2 from whole cell extract prepared from adult mouse testis. Co-IP followed by Western blot were performed as described in Figure 3-35A, except using testis. The input represents 5% of extract used in the IP reaction. Approximately 25% of each IP reaction was used in each Western blot analysis. The molecular weight standards are shown as kDa on the left side of the figure. (Kenji Rowel Lim, unpublished).

Aim 2: Finding the binding sites/genes regulated by Cecr2 using ChIP-Seq

3.11. Chromatin binding targets of CECR2

The CERF CECR2-containing complexes have been shown to possess chromatin-remodeling activity (Banting 2004), but this activity is likely to be limited to specific genomic regions as shown for different ATP-dependent remodelers in previous studies (Ho et al. 2009a, Mikkelsen et al. 2007). In order to gain information about the genomic regions targeted by CECR2 in ES cells and testis, chromatin immunoprecipitation coupled to massively parallel high-throughput DNA sequencing (ChIP-seq) was used. Similar experiments have been done for other chromatin remodellers. For example, it has been shown that esBAF, a SWI/SNF chromatin remodeling complex, is enriched at transcription start sites of genes involved in the maintenance of self-renewal and pluripotency of mouse ES cells. To increase specificity, I performed ChIP-seq using antibodies for CECR2, SNF2H and LUZP1 in ES cells and CECR2 and SNF2H in testis, seeking to determine the overlapping distribution across the genome. The results of this analysis should lead to identification of the important target sites for the CECR2-containing complexes and provide a foundation for studying the regulation of specific candidate genes.

3.11.1. Optimizing ChIP-seq technique

Firstly I optimized the ChIP technique using a commercial RNA pol II mouse monoclonal antibody (Active motif, cat. no. 39097) with a known binding site on the genome. The DNA-binding proteins in ES cells were crosslinked to DNA *in vivo* by treating cells with formaldehyde and the chromatin was sheared by sonication into small fragments (200-500 bp range) using a Bioruptor UCD-200. Sheared chromatin was also prepared from adult mouse testis as described in section 2.20.2. To check the efficiency of the sonication, 5 μl of the sonicated chromatin was decross-linked and subjected to agarose gel electrophoresis. Figure 3-43A shows the sheared DNA after sonication in the optimal conditions that resulted in fragments 200-500 bp in mouse ES cells. RNA pol II mouse monoclonal antibody was used to perform ChIP using sheared ES cell and testis chromatin. The ChIPed DNA amplified a known target of RNA pol II on the gene *Eef1a1* (Figure 3-43B), confirming that the ChIP procedure was successfully optimized. The

same ChIP technique conditions were used in the main experiment to find the CECR2-containg complex binding sites.

The quality of ChIP experiments depends on the specificity of the CECR2 antibody. I already had shown that the CECR2 antibody can immunoprecipitate CECR2 protein from ES cells and adult testis lysates (Figure 3-16). I then showed that the CECR2 antibody was able to immunoprecipitate the crosslinked DNA protein complexes in the optimized ChIP reaction (Figure 3-43C) and was therefore appropriate for use in ChIP-seq experiments.
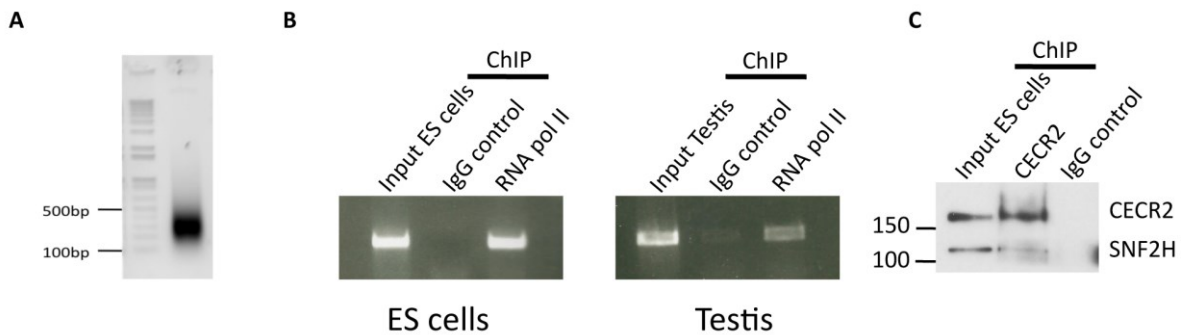
Figure 3-43 ChIP-seq optimization. A) Agarose gel electrophoresis analysis of sonication shearing of chromatin. TT2 ES cells were fixed with 1% formaldehyde and cross-linked chromatin was extracted from the cells and sonicated for 40 cycles of 30 sec ON/ 30 sec OFF at high setting with Bioruptor UCD-200 ultrasound sonicator. The sheared chromatin was decross-linked and run on a 2% agarose gel. A DNA smear of 200-500 bp indicates an optimized sonication condition. The left lane shows the DNA molecular weight ladder (MW). B) A control ChIP-PCR experiment showing the successful ChIP technique. After preparing cross-linked chromatin from ES cells and mouse testis, ChIP was performed using RNA pol II antibody or normal mouse IgG as a negative control. The end point PCR was performed on ChIPed DNA and the control Input DNA using EF1-alpha control primers. The PCR product was run on a 2% agarose gel. C) CECR2 antibody successfully immunoprecipitates cross-linked CECR2-containing complexes in mouse ES cells. Western blot for CECR2 and SNF2H following ChIP with anti-CECR2 antibody or rabbit IgG (CTRL) using chromatin extracts prepared from mouse ES cells. The input represents 10% of extract used in the IP reaction. Approximately 20% of each IP reaction was used in each Western blot analysis. The molecular weight standards are shown as kDa on the left side of the figure.

3.11.2. Chromatin immunoprecipitation followed by sequencing (ChIP-seq)

To compare chromatin binding sites of CECR2, SNF2H and LUZP1, I did ChIP experiments with our CECR2 antibody and commercial antibodies for SNF2H and LUZP1 using

chromatin extracts prepared from mouse ES cells. For the experiment I used two different negative controls.

1) "Mock" ChIP control using normal rabbit IgG, to detect nonspecific binding.

2) "Input" DNA, which is a sample of the original crosslinked and sheared chromatin before the IP step.

Two biological replicate experiments from independent cell cultures were used to assess reproducibility.

A similar procedure was used for adult testis. Because LUZP1 is not part of the CECR2 complex in testis (Figure 3-33), the ChIP experiments were performed using only CECR2 and SNF2H antibodies in testis. Two biological replicates and two technical replicates were used for ES cells and testis. Table 3-6 shows the total number of reads per sample obtained from the paired-end sequencing. For each biological replicate experiment, I obtained more than 10 million mapped reads, which is the minimum number of reads suggested by ENCODE for mammalian genomes (Landt et al. 2012). Compared to published data with similar technology for LUZP1, my sequencing gave more number of reads per sample and also a higher mapping rate indicating a better quality of the sequencing (Table 3-6).

Table 3-6 Overview of sequenced and mapped reads per data set. The data are the combination of reads from two technical replicates. Krebs et al (2010) read numbers and mapping rates (grey highlight) are given for comparison to a study with similar technology.

	Sample name	Number of clusters	Number of Reads	Mapping rate (%)
ES cells	CECR2_Rep2	45,848,001	91,696,002	76.7
	CECR2_Rep2	35,501,098	71,002,196	63.7
	SNF2H_Rep1	47,437,644	94,875,288	71.8
	SNF2H_Rep2	19,829,946	39,659,892	70.48
	LUZP1_Rep1	40,766,491	81,532,982	70.48
	LUZP1_Rep2	11,802,202	23,604,404	44.54
	Input	39,632,629	79,265,258	85.6
	IgG	13,544,343	27,088,686	61.9
Testis	Cecr2_Rep1	49,257,949	98,515,898	84.7
	Cecr2_Rep2	31,020,553	62,041,106	61.3
	SNF2H_Rep1	18,729,225	37,458,450	69.4
	SNF2H_Rep2	42,988,651	85,977,302	73
	IgG	32,872,008	65,744,016	68.9
(Krebs et al. 2010)	LUZP1 (ESC)		16,600,000	41%
	GCN5 (ESC)		14,700,000	56%

3.11.3. Read mapping and peak calling

In ES cells, I employed input as the control in peak calling using the program MACS. A P-value cutoff of 1e-05, gave 24,183, 12,838 and 13,486 peaks (merged peak sets of biological replicates) for CECR2, SNF2H and LUZP1, respectively. For testis, I employed IgG-ChIP as the control in peak calling using MACS. A P-value cutoff of 1e-03 (P-value cutoff of 1e-05 resulted in too few peaks), gave 16,699 and 18,657 peaks (merged peak sets of biological replicates) for CECR2 and SNF2H respectively. To assess the quality of the ChIP-seq data, the evolutionary conservation of the identified peaks was evaluated by PhastCons analysis using the Cistrome/Galaxy program (Liu et al. 2011). Cis-regulatory elements that are binding targets of transcription factors are more evolutionary conserved. In a successful ChIP-seq experiment, the average PhastCons score in the summit vicinity is higher than the flanking region. Figure 3-44 shows that the peaks obtained from my ES cells and testis ChIP-seq experiments have higher conservation scores near the summit.
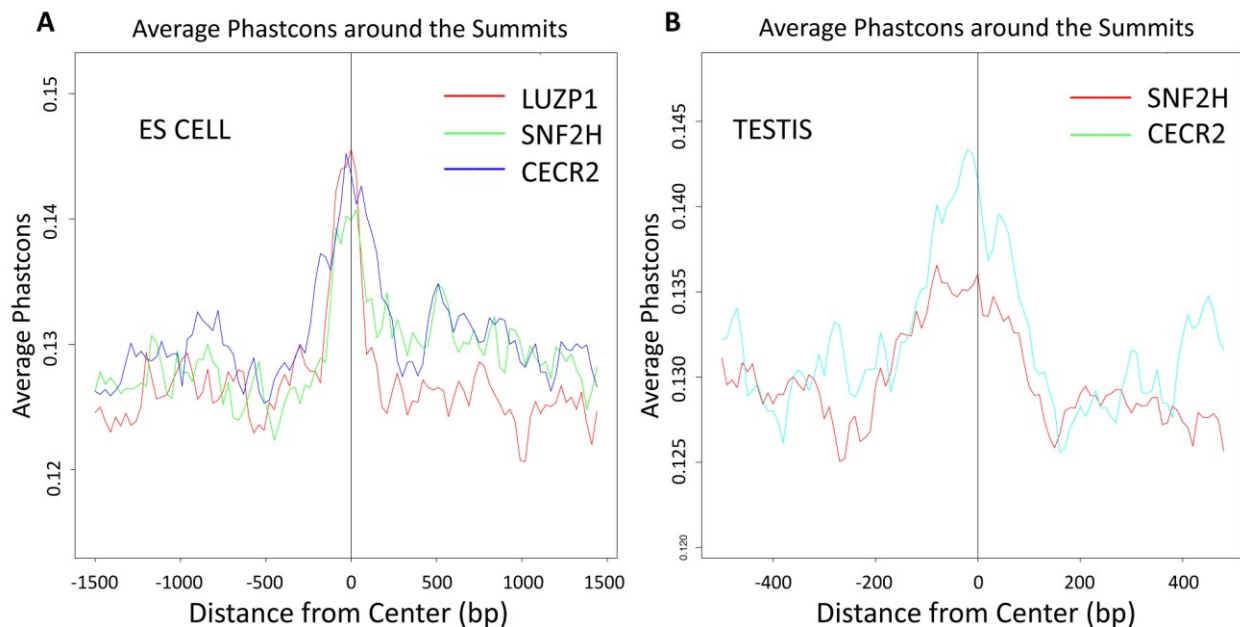
Figure 3-44 The average PhastCons score around the peak summits for ES cell and testes experiments show the acceptable quality of ChIP-seq peak calls. The 500 bp region of binding sites with the peak summits in the center were created from peaks obtained from ChIP-seq experiment on ES cells for CECR2, SNF2H and LUZP1 (A) and on testis for CECR2 and SNF2H (B) and used to plot the average PhastCon scores. The graph shows that the summit is more evolutionarily conserved than the surrounding region. A PhastCon score of 1.0 indicates that the sequence is 100% identical among placental mammals.

To determine whether common binding sites were shared by CECR2, SNF2H and LUZP1 in ES cells, I performed pairwise comparisons of the binding sites occupied by each protein. The majority of binding sites of CECR2 and SNF2H were not in common between these two proteins. Only 12.25% of CECR2 sites overlapped with 23% of SNF2H binding sites in ES cells (Figure 3-45A). Comparing CECR2 binding sites with LUZP1 sites in ES cells showed that 11.63% of CECR2 binding sites were co-occupied by 20.87% of LUZP1 binding sites. Comparison between LUZP1 and SNF2H showed that 7.81% of SNF2H binding sites were bound by LUZP1, and similarly 7.44% of LUZP1 binding sites were co-occupied by SNF2H in ES cells. Comparing all three proteins' binding sites revealed that 20.76% of CECR2, 39.13% of SNF2H and 37.25% of LUZP1 binding sites were shared by the other two proteins in ES cells. The same calculation for testis dataset showed that 8.18% of CECR2 binding sites were occupied by SNF2H, whereas 7.3% of SNF2H binding sites were also bound by SNF2H (Figure 3-45B). Since I have shown the interaction of CECR2 with LUZP1 and SNF2H using IP experiments, the colocalization of these proteins could indicate sites that are more likely CECR2 binding sites.

Distances from the center of the peaks to the nearest transcription start site (TSS) were determined by GREAT version 3.0.0 (McLean et al. 2010) for ES cells and testis (Figure 3-46 to 50). The least percentage of peaks was distributed within the -5 to + 5 kb region relative to TSS of the associated genes, which are likely to contain enhancer regions in both ES cells and testis for all three proteins. Most of the peaks mapped between 50kb and 500kb upstream and downstream of transcriptional start sites. This is similar to the binding profile of both a general enhancer-associated protein such as p300 in mouse ES cells, midbrain and forebrain (Visel et al. 2009), sequence-specific transcription factors such as neuron-restrictive silencer factor (NRSF) in the human Jurkat cell line, STAT3 in mouse ES cells (McLean et al. 2010) and chromatin remodeling factor such as BAF170, BAF155 and BRG1 in HeLa cells (Wang et al. 2014). This pattern differs from those of transcription factors such as RNA pol II in HeLa cells (Wang et al. 2014) and growth-associated binding protein (GABP) in the human Jurkat cell line with the highest number of the peaks close to the promoters (Valouev et al. 2008). My data suggests that the CECR2 complex regulates genes over long distances.
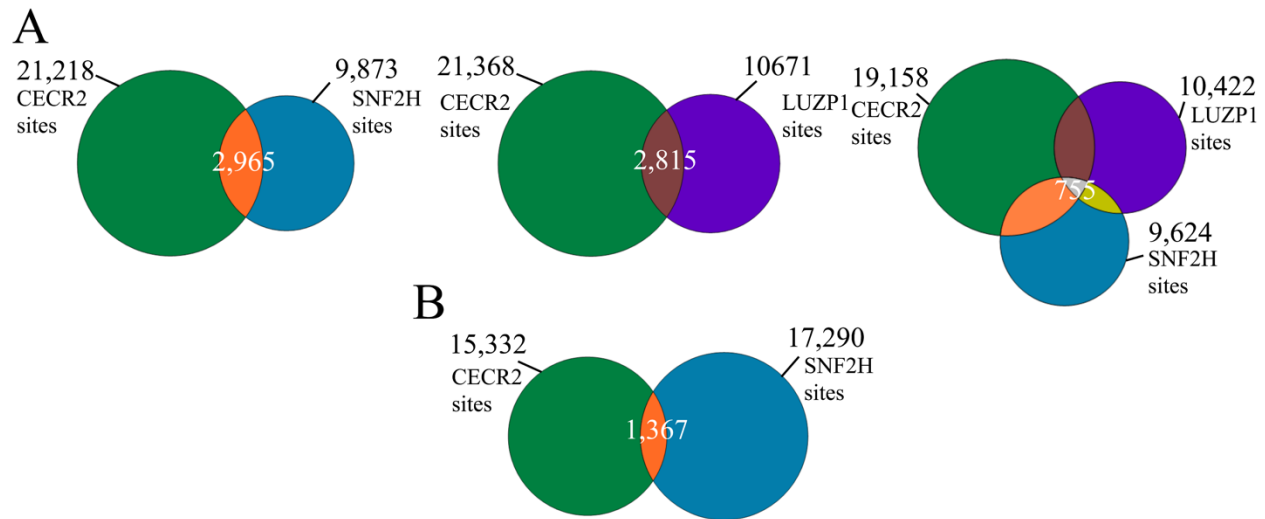
Figure 3-45 CECR2 co-occupy some genomic regions with SNF2H and LUZP1. A) Venn diagrams displaying number of overlapping binding sites of CECR2, SNF2H and LUZP1 in ES cells. B) Venn diagrams displaying overlaps in adult testis. The numbers indicate the number of binding sites for each protein and the number of overlapping binding sites.

To determine the location of the peaks relative to the genes globally, the distribution of the peaks was annotated according to the genomic regions (promoter, exons, introns and intergenic regions) using the Cis-Regulatory Element Annotation System (CEAS) web application (Shin et al. 2009) (Figure 3-46 to 50). Approximately 5% and 4.5% of CECR2 peaks were located in the promoter region (defined as - 5 kb upstream of TSS), 35% and 33.4% reside in gene bodies (exons and introns) and around 55% and 58.4% were localized in the distal intergenic regions in ES cells and testis, respectively (Figure 3-46B and 3-50B). I integrated ChIP-seq peak signals in a meta-gene (a hypothetical gene) plot to summarize ChIP-seq data complexity and to show a simple visual representative of the general occupancy pattern of the proteins. Examining CECR2 binding patterns near genomic features by computing the average CECR2 signal profile around the TSS (± 10 kb from TSS) and over the 3-kb meta-gene showed very low signal within ± 500 bp from the TSS and transcription termination site (TTS) compared to the body of the meta-gene in both ES cells and testis (Figure 3-46 and 50C,D). For LUZP1 in ES cells the pattern was different than CECR2. The ChIP enrichment for LUZP1 was low near the TSS and there was an increase after the transcription termination site (TTS) (Figure 48D). Figure 3-48C shows the occupancy pattern of LUZP1 around the TSS in a higher resolution. Close to the TSS (± 1000 bp) has a low signal and the highest increase in signal can be seen between ~1000-1500 bp upstream and downstream of the TSS following with a sudden decrease

136

in the signal. The signal for SNF2H, similar to CECR2, was low around the TSS and TTS and a high signal in the gene body (Figure 3-47D). Unlike to CECR2, the signals increase upstream of the TSS and downstream of the TTS and reaches a higher level compared to the gene body. The most obvious difference between the ChIP-seq enrichment for SNF2H in ES cells and testis was a higher signal between 500-1000 bp upstream of TSS in Testis (Figure 50C,D).
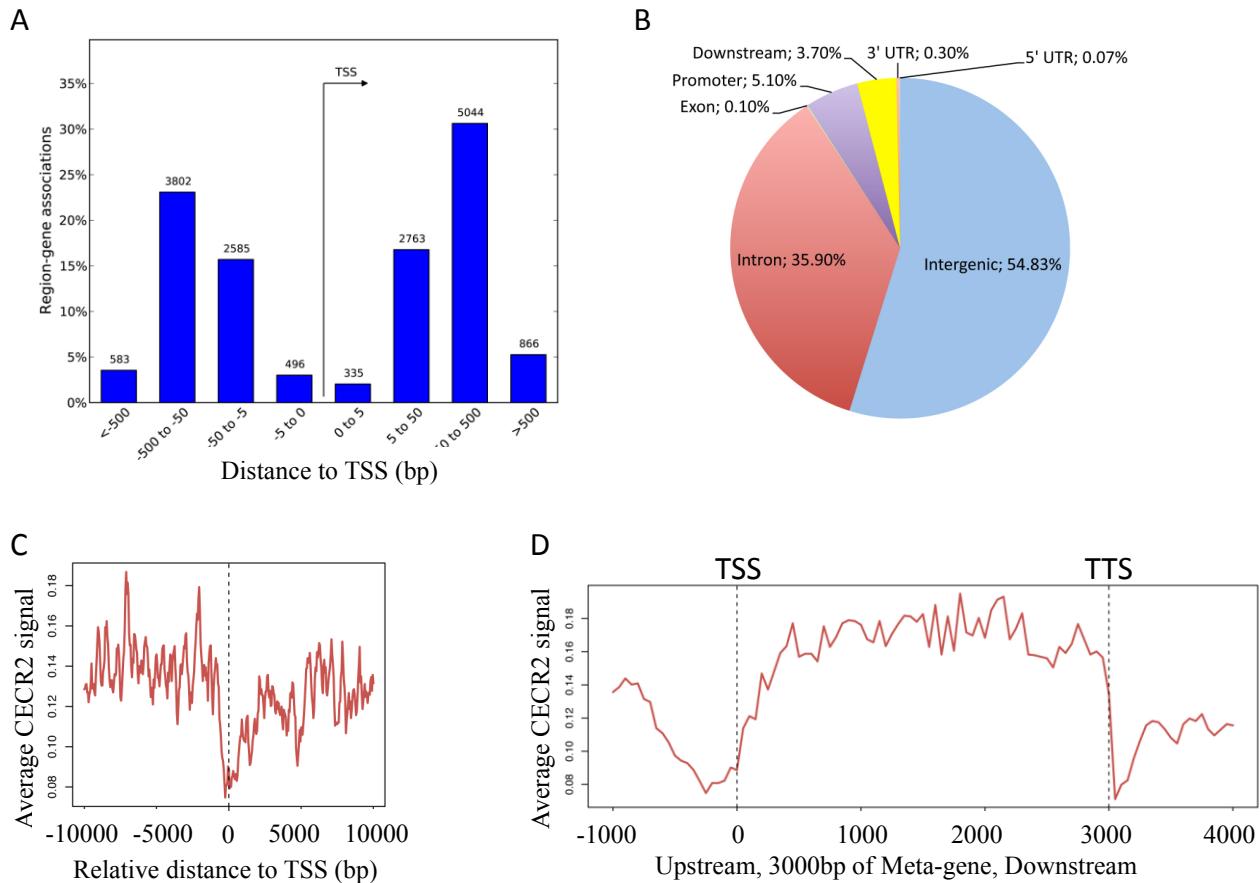
Figure 3-46 Genomic distribution of CECR2 in ES cells analyzed by CEAS. A) Number of CECR2 peaks distributed on the genome relative to the distance from transcription start sites (TSS). B) Pie chart representing the genomic distribution of CECR2 peaks in ES cells relative to gene annotations: Exon, Intron, Promoter (- 5 kb), Downstream (+ 5 kb), 3' UTR and 5' UTR. C) Average signal profile of CECR2 around a TSS and D) over the meta-gene. Numbers on the bar graph represent the number of peaks detected on the distance shown on the x-axis.
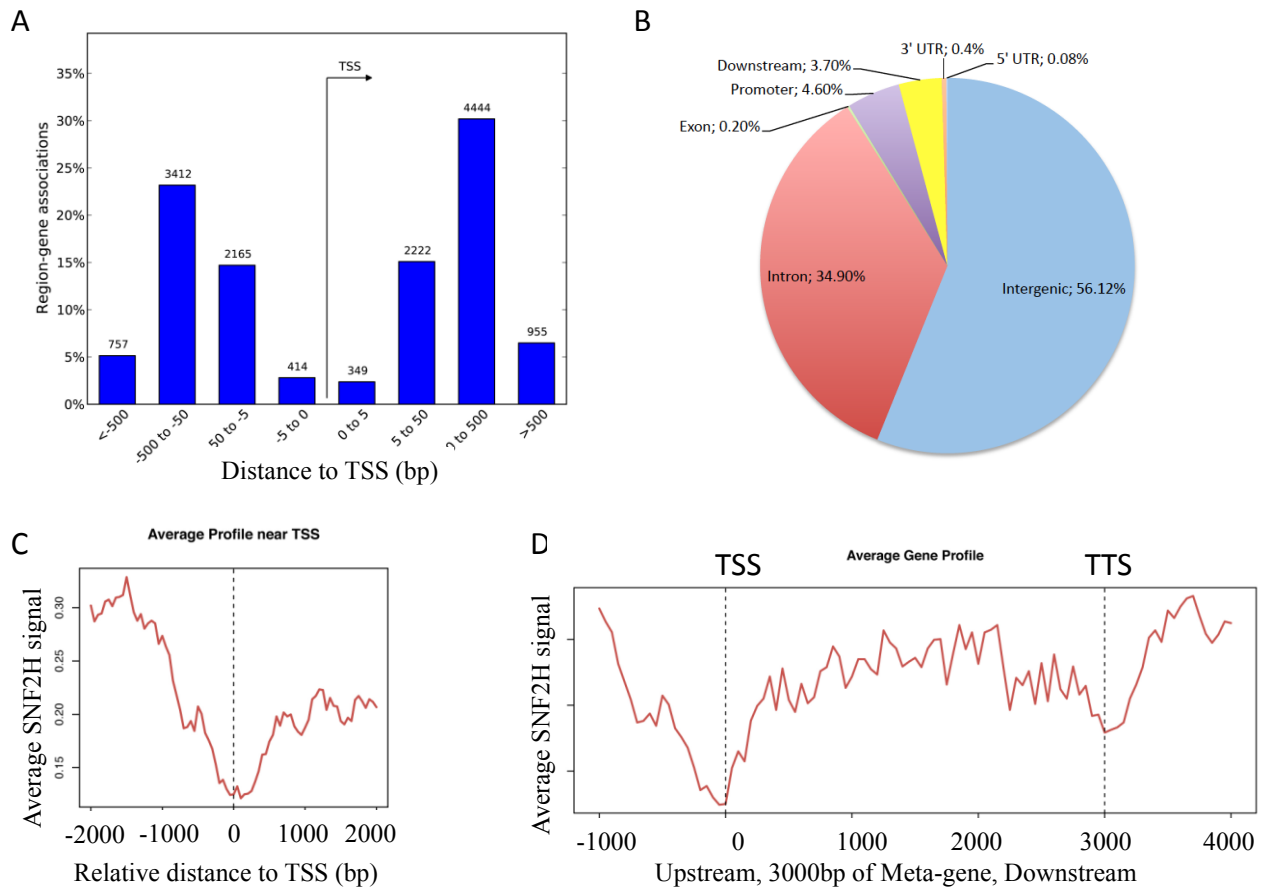
Figure 3-47 Genomic distribution of SNF2H in ES cells analyzed by CEAS. A) Number of SNF2H peaks distributed in the genome relative to the distance from transcription start sites (TSS). B) Pie chart representing the genomic distribution of SNF2H peaks in ES cells relative to gene annotations: Exon, Intron, Promoter (- 5 kb), Downstream (+ 5 kb), 3' UTR and 5' UTR. C) Average signal profile of SNF2H around a TSS and D) over meta-gene. Numbers on the bar graph represent the number of peaks detected on the distance shown on the x-axis.
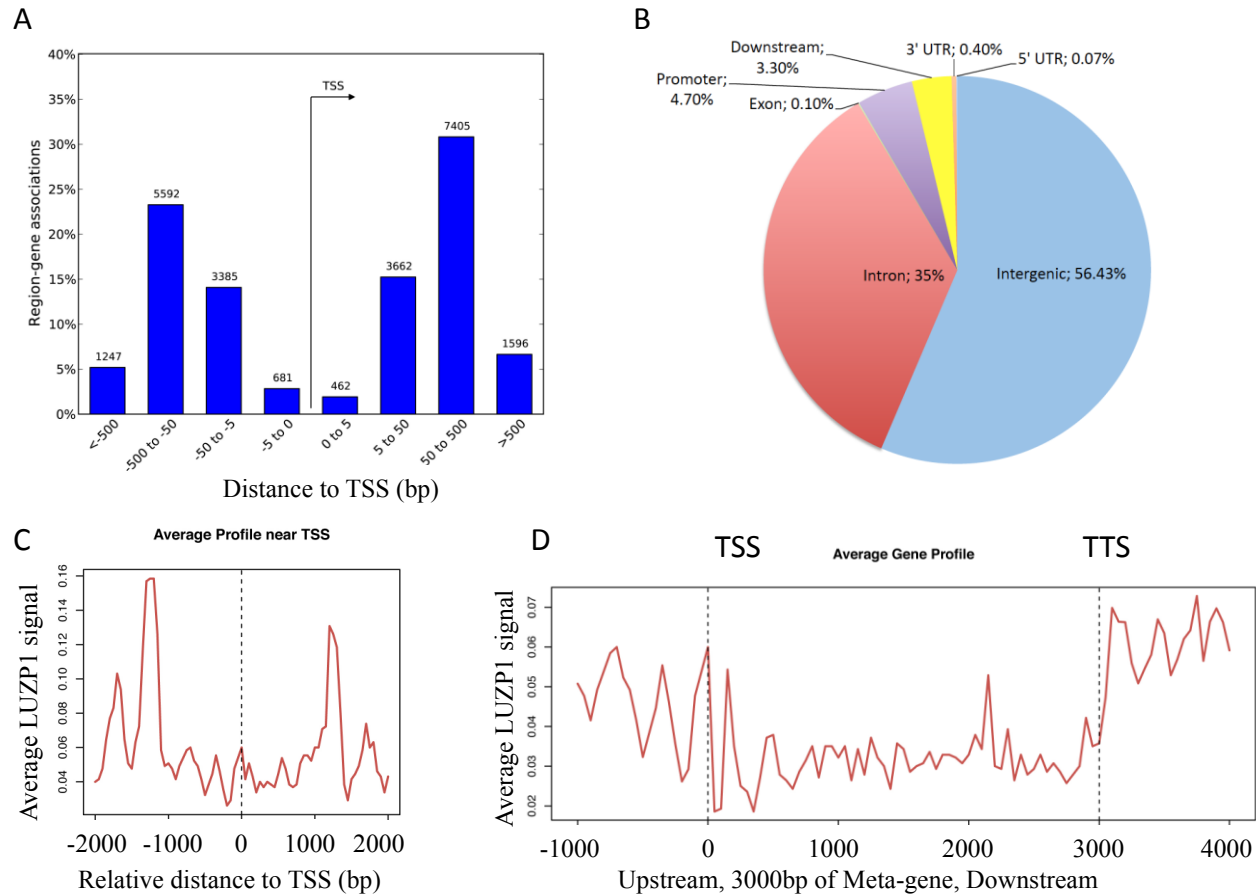
Figure 3-48 Genomic distribution of LUZP1 in ES cells analyzed by CEAS. A) Number of LUZP1 peaks distributed on the genome relative to the distance from transcription start sites (TSS). B) Pie chart representing the genomic distribution of LUZP1 peaks in ES cells relative to gene annotations: Exon, Intron, Promoter (- 5 kb), Downstream (+ 5 kb), 3' UTR and 5' UTR. C) Average signal profile of LUZP1 around a TSS and D) over a meta-gene. Numbers on the bar graph represent the number of peaks detected on the distance shown on the x-axis.
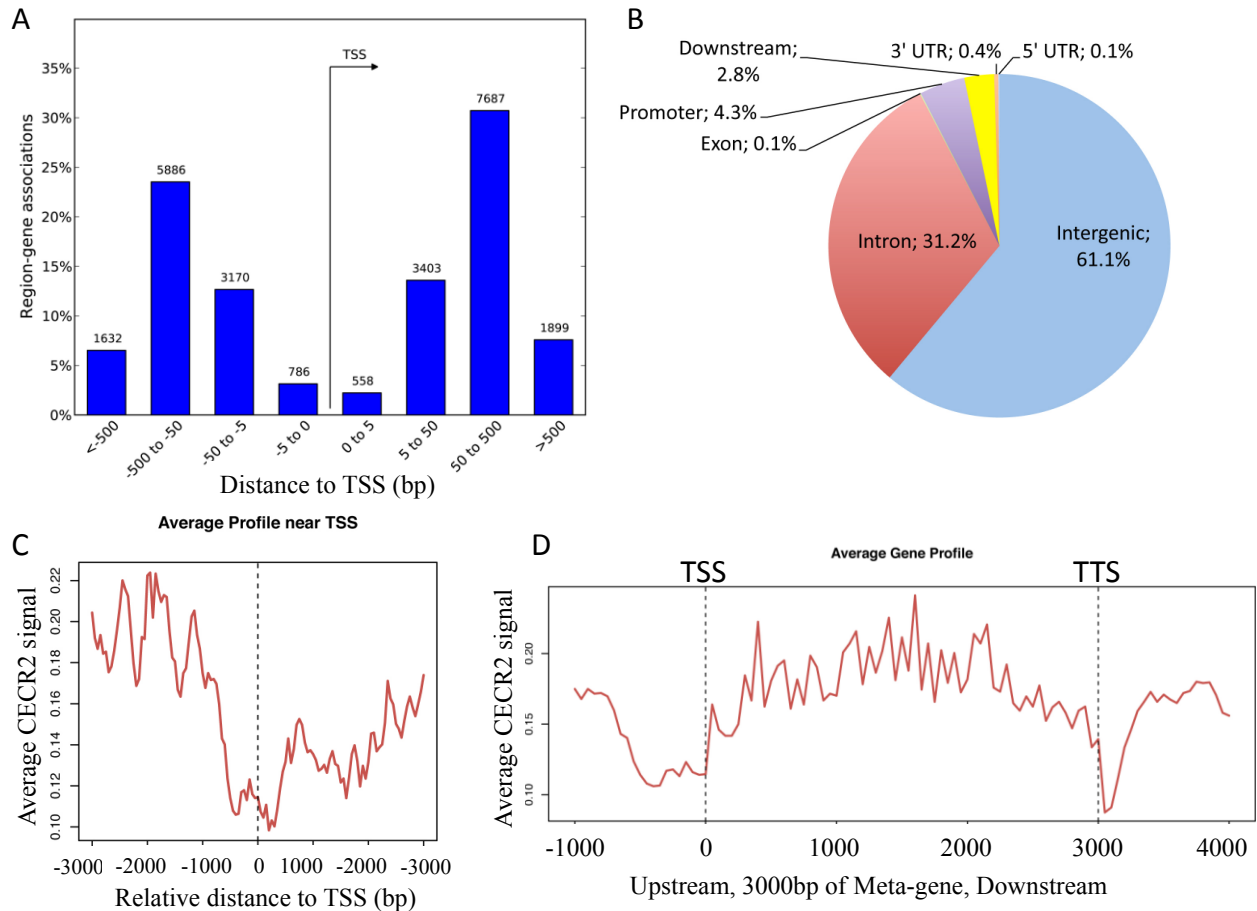
Figure 3-49 Genomic distribution of CECR2 in testis analyzed by CEAS. A) Number of CECR2 peaks distributed on the genome relative to the distance from transcription start sites (TSS). B) Pie chart representing the genomic distribution of CECR2 peaks in ES cells relative to gene annotations: Exon, Intron, Promoter (- 5 kb), Downstream (+ 5 kb), 3' UTR and 5' UTR. C) Average signal profile of CECR2 around a TSS and D) over a meta-gene. Numbers on the bar graph represent the number of peaks detected on the distance shown on the x-axis.
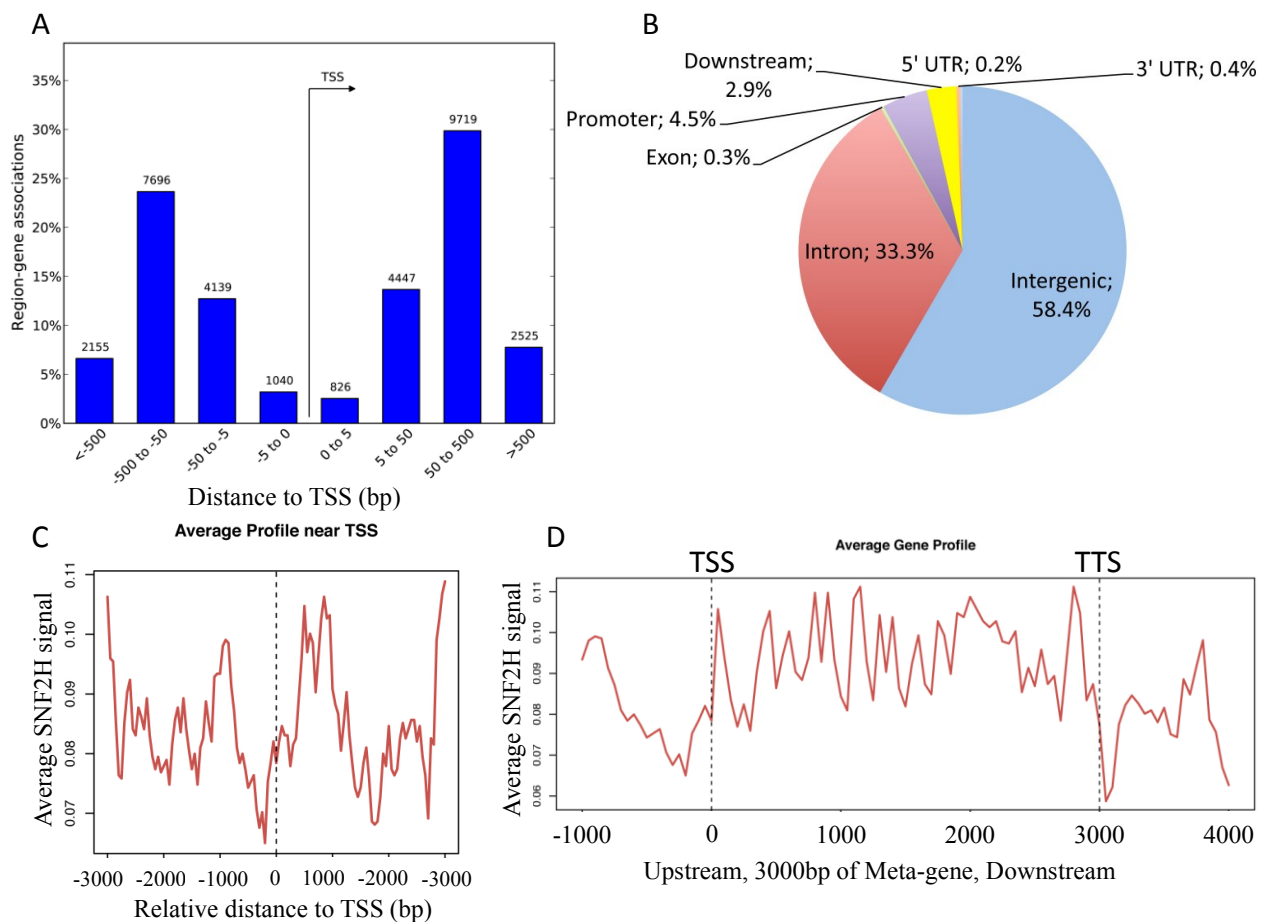
Figure 3-50 Genomic distribution of SNF2H in testis analyzed by CEAS. A) Number of SNF2H peaks distributed on the genome relative to the distance from transcription start sites (TSS). B) Pie chart representing the genomic distribution of SNF2H peaks in ES cells relative to gene annotations: Exon, Intron, Promoter (- 5 kb), Downstream (+ 5 kb), 3' UTR and 5' UTR. C) Average signal profile of SNF2H around a TSS and D) over a meta-gene. Numbers on the bar graph represent the number of peaks detected on the distance shown on the x-axis.

3.11.4. Gene ontology terms (ES cells)

To investigate the molecular and cellular processes of the genes downstream of CECR2 and possible mechanisms of CECR2 involvement in neural tube development and infertility, the Genomic Regions Enrichment of Annotation Tool (GREAT) was used to carry out gene ontology enrichment analysis. GREAT associates submitted genomic regions (ChIP-seq peaks) with the nearby genes by specifying a regulatory domain consisting of a basal domain (- 5 kb to + 1 kb of TSS) and an extension (up to 1Mb) domain for each gene. By analyzing the annotations of the genes located close to the submitted genomic regions, GREAT assigns biological meanings to them (McLean et al. 2010). The peaks obtained from the MACS program were sorted by - 10*log10(pvalue) and filtered to retain only peaks with a score >75th percentile of 10*log10(pvalue). The selected peaks were analyzed by GREAT to identify cellular components, biological processes and molecular functions. The enriched molecular functions for the CECR2 data set obtained from ES cells included "Wnt-protein binding" and "Wnt-activated receptor activity" (Table 3-7). The "Wnt-protein binding" term, the most abundant group, contains 55 binding regions located on 19 genes. "Wnt-activated receptor activity" was the second most abundant group with 35 regions in 13 genes. Appendix p shows the list of the genes included in these GO terms and their associated genomic regions. The data suggest that Wnt-signaling genes are modulated by CECR2. Biological processes involved in neural development were significantly over-represented in the data obtained from ES cell experiment, including "cerebellum development", "metencephalon development", "cerebellar cortex development", "hindbrain morphogenesis" and "cerebellum morphogenesis". Other GO terms related to embryonic development included heart and kidney development (Table 3-7). Appendices Q and R show the list of the genes included in these GO terms and their associated genomic regions. The data indicated that CECR2 binds to the cis-regulatory elements of genes that are involved in brain, heart and kidney development.

142

Table 3-7 GREAT gene ontology terms from analysis of CECR2 peaks from ES cells data set at a false discovery rate of 0.05. GO terms related to brain development are highlighted in yellow.

	Ontology	Term name	Binomial raw P-Value	Binomial FDR Q-Value	Binomial fold enrichment	#Hit genes
ES Cells	GO Molecular Function	Wnt-protein binding	7.74E-8	3.24E-6	2.11	19
		Wnt-activated receptor activity	2.25E-5	5.17E-4	2.08	13
	GO Biological Process	lymph vessel morphogenesis	8.71E-31	3.81E-28	6.1	9
		cerebellum development	1.02E-26	2.28E-24	2.44	44
		positive regulation of stem cell differentiation	6.53E-25	1.15E-22	3.72	9
		metencephalon development	1.05E-24	1.81E-22	2.26	51
		regulation of timing of neuron differentiation	1.00E-19	8.18E-18	4.96	5
		cerebellar cortex development	4.72E-19	3.62E-17	2.55	25
		hindbrain morphogenesis	1.71E-18	1.21E-16	2.55	23
		cardiac muscle tissue growth	5.38E-18	3.68E-16	3.13	13
		cerebellum morphogenesis	5.76E-18	3.92E-16	2.62	20
		muscle cell proliferation	1.21E-17	7.94E-16	3.19	13
		heart growth	1.43E-15	7.76E-14	2.8	14
		regulation of integrin activation	1.56E-15	8.43E-14	4.53	8
		ectodermal placode formation	3.03E-15	1.59E-13	2.68	10
		mesoderm formation	4.29E-15	2.23E-13	2.16	29
		positive regulation of developmental growth	5.68E-15	2.89E-13	2.23	24
		organ growth	1.13E-14	5.46E-13	2.2	23
		cardiac muscle cell development	6.82E-14	3.01E-12	2.15	26
		regulation of development, heterochronic	1.33E-13	5.58E-12	3.11	10
		regulation of timing of cell differentiation	2.83E-13	1.16E-11	3.09	9
		epithelial cell differentiation involved in kidney development	3.48E-13	1.42E-11	2.6	14

From the ES cell data sets, the CECR2 peaks were intersected with the SNF2H peaks to obtain the overlapping genomic targets of both proteins. Minimum overlap of 1 bp between the peaks was taken into account. From 2965 overlapping peaks, only peaks with a score >75th percentile of 10*log10(pvalue) were selected. The 642 unique overlapping genomic regions were used to perform gene ontology enrichment analysis using GREAT to identify biological processes related to the genomic regions shared with CECR2 and SNF2H proteins (Table 3-8). The enriched biological processes were regulation of "protein complex assembly" with 25 genes, "cerebellum development" with 14 genes, "metencephalon development" with 15 genes,

"hindbrain development" with 19 genes and "positive regulation of protein complex assembly" with 13 genes (Table 3-9). Three out of five enriched biological processes were involved in brain development. The GO terms associated with overlapping peaks of CECR2 and SNF2H showed that these two proteins are involved in regulating the genes that are important in brain development and reproduction. *Shh, Gli2, Map3k7, Sema4c and Sec24b* were among the genes that were associated with the enriched biological processes GO terms for overlapping binding targets of CECR2 and SNF2H (Table 3-9). All of these cause exencephaly when mutated (Echelard et al. 1993, Maier et al. 2011, Merte et al. 2010, Pan et al. 2009, Shim 2005). *Sec24b* mutants die with craniorachischisis, a completely open neural tube. *Cyp11a1*, *Lmx1a*, *Cdkn1b*, *Csf2* and *Fmn2* were among the genes containing overlapping binding sites of CECR2 and SNF2H in ES cells and all of them have a role in reproduction (Table 3-9). Intersecting data sets from the three proteins (CECR2, SNF2H and LUZP1) resulted in 755 overlapping peaks. GREAT assigned biological process GO terms to 436 unique overlapping genomic regions. All of the GO terms were related to kidney development (Table 3-10).

Table 3-8 GREAT gene ontology terms from analysis of shared genomic regions of CECR2 and SNF2H from ES data sets at a false discovery rate of 0.05. GO terms related to brain development are highlighted in yellow.

	Ontology	Term name	Binomial raw P-Value	Binomial FDR Q-Value	Binomial fold enrichment	#Hit genes
ES Cells	GO Biological Process	regulation of protein complex assembly	3.62E-06	0.03	2.63	25
		cerebellum development	3.95E-06	0.01	3.49	14
		metencephalon development	8.42E-06	0.02	3.18	15
		hindbrain development	1.30E-05	0.03	2.61	19
		positive regulation of protein complex assembly	3.38E-05	0.04	3.37	13

Table 3-9 Genes associated with "biological process" GO terms assigned for CECR2 and SNF2H overlapped peaks in ES cells. Green – mouse mutations associated with exencephaly. Red – mouse mutations associated with reproduction defects.

	Gene	GO Terms
ES cells	*Cdk5r1*	cerebellum development, metencephalon development, hindbrain development
	Cyp11a1	cerebellum development, metencephalon development, hindbrain development
	Ezh2	cerebellum development, metencephalon development, hindbrain development
	Gli2	cerebellum development, metencephalon development, hindbrain development
	Ldb1	cerebellum development, metencephalon development, hindbrain development
	Lhx1	cerebellum development, metencephalon development, hindbrain development
	Lmx1a (infertile, inner ear)	cerebellum development, metencephalon development, hindbrain development
	Lmx1b	cerebellum development, metencephalon development, hindbrain development
	Myo16	cerebellum development, metencephalon development, hindbrain development
	Prox1	cerebellum development, metencephalon development, hindbrain development,
	Sema4c	cerebellum development, metencephalon development, hindbrain development
	Skor2	cerebellum development, metencephalon development, hindbrain development
	Sstr2	cerebellum development, metencephalon development, hindbrain development
	Zbtb18	cerebellum development, metencephalon development, hindbrain development
	Sec24b	metencephalon development, hindbrain development, regulation of protein complex assembly
	Gata2	hindbrain development
	Mafb	hindbrain development
	Plxna2	hindbrain development
	Shh	hindbrain development
	Bid	positive regulation of protein complex assembly, regulation of protein complex assembly
	Cd24a	positive regulation of protein complex assembly, regulation of protein complex assembly
	Cdkn1b	positive regulation of protein complex assembly, regulation of protein complex assembly
	Csf2	positive regulation of protein complex assembly, regulation of protein complex assembly
	Dlg1	positive regulation of protein complex assembly, regulation of protein complex assembly
	Fmn2	positive regulation of protein complex assembly, regulation of protein complex assembly
	Foxc2	positive regulation of protein complex assembly, regulation of protein complex assembly
	Grb2	positive regulation of protein complex assembly, regulation of protein complex assembly
	Il5	positive regulation of protein complex assembly, regulation of protein complex assembly
	Irf8	positive regulation of protein complex assembly, regulation of protein complex assembly
	Nck2	positive regulation of protein complex assembly, regulation of protein complex assembly
	Pmaip1	positive regulation of protein complex assembly, regulation of protein complex assembly
	Snx9	positive regulation of protein complex assembly, regulation of protein complex assembly

Arfgef1	regulation of protein complex assembly	
Eif2ak2	regulation of protein complex assembly	
Eln	regulation of protein complex assembly	
Eps8	regulation of protein complex assembly	
Hjurp	regulation of protein complex assembly	
Lefty1	regulation of protein complex assembly	
Map3k7	regulation of protein complex assembly	
Pak3	regulation of protein complex assembly	
Ralb	regulation of protein complex assembly	
Sptbn1	regulation of protein complex assembly	
Trim30a	regulation of protein complex assembly	

Table 3-10 GREAT gene ontology terms from analysis of shared genomic regions of CECR2, SNF2H and LUZP1 from ES data sets.

	Ontology	Term name	Binomial raw P-Value	Binomial FDR Q-Value	Binomial fold enrichment	#Hit genes
ES cell	**GO Biological Process**	glomerular endothelium development	1.26E-06	0.01	28.07	3
		glomerular mesangial cell differentiation	3.15E-06	0.01	23.22	3
		glomerular mesangium development	5.41E-06	0.01	20.76	3
		vascular endothelial growth factor receptor signaling pathway	4.72E-05	0.03	5.55	6
		mesangial cell differentiation	9.62E-05	0.04	11.29	3

Then I focused on genes with detected peaks in the proximal promoter regions (from 1 kb upstream of the TSS to TSS). First the sequence of 1 kb upstream of transcription start sites for the *Mus musculus* (mm9) assembly were obtained from the UCSC website. Then, the CECR2 peaks were intersected with the UCSC known genes database (containing the sequence of 1 kb upstream of TSS of all known genes) to identify the genes bearing CECR2 peaks in ES cells. There are 334 genes (coding and non-coding genes) with at least one CECR2 peak in their proximal promoter region (Appendix S). To understand the general function of these genes, I used the PANTHER classification system (PANTHER GO-Slim Biological Process, PANTHER Protein Class, PANTHER Pathways) to identify GO terms over-represented by these genes (Mi et al. 2016). The genes were classified into different molecular functions including "catalytic activity" (23% of the genes), "nucleic acid binding transcription factor activity" (11% of the genes) and "receptor activity" (10.30% of the genes). Table 3-11 shows all of the GO terms enriched in 234 coding genes with CECR2 peaks in their proximal promoter region. PANTHER

analysis of these genes also showed 37 signal pathways for these genes. "Heterotrimeric G-protein signaling pathway" (genes: *Gng7, Creb3l1, Adora3, Ssr2, Rcvrn & Cnga3*), "Inflammation mediated by chemokine and cytokine signaling" (genes: *Ccr10, Actg2, Plcg2 & Ccr6*), "Wnt signaling pathway" (genes: *Actg2, Gng7 & Cdh7*) and "Cadherin signaling pathway" (genes: *Actg2, Ctnnd1 & Cdh7*) were the pathways with the highest number of gene hits. This suggests that CECR2 may be involved in a broad range of biological processes as expected for chromatin remodellers.

Table 3-12 shows the list of the 103 coding and 33 non-coding RNA genes bearing overlapping CECR2 and SNF2H peaks within their promoter region (up to 5 kb upstream of TSS) in ES cells. GO analysis using the PANTHER classification system for 103 coding genes classified them into different molecular functions including "binding", "catalytic activity",

Table 3-11 Gene ontology terms enriched in 234 ES cell coding genes with CECR2 peaks in their proximal promoter region (up to 1 kb upstream of TSS).

	Ontology	Category name (Accession)	Number of genes	Percent of gene hit[1]
ES Cells	GO Molecular Function	binding (GO:0005488)	66	28.20%
		catalytic activity (GO:0003824)	54	23.10%
		nucleic acid binding transcription factor activity (GO:0001071)	26	11.10%
		receptor activity (GO:0004872)	24	10.30%
		transporter activity (GO:0005215)	21	9.00%
		structural molecule activity (GO:0005198)	14	6.00%
		enzyme regulator activity (GO:0030234)	12	5.10%
		protein binding transcription factor activity (GO:0000988)	2	0.90%
	GO Biological Process	metabolic process (GO:0008152)	91	38.90%
		cellular process (GO:0009987)	82	35.00%
		biological regulation (GO:0065007)	54	23.10%
		response to stimulus (GO:0050896)	41	17.50%
		multicellular organismal process (GO:0032501)	38	16.20%
		localization (GO:0051179)	32	13.70%
		developmental process (GO:0032502)	31	13.20%
		immune system process (GO:0002376)	23	9.80%
		reproduction (GO:0000003)	9	3.80%
		cellular component organization or biogenesis (GO:0071840)	8	3.40%
		biological adhesion (GO:0022610)	8	3.40%
		apoptotic process (GO:0006915)	5	2.10%
	Pathway	Inflammation mediated by chemokine and cytokine signaling pathway (P00031)	4	1.70%
		Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway (P00026)	4	1.70%
		Heterotrimeric G-protein signaling pathway-rod outer segment phototransduction (P00028)	3	1.30%
		Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway (P00027)	3	1.30%
		Wnt signaling pathway (P00057)	3	1.30%
		Cadherin signaling pathway (P00012)	3	1.30%

			Huntington disease (P00029)	2	0.90%
			FGF signaling pathway (P00021)	2	0.90%
			FAS signaling pathway (P00020)	2	0.90%
			EGF receptor signaling pathway (P00018)	2	0.90%
			Axon guidance mediated by netrin (P00009)	1	0.40%
			Apoptosis signaling pathway (P00006)	1	0.40%
			Angiogenesis (P00005)	1	0.40%
			Alzheimer disease-presenilin pathway (P00004)	1	0.40%
			5HT2 type receptor mediated signaling pathway (P04374)	1	0.40%
			Integrin signalling pathway (P00034)	1	0.40%
			Adrenaline and noradrenaline biosynthesis (P00001)	1	0.40%
			GABA-B_receptor_II_signaling (P05731)	1	0.40%
			Endogenous_cannabinoid_signaling (P05730)	1	0.40%
			Androgen/estrogene/progesterone biosynthesis (P02727)	1	0.40%
			VEGF signaling pathway (P00056)	1	0.40%
			O-antigen biosynthesis (P02757)	1	0.40%
			Transcription regulation by bZIP transcription factor (P00055)	1	0.40%
			Thyrotropin-releasing hormone receptor signaling pathway (P04394)	1	0.40%
			T cell activation (P00053)	1	0.40%
			TGF-beta signaling pathway (P00052)	1	0.40%
			Oxytocin receptor mediated signaling pathway (P04391)	1	0.40%
			Parkinson disease (P00049)	1	0.40%
			Cytoskeletal regulation by Rho GTPase (P00016)	1	0.40%
			PDGF signaling pathway (P00047)	1	0.40%
			Histidine biosynthesis (P02747)	1	0.40%
			Histamine H1 receptor mediated signaling pathway (P04385)	1	0.40%
			Notch signaling pathway (P00045)	1	0.40%
			Nicotinic acetylcholine receptor signaling pathway (P00044)	1	0.40%
			B cell activation (P00010)	1	0.40%
			Angiotensin II-stimulated signaling through G proteins and beta-arrestin (P05911)	1	0.40%
			Cortocotropin releasing factor receptor signaling pathway (P04380)	1	0.40%

[1]Percent of gene hit: percentage of the total genes of enriched for the specific GO term

"receptor activity", "transporter activity", "nucleic acid binding transcription factor activity", "structural molecule activity" and "enzyme regulator activity". Table 3-13 also shows molecular pathways within which these genes participate. The pathways containing the most number of gene hits included "Integrin signalling pathway", "Inflammation mediated by chemokine and cytokine signaling pathway" with three genes, "EGF receptor signaling pathway", "Cytoskeletal regulation by Rho GTPase" and "FGF signaling pathway" with two genes and "Wnt signaling pathway" with one hit. Several genes in this list were related to brain development and reproduction according to the phenotypes that they produce in mutant mice (http://www.informatics.jax.org/). *Hsd17b2*, *Lpar1* and *Nf1* cause neural tube closure failure in mice and *Elmo1*, *Fgfr4*, *Ggt1*, *Insr*, *Itgb3*, and *Schip1* are the genes that lead to fertility abnormalities in mutant mice.

Table 3-12 List of the genes bearing overlapping CECR2 and SNF2H peaks within their promoter region (up to 5 kb upstream of TSS) in ES cells. Green – mouse mutations associated with exencephaly. Red – mouse mutations associated with reproduction defects.

Coding genes		
Gene ID	**MGI_ID**	**Gene Name**
Abca13	2388707	ATP-binding cassette, sub-family A (ABC1), member 13
Arpc5	3648102	predicted gene 16372; actin related protein 2/3 complex, subunit 5
Atp6v0a1	103286	ATPase, H+ transporting, lysosomal V0 subunit A1
Bc021891	2385307	cDNA sequence BC021891
C1qa	88223	complement component 1, q subcomponent, alpha polypeptide
C1ql2	3032521	complement component 1, q subcomponent-like 2
C78339	2145496	expressed sequence C78339
Ccdc57	1918526	coiled-coil domain containing 57
Cd300ld	2442358	similar to RIKEN cDNA 4732429D16 gene; RIKEN cDNA 4732429D16 gene
Cd63	99529	CD63 antigen
Cdr2l	2684867	cerebellar degeneration-related protein 2-like
Cnga3	1341818	cyclic nucleotide gated channel alpha 3
Cntnap4	2183572	contactin associated protein-like 4
Creg1	1344382	cellular repressor of E1A-stimulated genes 1
Cst11	1925490	cystatin 11
Dpep2	2442042	dipeptidase 2
Drg2	1342307	hypothetical protein LOC674305; predicted gene 8918; developmentally regulated GTP binding protein 2
Dux = Duxbl	1921649	double homeobox; RIKEN cDNA 4933403O03 gene; predicted gene, EG664783
Elmo1(oligozoospermia)	2153044	engulfment and cell motility 1, ced-12 homolog (C. elegans)
Eln	95317	elastin
Fam178b	3026913	family with sequence similarity 178, member B
Fcgr3	95500	Fc receptor, IgG, low affinity III
Fcrla	2138647	Fc receptor-like A
Fgfr4 (infertility)	95525	fibroblast growth factor receptor 4
Gabrp	2387597	gamma-aminobutyric acid (GABA) A receptor, pi
Gbp4	97072	guanylate binding protein 4
Gcm2	1861438	glial cells missing homolog 2 (Drosophila)
Ggt1 (infertility)	95706	gamma-glutamyltransferase 1
Ggt6	1918772	gamma-glutamyltransferase 6
Gm11595	3652308	predicted gene 11595; similar to 2300006N05Rik protein; keratin associated protein 4-8
Gm9766 =Cep85l	3642684	predicted gene 9766
Gpr55	2685064	G protein-coupled receptor 55
Gucy2e	105123	guanylate cyclase 2e
Hibadh	1889802	predicted gene 11225; 3-hydroxyisobutyrate dehydrogenase
Hint3	1914097	histidine triad nucleotide binding protein 3
Hsd17b2(exencephaly)	1096386	hydroxysteroid (17-beta) dehydrogenase 2
Hspb8	2135756	heat shock protein 8

Htr6	1196627	5-hydroxytryptamine (serotonin) receptor 6
Inf2	99839	subacute ozone induced inflammation; RIKEN cDNA 2610204M08 gene
Insr (subfertility)	96575	insulin receptor
Irf8	96395	interferon regulatory factor 8
Itgb3(subfertility)	96612	integrin beta 3
Kcna6	96663	potassium voltage-gated channel, shaker-related, subfamily, member 6
Kctd17	1920094	potassium channel tetramerisation domain containing 17
Klhdc8a	2442630	kelch domain containing 8A
Krt42	1915489	keratin 42
Krtap17-1	1925164	keratin associated protein 17-1
Lcmt2	1353659	leucine carboxyl methyltransferase 2
Lcn3	102669	lipocalin 3
Lefty1	107405	left right determination factor 1
Lpar1 (exencephaly)	108429	lysophosphatidic acid receptor 1
Lpar5	2685918	lysophosphatidic acid receptor 5
Lrfn3	2442512	leucine rich repeat and fibronectin type III domain containing 3
Ltf	96837	lactotransferrin
Lyg1	1916791	lysozyme G-like 1
Man1c1	2446214	mannosidase, alpha, class 1C, member 1
Mfsd4	2442786	major facilitator superfamily domain containing 4
Mlf2	1353554	similar to Myeloid leukemia factor 2; myeloid leukemia factor 2
Myh3	1339709	myosin, heavy polypeptide 3, skeletal muscle, embryonic
Myl2	97272	myosin, light polypeptide 2, regulatory, cardiac, slow
Naca	106095	nascent polypeptide-associated complex alpha polypeptide
Nacc2	1915241	nucleus accumbens associated 2, BEN and BTB (POZ) domain containing
Naglu inner ear	1351641	alpha-N-acetylglucosaminidase (Sanfilippo disease IIIB)
Nefh	97309	similar to neurofilament protein; neurofilament, heavy polypeptide
Nf1 (exencephaly)	97306	neurofibromatosis 1
Nfasc	104753	neurofascin
Nfe2	97308	nuclear factor, erythroid derived 2
Nfe2l3	1339958	nuclear factor, erythroid derived 2, like 3
Nlrp2	3041206	NLR family, pyrin domain containing 2
Nudt5	1858232	nudix (nucleoside diphosphate linked moiety X)-type motif 5
Oas3	2180850	2'-5' oligoadenylate synthetase 3
Oasl1	2180849	2'-5' oligoadenylate synthetase-like 1
Pde2a	2446107	phosphodiesterase 2A, cGMP-stimulated
Pla2g10	1347522	phospholipase A2, group X
Plb1	1922406	phospholipase B1
Plcg2	97616	phospholipase C, gamma 2
Pnmt	97724	phenylethanolamine-N-methyltransferase
Prss39	1270856	testicular serine protease 1
Ptpn7	2156893	protein tyrosine phosphatase, non-receptor type 7
Rasgrp4	2386851	RAS guanyl releasing protein 4
Rbm47	2384294	RNA binding motif protein 47
Rimbp2	2443235	RIMS binding protein 2
Schip1 (fertility)	1353557	schwannomin interacting protein 1

Six3	102764	sine oculis-related homeobox 3 homolog (Drosophila)
Slc26a11	2444589	solute carrier family 26, member 11
Slc36a1	2445299	solute carrier family 36 (proton/amino acid symporter), member 1
Slc45a2 (fertility)	2153040	solute carrier family 45, member 2
Slc47a2	3588190	solute carrier family 47, member 2; hypothetical protein LOC100048051
Slfn4	1329010	schlafen 4
Spata19	1922719	spermatogenesis associated 19
Tcap	1330233	titin-cap
Thap11	1930964	THAP domain containing 11
Tmem125	1923409	transmembrane protein 125
Tmem63a	2384789	transmembrane protein 63a
Tnfrsf1a	1314884	tumor necrosis factor receptor superfamily, member 1a
Tnxb	1932137	tenascin XB
Tomm7	1913419	similar to translocase of outer mitochondrial membrane 7 homolog; translocase of outer mitochondrial membrane 7 homolog (yeast)
Trhr2	2177284	thyrotropin releasing hormone receptor 2
Trpm8	2181435	transient receptor potential cation channel, subfamily M, member 8
Ttf1	105044	transcription termination factor, RNA polymerase I
Zfp354c	1353621	zinc finger protein 354C
1190005i06rik	1916168	RIKEN cDNA 1190005I06 gene
1700018b08rik	1923655	RIKEN cDNA 1700018B08 gene
1700029f09rik	1922873	RIKEN cDNA 1700029F09 gene
1700101e01rik	2685669	RIKEN cDNA 1700101E01 gene
2210011c24rik	1917384	RIKEN cDNA 2210011C24 gene

Non-coding genes		
NCBI_ID	**Gene name**	
NR_002841	*Rn4.5s*	ribosomal RNA
NR_046127	*Rn5s10*	ribosomal RNA
NR_046140	*Rn5s17*	ribosomal RNA
NR_046147	*Rn5s23*	ribosomal RNA
NR_046152	*Rn5s28*	ribosomal RNA
NR_046154	*Rn5s30*	ribosomal RNA
NR_046155	*Rn5s31*	ribosomal RNA
NR_035431	*Mir1199*	microRNA
NR_039583	*Mir3473d*	microRNA
NR_027666	*Fcrla*	non-coding RNA
NR_040672	*Fhad1os1*	long non-coding RNA
NR_045902	*Gm11747*	long non-coding RNA
NR_037980	*Gm16894*	long non-coding RNA
NR_038012	*Gm17751*	long non-coding RNA
NR_040296	*Gm19434*	long non-coding RNA
NR_040303	*Gm20743*	long non-coding RNA
NR_110357	*Kctd17*	non-coding RNA
NR_110990	*Pla2g10*	non-coding RNA
NR_003368	*Pvt1*	non-coding RNA
NR_132746	*Pvt1*	long non-coding RNA
NR_132747	*Pvt1*	long non-coding RNA

NR_038082	*Six3os1*	long non-coding RNA
NR_131046	1700123J17 Rik	long non-coding RNA
NR_038189	1700125H03Rik	long non-coding RNA
NR_040520	2310069B03Rik	long non-coding RNA
NR_028428	2610005L07Rik	non-coding RNA
NR_045042	2700038G22Rik	long non-coding RNA
NR_045942	4930401O10Rik	long non-coding RNA
NR_130997	4930557K07Rik	long non-coding RNA
NR_040480	4930568G15Rik	long non-coding RNA
NR_046037	4933406G16Rik	long non-coding RNA
NR_030708	6820431F20Rik	non-coding RNA
NR_015618	A930015D03Rik	non-coding RNA

Table 3-13 Gene ontology terms enriched in 103 coding genes with CECR2 and SNF2H overlapping peaks in their promoter region (up to 5 kb upstream of TSS) in ES cells.

	Ontology	Category name (Accession)	Number of genes	Percent gene hit[1]
ES Cell	**GO Molecular Function**	binding (GO:0005488)	32	30.50%
		catalytic activity (GO:0003824)	30	28.60%
		receptor activity (GO:0004872)	13	12.40%
		transporter activity (GO:0005215)	11	10.50%
		nucleic acid binding transcription factor activity (GO:0001071)	10	9.50%
		structural molecule activity (GO:0005198)	8	7.60%
		enzyme regulator activity (GO:0030234)	7	6.70%
		protein binding transcription factor activity (GO:0000988)	1	1.00%
	GO Biological Process	metabolic process (GO:0008152)	47	44.80%
		cellular process (GO:0009987)	33	31.40%
		response to stimulus (GO:0050896)	20	19.00%
		biological regulation (GO:0065007)	17	16.20%
		localization (GO:0051179)	16	15.20%
		developmental process (GO:0032502)	15	14.30%
		multicellular organismal process (GO:0032501)	11	10.50%
		immune system process (GO:0002376)	10	9.50%
		cellular component organization or biogenesis (GO:0071840)	4	3.80%
		apoptotic process (GO:0006915)	3	2.90%
		biological adhesion (GO:0022610)	2	1.90%
	Pathway	Integrin signalling pathway (P00034)	3	2.90%
		Inflammation mediated by chemokine and cytokine signaling pathway (P00031)	3	2.90%
		EGF receptor signaling pathway (P00018)	2	1.90%

152

Cytoskeletal regulation by Rho GTPase (P00016)		2	1.90%
FGF signaling pathway (P00021)		2	1.90%
Axon guidance mediated by netrin (P00009)		1	1.00%
Histamine H1 receptor mediated signaling pathway (P04385)		1	1.00%
Apoptosis signaling pathway (P00006)		1	1.00%
Angiogenesis (P00005)		1	1.00%
Insulin/IGF pathway-protein kinase B signaling cascade (P00033)		1	1.00%
Insulin/IGF pathway-mitogen activated protein kinase kinase/MAP kinase cascade (P00032)		1	1.00%
Adrenaline and noradrenaline biosynthesis (P00001)		1	1.00%
Coenzyme A biosynthesis (P02736)		1	1.00%
Gonadotropin releasing hormone receptor pathway (P06664)		1	1.00%
PI3 kinase pathway (P00048)		1	1.00%
PDGF signaling pathway (P00047)		1	1.00%
Thyrotropin-releasing hormone receptor signaling pathway (P04394)		1	1.00%
Nicotinic acetylcholine receptor signaling pathway (P00044)		1	1.00%
Oxytocin receptor mediated signaling pathway (P04391)		1	1.00%
Blood coagulation (P00011)		1	1.00%
B cell activation (P00010)		1	1.00%
CCKR signaling map (P06959)		1	1.00%
Huntington disease (P00029)		1	1.00%
Heterotrimeric G-protein signaling pathway-rod outer segment phototransduction (P00028)		1	1.00%
Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway (P00027)		1	1.00%
Wnt signaling pathway (P00057)		1	1.00%
Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway (P00026)		1	1.00%
5HT2 type receptor mediated signaling pathway (P04374)		1	1.00%
VEGF signaling pathway (P00056)		1	1.00%
General transcription by RNA polymerase I (P00022)		1	1.00%
TGF-beta signaling pathway (P00052)		1	1.00%
Androgen/estrogene/progesterone biosynthesis (P02727)		1	1.00%

[1]Percent of gene hit: percentage of the total genes of enriched for the specific GO term

To understand the functions of the genes with an overlapping peak for CECR2, SNF2H and LUZP1 in their promoter regions (5 kb upstream of TSS), overlapping peaks (at least 1 bp) of the three proteins obtained from ES cell ChIP-seq were intersected with the known genes promoter genomic intervals to identify the genes containing the peaks from all three proteins. There were 23 protein coding genes and 7 non-coding genes with at least one peak for each of the three proteins in their 5 kb promoter region (Table 3-14). Using the PANTHER classification system identified coding genes were categorized based on their molecular function. The most enriched molecular functions included "binding", "catalytic activity", "transporter activity" and "nucleic acid binding transcription factor activity" (Table 3-15). Biological process terms showed involvement of these genes in embryonic development. Three of the genes (*Gcm2, Nfasc and*

Trpm8) are involved in nervous system development and physiology (Dhaka et al. 2007, Hitoshi et al. 2011, Pillai et al. 2009). 4 genes (*Lefty1, Naca, and Tcap*) have a role in heart development (Knoll et al. 2011, Meno et al. 1998, Park et al. 2010). This is interesting, because according to a previous study cardiovascular deficits were the main cause of perinatal death of *Luzp1* mutant mice with full penetrance (Hsu et al. 2008). *Fgfr4* and *Insr* cause fertility phenotypes and contain the overlapping binding sites of the three proteins in the higher P value (10^{-3} instead of 10^{-5}). The GO terms obtained from the overlapping binding sites of CECR2, SNF2H and LUZP1 was a subset of the GO terms for overlapping CECR2 and SFN2H binding sites. This suggests that some of the functions of CECR2 probably involve the complex that only contains CECR2 and SNF2H.

Table 3-14 List of the genes bearing overlapping CECR2-SNF2H-LUZP1 peaks within their promoter region (up to 5 kb upstream of TSS) in ES cells.

Coding genes		
Gene ID	MGI_ID	Gene Name
4933403o03rik	3704116	double homeobox; RIKEN cDNA 4933403O03 gene; predicted gene, EG664783
Abca13	2388707	ATP-binding cassette, sub-family A (ABC1), member 13
C78339	2145496	expressed sequence C78339
Creg1	1344382	cellular repressor of E1A-stimulated genes 1
Dux (duxbl)	3704116	double homeobox; RIKEN cDNA 4933403O03 gene; predicted gene, EG664783
Gcm2	1861438	glial cells missing homolog 2 (Drosophila)
Kcna6	96663	potassium voltage-gated channel, shaker-related, subfamily, member 6
Klhdc8a	2442630	kelch domain containing 8A
Lefty1	107405	left right determination factor 1
Naca	106095	nascent polypeptide-associated complex alpha polypeptide
Nfasc	104753	neurofascin
Nfe2l3	1339958	nuclear factor, erythroid derived 2, like 3
Pnmt	97724	phenylethanolamine-N-methyltransferase
Pvt1	97824	plasmacytoma variant translocation 1
Rbm47	2384294	RNA binding motif protein 47
Rimbp2	2443235	RIMS binding protein 2
Sap130	1919782	Sin3A associated protein
Slc45a2	2153040	solute carrier family 45, member 2
Tcap	1330233	titin-cap
Tcfeb	103270	transcription factor EB
Tmem125	1923409	transmembrane protein 125
Tmem63a	2384789	transmembrane protein 63a

Tomm7	1913419	similar to translocase of outer mitochondrial membrane 7 homolog; translocase of outer mitochondrial membrane 7 homolog (yeast)
Trpm8	2181435	transient receptor potential cation channel, subfamily M, member 8
Non-coding genes		
NCBI_ID	**Gene name**	**Description**
NR_003368	Pvt1	long non-coding RNA
NR_040672	Fhad1os1	long non-coding RNA
NR_046037	4933406G16Rik	non-coding RNA
NR_045040	2700038G22Rik	non-coding RNA
NR_035431	Mir1199	microRNA
NR_040303	Gm20743	predicted gene, unclassified non-coding RNA gene
NR_028428	2610005L07Rik	cadherin 11 pseudogene

Table 3-15 Gene ontology terms enriched in 23 coding genes with CECR2-SNF2H-LUZP1 overlapping peaks in their promoter region (up to 5 kb upstream of TSS) in ES cells.

	Ontology	Category name (Accession)	Number of genes	Percent of gene hit[1]
ES Cells	**GO Molecular Function**	binding (GO:0005488)	5	22.70%
		catalytic activity (GO:0003824)	4	18.20%
		transporter activity (GO:0005215)	4	18.20%
		nucleic acid binding transcription factor activity (GO:0001071)	3	13.60%
		receptor activity (GO:0004872)	2	9.10%
		structural molecule activity (GO:0005198)	2	9.10%
	GO Biological Process	metabolic process (GO:0008152)	10	45.50%
		cellular process (GO:0009987)	6	27.30%
		developmental process (GO:0032502)	5	22.70%
		localization (GO:0051179)	4	18.20%
		biological regulation (GO:0065007)	2	9.10%
		response to stimulus (GO:0050896)	2	9.10%
		multicellular organismal process (GO:0032501)	2	9.10%
		cellular component organization or biogenesis (GO:0071840)	1	4.50%
		apoptotic process (GO:0006915)	1	4.50%
	Pathway	TGF-beta signaling pathway (P00052)	1	4.50%
		Adrenaline and noradrenaline biosynthesis (P00001)	1	4.50%

[1]Percent of gene hit: percentage of the total genes of enriched for the specific GO term

3.11.5. Gene ontology terms (Testis)

I used the GREAT program to identify the functional significance of all the peaks obtained from testis ChIP-seq data for CECR2 (Table 3-16). The GO terms with the highest number of hits was "acrosome reaction" with 30 region hits related to 11 genes (Table- 3-17). These genes are involved in different aspects of fertilization. The overlapped peaks of CECR2

and SNF2H (1366 peaks) were not enough to be associated with GO terms by the GREAT program. GREAT related 1959 genes to these overlapping peaks.

Table 3-16 GREAT gene ontology terms from analysis of CECR2 peaks from testis data sets at a false discovery rate of 0.05. GO terms related to reproduction are highlighted in yellow

	Ontology	Term name	Binomial raw P-Value	Binomial FDR Q-Value	Binomial fold enrichment	#Hit genes
Testis	GO Biological Process	activation of Rac GTPase activity	0.0001	0.014	2.07	9
		antifungal humoral response	0.0002	0.017	14	1
		acrosome reaction	0.0003	0.021	2	11
		negative regulation of phosphatidylinositol 3-kinase activity	0.0003	0.022	3.3	2
		gamma-aminobutyric acid metabolic process	0.0005	0.032	2.86	3
		cell motility involved in cerebral cortex radial glia guided migration	0.0005	0.031	3.03	2
		negative regulation of lipid kinase activity	0.0006	0.034	2.84	3
		gamma-aminobutyric acid biosynthetic process	0.0008	0.044	3.44	1

Table 3-17 Genes associated with "Acrosome reaction" GO term assigned for CECR2 peaks in testis. The Gene ontology enrichment analysis was carried out by GREAT. The √ shows the overlap between the binding targets of CECR2 with SNF2H. The shared genomic regions are highlighted.

Gene ID	Gene Name	Region (distance to TSS)	Overlap with SNF2H
Acr	Acrosin prepropeptide	chr15:89414919-89415419 (+16412)	√
Eqtn	Equatorin, sperm acrosome associated	chr4:94569399-94569899 (+25836)	
Glra1	Glycine receptor subunit alpha-1	chr11:55823303-55823803 (-401853), chr11:55803260-55803760 (-381810), chr11:55590999-55591499 (-169549), chr11:55537854-55538354 (-116404), chr11:55398359-55398859 (+23091)	
Glrb	Glycine receptor subunit beta	chr3:80652218-80652718 (+65114), chr3:80631786-80632286 (+85546)	
Gnpda1	Glucosamine-6-phosphate isomerase 1	chr18:38567733-38568233 (-69326), chr18:38471495-38471995 (+26912)	
Plcd4	Phospholipase C, delta 4	chr1:74602923-74603423 (+12235)	
Spaca7	sperm acrosome associated 7	chr8:12347743-12348243 (-225056), chr8:12398556-12399056 (-174243), chr8:12419543-12420043 (-153256), chr8:12436723-12437223 (-136076), chr8:12500119-12500619 (-72680), chr8:12577797-12578297 (+4998),	

		chr8:12595382-12595882 (+22583)	√
Spesp1	Sperm equatorial segment protein 1	chr9:62178524-62179024 (-48788), chr9:62147259-62147759 (-17523)	
Stx2	Syntaxin-2	chr5:129483234-129483734 (+30965)	
Syt6	Synaptotagmin-6	chr3:103084363-103084863 (-295452), chr3:103092400-103092900 (-287415), chr3:103209686-103210186 (-170129), chr3:103217562-103218062 (-162253), chr3:103248205-103248705 (-131610), chr3:103329420-103329920 (-50395), chr3:103532844-103533344 (+153029)	
Syt8	Synaptotagmin-8	chr7:149607974-149608474 (-12658)	

Then I made a list of the genes with the CECR2 peaks (P-value < 10-5) in the promoter region (5 kb upstream to TSS) (Table 3-18). The PANTHER GO analysis associated the 99 coding genes containing CECR2 peaks in their promoters to important GO terms (Table 3-19). The genes were classified into categories based on the molecular functions including "binding" (27% of genes) "catalytic activity" (22%), "nucleic acid binding transcription factor activity" (9%), "transporter activity" (8%), "receptor activity" (6%), "enzyme regulator activity" (5%), "structural molecule activity" (5%) and "protein binding transcription factor activity" (2%). Biological processes involved in different aspects of embryogenesis were over-represented in the testis data set. Thirteen genes were included in developmental process (GO:0032502) GO term including *Actl6a*, *Tmeff1*, *Cdc42bpa*, *Tmprss11d*, *Ear11*, *Pabpc4*, *Olig1*, *Fadd*, *Cit*, *Evx2*, *Barhl1*, *Phactr4* and *Phactr3*. Table 3-19 shows the most enriched pathways for the same set of the genes. There were a variety of pathways including "apoptosis signaling pathway", "gonadotropin releasing hormone receptor pathway" and "general transcription regulation". The data suggest that CECR2 is involved in modulating the genes that are involved in a broad range of functions in the testis.

Table 3-18 List of the genes bearing CECR2 peaks (P-value < 10-5) within their promoter region (up to 5 kb upstream of TSS) in testis.

Coding genes		
Gene ID	**MGI_ID**	**Gene Name**
Abcc2	1352447	ATP-binding cassette, sub-family C (CFTR/MRP), member 2
Actl6a	1861453	actin-like 6A
Alkbh6	2142037	alkB, alkylation repair homolog 6 (E. coli)
Angptl4	1888999	angiopoietin-like 4
Aqp7	1314647	aquaporin 7
Armcx5	2148026	armadillo repeat containing, X-linked 5
Asb14	2655107	ankyrin repeat and SOCS box-containing 14

Barhl1	1859288	BarH-like 1 (Drosophila)
C8b	88236	complement component 8, beta polypeptide
Caly	1915816	calcyon neuron-specific vesicular protein
Cbln3	1889286	cerebellin 3 precursor protein
Ccdc81	1918134	coiled-coil domain containing 81
Cdc20	1859866	cell division cycle 20 homolog (S. cerevisiae)
Cdc42bpa	2441841	CDC42 binding protein kinase alpha
Cdk20	2145349	cell cycle related kinase
Cfap43 (Wdr96)	1289258	cilia and flagella associated protein 43
Cit	105313	citron
Cmah	103227	cytidine monophospho-N-acetylneuraminic acid hydroxylase
Dpyd	2139667	dihydropyrimidine dehydrogenase
Ear11	1890465	eosinophil-associated, ribonuclease A family, member 11
Esyt1	1344426	family with sequence similarity 62 (C2 domain containing), member A
Evx2	95462	even skipped homeotic gene 2 homolog
Fadd	109324	Fas (TNFRSF6)-associated via death domain
Fbxw20	3584372	F-box and WD-40 domain protein 12
Fntb	1923684	similar to farnesyltransferase, CAAX box, beta; farnesyltransferase, CAAX box, beta
Frem2	2444465	Fras1 related extracellular matrix protein 2
Galnt6	1891640	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 6
Galnt9	2677965	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 9
Gemin8	2384300	similar to gem (nuclear organelle) associated protein 8; predicted gene 5455; gem (nuclear organelle) associated protein 8
Gm11554	3705237	predicted gene 11554
Gm14625	3705849	predicted gene 14625
Gm3409	3781580	hypothetical protein LOC100048784; RIKEN cDNA 4930449I24 gene; predicted gene 3415; predicted gene 6408; predicted gene 3402; hypothetical protein LOC100044549; predicted gene 3404; predicted gene 6309; predicted gene 3409; predicted gene 6370; hypothetical protein LOC100045734
Gprasp1	1917418	G protein-coupled receptor associated sorting protein 1
Gsr	95804	similar to Glutathione reductase, mitochondrial precursor (GR) (GRase); glutathione reductase
Gtf2a1l	1919078	general transcription factor IIA, 1-like
Hdgfrp2	1194492	hepatoma-derived growth factor, related protein 2
Hexim2	1918309	similar to hexamthylene bis-acetamide inducible 2; hexamthylene bis-acetamide inducible 2
Hs3st3a1	1333861	heparan sulfate (glucosamine) 3-O-sulfotransferase 3A1
Htatip2	1859271	HIV-1 tat interactive protein 2, homolog (human)
Irgm1	107567	immunity-related GTPase family M member 1
Itgbl1	2443439	integrin, beta-like 1
Khnyn	2451333	RIKEN cDNA 9130227C08Rik gene
Kpnb1	107532	karyopherin (importin) beta 1
Krtap4-13	1916714	keratin associated protein 4-13
Mfap2	99559	microfibrillar-associated protein 2
Mgat4e	1918251	MGAT4 family, member E

Mios	2182066	missing oocyte, meiosis regulator, homolog (Drosophila)
Mrpl4	2137210	mitochondrial ribosomal protein L4
Mtmr12	2443034	myotubularin related protein 12
Nedd9	97302	neural precursor cell expressed, developmentally down-regulated gene 9
Nefh	97309	similar to neurofilament protein; neurofilament, heavy polypeptide
Nfia	108056	nuclear factor I/A
Nisch	1928323	nischarin
Nr3c1	95824	nuclear receptor subfamily 3, group C, member 1
Olfr609	3030443	olfactory receptor 609
Olfr936	3646521	predicted gene, EG628171; olfactory receptor 936
Olig1	1355334	oligodendrocyte transcription factor 1
Pabpc4	3650566	similar to Poly A binding protein, cytoplasmic 4; poly(A) binding protein, cytoplasmic 4; predicted gene 12623; predicted gene 5088; hypothetical protein LOC100044219
Paip1	2384993	polyadenylate binding protein-interacting protein 1; similar to poly(A) binding protein interacting protein 1
Pappa2	3051647	pappalysin 2
Pcyox1	1914131	prenylcysteine oxidase 1
Peg3	104748	paternally expressed 3; antisense transcript gene of Peg3
Pfdn2	1276111	prefoldin 2
Pgd	97553	phosphogluconate dehydrogenase
Phactr3	1921439	phosphatase and actin regulator 3
Phactr4	2140327	phosphatase and actin regulator 4
Pla2g15	2178076	phospholipase A2, group XV
Pmm1	1353418	phosphomannomutase 1
Prdm14	3588194	PR domain containing 14
Pros1	1095733	protein S (alpha)
Rarb	97857	retinoic acid receptor, beta
Rfc3	1916513	replication factor C (activator 1) 3
Rfx3	106582	regulatory factor X, 3 (influences HLA class II expression); similar to Regulatory factor X, 3 (influences HLA class II expression)
Rfxap	2180854	regulatory factor X-associated protein
Rnft2	2442859	ring finger protein, transmembrane 2
Rpl27	98036	predicted gene 6599; predicted gene 6199; predicted gene 6341; predicted gene 6301; predicted gene 11518; similar to ribosomal protein L27; ribosomal protein L27; predicted gene 7053; predicted gene 11552; predicted gene 15730
Rspo1	2183426	R-spondin homolog (Xenopus laevis)
Scd4	2670997	stearoyl-coenzyme A desaturase 4
Serf2	1337041	small EDRK-rich factor 2
Sertad3	2180697	SERTA domain containing 3
Slc12a5	1862037	solute carrier family 12, member 5
Slc16a4	2385183	solute carrier family 16 (monocarboxylic acid transporters), member 4
Slc25a19	1914533	solute carrier family 25 (mitochondrial thiamine pyrophosphate carrier), member 19
Slc46a1	1098733	solute carrier family 46, member 1
Slc4a4	1927555	solute carrier family 4 (anion exchanger), member 4
Smu1	1915546	smu-1 suppressor of mec-8 and unc-52 homolog (C. elegans)
Stard10	1860093	START domain containing 10

Stx1a	109355	syntaxin 1A (brain)
Tcaf3 (Fam115e)	3042585	TRPM8 channel-associated factor 3
Thop1	1354165	thimet oligopeptidase 1
Tia1	107914	cytotoxic granule-associated RNA binding protein 1
Tmeff1	1926810	transmembrane protein with EGF-like and two follistatin-like domains 1; similar to Transmembrane protein with EGF-like and two follistatin-like domains 1
Tmprss11d	2385221	similar to airway trypsin-like protease; transmembrane protease, serine 11d
Tnfaip6	1195266	similar to TNF-stimulated gene 6 protein; tumor necrosis factor alpha induced protein 6
Tomm40	1858259	translocase of outer mitochondrial membrane 40 homolog (yeast)
Trim29	1919419	tripartite motif-containing 29
Ubxn6	1913780	UBX domain protein 6
Usp8	1934029	ubiquitin specific peptidase 8
Zcchc9	1916335	zinc finger, CCHC domain containing 9
Zdhhc4	1920131	zinc finger, DHHC domain containing 4
Zfp930	2675306	zinc finger protein 930
1110059G10RIK	1913452	RIKEN cDNA 1110059G10 gene
1700003E16RIK	1919087	RIKEN cDNA 1700003E16 gene
4931440L10RIK	1918251	RIKEN cDNA 4931440L10 gene

Non-coding genes

NCBI_ID	Gene symbol	Gene name
NR_029558	Mir146	microRNA
NR_029876	Mir375	microRNA
NR_039555	Mir5046	microRNA
NR_049198	Mir5624	microRNA
NR_105922	Mir6957	microRNA
NR_106019	Mir7052	microRNA
NR_104449	Galnt9	non-coding RNA
NR_030771	Gm13034	non-coding RNA
NR_110986	Serf2	non-coding RNA
NR_133683	Stx1a	non-coding RNA
NR_038079	Gm2447	long non-coding RNA
NR_110431	Lncenc1	long non-coding RNA
NR_037959	1600010M07Rik	long non-coding RNA
NR_045479	1700024B18Rik	long non-coding RNA
NR_045825	1700025K24Rik	long non-coding RNA
NR_038180	1700061F12Rik	long non-coding RNA
NR_045185	1700123O12Rik	long non-coding RNA
NR_040543	4930406D18Rik	long non-coding RNA
NR_073462	5730488B01Rik	long non-coding RNA
NR_045776	9530080O11Rik	long non-coding RNA
NR_045162	A330048O09Rik	long non-coding RNA
NR_040353	E030013I19Rik	long non-coding RNA

Table 3-19 Gene ontology terms enriched in 99 coding genes with CECR2 peaks in their proximal promoter region (up to 5 kb upstream of TSS) in testis.

	Ontology	Category name (Accession)	Number of genes	Percent of gene hit[1]
Testis	**GO Molecular Function**	binding (GO:0005488)	27	27.30%
		catalytic activity (GO:0003824)	22	22.20%
		nucleic acid binding transcription factor activity (GO:0001071)	9	9.10%
		transporter activity (GO:0005215)	8	8.10%
		receptor activity (GO:0004872)	6	6.10%
		enzyme regulator activity (GO:0030234)	5	5.10%
		structural molecule activity (GO:0005198)	5	5.10%
		protein binding transcription factor activity (GO:0000988)	2	2.00%
	GO Biological Process	metabolic process (GO:0008152)	39	39.40%
		cellular process (GO:0009987)	31	31.30%
		biological regulation (GO:0065007)	18	18.20%
		localization (GO:0051179)	15	15.20%
		developmental process (GO:0032502)	13	13.10%
		cellular component organization or biogenesis (GO:0071840)	8	8.10%
		multicellular organismal process (GO:0032501)	6	6.10%
		response to stimulus (GO:0050896)	5	5.10%
		biological adhesion (GO:0022610)	4	4.00%
		immune system process (GO:0002376)	4	4.00%
		apoptotic process (GO:0006915)	2	2.00%
		reproduction (GO:0000003)	1	1.00%
	Pathway	Mannose metabolism (P02752)	1	1.00%
		Metabotropic glutamate receptor group III pathway (P00039)	1	1.00%
		Apoptosis signaling pathway (P00006)	1	1.00%
		Integrin signalling pathway (P00034)	1	1.00%
		Parkinson disease (P00049)	1	1.00%
		Pentose phosphate pathway (P02762)	1	1.00%
		Gonadotropin releasing hormone receptor pathway (P06664)	1	1.00%
		DNA replication (P00017)	1	1.00%
		Nicotinic acetylcholine receptor signaling pathway (P00044)	1	1.00%
		Muscarinic acetylcholine receptor 2 and 4 signaling pathway (P00043)	1	1.00%
		Blood coagulation (P00011)	1	1.00%
		Muscarinic acetylcholine receptor 1 and 3 signaling pathway (P00042)	1	1.00%
		Metabotropic glutamate receptor group II pathway (P00040)	1	1.00%
		De novo pyrmidine ribonucleotides biosythesis (P02740)	1	1.00%

		Pyrimidine Metabolism (P02771)	1	1.00%
		Synaptic_vesicle_trafficking (P05734)	1	1.00%
		Transcription regulation by bZIP transcription factor (P00055)	1	1.00%
		General transcription regulation (P00023)	1	1.00%
		FAS signaling pathway (P00020)	1	1.00%

[1]Percent of gene hit: percentage of the total genes of enriched for the specific GO term

The list of the genes with overlapping peaks of CECR2 and SNF2H (P-value $< 10^{-3}$) in their promoters (5 kb upstream to TSS) from the testis data set were examined for over-represented GO terms using the PANTHER database system. Table 3-20 shows the list of coding and non-coding genes in this category. GO terms enriched for this set of the genes are shown in Table 3-21. Biological processes involved in development and reproduction were among the significantly over-presented GO terms. *Lrp6* and *Phactr4* are two of the genes associated with the developmental process term. Figure 3-51 shows the binding sites of the CECR2 complex upstream of these genes. About 50% of *Lrp6* mutants exhibit exencephaly and open eye lid phenotypes (Pinson et al. 2000, Zhou et al. 2009) and 100% of the *Phactr4* embryos die from exencephaly (Kim et al. 2007). Cell cycle GO terms were one of the sub-categories of cellular process associated with 5 genes in the list. *Cdc14b* was one the genes related with the Cell cycle GO term. *Cdc14b* mutant mice are sub-fertile, having a smaller litter size in both sexes (Wei et al. 2011). This is important because *Cecr2* mutant mice also are sub-fertile (Thompson et al. 2012). *Phactr4* and *Lrp6* also contain overlapped peaks of CECR2 and SNF2H in ES cells when the peaks were called using a bigger P-value (10^{-3} instead of 10^{-5}) (Figure 3-51). *Nfia*, *Pcsk1* and *Styx* also have a role in reproduction (Figure 3-52).

Table 3-20 List of the genes bearing overlapping CECR2 and SNF2H peaks (P-value $< 10^{-3}$) within their promoter region (up to 5 kb upstream of the TSS) in testis. Green – mouse mutations associated with NTD. Red – mouse mutations associated with reproduction defects.

Coding genes		
Gene ID	**MGI_ID**	**Gene Name**
Aars	2384560	alanyl-tRNA synthetase
Apbb1 (abnormal brain development)	107765	amyloid beta (A4) precursor protein-binding, family B, member 1
Aqp7	1314647	aquaporin 7

162

BC021891	2385307	cDNA sequence BC021891
C1qtnf5	2385958	C1q and tumor necrosis factor related protein 5; membrane-type frizzled-related protein
Cdc14b (subfertility)	2441808	CDC14 cell division cycle 14 homolog B (S. cerevisiae)
Cox17	1333806	cytochrome c oxidase, subunit XVII assembly protein homolog (yeast)
Egam1=Crxos1 (Male infertility)	3647299	Egam-1C; Crx opposite strand transcript 1; predicted gene 8743; predicted gene 8282
Crybb1	104992	crystallin, beta B1
Fam91a1 =D1Ertd622e	1277178	DNA segment, Chr 15, ERATO Doi 621, expressed
D15Ertd622e	1277184	DNA segment, Chr 1, ERATO Doi 622, expressed
Ddx19b	2148251	DEAD (Asp-Glu-Ala-Asp) box polypeptide 19b
Dgkb	2442474	diacylglycerol kinase, beta
Dpyd	2139667	dihydropyrimidine dehydrogenase
Eif3h	3647048	eukaryotic translation initiation factor 3, subunit H; predicted gene 6552
Fam178b	3026913	family with sequence similarity 178, member B
Fhad1	1920323	forkhead-associated (FHA) phosphopeptide binding domain 1
Galnt5	2179403	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 5
Gm10094	1277978	similar to SAP18; predicted gene 10094; Sin3-associated polypeptide 18
Gm13154	3651986	predicted gene 13154
Gm14625	3705849	predicted gene 14625(pseudogene)
Gm1966	3644223	predicted gene 1966; predicted gene 8995; GTPase, very large interferon inducible 1, (pseudogene)
Cep85l=Gm9766	3642684	predicted gene 9766
Gyg	1351614	glycogenin
H2bfm	1916639	H2B histone family, member M
Hilpda	1916823	hypoxia inducible lipid droplet associated
Ifih1	1918836	interferon induced with helicase C domain 1
Irgm1	107567	immunity-related GTPase family M member 1
Kcnj2	104744	potassium inwardly-rectifying channel, subfamily J, member 2
Ppp1r21 (Klraq1)	1921075	KLRAQ motif containing 1
Krtap4-13	1916714	keratin associated protein 4-13
Lrp6 (exencephaly)	1298218	similar to LDL receptor-related protein 6; low density lipoprotein receptor-related protein 6
Mfrp	2385958	C1q and tumor necrosis factor related protein 5; membrane-type frizzled-related protein
Mtmr12	2443034	myotubularin related protein 12
Muc5ac	104697	mucin 5, subtypes A and C, tracheobronchial/gastric
Mycbp2	2179432	MYC binding protein 2
Myh1	1339710	myosin, heavy polypeptide 2, skeletal muscle, adult; myosin, heavy polypeptide 1, skeletal muscle, adult
Nedd9	97302	neural precursor cell expressed, developmentally down-regulated gene 9
Nfia (infertility)	108056	nuclear factor I/A
Olfr1511	3031345	olfactory receptor 1511
Pabpc1	1349722	poly(A) binding protein, cytoplasmic 1
Pccb	1914154	propionyl Coenzyme A carboxylase, beta polypeptide
Pcsk1 (subfertility)	97511	proprotein convertase subtilisin/kexin type 1

Phactr4 (exencephaly)	2140327	phosphatase and actin regulator 4
Plin5	1914218	RIKEN cDNA 2310076L09 gene
Pros1	1095733	protein S (alpha)
Psmd3	98858	proteasome (prosome, macropain) 26S subunit, non-ATPase, 3
Rbm20	1920963	RNA binding motif protein 20
Rcan2	1858219	regulator of calcineurin 2
Rcc1	1913989	regulator of chromosome condensation 1
Rpl27	3650306	predicted gene 6599; predicted gene 6199; predicted gene 6341; predicted gene 6301; predicted gene 11518; similar to ribosomal protein L27; ribosomal protein L27; predicted gene 7053; predicted gene 11552; predicted gene 15730
Sap130	1919782	Sin3A associated protein
Sap18	1277978	similar to SAP18; predicted gene 10094; Sin3-associated polypeptide 18
Serf2	1337041	small EDRK-rich factor 2
Slc4a4	1927555	solute carrier family 4 (anion exchanger), member 4
Styx (infertility)	3709613	serine/threonine/tyrosine interaction protein; predicted gene 14698
Swap70	1298390	SWA-70 protein
Tmem180	1922396	transmembrane protein 180
Tmem55a	1919769	transmembrane protein 55A
Tmem59l	1915187	transmembrane protein 59-like
Tox3	3039593	TOX high mobility group box family member 3
p63=Trp63	1330810	transformation related protein 63
Utp20	1917933	UTP20, small subunit (SSU) processsome component, homolog (yeast)
Yars2	1917370	tyrosyl-tRNA synthetase 2 (mitochondrial)
Zcchc9	1916335	zinc finger, CCHC domain containing 9
Zdhhc5	1923573	zinc finger, DHHC domain containing 5
Zfp36l1 (open neural tube)	107946	zinc finger protein 36, C3H type-like 1
Zfp960	3052731	zinc finger protein 960
Zscan2	99176	zinc finger and SCAN domain containing 2
4931408C20RIK	3588222	RIKEN cDNA 4931408C20 gene
4932425I24RIK	2443598	RIKEN cDNA 4932425I24 gene

Non-coding genes

NCBI_ID	Gene name	
NR_037238	Rn4.5s	ribosomal RNA
NR_029728	Mir99a	microRNA
NR_002841	Mirlet7c-1	microRNA
NR_110986	Mir3077	microRNA
NR_037977	Yars2	non-coding RNA
NR_131014	Serf2	non-coding RNA
NR_038046	Gm53	long non-coding RNA
NR_038037	Smkr-ps	long non-coding RNA
NR_046040	4930433B08Rik	long non-coding RNA
NR_040544	1700009C05Rik	long non-coding RNA
NR_040263	E130307A14Rik	long non-coding RNA
NR_015465	4930455G09Rik	long non-coding RNA
NR_029535	C920009B18Rik	long non-coding RNA

Table 3-21 Gene ontology terms enriched in 65 coding genes with CECR2 and SNF2H peaks (P-value < 10^{-3}) in their proximal promoter region (up to 5 kb upstream of TSS) in testis. GO terms related to embryonic development and reproduction are highlighted in yellow.

	Ontology	Category name (Accession)	Number of genes	Percent of gene hit
Test is	**GO Molecular Function**	binding (GO:0005488)	22	33.80%
		catalytic activity (GO:0003824)	13	20.00%
		nucleic acid binding transcription factor activity (GO:0001071)	7	10.80%
		enzyme regulator activity (GO:0030234)	4	6.20%
		transporter activity (GO:0005215)	3	4.60%
		structural molecule activity (GO:0005198)	2	3.10%
		translation regulator activity (GO:0045182)	1	1.50%
		receptor activity (GO:0004872)	1	1.50%
	GO Biological Process	metabolic process (GO:0008152)	32	49.20%
		cellular process (GO:0009987)	20	30.80%
		biological regulation (GO:0065007)	17	26.20%
		localization (GO:0051179)	6	9.20%
		developmental process (GO:0032502)	5	7.70%
		cellular component organization or biogenesis (GO:0071840)	4	6.20%
		response to stimulus (GO:0050896)	4	6.20%
		multicellular organismal process (GO:0032501)	3	4.60%
		reproduction (GO:0000003)	2	3.10%
		apoptotic process (GO:0006915)	1	1.50%
		biological adhesion (GO:0022610)	1	1.50%
		immune system process (GO:0002376)	1	1.50%
	Pathway	Wnt signaling pathway (P00057)	2	3.10%
		Alzheimer disease-presenilin pathway (P00004)	2	3.10%
		Hedgehog signaling pathway (P00025)	1	1.50%
		Cell cycle (P00013)	1	1.50%
		Inflammation mediated by chemokine and cytokine signaling pathway (P00031)	1	1.50%
		p53 pathway (P00059)	1	1.50%
		p53 pathway feedback loops 2 (P04398)	1	1.50%
		p53 pathway by glucose deprivation (P04397)	1	1.50%
		P53 pathway feedback loops 1 (P04392)	1	1.50%
		Cytoskeletal regulation by Rho GTPase (P00016)	1	1.50%
		Nicotinic acetylcholine receptor signaling pathway (P00044)	1	1.50%
		Methylmalonyl pathway (P02755)	1	1.50%
		Blood coagulation (P00011)	1	1.50%
		Huntington disease (P00029)	1	1.50%
		Alzheimer disease-amyloid secretase pathway (P00003)	1	1.50%
		De novo pyrmidine ribonucleotides biosythesis (P02740)	1	1.50%
		Pyrimidine Metabolism (P02771)	1	1.50%
		Ubiquitin proteasome pathway (P00060)	1	1.50%
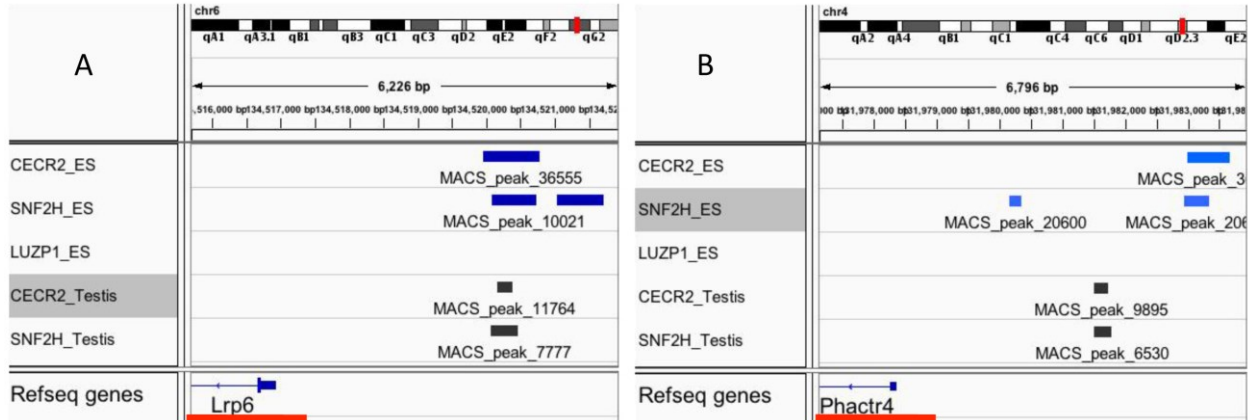
Figure 3-51 CECR2 and SNF2H co-occupy the promoter region of *Lrp6* and *Phctr4* in both ES cells and testis. The Integrative Genomics Viewer (IGV) browser was used to visualize the binding sites of CECR2, SNF2H and LUZP1 upstream of *Lrp6*. The binding sites in ES cells are shown as blue bars and in testis as black bars.
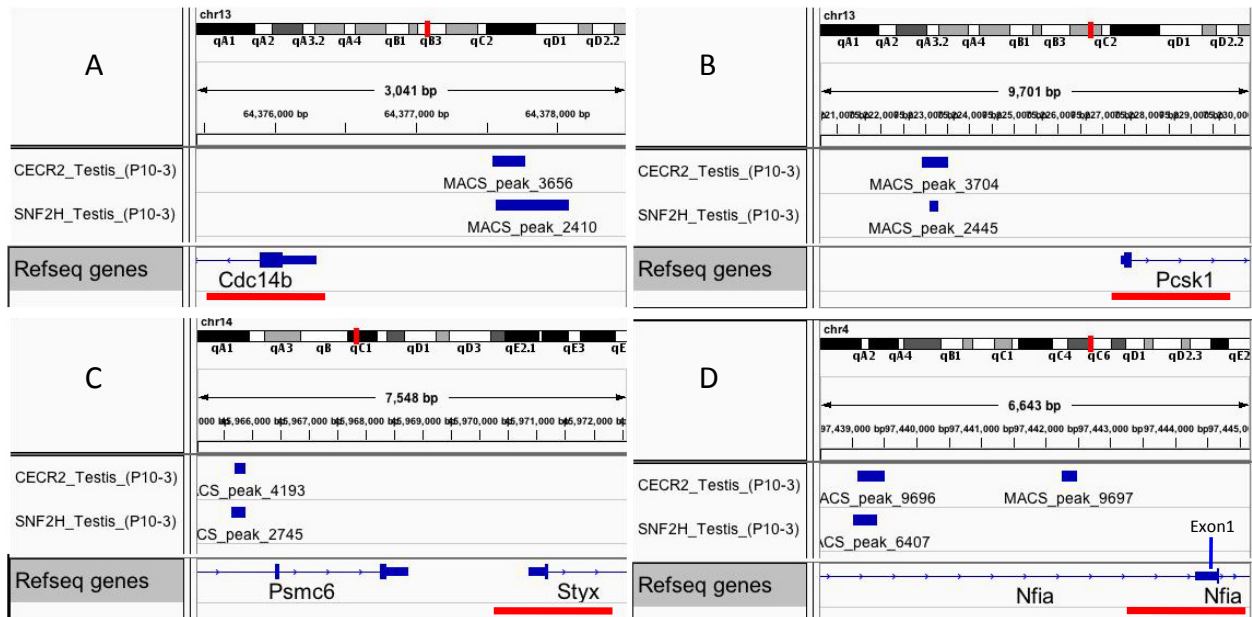
Figure 3-52 CECR2 and SNF2H co-occupy the promoter regions of genes associated with fertility in testis. Peak calling was performed using MACS and P-value cutoff=10^{-3}. The Integrative Genomics Viewer (IGV) browser was used to visualize the binding sites of CECR2 and SNF2H on the genomic regions of *Cdc14b* (A), *Pcsk1* (B), *Psmc6* (C) and *Nfia* (D). The binding sites are shown as blue bars. The genes are underlined with red line.

3.11.6. Genes containing the CECR2 binding sites in both ES and testis

Table 3-22 shows the list of the genes that in both ES cells and adult testis contain CECR2 binding sites within 5 kb upstream of their TSS. The list contains 14 genes that are involved in different biological processes. Four genes, *Caly*, *Nefh*, *Olig1, Nfia* and *Slc25a19* are involved in brain development. This analysis showed that these genes are probably the direct targets of CECR2 and mediate the involvement of this protein in embryonic brain development and reproduction.

Table 3-22 List of the genes bearing CECR2 peaks (P-value < 10-5) within their promoter region (up to 5 kb upstream of TSS) in both ES cells and testis. Green – Genes involved in the brain development. Red – Genes involved in fertility.
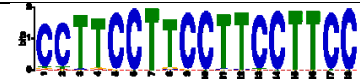
Coding genes		
Gene ID	**Gene name**	**Mutant phenotypes**
Aqp7	aquaporin 7	abnormal sperm physiology, abnormal kidney physiology
C8b	complement component 8	abnormal hematopoietic system physiology
Caly	calcyon neuron-specific vesicular protein	abnormal synaptic vesicle recycling
Cmah	cytidine monophospho-N-acetylneuraminic acid hydroxylase	cochlear outer hair cell degeneration,abnormal inner ear morphology,
Galnt6	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 6	
Galnt9	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 9	
Nedd9	neural precursor cell expressed, developmentally down-regulated gene 9	
Nefh	neurofilament, heavy polypeptide	abnormal neuron morphology
Nfia	nuclear factor I/A	male-sterility, low female fertility, lack of corpus callosum, hydrocephalus.
Olig1	oligodendrocyte transcription factor 1	adult cortical interneuron numbers, Fewer Numbers of Oligodendrocytes,
Prdm14	PR domain containing 14	female and male sterility
Slc25a19	solute carrier family 25 (mitochondrial thiamine pyrophosphate carrier), member 19	exencephaly
Slc46a1	solute carrier family 46, member 1	cell division cycle 20 homolog
Tomm40	translocase of outer mitochondrial membrane 40 homolog	

3.11.7. Motif enrichment in CECR2-associated genomic regions in ES cells

Discovery of significantly enriched binding motifs is an indicator of a successful ChIP-seq experiment (Shin et al. 2013). To test this and also to determine whether CECR2 co-localizes with site specific transcription factors, a motif enrichment analysis was performed using MEME-ChIP suite 4.11.2 (Machanick & Bailey 2011). MEME-ChIP uses two different motif discovery algorithms (MEME and DREME) to identify novel sequence motifs. The newly identified motifs are then used in a motif enrichment analysis algorithm (CentriMo) to be compared with databases containing previously characterized transcription factor motifs for similarity (Ma et al. 2014). The genomic DNA sequences of CECR2 peaks spanning the peak summits for 500 bp were extracted and were uploaded to MEME-ChIP and the default setting was used to identify common features of ChIP binding sites. 61 unique motifs were identified in the ES cells data set for CECR2, 11 of which did not match any annotated motifs in the JASPAR or UniPROBE mouse repositories but were highly significant (Table 3-23 & 3-24).

Table 3-23 Transcription factor binding motifs over-represented within the CECR2 peaks in ES cells. MEME-ChIP motif analysis was used to identify over-presented motifs in the central 500 bp of CECR2 peaks.

ID	Motif	Name (TF)	*E*-value	Region Matches
MA0472.1		EGR2	1.1e-390	7325
UP00042_2		Gm397_secondary	3.10E-281	5089
MA0073.1		RREB1	1.90E-267	3858
MA0006.1		Arnt::Ahr	1.40E-165	5491
UP00101_2		Sox12_secondary	5.40E-162	12732
MA0469.1		E2F3	5.40E-136	4462
MA0050.2		IRF1	5.90E-136	8004
MA0528.1		ZNF263	3.00E-132	9174
MA0508.1		PRDM1	9.30E-130	7835

MA0149.1		EWSR1-FLI1	6.40E-126	2948
UP00065_1		Zfp161_primary	7.80E-93	2617
MA0471.1		E2F6	3.90E-92	5403
UP00026_2		Zscan4_secondary	3.80E-85	5093
MA0162.2		EGR1	7.30E-81	7598
UP00026_1		Zscan4_primary	1.00E-72	5525
MA0513.1		SMAD2::SMAD3::SMAD4	3.50E-67	5546
UP00011_2		Irf6_secondary	4.20E-67	5649
MA0482.1		Gata4	4.70E-66	7346
UP00042_1		Gm397_primary	3.30E-51	3653
UP00036_2		Myf6_secondary	6.00E-35	6062
MA0516.1		SP2	1.10E-30	5666
UP00050_2		Bhlhb2_secondary	1.90E-29	4011
UP00007_1		Egr1_primary	1.20E-26	5040
MA0106.2		TP53	8.60E-26	2494
MA0147.2		Myc	1.50E-24	2756
MA0080.3		Spi1	5.40E-24	6511
UP00040_2		Irf5_secondary	5.70E-23	7356
MA0473.1		ELF1	2.00E-20	5712

ID	Logo	Name	p-value	Count
MA0598.1		EHF	1.30E-19	2418
MA0464.1		Bhlhe40	3.20E-19	3912
MA0259.1		HIF1A::ARNT	5.00E-19	2404
MA0525.1		TP63	6.90E-19	6589
MA0079.3		SP1	1.80E-18	5451
UP00002_1		Sp4_primary	4.90E-18	4761
MA0058.2		MAX	1.20E-17	4026
UP00097_1		Mtf1_primary	2.70E-17	1732
MA0506.1		NRF1	2.10E-14	1246
UP00014_2		Sox17_secondary	3.50E-13	3211
UP00021_1		Zfp281_primary	6.60E-13	5259
UP00008_2		Six6_secondary	2.70E-12	7099
MA0514.1		Sox3	4.60E-11	11368
MA0599.1		KLF5	1.20E-09	4832
UP00084_1		Gmeb1_primary	2.90E-09	1835
MA0474.1		Erg	2.30E-08	6308
MA0098.2		Ets1	1.50E-07	6348
UP00094_1		Zfp128_primary	2.60E-06	2542
UP00060_1		Max_primary	7.40E-06	2770
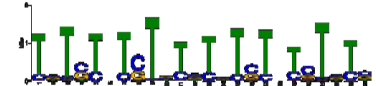
MA0517.1		STAT2::STAT1	8.30E-06	8196
UP00041_1		Foxj1_primary	7.10E-04	2974
MA0475.1		FLI1	3.80E-03	6121

Table 3-24 Unannotated identified motifs over-presented within DNA sequences under CECR2 peaks in ES cells. The DNA sequences of 500 bp summit region of CECR2 peaks were used in MEME-ChIP motif discovery analysis.

ID	Motif	Program	*E*-value	Region Matches (out of 21501)
2		MEME	6.1e-1014	11846
1		MEME	7.1e-790	10241
CACACWCA		DREME	2.7e-474	5957
3		MEME	4.1e-372	7720
CTKCCTYC		DREME	4.10E-118	4596
CRCRYGC		DREME	3.60E-113	4352
CASABAAA		DREME	5.20E-68	3786
GAMAGAAA		DREME	1.60E-45	3713
GGKCTAYA		DREME	6.30E-07	1317
GGCTGDCC		DREME	3.90E-05	1455
TATRTGAR		DREME	8.30E-04	1484

171

Fifty of these identified motifs were in the JASPER or UniPROBE Mouse repositories. Table 3-23 shows the identified motifs and the transcription factors binding to these motifs. The identified 50 motifs were associated with 33 transcription factors. These transcription factors corresponding to the identified motifs with P-value $< 10^{-5}$ were subjected to the PANTHER database system to identify Gene Ontology (GO) Biological Process categories and the pathways related with these transcription factors (Mi et al. 2016). Biological process analysis showed that these transcription factors are involved in different biological functions including "developmental process" (33.3% of the genes), "biological regulation" (72.70% of the genes), "apoptotic process" (6.1% of the gens). Table 3-25 shows all the biological functions identified for the transcription factors. The PANTHER analysis showed that the 33 transcription factors represented several signaling pathways such as "Wnt pathway" with 2 genes (*Trp53* and *Myc*), "P53 Pathway" (containing 11 genes) and "apoptosis signaling pathway" (1 gene).

Table 3-25 Gene Ontology of transcription factors sharing the same binding motifs with CECR2 in ES cells.

	Ontology	Category name (Accession)	Number of genes	Percent of gene hit
ES cell	GO Biological Process	biological regulation (GO:0065007)	24	72.70%
		metabolic process (GO:0008152)	24	72.70%
		cellular process (GO:0009987)	14	42.40%
		developmental process (GO:0032502)	10	30.30%
		response to stimulus (GO:0050896)	5	15.20%
		immune system process (GO:0002376)	3	9.10%
		cellular component organization or biogenesis (GO:0071840)	2	6.10%
		apoptotic process (GO:0006915)	2	6.10%
		multicellular organismal process (GO:0032501)	1	3.00%
	Pathway	PDGF signaling pathway (P00047)	6	18.20%
		p53 pathway feedback loops 2 (P04398)	4	12.10%
		Huntington disease (P00029)	3	9.10%
		p53 pathway (P00059)	3	9.10%
		Interleukin signaling pathway (P00036)	2	6.10%
		p53 pathway by glucose deprivation (P04397)	2	6.10%
		Gonadotropin releasing hormone receptor pathway (P06664)	2	6.10%
		Oxidative stress response (P00046)	2	6.10%
		P53 pathway feedback loops 1 (P04392)	2	6.10%
		CCKR signaling map (P06959)	2	6.10%
		Wnt signaling pathway (P00057)	2	6.10%
		Apoptosis signaling pathway (P00006)	1	3.00%
		Angiogenesis (P00005)	1	3.00%
		EGF receptor signaling pathway (P00018)	1	3.00%
		Asparagine and aspartate biosynthesis (P02730)	1	3.00%
		Ras Pathway (P04393)	1	3.00%
		Cell cycle (P00013)	1	3.00%
		VEGF signaling pathway (P00056)	1	3.00%

3.11.8. Identified uncharacterized motifs for CECR2 in ES cells

DREME and MEME discovered 11 short binding motifs by processing CECR2 peaks from ES cells. However, the CentriMo algorithm did not find significant similarities of these novel putative CECR2 binding motifs in transcription factor databases. I therefore used the CECR2 binding regions for three of the novel motifs and subjected them to GREAT gene ontology analysis to look for any functional clustering. Table 3-26 shows GO terms for genes associated with the CECR2 binding sites containing two of the discovered binding motifs by DREME. CACACWCA was identified within 48% and CRCRYGC in 24% of all CECR2 peaks throughout the genome from ES cells. Interestingly, GREAT connected the genes associated with these peaks with several GO terms related to brain development. This included genes causing exencephaly in mice when mutated, such as *itgb1*, *shh* and *Nog*. GREAT also associated CRCRYGC with the molecular function term of "RNA polymerase II distal enhancer sequence-specific DNA binding transcription factor activity" with 127 peaks over 36 genes.

3.11.9. Motif enrichment in overlapping binding sites of CECR2 with SNF2H and LUZP1 in ES cells

I identified the significantly enriched binding motifs shared with CECR2 and SNF2H in ES cells. The common unique peaks (P-value < 10-5) of CECR2 and SNF2H (2131 peaks) in ES cells were subjected to MEME-ChIP with the same setting as above. Thirteen unique motifs (E-value ≤0.05) were identified which were a subset of motifs identified for CECR2. Eight of the motifs were associated with 8 transcription factors (Table 3-26). The PANTHER classification system associated 4 of the transcription factors with 5 different pathways: "cell cycle" (transcription factor: E2F3), "gonadotropin releasing hormone receptor pathway" (GATA4), "PDGF signaling pathway" (FLI1), "p53 pathway feedback loops 2" (E2F3), "p53 pathway" (E2F3). The data showed the transcription factors that may cooperate with CECR2/SNF2H containing complexes to regulate target genes.

Six of the identified shared motifs for CECR2 and SNF2H were not significantly similar to known transcription factors motifs (Table 3-27). The most significant unannotated motif was discovered by MEME in 43% of the CECR2 and SNF2H overlapping peaks (923 of 2131 peaks). GREAT associated this set of peaks containing the most significant motifs with 1421 genes. GO terms were related to "cerebellum development", "metencephalon development"

173

Table 3-26 GREAT gene ontology terms from analysis using CECR2 peaks of two novel motifs from ES cells. GO terms associated with the brain development are highlighted in yellow.

	Ontology	Term name	Binomial raw P-Value	Binomial FDR Q-Value	Binomial fold enrichment	#Hit genes
Motif: CACACWCA (in 48% of all CECR2 peaks)						
ES Cell	GO Biological Process	hindbrain development	5.37E-44	5.40E-40	2.694412	48
		cerebellum development	1.83E-33	3.07E-30	3.051121	28
		lens fiber cell differentiation	1.10E-24	3.82E-22	4.371713	12
		patterning of blood vessels	5.69E-21	8.41E-19	2.742564	20
		cardiac muscle cell proliferation	1.21E-20	1.63E-18	4.075217	10
		cerebellar granular layer morphogenesis	9.24E-18	6.88E-16	5.255555	7
		cell chemotaxis	2.48E-17	1.71E-15	2.281085	38
		cerebellar granule cell differentiation	7.84E-17	5.05E-15	5.477791	6
		positive regulation of neural precursor cell proliferation	9.45E-17	5.98E-15	2.653874	17
		forelimb morphogenesis	2.10E-16	1.27E-14	2.532319	18
		cell differentiation in hindbrain	3.13E-16	1.84E-14	3.080112	12
		midbrain development	1.66E-15	8.40E-14	2.87817	15
		embryonic pattern specification	8.91E-15	4.11E-13	2.257061	25
		lens development in camera-type eye	5.10E-13	1.86E-11	2.078833	24
		anterior/posterior axis specification	1.11E-11	3.34E-10	2.399812	18
		branching involved in salivary gland morphogenesis	1.65E-10	4.23E-09	2.725551	10
		neural tube patterning	9.82E-10	2.26E-08	2.313085	17
		exocrine system development	1.45E-09	3.24E-08	2.136715	19
		response to sterol depletion	6.24E-09	1.23E-07	4.894456	6
		stem cell division	5.69E-08	9.61E-07	2.371818	13
Motif: CRCRYGC (in 24% of all CECR2 peaks)						
ES Cell	GO Biological Process	respiratory system development	4.29E-27	2.40E-24	2.007334	91
		lung development	1.50E-26	7.56E-24	2.044259	82
		respiratory tube development	5.53E-26	1.99E-23	2.019181	84
		hindbrain development	1.04E-23	2.13E-21	2.304111	54
		cerebellum development	1.14E-21	1.71E-19	2.755001	30
		metencephalon development	3.86E-21	5.24E-19	2.588172	36
		cerebellum morphogenesis	1.52E-20	1.76E-18	3.466931	18
		cardiac muscle cell development	8.45E-20	8.95E-18	3.006394	21
		hindbrain morphogenesis	6.88E-19	6.79E-17	3.185961	20
		regulation of myotube differentiation	8.92E-18	7.66E-16	2.673359	22
		regulation of muscle cell differentiation	1.37E-17	1.14E-15	2.121481	36
		cardiac cell development	2.16E-17	1.75E-15	2.732636	22
		cerebellar cortex development	4.83E-17	3.83E-15	3.00113	19
		regulation of striated muscle cell differentiation	2.39E-16	1.68E-14	2.18828	31
		cerebellar cortex morphogenesis	3.52E-16	2.41E-14	3.368773	14
		mesoderm morphogenesis	3.61E-16	2.44E-14	2.592851	25
		striated muscle cell development	1.77E-15	1.16E-13	2.254768	34
		regulation of development, heterochronic	6.81E-15	4.12E-13	4.158188	9
		regulation of timing of cell differentiation	2.06E-14	1.19E-12	4.105858	8
		neural tube patterning	2.70E-14	1.54E-12	2.910095	18

Table 3-27 Unique motifs over-represented within the CECR2 and SNF2H overlapping peaks (P-value < 10-5) in ES cells. MEME-ChIP motif analysis was used to identify over-presented motifs in the central 500 bp of peaks shared with CECR2 and SNF2H.

ID	Known Motif	Name	*E*-value	Region Matches (out of 2131)
UP00101_2		Sox12_secondary	5.90E-11	1265
MA0469.1		E2F3	4.90E-08	370
MA0149.1		EWSR1-FLI1	3.30E-07	218
MA0050.2		IRF1	5.10E-04	778
MA0482.1		Gata4	1.20E-03	529
MA0508.1		PRDM1	2.60E-03	703
MA0472.1		EGR2	2.70E-03	228
MA0471.1		E2F6	1.10E-02	678

ID	Novel Motif	Program	*E*-value	Region Matches (out of 2131)
1		MEME	9.50E-32	885
2		MEME	8.80E-23	616
3		MEME	1.40E-07	196
ATKCAYAC		DREME	6.30E-05	480
CCTTCCTW		DREME	2.80E-04	309
CCTCCTCC		DREME	1.30E-02	301

175

and "hindbrain development" (Table 3-28). Many of the remaining GO terms were related to kidney development. GREAT also associated the second most significant unannotated motif with 4 GO terms including "cerebellum development", "endothelial cell proliferation", "positive regulation of developmental growth" and "positive regulation of nephron tubule epithelial cell differentiation". The DREME algorithim identified 3 unannotated motifs shared between CECR2 and SNF2H in ES cells (Table 3-27). GREAT analysis only defined GO terms for the ATKCAYAC motif (Table 3-28). This motif was associated with the molecular function of "transcription regulatory region DNA binding". There were several GO biological processes terms related to embryonic development including "embryonic axis specification", "embryonic pattern specification" and "formation of primary germ layer" suggesting that the genes associated with these GO terms are probably regulated by CECR2/SNF2H-containing complexes.

Table 3-29 shows the motifs identified using overlapped peaks between three proteins, CECR2, SNF2H and LUZP1 in ES cells (436 peaks). There was no known motif identified for this data set. All 5 motifs were unannotated. GREAT associated the peaks of the motifs to kidney development including "glomerular endothelium development", "glomerular mesangial cell differentiation" and "glomerular mesangium development" (Table 3-30).

Table 3-28 GREAT gene ontology terms from an analysis using overlapping peaks (P-value < 10^{-5}) of CECR2 and SNF2H from ES cells containing novel motifs. GO terms related to brain development are highlighted in yellow.

Motif: MEME1 (in 48% of all peaks)						
	Ontology	Term name	Binomial raw P-Value	Binomial FDR Q-Value	Binomial fold enrichment	#Hit genes
ES Cell	GO Biological Process	hindbrain development	7.93E-07	0.003	2.51	26
		regulation of myeloid cell differentiation	0.0002	0.04	2.02	25
		nephron development	9.72E-06	0.01	2.38	22
		kidney epithelium development	1.04E-05	0.01	2.60	18
		metencephalon development	6.17E-05	0.03	2.55	18
		positive regulation of myeloid cell differentiation	0.0001	0.04	2.65	16
		cerebellum development	0.0001	0.05	2.55	16
		ureteric bud morphogenesis	0.0003	0.05	2.30	14
		glomerulus development	3.49E-05	0.02	2.88	13
		regulation of generation of precursor metabolites and energy	0.0001	0.04	3.30	12
		metanephric nephron development	0.0002	0.04	2.89	11
		nephron epithelium morphogenesis	0.0003	0.04	2.90	10
		regulation of polysaccharide metabolic process	0.0001	0.04	4.48	8

		regulation of epithelial cell differentiation involved in kidney development	0.0002	0.05	4.38	7
		regulation of glycogen metabolic process	0.0002	0.04	4.36	7
		epithelial cell differentiation involved in kidney development	0.0002	0.04	3.64	7
		endothelial cell proliferation	2.92E-05	0.02	5.26	5
		positive regulation of glycogen metabolic process	0.0002	0.04	5.65	5
		positive regulation of nephron tubule epithelial cell differentiation	2.81E-06	0.007	23	3
		ureteric bud elongation	0.0001	0.05	7	3

Motif: MEME2 (in 31% of all peaks)

	GO Biological Process	cerebellum development	6.49E-06	0.065	3.375455	13
		positive regulation of developmental growth	1.46E-05	0.049	3.823745	11
		endothelial cell proliferation	1.30E-05	0.065	6.581207	5
		positive regulation of nephron tubule epithelial cell differentiation	1.90E-05	0.047	26.47592	3

Motif: ATKCAYAC (in 26% of peaks)

	GO Molecular Function	regulatory region DNA binding	5.35E-08	0.0001	2.2112	36
		transcription regulatory region DNA binding	9.34E-08	0.0001	2.1924	35
	GO Biological Process	regulation of cellular component biogenesis	2.31E-05	0.0043	2.0000	30
		cell fate commitment	1.12E-06	0.0012	2.2702	27
		cellular response to cytokine stimulus	5.18E-05	0.0066	2.3111	26
		regulation of anatomical structure size	0.0001	0.0109	2.1392	23
		anterior/posterior pattern specification	1.70E-05	0.0037	2.3755	20
		transcription from RNA polymerase II promoter	7.97E-05	0.0088	2.2933	18
		regulation of muscle cell differentiation	8.70E-07	0.0014	3.1315	14
		nephron development	7.54E-05	0.0086	2.6379	13
		regulation of striated muscle cell differentiation	4.70E-07	0.0011	3.4524	12
		regulation of epithelial cell differentiation	3.28E-05	0.0054	2.8887	12
		embryonic pattern specification	1.50E-05	0.0036	3.6102	11
		segmentation	2.41E-05	0.0043	2.9537	11
		formation of primary germ layer	5.03E-05	0.0065	3.1197	11
		kidney epithelium development	6.76E-05	0.0080	2.9289	11
		regulation of stem cell differentiation	0.0001	0.0117	2.7920	11
		anterior/posterior axis specification	3.66E-05	0.0057	4.2623	8
		embryonic axis specification	1.17E-05	0.0035	4.7985	7
		regulation of systemic arterial blood pressure by baroreceptor feedback	3.69E-06	0.0020	9.1703	4
		positive regulation of the force of heart contraction by epinephrine-norepinephrine	1.26E-05	0.0037	17.394	3
		positive regulation of heart rate by epinephrine-norepinephrine	6.84E-05	0.0081	12.163	3

Table 3-29 Unique motifs over-represented within the CECR2, SNF2H and LUZP1 overlapping peaks (P-value < 10-5) in ES cells. The significant motifs (E-value ≤ 0.05) found by the programs MEME, DREME and CentriMo; clustered by similarity and ordered by E-value.

ID	Known Motif	Program	*E*-value	Region Matches (out of 436)
1		MEME	1.4e-1353	246
2		MEME	3.8e-1313	205
3		MEME	1.9e-493	230
ASAAACA		DREME	9.8e-006	70
MATGCACA		DREME	3.7e-002	54

Table 3-30 GREAT gene ontology terms from an analysis using overlapping peaks (P-value < 10-5) of CECR2, SNF2H and LUZP1 from ES cells containing novel motifs.

	Ontology	Term name	Binomial raw P-Value	Binomial FDR Q-Value	Binomial fold enrichment	#Hit genes
ES Cell	GO Biological Process	glomerular endothelium development	3.82E-06	0.01	39.81	3
		glomerular mesangial cell differentiation	8.04E-06	0.02	32.93	3
		glomerular mesangium development	1.24E-05	0.02	29.44	3

3.11.10. Motif enrichment in CECR2-associated genomic regions in testis

I processed putative CECR2 peaks from testis ChIP-seq data (p value 10^{-5}) for motif discovery with MEME-ChIP. The DNA sequence extracted from 500 bp summit regions of CECR2 peaks were used in MEME-ChIP and the default setting was used to find binding motifs. 26 unique motifs were identified in testis data set for CECR2, 21 of which were in the JASPAR or UniPROBE data base and 5 of the discovered motifs were unannotated (Table 3-31). All of the motifs are among the ones that were identified for ES cells except for UP00018_2 (Irf4_secondary) (Table 3-23). The identified known motifs belong to 25 transcription factors. According to the PANTHER classification system, 73% of the TFs (17 genes) were associated with molecular function term of "nucleic acid binding transcription factor activity" (GO:0001071) and ~9% (EWSR1 and GATA4) were associated with "catalytic activity" (GO:0003824). The biological processes and pathways that are carried out by these transcription factors are displayed in Table 3-32. The most enriched pathways included "Gonadotropin releasing hormone receptor pathway" (4 TFs: GATA4, EGR1, SMAD2 and SMAD4) and "Interleukin signaling pathway" (3 TFs: STAT1, STAT2 and SPI1). Similar to ES cells, there was enrichment of "Wnt pathway" with 1 gene (SMAD4) and "P53 Pathway" (1 TF: E2F3). From the enriched pathways "gonadotropin releasing hormone receptor pathway" and "TGF-beta signaling pathway" have roles in reproduction.

To investigate the enrichment of binding motifs co-occupied by both CECR2 and SNF2H in testis, I used putative binding peaks for CECR2 and SNF2H (P-value <0.001) obtained from the testis experiment. To find the overlapping peaks I intersected CECR2 peaks with SNF2H peaks using bedtools2 (Quinlan & Hall 2010). The unique common peaks (1366 peaks) were used in the MEME-ChIP website to identify the possible binding motifs. The identified motifs were a subset of known motifs I had already identified for CECR2 (Table 3-31 yellow highlighted). The same GO terms enriched for CECR2 were also enriched for CECR2 and SNF2H common peaks.

The ChIP-seq experiment analysis revealed the genes that are probably regulated directly by CECR2-containing complexes in ES cells and adult testis. As expected from a chromatin remodeler, CECR2 bound to many genomic regions. In both ES cells and testis, CECR2 bound to the promoter region or the cis-regulatory regions of the genes that are involved in reproduction

and brain development. Motif analysis of the CECR2 complex binding sites suggests that this complex may interact with other regulatory factors as a mechanism of targeting to chromatin.

Table 3-31 Identified motifs over-represented within the CECR2 peaks in testis. MEME-ChIP motif analysis was used to identify over-presented motifs in central 500 bp of CECR2 peaks. Highlighted motifs are CECR2 and SNF2H common motifs.

ID	Motif	Name	*E*-value	Region Matches
1		MEME	3.30E-258	1211
2		MEME	7.40E-245	1425
GAGAGAGA		DREME	7.10E-218	1006
3		MEME	1.70E-177	1477
MA0508.1		PRDM1	2.30E-97	1317
MA0050.2		IRF1	5.90E-92	1280
MA0528.1		ZNF263	6.70E-78	1276
UP00011_2		Irf6_secondary	6.00E-62	1309
MA0513.1		SMAD2::SMAD3::SMAD4	1.30E-52	989
AYACACAC		DREME	8.70E-47	622
UP00040_2		Irf5_secondary	9.50E-34	1109
MA0469.1		E2F3	1.30E-28	484
MA0482.1		Gata4	1.00E-24	839
MA0149.1		EWSR1-FLI1	4.70E-23	327
MA0471.1		E2F6	1.10E-20	865
UP00101_2		Sox12_secondary	1.90E-20	1150

Accession	Motif	TF	p-value	Count
MA0472.1		EGR2	2.00E-19	980
MA0073.1		RREB1	2.50E-15	645
MA0080.3		Spi1	4.00E-10	953
GGAAGSAA		DREME	1.60E-09	258
UP00042_2		Gm397_secondary	2.20E-09	928
MA0514.1		Sox3	3.90E-09	570
MA0162.2		EGR1	1.70E-07	742
MA0006.1		Arnt::Ahr	7.90E-05	1048
UP00018_2		Irf4_secondary	7.90E-05	976
MA0517.1		STAT2::STAT1	1.30E-03	792

Table 3-32 Gene Ontology of transcription factors sharing the same binding motifs with CECR2 in testis.

	Ontology	Category name (Accession)	Number of genes	Percent of gene hit
Testis	GO Biological Process	metabolic process (GO:0008152)	18	78.30%
		biological regulation (GO:0065007)	17	73.90%
		cellular process (GO:0009987)	10	43.50%
		response to stimulus (GO:0050896)	7	30.40%
		immune system process (GO:0002376)	6	26.10%
		developmental process (GO:0032502)	5	21.70%
		apoptotic process (GO:0006915)	2	8.70%
		multicellular organismal process (GO:0032501)	1	4.30%
	Pathway	Gonadotropin releasing hormone receptor pathway	4	17.40%
		Interleukin signaling pathway	3	13.00%
		TGF-beta signaling pathway	3	13.00%
		PDGF signaling pathway	3	13.00%
		JAK/STAT signaling pathway	2	8.70%
		EGF receptor signaling pathway	2	8.70%
		p53 pathway	1	4.30%
		p53 pathway feedback loops 2	1	4.30%

		Wnt signaling pathway	1	4.30%
		Cell cycle	1	4.30%
		Interferon-gamma signaling pathway	1	4.30%
		Angiogenesis	1	4.30%
		Hypoxia response via HIF activation	1	4.30%
		Asparagine and aspartate biosynthesis	1	4.30%
		Inflammation mediated by chemokine and cytokine signaling pathway	1	4.30%
		Ras Pathway	1	4.30%
		Oxidative stress response	1	4.30%

4. Discussion

4.1. Raising an antibody to CECR2 allowed extensive biochemical and molecular analysis

In my project, to address the role of *Cecr2* in development I needed to do many biochemical/molecular analyses on wild-type CECR2 protein. In order to perform these studies in the mouse, a specific and versatile antibody against CECR2 was required. There had been many attempts to raise antibodies against mouse CECR2 previously in the McDermid lab with no success (Figure 3-1). For that reason previous biochemical analyses on the CECR2 protein were performed using a β-galactosidase antibody against the CECR2^{Gt45Bic} fusion protein in heterozygous or mutant mice. The CECR2^{Gt45Bic} fusion protein was very informative in determining the expression pattern of CECR2 in mice and also identifying the presence of CECR2-containing complexes (Banting 2004, Thompson et al. 2012). However, the CECR2^{Gt45Bic} fusion protein is missing a big segment of CECR2 including the conserved bromodomain, which binds to acetylated lysine residues on histones and found in the structure of all of the BAZ-like protein components of the ISWI complexes except for RSF-1 (Figure 1-4) (Muller et al. 2011). For instance, the bromodomain of TIP5 (the binding member of NoRC complex) binds to acetylated H3 and H4 and its binding is a prerequisite for remodeling function of the NoRC complex (Zhou & Grummt 2005). Therefore, since CECR2^{Gt45Bic} lacks the bromodomain, it is very likely that its targeting and remodeling function has been affected and its protein binding is not representative of wild-type CECR2. The presence of several abnormal phenotypes in *Cecr2*^{Gt45Bic} mutant animals including exencephaly and subfertility indicates that CECR2^{Gt45Bic} protein function is impaired. Therefore, to test all the data obtained from biochemical analysis of the fusion protein and to rule out artifacts related to the βgeo cassette in future experiments, I decided to study the wild-type CECR2.

Since all of the attempts to produce antibody to CECR2 using small peptides had failed, including an attempt at chicken polyclonal antibodies, I decided to use a 35 kDa fragment of the

CECR2 protein to produce antibodies in rabbits. The polyclonal antibody produced using this approach was highly specific towards both the denatured and native forms of CECR2. Specificity was confirmed by detecting a band in mouse wild-type embryo extracts that was absent in the $Cecr2^{tm1.1Hemc}$ mutant embryo extract (Figure 3-6). The band was approximately 170 kDa, consistent with the calculated size of mouse CECR2. The antibody also confirmed that $Cecr2^{Gt45Bic}$ allele produced a small amount of native protein detectable by Western blot. In order to determine the composition of CECR2 complexes and their binding sites *in vivo*, the antibody must detect the native form of CECR2. An epitope could be hidden in the native structure of CECR2 or under interacting protein binding sites. However, the fact that the antibody was able to successfully immunoprecipitate CECR2 from protein extracts (Figure 3-16) and detect CECR2 by immunofluorescence (Figure 4-1) suggested it could be ChIP-grade. This was supported by its ability to bind to CECR2-containing complexes bound to DNA as part of the ChIP procedure (Figure 3-43C). ChIP is a demanding technique for antibodies and can cause alteration or destruction of epitopes by formaldehyde, leading to a decrease of their affinity. Epitopes can also be occluded in the protein-DNA complex. The presence of SNF2H in the CECR2 ChIP immunoprecipitate (Figure 3-43C) confirmed that the CECR2 antibody immunoprecipitated CECR2-containing complexes after the cross-linking reaction. The fact that the antibody generated against mouse CECR2 protein cross-reacts with human CECR2 increases the versatility of this reagent (Figure 3-8). Taken together, the rabbit polyclonal antibody produced with a CECR2 protein fragment has been shown to be highly specific towards denatured and native forms of CECR2 *in vitro* and *in vivo* and to immunoprecipitate complexes. Therefore, the CECR2 polyclonal antibody I generated was well suited to my project goals.

4.2. CECR2 antibody leads to a refined understanding of CECR2 expression

All the information about the cellular localization and tissue distribution of CECR2 came from investigations with the Cecr2^{Gt45Bic} fusion protein, including that it is a nuclear protein (Banting 2004, Tate et al. 1998, Thompson et al. 2012), as suggested by its putative NLS (Figure 3-1). Cell fractionation and immunofluorescence microscopy using the new CECR2 antibody confirmed this (Figure 3-14 and 3-15). The ability of the CECR2 antibody to immunoprecipitate DNA with CECR2 protein in the ChIP-seq experiment indicated that CECR2 associates with chromatin as predicted by its AT hook and a bromodomain. The AT hook is a structural

characteristic of many DNA-binding proteins, which bind the minor groove of AT sequence (Aravind & Landsman 1998) and the bromodomain binds the specific histone.

Previously, expression of CECR2 was determined by X-gal staining in heterozygous *Cecr2*^{Gt45Bic} embryos and some adult tissue. Western blot analysis using the CECR2 antibody also confirmed CECR2 expression in the adult hippocampus (data not shown, high background).
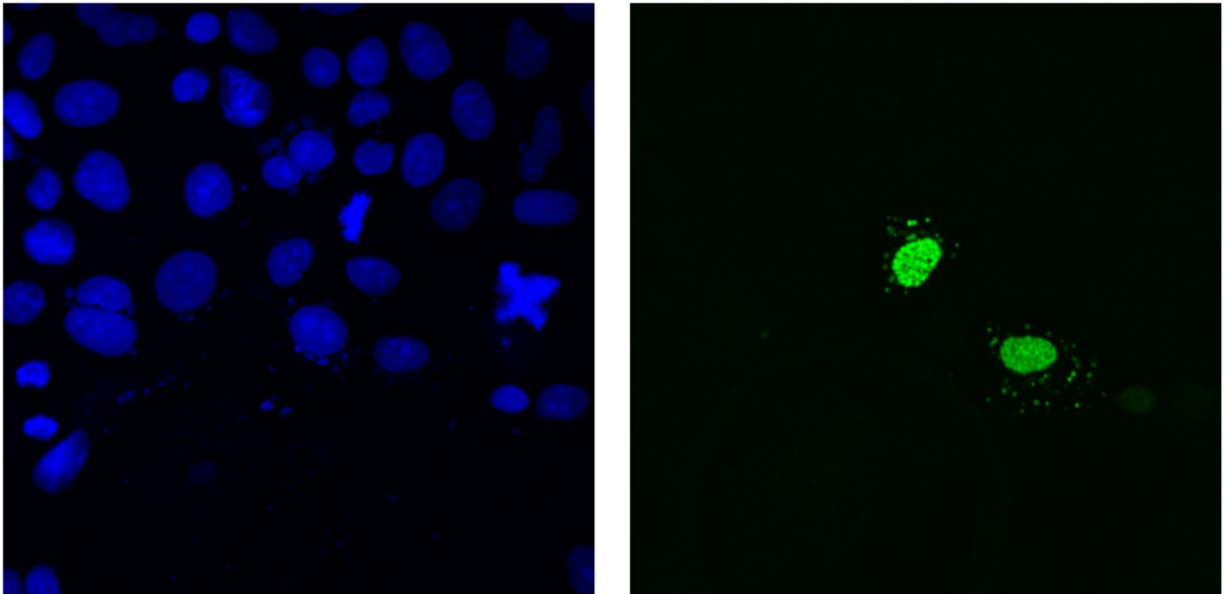
Figure 4-1 CECR2 antibody successfully detected human CECR2 in an immunofluorescence staining assay. CaCo-2 cells were transiently transfected with an expression vector containing human *CECR2* and then the cells were stained with CECR2 antibody (green) (right panel), and counterstained with DAPI (blue) (left panel). The presence of green cells indicates that 2 of the cells shown on the left panel are expressing *Cecr2* (Alaina Terpstra, unpublished).

It has been shown previously that the CECR2^{Gt45Bic} fusion protein is present in adult testis in spermatogonia and spermatocytes (Thompson et al. 2012). Interestingly, only spermatogonia type A cells located at the margins of seminiferous tubules showed *Cecr2* expression in immunofluorescence (Figure 3-12). No staining was detected in spermatocytes. This was supported by Western blot analysis of immortalized mouse cell lines GC1 and GC2 (Figure 3-11, only for GC1). GC1 are type B spermatogonial cell and GC2 are spermatocytes. CECR2 was not detected in either of the cell lines by Western blot. The immunofluorescence result that CECR2 is

185

only detectable in spermatogonia type A cells was not consistent with the CECR2^{Gt45Bic} fusion protein localization in testis, which stains all spermatogonia and spermatocytes. The fusion protein must form a different protein conformation that may be more proteolytically stable due to the LacZ tag (Kimple et al. 2013), leading to persistence of CECR2^{Gt45Bic} fusion protein in the spermatocytes. The ability to study protein expression with the CECR2 antibody has lead to a better understanding of CECR2 expression.

4.3. A unique CECR2 isoform

In Western analysis with the new CECR2 antibody, a smaller and fainter band for CECR2 was detected by Western blot analysis in embryonic and adult tissues including whole embryo, adult testis, midbrain, cerebellum and neurospheres, although not ES cells (Figure 3-8, 3-9 and 3-11). The size difference between the CECR2 band and this band is very small: ~170 kDa vs ~160 kDa. Surprisingly, Western blot analysis of protein extracts from *Cecr2*^{Gt45Bic} and *Cecr2*$^{Tm1.1Hemc}$ homozygous mutants also showed this smaller band, suggesting that the band is non-specific. To rule out the presence of any *Cecr2* transcript in the mutant, cDNA samples from the testis were used in RT-PCR from different exons and surprisingly, all exons were detected in the mutant except for exon 1. This confirmed that there was another transcript of *Cecr2* in the mutant mice, since no closely related protein is known to be in the genome. The presence of a *Cecr2* transcript in the wildtype and *Cecr2*^{Gt45Bic} mutant has been recently confirmed by RNA-seq in our lab (Kacie Norton, unpublished). The newly discovered transcript has an alternative exon 1 located 21 kb upstream of the canonical exon 1 and splices directly to exon 2. It is possible that the second band detected in Western blot is the protein product of this transcript.

As one of my controls for immunoprecipitated CECR2-containing complexes analyzed by LC-MS/MS, I used protein extract from testis of homozygous FVB/N *Cecr2*$^{Tm1.1Hemc}$ mutants (FVB/N mutants survive to adulthood). Four unique peptides belonging to CECR2 with the total percent coverage of 6.13% were detected in the homozygous *Cecr2*$^{Tm1.1Hemc}$ mutant control (Appendix M). The detected peptides include GPSQALRGAQGGESmmDSPEmIAmQQLSSR (exon16), ELPPELSHLDLnSPmREGK (exon 10), GSFQEVHRPPGLQMHPVQSQSLFPK (exon 17-18) and SnGELSLcRESERqK (exon 4-5). Ordinary Blast (basic local alignment search tool) searches only revealed CECR2 protein for the peptides indicating that these peptides were

unique. Taken together, the data suggest there is a non-canonical CECR2 protein present in wildtype and mutant mice that does not contain the canonical exon 1. To confirm the existence of a non-canonical CECR2 protein isoform, in mutant mice, further mass spectrometry analysis of the non-canonical CECR2 band would be very informative. If peptides belonging to CECR2 were found, then this would be evidence of a non-canonical CECR2 polypeptide.

$Cecr2^{tm2a(EUCOMM)Hmgu}$ is a newly produced $Cecr2$ mutation on the C57BL/6N strain. In this mutation, the lacZ reporter cassette has replaced exon 4 of $Cecr2$, which results in a frameshift in the rest of the gene. Penetrance of exencephaly appears to be 100% like the $Cecr2^{Tm1.1Hemc}$ mutation. So far studying the phenotype of 4 mutants has revealed that in addition to exencephaly, the mutants exhibit coloboma, which has never been detected with the other $Cecr2$ mutations (Kacie Norton, unpublished). Ocular coloboma, which results from the inability to close the choroid fissure during early eye development, is an associated symptom of cat eye syndrome. CECR2 is one of the 10 genes present in the cat eye syndrome critical region in mouse, however cat eye syndrome results from a triplication of the region rather than a loss. Although coloboma may have been missed in previous mutations due to the colourless iris in the BALB/c and FVB/N strains, it is also possible that these mutations do not result in the complete loss of function in all circumstances. It is possible that coloboma seen in the $Cecr2^{tm2a(EUCOMM)Hmgu}$ is caused by the lack of this $Cecr2$ isoform. Since there is a frameshift after exon 4 in $Cecr2^{tm2a(EUCOMM)Hmgu}$, the presence of any $Cecr2$ isoform in this mutant is improbable. To test this hypothesis, the presence of the $Cecr2$ isoform transcript in the $Cecr2^{tm2a(EUCOMM)Hmgu}$ mutant can be investigated by performing RT- PCR using the RNA extracts from the mutant tissues. The expression of the $Cecr2$ protein isoform can also be investigated using Western blot analysis.

Even if there is a non-canonical CECR2 protein in the $Cecr2^{Gt45Bic}$ and $Cecr2^{tm1.1Hemc}$, it does not compensate for canonical CECR2 loss of function, since they show exencephaly at 54% and 100% penetrance respectively.

4.4. The presence of SNF2H and SNF2L in CECR2-containing complexes

The ISWI complexes are multi-protein complexes containing SNF2H or SNF2L as the catalytic interacting partner in mammalian cells (Erdel & Rippe 2011). There have been eight

distinct complexes isolated from mammalian cells belonging to this family of ATP-dependent chromatin remodeling complexes including ACF (Bochar et al. 2000, LeRoy 2000), CHRAC (Poot et al. 2000), WICH (Bozhenok et al. 2002a), B-WICH (Cavellan et al. 2006), RSF (LeRoy et al. 1998), NoRC (Strohner et al. 2001), NURF (Barak et al. 2003) and CERF (Banting et al. 2005). There is also a study showing the probability of the presence of mammalian ToRC complex (Emelyanov et al. 2012). In the mammalian ISWI complexes one or more than one non-catalytic proteins bind to SNF2H or SNF2L. Human CECR2-containing remodeling factor (CERF) was first isolated from HEK293 cells and consists of CECR2 and SNF2L (Banting 2004). The presence of a CECR2-containing complex in mouse has also been reported in ES cells and adult testis using the $Cecr2^{Gt45Bic}$ fusion protein (Thompson et al. 2012). Interestingly, mouse CECR2 interacts with SNF2H instead of SNF2L. The reciprocal co-immunoprecipitation assays using the CECR2 antibody confirmed that wild type CECR2 forms a complex with SNF2H in mouse ES cells and adult testis (Figure 3-20 and 3-21). My results showed that CECR2 also interacts with SNF2L in both ES cells and adult testis (Figure 3-29 and 3-30). Comparing the amount of SNF2H and SNF2L co-immunoprecipitated with CECR2 antibody suggests that CECR2 interacts predominantly with SNF2H in both mouse ES cells and adult testis. Similar results have been previously shown for CHRAC complexes. In CHRAC complex isolated from HeLa cells, ACF1 interacts mainly with SNF2H, but minor amounts of SNF2L are also found to be interacting with ACF1 (Poot et al. 2000). Another experiment showed that in mouse ES cells, ACF1/SNF2H-containing complexes outnumber ACF1/SNF2L-containing complex (Bozhenok et al. 2002a). In my analysis of isolated CECR2-containing complexes with LC-MS/MS, SNF2L was detected as one of the proteins co-immunoprecipitated with CECR2 in ES cells (five independent experiments) but not in adult testis (SNF2H was detected in all experiments) (Table 3-4 and 3-5). The fact that SNF2L was not detected in the latter could be explained by the low amount of SNF2L interacting with CECR2 antibody. Taken together, the data suggest that CECR2 is associated mostly with SNF2H and with a minor amount of SNF2L in ES cells and the adult testis. Since CECR2 interacts with both these proteins, it is possible that CECR2 forms two distinct complexes by binding to SNF2L or SNF2H. Another possibility is that CECR2, SNF2H and SNF2L are components of the same complex. SNF2H and SNF2L show functional differences according to the phenotypes of their mutant mice and *Snf2l* cannot compensate for the

lack of *Snf2h* during embryogenesis (Yip et al. 2012b). Thus, it is possible that CECR2-containing complexes containing SNF2H or SNF2L may function differently.

Another adult mouse tissue with detectable CECR2 is ovary (Figure 3-8). It has been shown that SNF2L and SNF2H both are expressed in adult mouse ovary. While SNF2H is expressed more during the growth of follicles, SNF2L expression rapidly increases in granulosa cells after ovulation indicating distinct expression patterns for these two ISWI proteins (Lazzaro & Picketts 2001). My co-immunoprecipitation assay showed that CECR2 interacts with SNF2L in the adult ovary, while SNF2H was not detectable using Western blot analysis (Figure 3-30B). This suggests that most of the CECR2 is associated with SNF2L in adult ovary although a small amount of SNF2H below detection levels cannot be ruled out. Overall, my results suggest that CECR2 interacts predominantly with SNF2H in adult testis but with SNF2L in adult ovary.

4.4.1. CECR2 probably interacts physically with SNF2H

In ISWI complexes, SNF2H is thought to physically interact with the largest non-catalytic components (the BAZ-like proteins) by binding to the DDT domain (Figure 1-4). These specialized non-catalytic binding partners include ACF1, WSTF, RSF1, TIP5 and BPTF. CECR2 also shares three functional domains with these proteins including DDT, AT hook and a bromodomain. It has been shown that deletion of the DDT domain of ACF1 prevents its binding to the ISWI protein in Drosophila (Eberharter et al. 2004, Fyodorov & Kadonaga 2002). Isolating CECR2-containing complexes in harsh IP conditions including high salt concentration and stringent wash steps did not disturb interaction between CECR2 and SNF2H, suggesting a robust physical interaction between CECR2 and SNF2H. The co-immunoprecipitation of the CECR2^{Gt45Bic} fusion protein with SNF2H increases the probability that the DDT domain of CECR2 binds to SNF2H, since only 20% of CECR2 amino acids are present in the CECR2^{Gt45Bic} fusion protein including its DDT domain. Co-immunoprecipitation of SNF2H with CECR2 in the presence of ethidium bromide and RNase reduced the probability that this interaction is mediated through nucleic acids (Figure 3-23). Altogether, my data suggest that SNF2H may physically interact with the DDT domain of CECR2 and the interaction is nucleic acid independent. To determine if the DDT domain is responsible for the interaction between CECR2 and SNF2H, a truncation analysis can be performed. After cloning and expressing the truncated CECR2 (CECR2 with deleted DDT domain), its interaction with SNF2H could be analyzed using

the yeast two-hybrid assay. If the DDT domain is physically interacting with SNF2H, then the truncated CECR2 protein should be unable to bind to SNF2H in the yeast two-hybrid assay.

4.5. CECR2-containing complexes

Immunoprecipitation with Flag-tagged SNF2L indicated that human CECR2 forms a complex with SNF2L with the size of 600 kDa (Banting et al. 2005). Interestingly, a later study using the CECR2^{Gt45Bic} fusion protein and β-gal antibody showed that CECR2-containing complexes have different sizes in mouse ES cells and adult mouse testis (Thompson et al. 2012). The complex size was estimated to be 300-400 kDa in ES cells and 0.9-1 MDa in testis. If the complex only contains CECR2 and ISWI protein, its size should be ~300 kDa, the size that has been reported for ES cells. But the molecular size of ~1 MDa suggested that there should be other interacting proteins in the complex, at least in adult testis. One of the disadvantages of the LacZ tag is its tendency to form homotetramers in solution (Kimple et al. 2013), which could be occurring with the CECR2^{Gt45Bic}-containing complexes isolated from testis. Using the antibody against native CECR2, I showed that the size of CECR2-containing complexes is dependent on the condition of the experiments. First, freezing the protein extracts before gel filtration analysis led to smaller complex sizes, presumably due to dissociation. Therefore, I always used fresh protein extracts for all of my biochemical analysis. Secondly, the size of the CECR2-containing complexes is affected by the salt concentration used for protein extraction and gel filtration running buffer. Using the salt concentration (420 mM) that had been used previously by Thompson et al. (2012) to determine the CECR2^{Gt45Bic}-containing complexes led to several peaks in gel filtration profile of testes using the CECR2 antibody of native protein (data not shown). By decreasing the salt concentration to 150 mM (closer to physiological salt concentration) the CECR2-containing complexes were detected in fractions with high molecular weights in both ES cells and testis. Using 150 mM salt led to a single peak for adult testis with at the highest elution in a fraction corresponding to ~1.8 MDa in testis (Figure 3-18) and resulted in a change in the estimated complex size in ES cells from ~670 kDa to ~2 MDa (Figure 3-19). It has been shown that the energy of protein-protein interactions is dependent on salt (Zhang et al. 2011). Salt determines the ion strength of the solution and its concentration has a direct relationship with ion strength. In a study of the effect of NaCl concentration on five heterodimers and two

homodimers, both experimental data and the numerical calculations showed that the increase in salt concentration weakens the binding affinity of the components (Bertonati et al. 2007). My data suggest that using high salt concentration for protein extraction and during gel filtration weakens the protein-protein association and leads to smaller-sized CECR2-containing complexes. The size of the complex at 150 mM NaCl, which is closer to physiological salt concentration, is likely closer to the true size of the CECR2-containing complexes. I always used the same salt concentration during protein extraction and gel filtration. The dissociation of the complexes could occur during either of these steps. Preparing protein extracts under high salt concentration (420 mM) followed by gel filtration in high salt concentration (420 mM) would be helpful to answer whether the dissociation occurs during the protein extraction step or during the chromatography step.

4.6. Components of the CECR2-containing complexes

Determining the size of the CECR2-containing complexes in ES cells and testis showed different elution profiles and suggested that there could be different complexes based on their composition. The large size of the CECR2-complexes could be caused by the presence of multi-heterodimers of CECR2 and SNF2H, which are the known components of the complexes. In all of the known complexes belonging to ISWI family only one copy of each component has been reported. Another possibility is that there are components in these complexes other than SNF2H or SNF2L. Other ISWI complexes with multiple components have been isolated previously. The NURF complex in Drosophila and mouse consists of 4 interacting proteins and the CHRAC complex in Drosophila, yeast and mammals contains 4 interacting proteins (Poot et al. 2000, Barak et al. 2003, Eberharter et al. 2001, McConnell et al. 2004, Tsukiyama & Wu 1995). ISWI-D from *Xenopus Laevis* contains 5 interacting proteins with molecular weights of 200, 135, 70, 55 and 17 kDa (Guschin et al. 2000). The difference in the size of CECR2-containing complexes in ES cells and testis, 2 MDa and 1.8 MDa, may indicate the difference in the components of the complexes and tissue-specificity of the components. Identifying the components of the complexes could help us have a better idea on the function of these complexes. The presence of different non-catalytic components in ATP-dependent chromatin remodeling complexes results in different functions. Mammalian SWI/SNF like Brg1/Brm associated factors (BAF) have diverse

functions and are involved in various biological processes (Wu 2012). BAF complexes contain ~15 protein components consisting of a common core with tissue-specific components (Lessard et al. 2007). The core consists of the catalytic component, BRG1 or BRM, and a number of non-catalytic components, which are nonexchangeable. BAF complexes also contain exchangeable specificity components which change the function of the BAF complex (Brechalov et al. 2016). For example, the BAF complexes in neural stem cells (neuronal-progenitor-specific BAF or npBAF), and post-mitotic neurons (neuron-specific BAF or nBAF), share 7 nonexchangeable components but differ in three of their components, which result in the difference in their function. The npBAF complex is involved in the proliferation of neural stem cells (Lessard et al. 2007), whereas nBAF regulates genes that are involved in dendritic out-growth (Wu et al. 2007). There are also two other tissue-specific BAF complexes isolated from ES cells (Ho et al. 2009b) and cardiac progenitors (Ho & Crabtree 2010) with distinctive functions caused by their differences in their exchangeable components. Another example of combinatorial assembly of ATP-dependent complexes is the nucleosome-remodeling and histone deacetylase (NURD) complexes, which have tissue-specific components and resulting distinct tissue-specific functions (Denslow & Wade 2007).

Based on the sizes of the complexes I hypothesized that there must be other components in these complexes and probably tissue-specific components distinguishing the ES complexes and testis complexes.

4.7. Novel binding partners of CECR2

I used immunoprecipitation and LC-MS/MS analysis to investigate the other possible interacting partners of CECR2. I repeated this experiment 9 times with ES cell samples and 5 times with testis using different conditions (Appendix C). For some experiments, I only used 1 or 2 bands to check for the presence of CECR2 and SNF2H to confirm that the experiment works. Tables 3-2 and 3-3 show the list of proteins from the experiments that I analyzed all of the SDS-PAGE gel slices containing all of the CECR2 co-immunoprecipitated proteins. To make the list of the proteins all the nonspecific proteins that were also detected in the IgG control were removed from the list. I also removed the frequently detected background contaminants in mass spectrometric approaches (Mellacheruvu et al. 2013) (Table 4-1).

To create the list of the strong candidate proteins interacting with CECR2, I looked for proteins detected in more than one experiment (Tables 3-4 and 3-5). The candidate proteins from ES cells were categorized into 3 GO categories: 36% of the proteins were "nucleic binding proteins" (SNF2H, SNF2L, MATR3 and TOP2A). Other protein classes include: calcium-binding protein (CCAR2) and cytoskeletal protein (Tmod3). Unclassified proteins were LUZP1, RUVBL1, THRAP3, RIF1 and DNTTIP2. The candidate proteins from the testis were categorized in 3 groups: 45% "nucleic binding proteins" (SNF2H, MATR3, TDRD6 and ILF2), calcium-binding protein (CCAR2) and unclassified (SHCBP1L, RIF1, PABPC1 and PNN). Common to both lists are SNF2H, CCAR2, RIF1 and MATR3 while 14 proteins are unique. This suggests that the components of CECR2-containing complexes in ES cells and testis may be tissue-specific. Alternatively, some proteins may be common but below the limits of detection in one tissue or some proteins may be artifacts. Therefore, all protein interactions must be validated in both tissues.

Table 4-1 List of the protein families that are considered as background contaminants in the mass spectrometry analysis. All of these proteins were deleted from my list of proteins detected in the mass spectrometry experiments.	**Protein families (contaminants)**
	Heat shock proteins
	Keratins
	Tubulins
	Actins
	Elongation factors
	Ribonucleoproteins
	Ribosomal proteins

The most frequently detected protein for both tissues is SNF2H (also known as SMARCA5), as expected from my co-immunoprecipitation results following up on Thompson et al (2012). SNF2L (also known as SMARCA1) was the second most frequently detected protein in ES cells, again as expected from my co-immunoprecipitation results. Surprisingly, SNF2L was not detected in any of the testis samples. In fact, co-immunoprecipitation of SNF2L using the CECR2 antibody was difficult in testis, although it was confirmed in Figure 3-30A. According to my Western blot analyses SNF2H is more predominantly expressed in adult testis compared to SNF2L. Comparing the intensity of CECR2 band to SNF2L band in Western blot analysis of immunoprecipitated samples using CECR2 antibody suggests that the amounts of CECR2 and

SNF2L in input and IP lanes are not proportional (Figure 3-30A). This suggests that a very small proportion of SNF2L is incorporated in CECR2 complexes. The low amount of SNF2L immunoprecipitated with CECR2 antibody from testis samples could explain why it was not detected in mass spectrometry analyses. Interestingly, according to co-immunoprecipitation assays CECR2 only interacts with SNF2L in mouse adult ovary (Figure 3-30B).

4.7.1. LUZP1 is a novel component of CECR2-containing complex in mouse ES cells

Leucine Zipper Protein 1 (LUZP1) was a strong candidate proteins for interacting with CECR2 in ES cells because of its involvement in neural tube development. LUZP1 is a 1067 amino acid protein discovered first in a mouse bone marrow cDNA library (Sun et al. 1996). LUZP1 is detected in the neuroepithelium of mouse embryos (E9) (Hsu et al. 2008) and cerebellum of the adult rat (Sun et al. 1996). The expression of *Cecr2* and *Luzp1* in the same structures suggests the possibility of interaction between these two proteins. All *Luzp1* mutants die perinatally (Hsu et al. 2008). The main cause of death in *Luzp1* mutants is likely duo to cardiovascular defects. Interestingly, 42% of the mutants also exhibit exencephaly on the C57BL/6 strain background. This indicates that *Luzp1* is involved in neural tube closure. Exencephaly is seen in *Cecr2*^{Gt45Bic} mutant at ~69% penetrance on the C57 background (Kooistra 2009).

Reciprocal immunoprecipitation assays confirmed the interaction of LUZP1 with CECR2 and SNF2H in ES cells (Figures 3-31 and 3-32). These data suggest that these three proteins are components of the same complex. A possibility that cannot be ruled out by my data is that CECR2 could form two distinct complexes with SNF2H and LUZP1 and there is not a single complex containing all three proteins. To determine whether LUZP1 and SNF2H form a complex with CECR2, a sequential co-immunoprecipitation assay could be performed, first with the CECR2 antibody and followed by the LUZP1 or SNF2H antibody.

4.7.2. LUZP1 promotes stability of CECR2-containing complexes in ES cells.

Gel filtration assay on whole cell lysate prepared from ES cells showed that LUZP1 eluted over a very broad range of sizes (Figure 3-34). The immunoprecipitation of LUZP1 with CECR2 from the fractions corresponding to the molecular size of 1.8-2 MDa indicated the incorporation of LUZP1 in the big CECR2-containing complex. The broad elution profile of LUZP1 suggests its incorporation in other complexes with different sizes. LUZP1 has been

reported to be part of a very big complex (>1.5 MDa) called meta-coactivator complex (MECO) (Krebs et al. 2010). The MECO complex is the product of the association of two independent complexes, the Ada-Two-A-containing (ATAC) histone acetyltransferase and mediator coactivator (MED) complexes held together by a molecular bridge. The mammalian ATAC complexes alone have different sizes, varying from 600 kDa to 2 MDa (Nagy et al. 2010) and the mammalian MED complex is also an ~1.2 MDa complex with 28-30 known components (Malik & Roeder 2005, Taatjes 2010) (Malik & Roeder 2005). LUZP1 is part of the molecular bridge that connects the ATAC complex to the MED complex in mouse ES cells and the lack of LUZP1 leads to the partial dissociation of the MECO complex (Krebs et al. 2010). In *Luzp1* mutant ES cells there is a 50% decrease in level of the MECO complexes indicating that it is important for the stability of the complex in addition to other factors. I used homozygous *Luzp*-KO/lacZ-KI ES cells to reveal a decrease in the size range of the CECR2-containing complex in the absence of LUZP1 (Figure 3-34). There was no detectable CECR2 in 2 MDa fractions as we expected. The maximum size of the CECR2-containg complex in the *Luzp1* mutant ES cells is ~1.6 MDa, and the peak is approximately 1 MDa. The loss of the largest fraction could be due to a low level of dissociation seen in the wildtype ES cell control. Regardless, the larger sizes were much reduced with an increase in the smaller sizes. This suggest that the presence of LUZP1 is required for the stability of the CECR2-containing complexes in ES cells, although similar to MECO other proteins may also be involved since a small amount of big complex remained. It seems that the core part of the complex is CECR2-SNF2H, fraction 44. None of the known components of the ATAC or MED complexes were detected in the mass spectrometry analyses suggesting that LUZP1 is not mediating the interaction of one of these complexes with CECR2-containing complexes. Altogether the data indicate that LUZP1 is a novel interacting protein with CECR2 and is important for the stability of the ~1.8 MDa complex mouse ES cells.

4.7.3. CECR2 and LUZP1 may have different but overlapping roles in neural tube development

Luzp1 is expressed during embryonic development in the neuroepithelium of the neural plate and in the heart primordium (Hsu et al. 2008). I showed that CECR2 and LUZP1 expression is also detectable in neurospheres and that LUZP1 is a component of CECR2-containing complexes in neurospheres (Figure 3-36). It has been shown that failure of neural tube closure in

Luzp1 mutants is correlated with ectopic SHH expression and increased cell death level in the hindbrain (Hsu et al. 2008). SHH expression is limited to the floor plate of the neural plate in wild-type embryos during closure of the neural tube. In the *Luzp1* mutants, there is an increase in the expression of SHH in the floor plate. Also, its expression extends to the dorsolateral neuroepithelium and the head mesenchyme. Ectopic SHH signal transduction in the mid/hindbrain of the exencephalic embryos has been also shown for *Nog* mutants. In these mutants *Gli1* and *Ptch1* ectopic expression (but not SHH) was detected in cranial neural folds (Stottmann et al. 2006). In addition, the *Luzp1* mutants show an increased level of cell death in the neural tube. Elevated levels of cell death in the hindbrain and the surrounding mesenchyme has also been reported for other genetic mutations (e.g., *Cart1* (Zhao et al. 1996), *Apob* (Homanics et al. 1995), *Tcfap2a* (Schorle et al. 1996) and *Mekk4* (Chi et al. 2005)) that show failure of neural tube closure. Immunostaining of *Cecr2* wild-type and homozygous *Cecr2*$^{Tm1.1Hemc}$ mutant embryos during the time of neural tube closure using the SHH antibody did not show any difference in staining. In both genotypes SHH was detected in the floor plate (Figure 3-37 and 3-38). There was no detectable ectopic expression of SHH in the dorsolateral neuroepithelium in the *Cecr2* mutants. TUNEL staining of *Cecr2* wild-type and homozygous *Cecr2*$^{Tm1.1Hemc}$ mutant embryos during the time of neural tube closure did not show any detectable difference in the number of apoptotic cells between wild-types and mutants (Figure 3-39). However, only two embryos were examined for apoptosis in wild-type and *Cecr2* mutant embryos, so further analysis is warranted.

The exencephaly phenotype seen in both *Luzp1* and *Cecr2* mutants suggest a crucial role of both in brain development, however they do not share the ectopic expression of SHH and elevated level of apoptosis in the neural tube and the surrounding mesenchyme. In addition to exencephaly, *Luzp1* mutant mice also show complex cardiovascular malformations, omphalocele, and cleft palate. The differences in the phenotypes of *Luzp1* and *Cecr2* mutants suggest that LUZP1 is part of multiple complexes with multiple functions. Lack of LUZP1 leads to multiple defects in multiple organs. Perhaps LUZP1 functions as a component of multiple complexes during neural tube closure, which function in different aspects of the neural tube development. Lack of all LUZP1 complexes leads to failure of multiple functions, which leads to exencephaly and additional effects involving ectopic SHH and elevated apoptosis. Lack of CECR2 only impairs one of the LUZP1 complexes, which leads to exencephaly.

The genetic cause of exencephaly is very complex. Mutations in ~365 genes have been reported to lead to neural tube closure defects in mice (https://ntdwiki.wikispaces.com). In ~70% of the mutations the defect manifests only as exencephaly. The mutations in these genes result in various molecular events, including: alteration of cell death rate in the neural tube and the head mesenchyme that leads to failure of the neural fold elevation; failure of the fusion of the opposed neural folds; misregulation of actin function leading to failure of neural folds bending (Juriloff & Harris 2000). The exact cause of the exencephalic phenotype in *Cecr2* mutants is not known, although the hinges appear to form and then fall. Heterozygotes are slower to close (Dawe et al. 2011). Since the CECR2-containing complex is a chromatin remodeling factor, the mutation of this gene can affect the expression of other genes by altering the structure of chromatin. Exencephaly has been reported for other chromatin remodellers including *Smarca4 (Brg1)* (Bultman et al. 2000) and *Smarcc1* (Kim et al. 2001). To explore the mechanism underlying the exencephaly phenotype of *Cecr2* mutants, microarray analysis of the head region of exencephalic embryos (12-17 somite stage) would be very informative. The genes with altered expression levels in the mutants compared to wild-type can help to connect the exencephaly to the molecular events that are regulated by *Cecr2*.

4.7.4. LUZP1 may be a tissue-specific component of CECR2-containing complex

Interestingly, LUZP1 was not detected in any of the five mass spectrometry analyses that I performed using adult testis (Table 3-3 and 3-5, Appendix K, L, M and N) despite it being a similar size to SNF2H and SNF2L (and therefore should be present in the partial experiments as well). I showed that *Luzp1* expression is strong in adult testis as judged by Western blot analysis (Figure 3-35A, INPUT). I was never able to co-immunoprecipitate LUZP1 using the CECR2 antibody under various conditions (Figure 3-35). Preliminary immunofluorescence experiments by graduate student Alaina Terpstra supports a lack of interaction between CECR2 and LUZP1 in adult testis. CECR2 was localized to spermatogonia type A cells and there was no detectable CECR2 in other cell types in the seminiferous tubules (Figure 3-12). Interestingly, Alaina found that LUZP1 was detected in most of the cell types in the spermatogenesis cycle except for those in the periphery of the tubule, which includes spermatogonia (Figure 4-2). This suggests that LUZP1 does not form a complex with CECR2 in adult testis because the two proteins don't

colocalize. Thus LUZP1 appears to be a component of the CECR2-containing complex found in ES cells but not in testis.
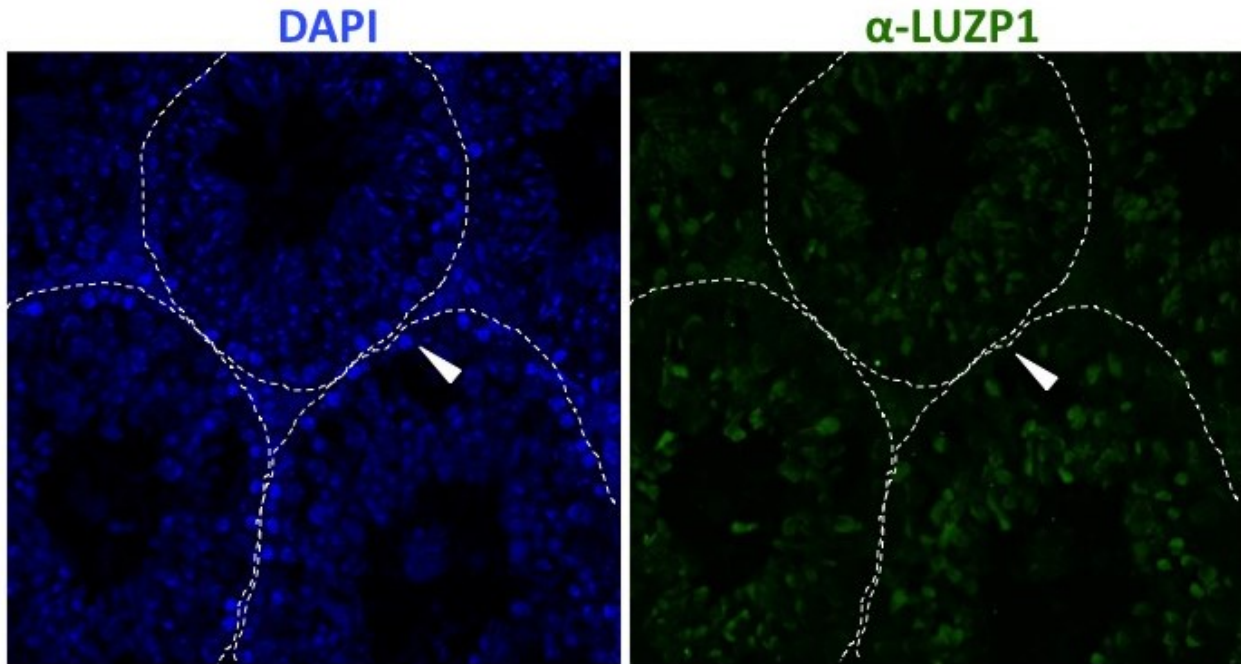
Figure 4-2 LUZP1 is not detected in peripheral cells, including spermatogonia, of adult mouse seminiferous tubules. Testis from wild-type adult mice were sectioned and the presence of LUZP1 was detected by immunoflouroscence staining using the rabbit LUZP1 antibody and counterstained with DAPI to visualize DNA. Arrowheads shows the spermatogonia located at the marginal regions of seminiferous tubule in wild-type testis. Dashed lines mark the boundary of the seminiferous tubule. (Alaina Terpstra, unpublished data).

4.7.5. CCAR2 is another novel component of CECR2-containing complexes

Cell Cycle And Apoptosis Regulator 2 (CCAR2) is another strong candidate for a novel CECR2-interacting protein. It was identified in three independent mass spectrometry analyses both in ES cells and testis (Table 3-4 and 3-5). The interaction with CECR2 was verified by reciprocal co-immunoprecipitation in ES cells, with robust reproducibility. The weak CCAR2 bands compare to the strong band for CECR2 indicates that most of the CCAR2 in the ES cells is not interacting with CECR2 (Figure 3-41). It is possible that CCAR2 is present as a single protein or incorporated into other complexes in the ES cells. CCAR2 (also known as DBC1) was first isolated from breast cancer biopsies (Hamaguchi et al. 2002). It contains a leucine zipper domain and a nuclear localization signal (Kim et al. 2008). Its leucine zipper domain is important for protein-protein interaction. It has been shown that CCAR2 physically binds to SIRT1 using its

leucine zipper domain and that deletion of this domain prevents the interaction. The CCAR2/SIRT1 complex is involved in apoptosis (Kim et al. 2008, Park et al. 2014). CCAR2 also forms a complex with BRCA1 and functions as a transcriptional repressor (Hiraike et al. 2010). CCAR2 is involved in many other cellular functions including cell cycle regulation and double-strand DNA damage repair (Tanikawa et al. 2013) and regulation of nuclear receptors function (Sakurabashi et al. 2015). SIRT1 and BRCA1 were not detected in any of my mass spectrometric analyses. The multiple functions of CCAR2 and its interaction with other proteins may explain why a small amount of CCAR2 was co-immunoprecipitated with CECR2.

While CCAR2 expression is also very strong in adult testis as judged by Western blot analysis (Figure 3-42, Input), the reciprocal co-immunoprecipitation with CCAR2 was difficult to reproduce. We tried multiple conditions and only a salt condition of 100 mM for the IP reaction (rather than 150 mM) and wash steps (rather than 420 mM) was successful. Lowering the salt concentration increases the probability of non-specific binding. Preliminary immunofluorescence indicates that CECR2 does not co-localize with CCAR2 in adult testis (Alaina Terpstra, unpublished). This suggests that the interaction between CCAR2 and CECR2 may be non-specific in the immunoprecipitation assay in the testes, and possibly a spurious post-lysis interaction. The latter could be tested by crosslinking proteins before lysis. Another possibility is that CCAR2 is not a core component of the CECR2-containing complex and its binding to CECR2 depends on unknown specific conditions. Other protein interaction methods can also be useful such as yeast-2-hybrid or BioID. BioID is a new technique that is used to show the interaction between proteins (Roux et al. 2013). The protein of interest is fused with a promiscuous biotin ligase and expressed in a cell line. The ligase biotinylates all the proteins interacting with the fusion protein. Identification of the biotinylated proteins in the cells will reveal the interacting partners of the protein of interest. The BioID advantage is the identification of weak or transient interactions. Using this method, the interacting proteins that disassociate during the lysate preparation are also identified.

I conclude that CCAR2 is a novel component of CECR2-containing complexes in mouse ES cells and probably it is not part of the complex in adult testis. CCAR2 control the expression of many genes by regulating the activity of selected chromatin remodeling enzymes. It directly binds the NAD$^+$-dependent histone deacetylase SIRT1 and prevents its binding to the substrates. It has been shown that CCAR2 hyperacetylates p53 through inactivating SIRT1, which leads to

activation of the apoptosis pathway (Zhao et al. 2008). CCAR2 also binds to other deacetylases including HDAC3 (Chini et al. 2010), HDAC5, HDAC9 and SIRT7 (Joshi et al. 2013) indicating that CCAR2 plays an important role in gene expression by interacting with the deacetylases. CCAR2 is involved in regulating transcription through its binding to nuclear receptors. CCAR2 is a component of the nuclear receptor co-regulator (NRC) interacting factor 1 (NIF-1) complex. This complex requires CCAR2 to upregulate genes by activating retinoic acid receptor (RARα) (Garapaty et al. 2009). CCAR2 functions as a transcriptional repressor by directly binding to estrogen receptor (ER)β and negatively affecting its transcriptional activity (Koyama et al. 2010). CCAR2 directly binds to Androgen Receptor (AR) in *Xenopus* oocytes and LnCAP prostate cancer cells and enhances its chromatin binding and consequently its transcriptional activity (Fu et al. 2009). These data show the importance of CCAR2 in the functioning of its interacting partners. CECR2 binds to chromatin and downregulates multiple genes (Fairbridge et al. 2010). It is possible that CCAR2 regulates the transcriptional function of the CECR2 complex.

CCAR2 is also involved in DNA repair. The siRNA-mediated depletion of *Ccar2* leads to an impaired DSB repair response in gamma irradiated U2OS cells (Tanikawa et al. 2013). CCAR2 is necessary for maintaining the genomic DNA integrity in the UV-irradiated Hs578T breast cancer cell line (Kim & Kim 2013). Depletion of CCAR2 results in the accumulation of DNA damage following UV irradiation in these cells. Similar function in DNA damage response has been reported for CECR2 (Lee et al. 2012). The siRNA knockdown of CECR2 in 293T cells leads to an impaired DSB repair following irradiation indicating a role for CECR2 in DSB repair. The interaction of CECR2 with CCAR2 suggests that CECR2 could be involved in the DNA repair response by its interaction with CCAR2. It is also possible that the interaction between these two proteins is transient and exists during the DNA damage responses. BioID assay would be helpful to investigate this interaction.

4.7.6. Other interacting candidates of the CECR2 complex

There are other candidates identified by mass spectrometry analysis that I did not test their interaction with CECR2 (Tables 3-4 and 3-5). Interestingly, there are different proteins identified for ES cells and testis. RIF1 (replication timing regulatory factor 1) is the only one of these untested candidates that was identified in both ES cells and testis samples. It was detected in three independent mass spectrometry analyses (Tables 3-4 and 3-5). RIF1 is involved in stem

cell pluripotency (Wang et al. 2006), double-stranded DNA damage repair (Feng et al. 2013), stalled replication forks repair (Xu et al. 2010) and maintaining genomic integrity (Kumar & Cheok 2014). RIF1 loss in mice results in embryonic lethality at reduced penetrance with the few survivors being less fertile compared to wild-type mice (Buonomo et al. 2009). *Rif1* mutation causes IR hypersensitivity of mouse embryonic fibroblasts (MEFs) (Chapman et al. 2013). The siRNA-mediated depletion of *Rif1* in HEK293 cells results in reduced 53BP1 foci at all time points following IR indicating a reduced efficiency of the DNA damage repair. Interestingly, Lee et al. (2012) showed similar result for *Cecr2* in 293T cells. *Cecr2* deficient cells showed reduced number of 53BP1 and γ-H2AX foci following IR, although formation of γ-H2AX foci is seen in *Rif1* deficient cells but fails to resolve. The data suggest that *Cecr2* and *Rif1* may cooperate during the DNA damage response.

The identification of different interacting proteins for ES cells and adult testis supports the idea that the CECR2-containing complexes contain tissue-specific components. From the list of the candidates in testis, TDRD6 and SHCBP1L are involved in spermatogenesis. Lack of TDRD6 results in infertility caused by affected spermiogenesis (Tanaka et al. 2011). TDRD6 is found in the chromatoid body (CB) in male germ cells and lack of TDRD6 leads to a disorganized CB (Fanourgakis et al. 2016). CBs localize close to the nucleus of spermatids and have translational repression and RNA silencing function. *Shcbp1l* gene knockout male mice show subfertility and produce reduced numbers of sperms compared to wild-type animal (Liu et al. 2014). SHCBP1L has a role in maintaining the stability of the meiotic spindle by forming a complex with heat shock 70 kDa protein (2HSPA2) in male germ cells. Disordered chromosome arrangement is caused during meiosis by Shcbp1l mutation, which leads to increased number of apoptotic cells in seminiferous tubules. Although TDRD6 and SHCBP1L show similar fertility abnormalities to CECR2, their expression in testis is different and none of them has been detected in spermatogonia (Liu et al. 2014, Vasileva et al. 2009).

A candidate protein in ES cells is TOP2A (Topoisomerase II Alpha) that was detected in four independent mass spectrometry analyses (Table 3-4). TOP2A is involved in DNA decatenation and chromosomal segregation (de Campos-Nebel et al. 2010). It has been shown that TOP2A interacts with the mammalian ATP-dependent chromatin remodeling BAF complex. In order for TOP2A to bind to its 12,000 chromatin binding sites in the genome, an interaction with the BAF complex is required (Dykhuizen et al. 2013), suggesting BAF is involved in

decatenating DNA. A confirmed interaction of CECR2 with TOP2A would suggest that CECR2 may also be involved in decatenating DNA.

4.8. Identifying binding targets of CECR2-containing complexes using ChIP-Seq

CECR2 has been shown to be required for normal neural tube closure and reproduction (Banting et al. 2005, Thompson et al. 2012). CECR2 is also involved in γ-H2AX formation and DSB repair, at least in HEK 293 cells (Lee et al. 2012). Similar to other ATP-chromatin remodellers, we hypothesize that CECR2 is involved in the regulation of transcription of many genes. The *Cecr2* mutation results in significant reduction of *Alx1*, *Dlx5*, *Ncapd2*, *Six1, Epha7*, *Eya1*, and *Lix1* mRNAs. From these genes, *Alx1*, *Dlx5* and *Epha7* have important roles in neural tube closure and cause NTDs when mutated (Acampora et al. 1999, Holmberg et al. 2000, Zhao et al. 1996). *Six1* and *Epha7* mutants share inner ear abnormality with *Cecr2* mutants (Bosman et al. 2009, Dawe et al. 2011, Zou et al. 2008). The abnormal phenotypes seen in *Cecr2* mutants may be associated with the deregulation of these genes. However, it is not known if CECR2 regulates the expression of these genes directly by binding to their regulatory elements. It is possible that CECR2 regulates the expression of these genes indirectly through other genes. The non-catalytic components of the ATP-dependent chromatin remodeling complexes play an important role in targeting these complexes to chromatin. I have also showed that CECR2-containing complexes have tissue specific components such as LUZP1. To investigate general and tissue-specific binding sites of the CECR2 complexes, I identified overlapping binding targets of CECR2, SNF2H and LUZP1 in ES cells and CECR2 and SNF2H in testis.

ChIP-seq analysis makes it possible to map the genome wide binding sites of chromatin-binding proteins. Using my CECR2 antibody I was able to identify putative gene regulatory sites associated with brain development and reproduction. This thesis presents genome wide analysis of binding targets of the CECR2-containing complexes in ES cells and adult testis. Determining the position of the binding sites relative to TSS of genes in ChIP-seq data sets for three proteins showed that only a small proportion of the binding sites are close to TSSs and within promoters, and the majority of the binding sites for all proteins are found at long distances from annotated genes inside the intergenic regions or inside gene bodies (exons and introns) in both ES cells and testis. Since the CECR2 binding sequences show a higher conservation score near the summit (Figure 3-44), it suggests that the CECR2-containing complexes bind to cis-regulatory elements

that exert their regulation on promoters from long distances. Another possibility is that these binding regions are regulatory elements of nearby genes that have not been annotated. Some of the sites could also be non-specific background. The finding that the majority of the binding sites of CECR2-containing complexes are located at large distances from TSSs indicates that to come up with accurate conclusions on the effect of the CECR2 complexes on transcription, the investigations should not be restricted to promoter regions.

Firstly, the ChIP-seq data is required to be verified by ChIP-QPCR. None of the known genes that have been confirmed to be downregulated in *Cecr2* mutants contain the CECR2 complex binding sites in my datasets. This indicates that they are missed in my experiment or they are secondary effects rather than direct one. To confirm directly whether CECR2 distal binding sites are indicative of enhancer elements and identifying the affected transcripts *in vivo*, a large-scale enhancer reporter analysis can be performed. To do this, the selected regions of the genome can be amplified and cloned into enhancer reporter vectors. By screening in appropriate cell lines followed by sequencing of the screened genomic regions one can show whether the binding region is a cis-regulatory region. Secondly, I performed ChIP-seq on ES cells and testis and the data indicate that some of the CECR2 complexes occupy different genomic regions in ES cells and testis. It is likely that the chromatin activity of the CECR2 complexes affect different set of genes in ES cells and embryonic neural tube tissue. Comparison of transcript levels of the candidate genes in ES cells following CECR2 knockdown and wild-type ES cells can be used to confirm the regulation of these genes by CECR2 in ES cells.

4.9. Possible functions of the CECR2 complexes revealed by GO term analysis

GO term analysis is used to associate the identified genes in large-scale experiments with known biological processes. Overall, the GREAT analysis suggested that the putative cis-regulatory binding targets of CECR2 may regulate a wide variety of biological processes in both ES cells and testis (Tables 3-7 and 3-16). The genes bound by CECR2 in ES cells are enriched for several terms involved in brain development. GREAT did not associate any brain development GO term for the CECR2 peaks in testis. The top Gene Ontology Biological Process enrichments for the CECR2 binding targets in ES cells include metencephalon development (51 genes) and cerebellum development (44 genes). Other GO terms related to the nervous system

development are cerebellar cortex development (25 genes), hindbrain morphogenesis (23 genes) and regulation of timing of neuron differentiation (5 genes). Several genes associated with the brain development GO terms cause exencephaly when mutated (Table 4-2). This suggests that CECR2 in involved in the central nervous system development by regulating the expression of these genes by binding to their regulatory regions. Many of these genes have a role in brain development after neural tube closure. Although a specific defect has not been detected in the brains of mice that close their neural tubes, we cannot rule out undetected subtle effects. GO analysis of the overlapping binding sites of CECR2 and SNF2H in ES cells also associated the target genes with the brain development (19 genes) (Table 3-9). Three of the five GO terms defined for the overlapping binding sites of CECR2 and SNF2H are related to brain development in ES cells and include the genes *Gli2*, *Sema4c*, *Sec24b*, *Shh* and *Map3k7*. All these genes contain overlapped binding sites for CECR2 and SNF2H and result in exencephaly or open neural tube in mice when mutated. Visualizing the peaks using Integrative Genomics Viewer (IGV) showed that the LUZP1 binding site also overlaps with CECR2 and SNF2H binding sites in several of the genes associated with GO terms related to the brain development including *Gli2* and *Sec24b* in ES cells. Although I have not yet validated any of the binding sites and have not shown that *Cecr2* regulates the expression of the genes, I will go through two major pathways for further study.

Table 4-2 Genes bearing CECR2 binding sites and associated with the brain development GO terms in ES cells that cause exencephaly when mutated.

Gene	Peak genomic region (distance to TSS)
Dlc1	chr8:37976198-37976698 (+25239), chr8:37890311-37890811 (+111126), chr8:37641833-37642333 (+359604), chr8:37553859-37554359 (+447578)
Gli2	chr1:121263271-121263771 (-313325), chr1:121224795-121225295 (-274849), chr1:121192633-121193133 (-242687), chr1:121144839-121145339 (-194893), chr1:121065380-121065880 (-115434), chr1:121060504-121061004 (-110558), chr1:121039145-121039645 (-89199), chr1:121027929-121028429 (-77983), chr1:120993910-120994410 (-43964), chr1:120834325-120834825 (+115621), chr1:120719709-120720209 (+230237), chr1:120672722-120673222 (+277224), chr1:120651819-120652319 (+298127), chr1:120648485-120648985 (+301461), chr1:120608845-120609345 (+341101)
Hes1	chr16:29697284-29697784 (-367892), chr16:30135498-30135998 (+70322), chr16:30192036-30192536 (+126860)
Hes3	chr4:151670209-151670709 (-4688)
Lhx1	chr11:84464837-84465337 (-125570), chr11:84440123-84440623 (-100856), chr11:84424059-84424559 (-84792), chr11:84419853-84420353 (-80586), chr11:84360709-84361209 (-21442), chr11:84347541-84348041 (-8274)

Lrp6	chr6:134566192-134566692 (-49511), chr6:134524166-134524666 (-7485), chr6:134520334-134520834 (-3653), chr6:134507188-134507688 (+9493), chr6:134493669-134494169 (+23012), chr6:134385101-134385601 (+131580)
Myh10	chr11:68647781-68648281 (+142213)
Otx2	chr14:49608344-49608844 (-321406), chr14:49081143-49081643 (+205795)
Ptprs	chr17:56497254-56497754 (+118402)
Rfx4	chr10:84314890-84315390 (+96333)
Rpgrip1l	chr8:93706988-93707488 (+129923)
Sec24b	chr3:129680019-129680519 (+83556)
Sema4c	chr1:36664248-36664748 (-49304), chr1:36654110-36654610 (-39166), chr1:36653260-36653760 (-38316)

4.9.1. Shh signaling pathway

Gli2 is a component of Shh pathway, which has been associated with neural tube defects (Murdoch & Copp 2010). Activation of the Shh signaling pathway leads to activation of *Gli2*, which functions as a transcription activator and a weak repressor (Pan et al. 2006). Over activation of the Shh pathway leads to NTDs in mice. Loss of function or gain of function of nine genes have been shown to cause exencephaly through increasing the activation of the Shh pathway (Murdoch & Copp 2010). It is possible that the CECR2 complex function as a transcription repressor of Gli2 and lack of the CECR2 protein removes this repression, which results in an increase in *Gli2* expression. The increased amount of the Gli2 protein could be responsible for the neural tube defect seen in the *Cecr2* mutant mice. *Gli2* is located downstream of Shh in the signaling pathway and its expression is downregulated during neural tube patterning (Cohen et al. 2015). *Gli2^{P1-4}* contains a mutation in its PKA sites, which doubles its stability compared to wild-type Gli2 protein (Pan et al. 2006). *Gli2^{P1-4}* homozygous embryos exhibit exencephaly, enlarged craniofacial structures, and polydactyly. Ectopic expression of *Shh* in the dorsolateral neural fold in exencephalic embryos has been observed in 42% of homozygous *Luzp1* mutant mice (C57BL/6J strain) (Hsu et al. 2008). My immunofluorescence experiment did not show any ectopic expression of *Shh* in homozygous *Cecr2* mutant mice (Figure 3-37). Although the data suggest that LUZP1 functions independent of CECR2 in regulating *Shh* expression during neural tube closure, it does not rule out the involvement of the CECR2 complexes in the Shh signaling pathway. The newly characterized *Cecr2* mutants, *Cecr2$^{tm2a(EUCOMM)Hmgu}$*, besides exencephaly and coloboma exhibit postaxial polydactyly, which is typical of Shh signaling involvement. Determining the Gli2 protein level using

immunofluorescence during neural tube closure in homozygous *Cecr2* mutant embryos would show whether CECR2 regulates expression of *Gli2* by binding to its cis-regulatory elements.

4.9.2. Wnt signaling

Examination of the molecular function categories identified two molecular function GO terms for CECR2 binding sites in ES cells: Wnt-protein binding and Wnt-activated receptor activity (Table 3-7). The data suggest that a cluster of genes with CECR2 binding sites inside their regulatory regions are connected to the Wnt signaling pathway, indicating that CECR2 has an important role in regulating this signal transduction pathway. Additionally, focusing on the CECR2 peaks that were bound to a close distance to the TSS of the genes (≤1000 bp upstream) in ES cells showed enrichment for the genes in the Wnt signaling pathway (Table 3-11). This pathway also was one of the signal pathways that PANTHER GO analysis showed for the genes bearing the overlapping peaks of CECR2 and SNF2H in their promoter region in ES cells (up to 5 kb upstream of TSS). Interestingly, GO analysis of the genes with overlapping peaks of CECR2 and SNF2H close to TSSs (up to 5 kb upstream of a TSS) in testis also showed enrichment for Wnt signaling pathway genes (P00057) (Table 3-21). The Wnt signaling pathway was the strongest hit according to the number of genes associated with the pathway.

Wnt signaling pathways are involved in regulation of various biological processes, including cell proliferation, cell polarity and programmed cell death, These pathways play a crucial role in embryonic development of various systems including the nervous system (Ciani & Salinas 2005). Wnt signaling pathways have been categorized into two groups, canonical and non-canonical (van Amerongen & Nusse 2009). Non-canonical signaling is involved in modulation of the cytoskeleton. This pathway is also called the planar cell polarity (PCP) pathway since it mediates cell polarity in epithelia (Ciani & Salinas 2005). The PCP pathway also regulates convergent extension of gastrulating embryo and the orientation of the sensory hair cells in the inner ear. The PCP pathway has been connected to neural tube defects in numerous studies (Greene & Copp 2014). Most PCP genes in mice cause craniorachischisis, the more severe form of NTD, when mutated. However, exencephaly has also been reported for some of the PCP genes. Combination of *Vangl2$^{Lp/+}$* with *Dvl3$^{+/-}$* leads to exencephaly in 9% of the mutant mice (Etheridge et al. 2008). Fifty six percent of *Vangl2$^{Lp/+}$;Fzd1$^{+/-}$* and 46% of *Vangl2$^{Lp/+}$;Fzd2$^{+/-}$* mutant mice exhibit exencephaly (Yu et al. 2010).

The *Cecr2*^{tm1.1Hemc} deletion mutation results in exencephaly in 100% of BALB/c mice (Fairbridge et al. 2010). These mutants also display disorientation of the sensory hair cells in the inner ear (Fairbridge et al. 2010). These phenotypes are similar to the phenotypes that have been reported for the PCP pathway genes, suggesting that CECR2 may be involved in modulation of PCP pathway genes. A genetic interaction between *Cecr2* and *Vangl2* has been shown previously, although this interaction was only confirmed for NTD occurrence and not the inner ear abnormality (Dawe et al. 2011). On the other hand, qRT-PCR analysis didn't show any change in the expression level of PCP pathway genes in the homozygous *Cecr2*^{tm1.1Hemc} embryos head during the neural tube closure (11–14 somite). However, there was a significant decrease in the *Lix1* transcript in post-closure embryos (16-18) and the *Frzb* transcript in the body of the embryos (Dawe et al. 2011). Although qRT-PCR analysis didn't show any detectable change in the expression of PCP pathway genes during the neural tube closure in mouse, the change in the localization of the proteins were not investigated (Leduc et al. 2016). My ChIP data, nevertheless, show enrichment of the CECR2 complex on the regulatory regions of the PCP pathway genes. Appendix P shows the list of PCP pathway genes with CECR2 binding sites in their regulatory regions. Visualization of the genomic regions of these genes using IGV revealed that some of these genes contain the overlapping binding sites of CECR2 with SNF2H and LUZP1: *Fzd10* (P value 10^{-3}), *Fzd9* (P value 10^{-5}), *Fzd7* (P value 10^{-5}), *Ptpro* (P value 10^{-3}) and *Ror2* (P value 10^{-3}). The *Lrp6* genomic region contains only overlapping binding site for CECR2 and SNF2H.

According to the ChIP-seq data, CECR2 is probably involved in regulating PCP pathways in ES cells and testis. The ChIP-seq data should be verified by ChIP-qPCR using neurulating neural tube. Then, the distal cis-regulatory elements that are bound by CECR2 and their effect on associated PCP pathway genes should be confirmed by the enhancer reporter analysis.

4.10. CECR2 binds to the promoter region of the genes associated with CNS development

Comparing the genes containing a binding site for the CECR2 complex with the Mouse Genome Informatics (MGI) database list of the genes loci associated with neural tube defects revealed several genes in both the ES cell and testis dataset (Table 3-12 and 3-20). *Hsd17b2*, *Lpar1* and *Nf1* are genes that contain the overlapping binding sites for CECR2 and SNF2H in ES

cells. Using a higher P value (10^{-3} instead of 10^{-5}) for peak calling with LUZP1 showed that this protein also overlaps with the two other proteins in the promoter region of *Hsd17b2* and *Lpar1*. None of these genes were detected using the testis dataset. The same analysis using testis dataset showed that *Lrp6* and *Phactr4* are the genes that result in neural tube defect when mutated. Interestingly, when I used the same analysis with ES cell with a higher P value (10^{-3} instead of 10^{-5}), there was an overlapping binding site of CECR2 and SNF2H in the promoter region of both genes (Figure 3-51).

Hsd17b2

Hsd17b2 contains the binding site of the three proteins just upstream of its TSS in ES cells (-319; chr8:120225277-120225777). 17-beta-dehydrogenase 2 (HSD17B2) converts estradiol, testosterone, and dihydrotestosterone to their less-active forms (Rantakari et al. 2008). Lack of *Hsd17b2* expression results in lethality of 76% of mice between E9 and first postnatal day. The embryos show abnormalities in placenta, brain and kidney. Some of the mutants exhibit exencephaly and the ones that reach adulthood exhibit progressive hydrocephalus. The mechanism of *Hsd17b2* involvement in the brain development is not known, but it is speculated that it functions through regulating sex steroids. Interestingly, *Hsd17b2* has been shown to play a role in spermatogenesis independent of sex steroids (Zhongyi et al. 2007). Transgenic mice expressing human *HSD17B2* exhibit growth retardation and male infertility. At early adulthood, 50% of males are subfertile and 50% are infertile. Later in life all of the males became infertile, suggesting a progressive degenerative process. The mutants did not show any difference with the wild-type mice in their sex steroids. Retinoic acid receptor agonist is able to partially rescue the disrupted spermatogenesis in the transgenic mice, indicating that HSD17B2 has a role in retinoic acid signaling. CECR2 could be involved in both brain development and spermatogenesis by regulating the expression of *Hsd17b2*.

Lpar1

Lpar1 contains the binding site of the CECR2 complex 2,592 bp upstream of the TSS in ES cells (chr4:58568571-58569071). *Lpar1* encodes the lysophosphatidic acid receptor 1 (LPA$_1$), which is one of the cell surface receptors for lysophosphatidic acid (LPA) (Yung et al. 2014). This receptor is one of the type I, rhodopsin-like G protein-coupled receptors (GPCRs) that is widely expressed in adult mice. Activation of LPA$_1$ leads to activation of three types of G proteins (G$_{\alpha i/o}$, G$_{\alpha q/11}$ and G$_{\alpha 12/13}$) which subsequently induce activation of several pathways

including mitogen-activated protein kinase (MAPK), phospholipase C, Akt, and Rho pathways (Choi et al. 2010). When activated, LPA_1 generates a range of cellular events including: cell proliferation, cell migration and cytoskeletal changes, and Ca^{2+} mobilization (Choi et al. 2010). Disruption of *Lpar1* results in perinatal lethality of 50% of offspring in a mixed background (C57Bl/6J x 129) (Contos et al. 2000). The mutants exhibit defective suckling behaviour caused by olfactory defects, craniofacial dysmorphism with blunted snouts, and elevated levels of apoptosis in sciatic nerve Schwann cells. A small percentage of mutant embryos (3 of 61) exhibit exencephaly. Comparing the transcription level of *Lpar1* in ES cells following CECR2 knockdown and wild-type ES cells will show whether the CECR2 complex regulates its expression. Regulation of *Lpar1* with the CECR2 complex will link the complex to G proteins.

Lrp6

LRP6 is the only Wnt signaling gene that contains the overlapping binding sites of CECR2 and SNF2H in both ES cells (-3,653; chr6:134520334-134520834 and -7,485; chr6:134524166-134524666) and testis (-3,349; chr6:134520030-134520530)(Figure 3-52). The binding site for ES cells and testis also overlap. LUZP1 does not have any peaks overlapping with the CECR peak in ES cells in the *Lrp6* promoter region. Testis CECR2 complex doesn't contain LUZP1. Low-density lipoprotein receptor related protein 6 (*Lrp6*) is an essential component of the canonical Wnt signaling pathway (Angers & Moon 2009). Binding of Wnt to LRP5/6 and the Frizzled (FZ) receptor complex starts a downstream signal transduction cascade that leads to stability of β-catenin and subsequent activation of target gene transcription (Ciani & Salinas 2005). Both homozygous $Lrp6^{Cd/Cd}$ (a gain of function mutation) and $Lrp^{-/-}$ gene-trap (a loss of function mutation) embryos exhibit exencephaly (Carter et al. 1991, Pinson et al. 2000). In addition to exencephaly, the *Lrp6* homozygous null mutants display ocular coloboma, open fetal eyelids, cleft lip and palate and limb defects including oligodactyly and polydactyly (Zhou et al. 2009). Open fetal eyelids are also found in *Cecr2* mutants (Banting et al. 2005). Interestingly, coloboma and polydactyly have been seen in our newly characterized *Cecr2* mutation, $Cecr2^{tm2a(EUCOMM)Hmgu}$ (preliminary unpublished data). In addition to my ChIP-seq data, the similarity in the phenotypes associated with the two genes makes *Lrp6* a good candidate gene for being directly regulated by *Cecr2*. Although *Lrp6* is a component of the canonical Wnt signaling pathway, its mutation also results in compromised non-canonical Wnt signaling in mice indicating that *Lrp6* is involved in neurulation through the non-canonical Wnt signaling pathway

(Gray et al. 2013). *Lrp6* modulates the PCP pathway by altering RhoA activity, which is a known component of the PCP pathway. Elevated and suppressed RhoA activity during neural tube closure has been detected in *Lrp6*$^{Cd/Cd}$ (a gain of function mutation) and *Lrp6*$^{-/-}$ gene-trap (a loss of function mutation) embryos, respectively. Interestingly, *Lrp6* does not alter RhoA activity by changing its transcription level (Gray et al. 2013). It is possible that CECR2 modulates the PCP pathways through regulating *Lrp6* transcription and subsequently impacts neurulation by the RhoA-dependent mechanism that has been shown for *Lrp6*.

Phactr4

Phactr4 contains the binding sites of CECR2 and SNF2H in both ES cells (-4,976; Chr4:131983006-131983669) and testis (-3,262; chr4:131981506-131981741). The binding site in ES cells does not overlap with the binding site in testis (Figure 3-51). Phosphatase and actin regulator 4 (PHACTR4) is a member of Phactr family of proteins that function by interacting with actin and protein phosphatase 1 (PP1) (Huet et al. 2013). Using fluorescence *in situ* hybridization (FISH) it has been shown that *Phactr4* is expressed in the central nervous system during embryonic development with the strongest expression in the ventricular zone (VZ) of developing forebrain, basal ganglia, and inferior colliculus (Kim et al. 2012). Its expression continues in the postnatal brain in the lateral ventricular wall, dentate gyrus (DG) of the hippocampal formation, olfactory bulb and cerebellum. The neural stem cells are located in the ventricular zone suggesting that *Phactr4* similar to *Cecr2* is expressed in neural stem cells. *In vitro* immunocytochemistry experiments have shown that PHACTR4 is detectable in neurospheres isolated from the subventricular zone tissue (Cho et al. 2014). In the neural stem cells, PHACTR4 binds to intermediate filament proteins including nestin and GFAP. *Humdy* mice have a missense mutation in *Phactr4* that disrupts its ability to interact with PP1 (Kim et al. 2007). The *Humdy* mutation results in embryonic lethality of most of the homozygous animals by E14.5. The rest of the embryos die shortly after birth. All of the mutants exhibit exencephaly and retinal coloboma. The neural tube defects in *Humdy* mice have been linked to increased cell proliferation in the neuroepithelium. PHACTR4 binding leads to activation of protein phosphatase 1, which controls the cell cycle by dephosphorylating and activating Rb. In the absence of PHACTR4, Rb becomes inactive and consequently E2F protein and its targets are activated resulting in excessive cell proliferation and exencephaly. It is possible that *Cecr2* regulates cell cycle by controlling the expression level of *Phactr4*. To investigate this hypothesis

the MCM2 expression level could be compared in *Cecr2* mutant and wild-type neural tube. MCM2 is an E2F target and shows a significant elevated expression in *humdy* neural tube and retina.

4.11. Fertility

Studying fertility in non-exencephalic BALB/c *Cecr2*[Gt45Bic] mutants has shown that the homozygous adults have smaller testes compared to the wild-type animals and the mutant males display subfertility (Thompson et al. 2012). Histological analyses of the adult testis showed severe morphological abnormalities in seminiferous tubes in the mutants (Kacie Norton, unpublished data). Testis is one of the few adult tissues expressing *Cecr2*. I previously showed that the CECR2 complex in testis does not contain LUZP1, which is a component of the complex in ES cells. I performed ChIP-seq data using CECR2 and SNF2H antibodies to identify the binding sites of the CECR2 complex in this tissue. Finding the genes directly regulated by CECR2 is essential to understand how this protein regulates fertility. Interestingly, functional classification of CECR2 peaks in testis based on gene ontology using GREAT revealed that the greatest enrichment was in genes involved in the "acrosome reaction" with 30 region hits related to 11 genes (Table 3-16). The list of the genes associated with "acrosome reaction" GO term is shown in Table 3-17. All of the genes have been linked to reproduction. The only gene associated with this GO term that contains the overlapping peak of both CECR2 and SNF2H in testis is *Spaca7*. Visualization of the peaks in IGV showed that CECR2, SNF2H and LUZP1 peaks in ES cells also overlap with this genomic region, which is located at +22583 from TSS. This could be a regulatory region of the gene that should be confirmed. Sperm acrosome associated 7 (SPACA7) is only expressed in adult testis, released during the acrosome reaction and facilitates fertilization (Nguyen et al. 2014). The mechanism of SPACA7 function is unknown.

There are also reproduction related genes with overlapping binding sites for CECR2 and SNF2H in the ES cells dataset that GREAT has associated with "biological process" GO terms (Table 3-9). *Cyp11a1*, *Lmx1a*, *Cdkn1b*, *Csf2* and *Fmn2* are genes from ES cell dataset that result in fertility defects in mice, when mutated. The binding site of CECR2 and SNF2H on the *Lmx1a* regulatory region also overlaps with a LUZP1 binding site. *Lmx1a* null mutation causes sterility, deafness and neurological abnormalities (Chizhikov et al. 2006). The adult mutant mice do not

have any hair cells in the inner ear and show severe abnormalities in the organ of Corti (Nichols et al. 2008).

GO terms of the genes with CECR2 binding sites in their promoter region (up to 5 kb upstream of TSS) showed that CECR2 is involved in different biological events including embryonic development and reproduction in both ES cells and testis. I only focused on the genes that contained overlapping binding site for at least both CECR2 and SNF2H. Several genes were identified that cause fertility abnormalities in mice, when mutated. None of the genes related to reproduction were shared in ES cells and testis, suggesting the tissue specificity of the CECR2-containing complexes in regulating the genes. *Elmo1*, *Fgfr4*, *Ggt1*, *Insr*, *Itgb3* and *Schip1* are the genes in the ES cell list that have been reported to be involved in reproduction. *Fgfr4* and *Insr* also contain an overlapping binding site of LUZP1 when applying higher P value for peak calling step (10-3 instead of 10-5). Analyzing the overlapping peaks of CECR2 and SNF2H in the testis dataset also revealed several genes related to reproduction containing a shared peak in their promoter region. *Cdc14b*, *Nfia*, *Pcsk1* and *Styx* are the genes in testis that, in addition to *Phactr4*, produce fertility phenotype in mice.

Fgfr4

Fibroblast growth factor receptor 4 (*Fgfr4*) is one of a family of highly conserved tyrosine kinase receptors (Cotton et al. 2008). These receptors are components of the FGF signaling pathway. Activation of the receptor by fibroblast growth factors results in initiation of intracellular signaling events that eventually activates other well-known pathways including the Ras-dependent mitogen-activated protein kinase (MAPK), the phosphatidylinositol 3-kinase (PI3K) and the signal transducer and activator of transcription (STAT) signaling pathways (Turkington et al. 2014). The FGF signaling pathways regulate a variety of cellular processes including proliferation, apoptosis, differentiation and cell adhesion (Cotton et al. 2008) and therefore play important roles in embryogenesis. The $fgfr-3^{-/-}$ $fgfr-4^{-/-}$ double homozygous mutant mice are infertile, although either of single mutants do not show infertility (Weinstein et al. 1998). Immunohistochemical analysis in adult human testis showed that *fgfr-3* is mainly expressed in spermatogonia and *fgfr-4* is detectable in Sertoli cells, all germ cells and spermatids (Saucedo et al. 2015). In mice, FGFR4 is detected during testicular development and postnatal stages in testis (Lai et al. 2016). The protein is detected in Leydig cells and the seminiferous tubules at E14-17. Its expression is limited to spermatogonia and Leydig cells at P10 and its

expression is detected in the whole testis in older mice (P20, 35, and 65). An overlapping binding site of CECR2, SNF2H and LUZP1 was detected in the *Fgfr4* promoter region (up to 5 kb upstream of TSS) in ES cells suggesting that the CECR2 complex may be involved in regulating the expression of this gene directly and subsequently affect reproduction through modulating the FGF signaling pathway and its downstream pathways.

Insr

Insr encodes the insulin receptor (IR), which is a receptor tyrosine kinase and is found in most tissues of the body (Siddle 2012). Binding of insulin or insulin-like growth factor (IGF) leads to a structural conformation change of the IR receptor that results in phosphorylation and activation of downstream pathways including phosphoinositide- 3 kinase (PI3K)/Akt and the Ras/mitogen-activated protein kinase (MAPK) pathways (Tatulian 2015). By regulating these pathways, IR modulates diverse cellular processes including gene expression, glucose, lipid and protein metabolism, cell growth and differentiation. Mice lacking the insulin receptor gene die within hours after birth caused by diabetic ketoacidosis (Accili et al. 1996). The mutants do not exhibit any abnormal embryonic development except for a fatty degeneration of the liver. CNS-specific disruption of *Insr* using the Cre-loxP system resulted in reduced fertility in both sexes caused by defective spermatogenesis and ovarian follicle maturation (Brüning et al. 2000). The mutants also exhibited obesity, mild insulin resistance and hypertriglyceridemia. The impaired spermatogenesis and ovarian follicle maturation were the result of a reduction of circulatory luteinizing hormone (LH) caused by hypothalamic dysregulation of LH releasing hormone (LHRH) secretion. Reproductive hormone levels have not been investigated in *Cecr2* mutant mice. Since my ChIP-seq data show that the CECR2 complex may regulate the expression of *Insr*, it is possible that the *Cecr2* mutants produce an abnormal level of these hormones. The serum concentration of testosterone, LH and follicle-stimulating hormone (FSH) can be measured using immunofluorometric assays in the *Cecr2* mutant mice to find out whether the lack of CECR2 affects fertility through the reproductive hormones.

Elmo1

Elmo1 is widely expressed in many tissues (Gumienny et al. 2001). Loss of ELMO1 protein does not result in gross abnormalities (Elliott et al. 2010). The only severe defect was detected in testis. The cellular organization of the seminiferous epithelium of the mutant mice was strikingly disorganized. The presence of syncytial giant cells in the seminiferous epithelium

was easily detected in histological preparations. The same disorganization and giant cells have been observed in the *Cecr2* mutants as well. *Elmo1* has an important role in clearance of apoptotic germ cells from the seminiferous epithelium by Sertoli cells. The phenotype of the seminiferous epithelium in *Elmo1* mutants is very similar to what we have seen in *Cecr2* mutants. Although my ChIP-seq data revealed a binding site for the CECR2 complex only in ES cells, the similarity of the seminiferous epithelium in *Elmo1* and *Cecr2* mutants makes it worth investigating as a candidate gene that is regulated by CECR2 directly.

Cdc14b

Cell division cycle homolog 14B (CDC14B) is a phosphatase that is involved in different cellular processes including microtubule bundling and stability and the G2/M DNA damage check- point (Wei et al. 2011). It has been shown that this protein is required for DNA repair and lack of CDC14B in mouse embryonic fibroblasts leads to defects in repairing DNA double-strand breaks induced by ionizing radiation (Lin et al. 2015). Human CDC14B plays an important role in mitotic progression through regulation of Cdk1/cyclin B activity (Tumurbaatar et al. 2011). Knock down of human CDC14B using siRNA leads to increased Cdk1 activity and eventually mitotic arrest that results in cell death. *Cdc14b* deletion mice display premature aging and subfertility in both sexes (Wei et al. 2011). Depletion of CDC14B in oocytes using RNAi leads to premature meiotic resumption indicating that CDC14B regulates the meiotic cell cycle in mouse oocytes (Schindler & Schultz 2009). Overall, the data indicate that this protein serves as an important regulator of the cell cycle. There is some evidence indicating that CECR2 is involved in DNA repair, at least in HEK 293 cells (Lee et al. 2012). It is possible that CECR2 functions in the cell cycle through regulation of the expression of *Cdc14b* in the testis. CECR2 and SNF2H bind to 1382 bp upstream of *Cdc14b* in testis suggesting that the CECR2 complex may regulate its expression directly by binding to its promoter. Comparing the transcription level of *Cdc14b* in *Cecr2* mutant and wild-type testis would reveal whether *Cecr2* is involved in regulation of this gene and subsequently in regulating cell cycle.

Taken together, the ChIP-seq data showed that the CECR2 complex could be involved in different aspects of reproduction through regulating different genes and signal transduction pathways. My candidate genes participate in various pathways that control a broad range of cellular processes including cell cycle, apoptosis, cell proliferation and differentiation and cell adhesion. Spermatogenesis and oogenesis are very complex processes and many genes are

involved in controlling these processes during different stages. Some genes regulate spermatogenesis and oogenesis directly by being expressed in testis and ovary. The CECR2 complex can regulate the expression of these genes directly by binding to their regulatory regions or through another gene indirectly. It is also possible that the CECR2 complex regulates fertility indirectly through reproductive hormones by regulating the expression of genes such as *Insr* in the CNS that control the hypothalamic-pituitary-gonadal axis.

4.12. Novel functions

GREAT analysis of CECR2 binding sites also associated this protein to biological process GO terms that had never been reported for CECR2. For example, there are three GO terms related to the heart development and a term related to kidney development in ES cells (Table 3-7). There is evidence of involvement of CECR2 in kidney development in the FVB *Cecr2*$^{Tm1.1Hemc}$ mutants, including absent kidneys and double ureters (Fairbridge 2013). The role of CECR2 in heart development has not been experimentally verified. I have shown that CECR2 is not detectable in adult heart by Western blot analysis, however that does not exclude CECR2 from functioning in the early stages of heart development. On the other hand, it has been shown that LUZP1, which is a component of the CECR2 complex in ES cells, has an important role in heart development. The main cause of lethality in *Luzp1* mutants is the cardiovascular deficits (Hsu et al. 2008). Ventricular septal defect (VSD), double outlet of right ventricle (DORV) and transposition of great arteries (TGA) are heart defects that have been observed in *Luzp1* mutants. Heart and kidney defects are two of characteristic features of cat eye syndrome (Schinzel et al. 1981). Interestingly, among the genes that are associated with heart development GO terms are the genes bearing overlapping binding sites of CECR2 and LUZP1 in their distal cis-regulatory regions in the ES cell dataset including *Notch1*, *Tgfbr2*, *smad1*, *Ctnnb1*, *Myocd* and *Prkar1a*. Figure 4-3 shows the overlapping binding sites of CECR2 and LUZP1 in some of these genes. SNF2H also occupies the same genomic region as CECR2 and LUZP1 in the regulatory region of *Notch1* and *Smad1*, indicating that these proteins bind to these genomic regions as components of the CECR2-containing complex. *Notch1* is one of the components of the Notch pathway that plays a crucial role during mammalian cardiogenesis (High & Epstein 2008). The mice with a conditional mutation of *Tgfbr2* in neural crest stem cells die perinatally displaying cardiovascular

defects including a truncus arteriosus, VSD and abnormal patterning of the arteries arising from the aortic arch (Wurdak et al. 2005). These mutants also display mid/hindbrain abnormalities (Falk et al. 2008). Loss of TGFBR2 in the mid/hindbrain leads to expansion of the mutant isthmal and inferior tectal neuroepithelium starting at E13.5 caused by an increase in the proliferation of neural progenitors. *Myocd* is involved in cardiovascular development by regulating the bone morphogenetic protein 10 (BMP10) signaling pathway, which in turn, regulates proliferation and apoptosis of cardiomyocites during heart development (Huang et al. 2012). Deletion of *Myocd* results in embryonic lethality by E10.5 caused by heart failure. VSD was also seen in embryos with a cardiomyocyte-restricted mutation of the *Myocd*. *Lefty1*, *Naca* and *Tcap* are also among the ES cell genes that contain the overlapping binding sites of CECR2, LUZP1 and SNF2H in their promoter region (up to 5 kb upstream of TSS) and are involved in embryonic heart development (Figure 4-4). It has been reported that the homozygous *Lefty1* mutant mice die before weaning (Meno et al. 1998). VSD, DORV and TGA, the same defects seen in *Luzp1* mutants, were among the phenotypes seen in the *Lefty1* mutants. Heart defects including ventricular hypoplasia, poorly formed interventricular septum and decreased cardiomyocyte proliferation are also seen in the *Naca* mutant mice (Park et al. 2010).

Altogether these results suggest the possible role of the CECR2 complexes in regulating genes that are involved in kidney and heart development in ES cell dataset. To confirm these roles the mRNA levels of the candidate genes involved in heart and kidney development can be quantified with qRT-PCR in the *Cecr2* mutant embryos to determine any changes compared to wild-type embryos.
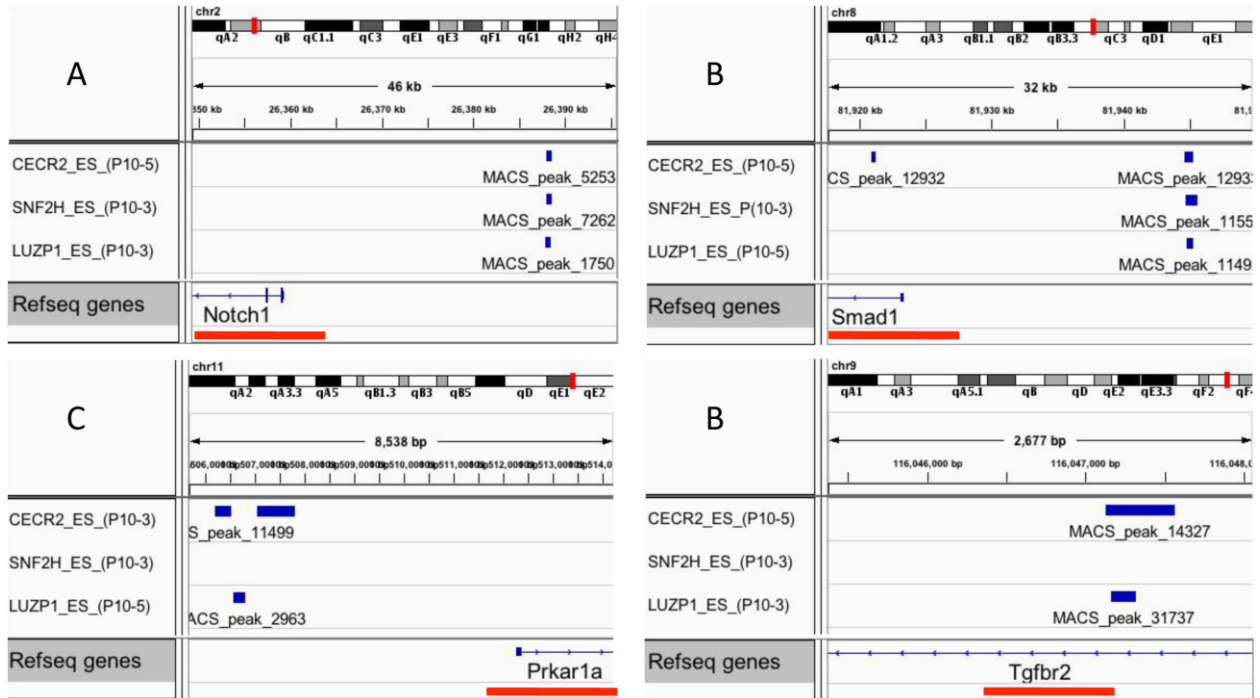
Figure 4-3 CECR2 and LUZP1 co-occupy the regulatory regions of genes associated with GO terms related to heart development. The Integrative Genomics Viewer (IGV) browser was used to visualize the binding sites of CECR2, SNF2H and LUZP1 on the genomic regions of Notch1 (A), Smad1 (B), Prkar1a (C) and Tgfbr2 (D). The binding sites are shown as blue bars.

217

Figure 4-4 CECR2, SNF2H and LUZP1 co-occupy the promoter regions of genes associated with heart development in ES cells. Peak calling was performed using MACS and P-value cutoff=10^{-5}. The Integrative Genomics Viewer (IGV) browser was used to visualize the binding sites of CECR2, SNF2H and LUZP1 on the genomic regions of Lefty (A), Naca (B) and Tcap (C). The binding sites are shown as blue bars. The genes are underlined with red line.

4.13. CECR2 function is regulated by co-factors

Combinations of transcription factors bind regulatory elements to control gene expression (Won et al. 2009). Several studies show that interacting with regulatory factors is one of the mechanisms that chromatin remodeling complexes use to target their binding sites on chromatin (Goldmark et al. 2000, Morris et al. 2013). Identification of co-factors that function with CECR2 is crucial to understand how the CECR2 complex is involved in neurulation and reproduction as well as other possible functions. DNA motif discovery using the ChIP-seq data is a very useful approach to investigate the presence of co-factors in the binding sites of the CECR2 complexes. Analysis of DNA sequence motifs at CECR2 peaks identified binding sites for a large set of transcription factors in ES cells and testis datasets ($P < 10^{-5}$). The enriched primary motifs identified for testis was a subset of the motifs identified in ES cells. It suggests that CECR2 recognizes specific DNA sequences in the genome or CECR2 interacts with specific transcription factor families to regulate function. The identified large set of transcription factors suggests that

CECR2 is recruited by many transcription factors to its binding targets in ES cells and testis. These transcription factors are involved in different biological processes, which is consistent with the large range of possible functions that has been predicted for CECR2 using my ChIP-seq data.

DNA motifs provide binding sites at enhancers and promoters for transcription factors that could, in turn, recruit remodeling complexes (Morris et al. 2013). Transcription factors can recruit the chromatin remodeling proteins through physical protein-protein interactions (Debril et al. 2004, Kowenz-Leutz & Leutz 1999). Transcription factors can also recruit chromatin remodellers to the chromatin without any direct interaction with them. In this case a transcription factor recruits proteins that change the histone modifications of the binding sites that leads to the subsequent binding of the remodeling complexes to those specific histone marks (Morris et al. 2013).

Integrating motifs identified for CECR2 and SNF2H showed the transcription factors that may co-function with the CECR2 complex. The motifs were almost the same for ES cells and testis, the most abundant of which are E2F3, IRF1, Gata4, PRDM1 and EGR2. These are the motifs that are identified in both ES cells and testis. These transcription factors have been shown to play a role in various biological processes. E2F3 regulates the expression of genes with a function in the cell cycle (Humbert et al. 2000). GATA4 is another transcription factor that binds to the same sequences that CECR2 and SNF2H bind in ES cells and testis. The motif that is bound by GATA4 is present in the regulatory regions of many genes that are important in many aspects of embryogenesis including heart and testis development (Lourenço et al. 2011, Zhou et al. 2012). This is more evidence suggesting that the CECR2-containing complexes may be involved in heart development. A mutation in *Gata4* leads to embryonic lethality between E11.5 and E13.5 caused by heart abnormalities including DORV and VSD, which are also seen in cat eye syndrome patients (Crispino et al. 2001). PRDM1 has an essential role in different aspects of embryonic development (Morgan et al. 2009). It is important for primordial germ cell specification, placental and heart development. *Egr2* mutation in mice showed that it plays an important role in hindbrain development (Schneider-Maunoury et al. 1993). IRF1 has been shown to have a role in DNA repair regulation (Prost et al. 1998).

In addition to the known motifs, I discovered some unknown motifs in the binding sites of the CECR2/SNF2H complex in ES cells and testis. These motifs do not match with any of the known motifs in UniPROBE and JASPAR databases. These novel motifs are possible candidates

for being specific DNA binding sequences for the CECR2 complex, although experimental investigation of the motifs are required to confirm the result. Gene Ontology enrichment analysis to predict their transcriptional roles revealed that the genes associated with these sites in ES cells are enriched in several ontology terms associated with different aspects of embryonic development including brain, heart and kidney development. All of the unknown motifs identified in the overlapping binding sites of the three proteins were associated with the genes that are only involved in kidney development as defined by GREAT. GREAT didn't associate any of the novel motifs from the testis dataset to any GO terms.

Taken together, the motif discovery analysis suggests that CECR2-containing complexes in ES cells and testis co-occupy genomic regions together with several transcription factors and co-regulate the expression of genes that are involved in different aspects of biological processes. I also discovered some unknown motifs that might be functional elements but require validation.

4.14. Future work

4.14.1. Investigating function of the non-canonical *Cecr2* isoform

Biochemical analyses in this study provided some evidence that a non-canonical CECR2 isoform exists in the mouse. This isoform was detected in both wild type and the homozygous $Cecr2^{Tm1.1Hemc}$ mutants. There is evidence of the existence of this isoform both at the mRNA and protein level. These data suggest that the homozygous $Cecr2^{Tm1.1Hemc}$ mutation is not a null mutation and it is possible that the non-canonical isoform is functional. Recent study of our newly characterized *Cecr2* mutation, $Cecr2^{tm2a(EUCOMM)Hmgu}$, showed that there are additional phenotypes (coloboma, polydactyly) in these mutants compared to $Cecr2^{Tm1.1Hemc}$ mutants indicating that the non-canonical isoform may not be functional in the $Cecr2^{tm2a(EUCOMM)Hmgu}$, resulting in the additional phenotypes. RNA-seq and qRT-PCR data have confirmed the existence of the non-canonical CECR2 isoform at the mRNA level (Kacie Norton, unpublished). The existence of the isoform in its protein form can be confirmed by immunoprecipitation of this protein from the $Cecr2^{Tm1.1Hemc}$ mutant testis using the CECR2 antibody, followed by mass spectrometry analysis.

4.14.2. Other components of the CECR2 complex

The biochemical analyses in this study provide strong evidence that the CECR2-containing complexes exist in different sizes and there are novel components other than CECR2 and ISWI proteins. The results strongly suggest that LUZP1 is not part of the complex in testis and it interacts with CECR2 in ES cells. The large size of the complex in both ES cells and testis indicates that there must be other components in these complexes. The mass spectrometry analyses led to the identification of several strong candidate proteins interacting with CECR2 (Table 3-4 and 3-5). To confirm that these proteins are interacting with CECR2 in a complex, the BioID system can be performed using protein extracts from ES cells and testis. This technique is suitable for finding the protein-protein interactions, especially the weak interactions between the core complex and other components. It is possible that some of the proteins are tissue-specific.

RIF1 is one of the strong candidates for being a component of the CECR2 complex both in ES cells and testis. The depletion of Rif1 in HEK293 cells leads to the reduced 53BP1 foci following IR. Interestingly, it has been shown that CECR2 has a role in DNA repair by promoting the formation of 53BP1 and γ-H2AX foci following IR. It is possible that CECR2 is recruited to the DNA damage sites through its interaction with RIF1. The interaction of CECR2 and RIF can be investigated by the BioID system to find any interaction between these two proteins.

4.14.3. Validation of genomic regions targeted by the CECR2-containing complexes

The CECR2-containing complexes are chromatin remodeling complexes. We hypothesize that these complexes are able to regulate gene expression by altering the chromatin structure. My ChIP-seq data showed that the CECR2-containing complexes are targeted to specific genomic regions and there are similarities and differences between their binding targets between ES cells and testis. Only a small fraction of the binding sites are near TSS and the majority of them are located at a large distance from the TSS. My analysis showed that the binding sites have a higher conservation score near the summit suggesting that these genomic regions located far from the TSS might be cis-regulatory elements that exert their regulation on promoters from a distance. There are multiple genes that are associated with CNS development and reproduction. Since the CECR2 complex binds to many genomic regions it would be reasonable to focus on the genes that have been associated with those aspects of development. Performing a complementary RNA-

seq in testis (wild-type vs non-penetrant adult BALB/c $Cecr2^{Gt45Bic}$ mutant males) and neurulating plate (wild-type vs $Cecr2^{Tm1.1Hemc}$ mutant embryos) would be a great way to make a list of the genes that are regulated by the CECR2 complex. Comparing the ChIP-seq and RNA-seq data will show the genes that are directly regulated by the CECR2 complex. It also will reveal the genes that are the secondary targets of the complex.

4.14.4. Investigating the cardiovascular abnormalities in the *Cecr2* mutant embryos

I showed that LUZP1 is a novel component of the CECR2-containing complexes in ES cells. *Luzp1* mutants exhibit severe cardiovascular defects including VSD (Hsu et al. 2008), which is also observed in cat eye syndrome patients (Schinzel et al. 1981). The ChIP-seq data analysis revealed multiple genes that are involved in heart development and show similar heart defects when mutated. Any abnormalities in heart development can be easily detected using morphological analysis of the histological sections prepared from the heart of the exencephalic $Cecr2^{Tm1.1Hemc}$ mutants. Ventriculoarterial connection can be examined by casting blue dye into the right ventricle of the embryos. Dye cast will allow us to trace any deficit related to the ventricles, atriums and the main vessels. *Luzp1* mutant embryos exhibit multiple types of cardiovascular deficits including DORV, VSD and TGA that can be detected using dye cast.

The ChIP-seq experiment suggests that the CECR2 complex occupies the regulatory regions of multiple genes involved in the cardiovascular development including *Notch1*, *Tgfbr2*, *smad1*, *Ctnnb1*, *Myocd*, *Prkar1a*, *Myocd*. *Lefty1*, *Naca* and *Tcap*. These genes regulate different aspects of cardiovascular system development. Comparing the expression of these genes in the heart of wild-type and $Cecr2^{Tm1.1Hemc}$ mutant embryos in different embryonic stages would show if CECR2 is involved in heart development by controlling these genes.

4.14.5. Fertility abnormailies

Non-penetrant adult BALB/c $Cecr2^{Gt45Bic}$ mutant males have smaller testis sizes compared to the wild-type animals and are subfertile. Subfertility can be caused by abnormalities in different stages of reproduction. One of the candidate genes containing the CECR2 complex in its regulatory region is *Insr*. This gene is involved in fertility by controlling the circulatory LH level. It is possible that the reproductive abnormalities seen in *Cecr2* mutants are the result of hormonal events. To find out the reproductive hormones including LH, Follicle Stimulating

Hormone (FSH), Gonadotropin Releasing Hormone (GnRH) and Testosterone can be compared between wild-type and non-penetrant adult BALB/c *Cecr2*^{Gt45Bic} mutant males.

4.15. Concluding remarks

This work has focused on the CECR2-containing complexes in mice. In this work for the first time an anti-CECR2 antibody was produced that successfully works in various assays including Western blot, ELIZA, immunoprecipitation, immunofluorescence staining and ChIP-seq. Prior to this study, all data for the CECR2 protein complexes were obtained using a CECR2 fusion protein. The biochemical investigation of the CECR2 protein in its native form shows that it acts differently from the fusion protein.

This work for the first time provides evidence of two novel components of the CECR2-containing complexes in the mouse: LUZP1 and CCAR2. It also demonstrates that LUZP1 is an ES cell-specific component indicating the difference between the testis and ES cell complexes in mice. The lack of LUZP1 results in partial disassociation of the CECR2 complexes.

This is also the first attempt to find the chromatin binding sites of the CECR2 complexes in the mouse and it suggests that the CECR2-containing complexes bind the promoter and cis-regulatory regions of many genes. The genes containing the binding sites of the CECR2 complexes are involved in various aspects of embryonic development including heart, brain and kidney development and some are involved in reproduction. This work provides evidence that the CECR2 complex co-occupies genomic regions together with several transcription factors suggesting that they co-regulate the expression of genes. Finally, this work supports the hypothesis that *Cecr2* regulates the genes that are components of the Wnt-signaling pathways.

Literature Cited

Acampora D, Merlo GR, Paleari L, Zerega B, Postiglione MP, et al. 1999. Craniofacial,
vestibular and bone defects in mice lacking the Distal-less-related gene Dlx5.
Development. 126(17):3795–3809

Accili D, Drago J, Lee EJ, Johnson MD, Cool MH, et al. 1996. Early neonatal death in mice
homozygous for a null allele of the insulin receptor gene. *Nat. Genet.* 12(1):106–9

Aihara T, Miyoshi Y, Koyama K, Suzuki M, Takahashi E, et al. 1998. Cloning and mapping of
SMARCA5 encoding hSNF2H, a novel human homologue of Drosophila ISWI.
Cytogenet. Cell Genet. 81(3–4):191–93

Alexiadis V, Varga-Weisz PD, Bonte E, Becker PB, Gruss C. 1998. In vitro chromatin
remodelling by chromatin accessibility complex (CHRAC) at the SV40 origin of DNA
replication. *EMBO J.* 17(12):3428–3438

Alkhatib SG, Landry JW. 2011. The Nucleosome Remodeling Factor. *FEBS Lett.* 585(20):3197–
3207

Alvarez-Saavedra M, De Repentigny Y, Lagali PS, Raghu Ram EVS, Yan K, et al. 2014. Snf2h-
mediated chromatin organization and histone H1 dynamics govern cerebellar
morphogenesis and neural maturation. *Nat. Commun.* 5:

Angers S, Moon RT. 2009. Proximal events in Wnt signal transduction. *Nat. Rev. Mol. Cell Biol.*

Aravind L, Landsman D. 1998. AT-hook motifs identified in a wide variety of DNA-binding
proteins. *Nucleic Acids Res.* 26(19):4413–4421

Ashe A, Morgan DK, Whitelaw NC, Bruxner TJ, Vickaryous NK, et al. 2008. A genome-wide
screen for modifiers of transgene variegation identifies genes with critical roles in
development. *Genome Biol.* 9(12):1

Aydin OZ, Marteijn JA, Ribeiro-Silva C, Rodriguez Lopez A, Wijgers N, et al. 2014a. Human ISWI complexes are targeted by SMARCA5 ATPase and SLIDE domains to help resolve lesion-stalled transcription. *Nucleic Acids Res.* 42(13):8473–85

Aydin ÖZ, Vermeulen W, Lans H. 2014b. ISWI chromatin remodeling complexes in the DNA damage response. *Cell Cycle*. 13(19):3016–25

Badenhorst P, Xiao H, Cherbas L, Kwon SY, Voas M, et al. 2005. The Drosophila nucleosome remodeling factor NURF is required for Ecdysteroid signaling and metamorphosis. *Genes Dev.* 19(21):2540–2545

Bannister AJ, Kouzarides T. 2011. Regulation of chromatin by histone modifications. *Cell Res.* 21(3):381–395

Banting GS. 2004. CECR2, a protein involved in neurulation, forms a novel chromatin remodeling complex with SNF2L. *Hum. Mol. Genet.* 14(4):513–24

Banting GS, Barak O, Ames TM, Burnham AC, Kardel MD, et al. 2005. CECR2, a protein involved in neurulation, forms a novel chromatin remodeling complex with SNF2L. *Hum. Mol. Genet.* 14(4):513–24

Bao X, Lu CM, Liu F, Gu Y, Dalton ND, et al. 2007. Epinephrine Is Required for Normal Cardiovascular Responses to Stress in the Phenylethanolamine N-Methyltransferase Knockout Mouse. *Circulation*. 116(9):1024–31

Bao Y, Shen X. 2007. INO80 subfamily of chromatin remodeling complexes. *Mutat. Res. Mol. Mech. Mutagen.* 618(1–2):18–29

Barak O, Lazzaro MA, Lane WS, Speicher DW, Picketts DJ, Shiekhattar R. 2003. Isolation of human NURF: a regulator of Engrailed gene expression. *EMBO J.* 22(22):6089–6100

Barnett C, Krebs JE. 2011. WSTF does it all: a multifunctional protein in transcription, repair, and replicationThis paper is one of a selection of papers published in a Special Issue entitled 31st Annual International Asilomar Chromatin and Chromosomes Conference, and has undergone the Journal's usual peer review process. *Biochem. Cell Biol.* 89(1):12–23

Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, et al. 2007. High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell*. 129(4):823–37

Bartholomew B. 2014. Regulating the Chromatin Landscape: Structural and Mechanistic Perspectives. *Annu. Rev. Biochem.* 83(1):671–96

Bell AC, West AG, Felsenfeld G. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*. 98(3):387–396

Bertonati C, Honig B, Alexov E. 2007. Poisson-Boltzmann Calculations of Nonspecific Salt Effects on Protein-Protein Binding Free Energies. *Biophys. J.* 92(6):1891–99

Biswas D, Takahata S, Xin H, Dutta-Biswas R, Yu Y, et al. 2008. A Role for Chd1 and Set2 in Negatively Regulating DNA Replication in Saccharomyces cerevisiae. *Genetics*. 178(2):649–59

Biswas M, Voltz K, Smith JC, Langowski J. 2011. Role of Histone Tails in Structural Stability of the Nucleosome. *PLoS Comput. Biol.* 7(12):e1002279

Bochar DA, Savard J, Wang W, Lafleur DW, Moore P, et al. 2000. A family of chromatin remodeling factors related to Williams syndrome transcription factor. *Proc. Natl. Acad. Sci.* 97(3):1038–1043

Börner K, Jain D, Vazquez-Pianzola P, Vengadasalam S, Steffen N, et al. 2016. A role for tuned levels of nucleosome remodeler subunit ACF1 during Drosophila oogenesis. *Dev. Biol.* 411(2):217–30

Bosman EA, Quint E, Fuchs H, Hrabé de Angelis M, Steel KP. 2009. Catweasel mice: A novel role for Six1 in sensory patch development and a model for branchio-oto-renal syndrome. *Dev. Biol.* 328(2):285–96

Bossen C, Murre CS, Chang AN, Mansson R, Rodewald H-R, Murre C. 2015. The chromatin remodeler Brg1 activates enhancer repertoires to establish B cell identity and modulate cell growth. *Nat. Immunol.* 16(7):775–84

Bozhenok L, Wade PA, Varga-Weisz P. 2002a. WSTF-ISWI chromatin remodeling complex targets heterochromatic replication foci. *EMBO J.* 21(9):2231–41

Bozhenok L, Wade PA, Varga-Weisz P. 2002b. WSTF_ISWI chromatin remodeling complex targets heterochromatic replication foci. *The EMBO Journal*. 21(9):2231–41

Brechalov AV, Valieva ME, Georgieva SG, Soshnikova NV. 2016. PHF10 isoforms are phosphorylated in the PBAF mammalian chromatin remodeling complex. *Mol. Biol.* 50(2):278–83

Brehm A, Tufteland KR, Aasland R, Becker PB. 2004. The many colours of chromodomains. *BioEssays*. 26(2):133–40

Brüning JC, Gautam D, Burks DJ, Gillette J, Schubert M, et al. 2000. Role of Brain Insulin Receptor in Control of Body Weight and Reproduction. *Science*. 289(5487):2122–2125

Bultman S, Gebuhr T, Yee D, La Mantia C, Nicholson J, et al. 2000. A Brg1 null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. *Mol. Cell*. 6(6):1287–1295

Buonomo SBC, Wu Y, Ferguson D, de Lange T. 2009. Mammalian Rif1 contributes to

    replication stress survival and homology-directed repair. *J. Cell Biol.* 187(3):385–98

Burgio G, La Rocca G, Sala A, Arancio W, Di Gesù D, et al. 2008. Genetic Identification of a

    Network of Factors that Functionally Interact with the Nucleosome Remodeling ATPase

    ISWI. *PLoS Genet.* 4(6):e1000089

Carter M, Ulrich S, Oofuji Y, Williams DA, Ross E. 1991. Crooked tail (Cd) models human

    folate-responsive neural tube defects. *Hum. Mol. Genet.* 8(12):2199–2204

Cavellan E, Asp P, Percipalle P, Farrants A-KO. 2006. The WSTF-SNF2h Chromatin

    Remodeling Complex Interacts with Several Nuclear Proteins in Transcription. *J. Biol.*

    Chem. 281(24):16264–71

Chapman JR, Barral P, Vannier J-B, Borel V, Steger M, et al. 2013. RIF1 Is Essential for 53BP1-

    Dependent Nonhomologous End Joining and Suppression of DNA Double-Strand Break

    Resection. *Mol. Cell.* 49(5):858–71

Chen J, Morosan-Puopolo G, Dai F, Wang J, Brand-Saberi B. 2010. Molecular cloning of

    chicken Cecr2 and its expression during chicken embryo development. *Int. J. Dev. Biol.*

    54(5):925–29

Chen L, Conaway RC, Conaway JW. 2013a. Multiple modes of regulation of the human Ino80

    SNF2 ATPase by subunits of the INO80 chromatin-remodeling complex. *Proc. Natl.*

    Acad. Sci. 110(51):20497–502

Chen PB, Hung J-H, Hickman TL, Coles AH, Carey JF, et al. 2013b. Hdac6 regulates Tip60-

    p400 function in stem cells. *eLife*. 2:

Chi H, Sarkisian MR, Rakic P, Flavell RA. 2005. Loss of mitogen-activated protein kinase kinase kinase 4 (MEKK4) results in enhanced apoptosis and defective neural tube development. *Proc. Natl. Acad. Sci. U. S. A.* 102(10):3846–3851

Chini CCS, Escande C, Nin V, Chini EN. 2010. HDAC3 Is Negatively Regulated by the Nuclear Protein DBC1. *J. Biol. Chem.* 285(52):40830–37

Chizhikov V, Steshina E, Roberts R, Ilkin Y, Washburn L, Millen KJ. 2006. Molecular definition of an allelic series of mutations disrupting the mouse Lmx1a (dreher) gene. *Mamm. Genome.* 17(10):1025–32

Cho HM, Kim JY, Kim H, Sun W. 2014. Phosphatase and actin regulator 4 is associated with intermediate filaments in adult neural stem cells and their progenitor astrocytes. *Histochem. Cell Biol.* 142(4):411–19

Choi JH, Sheu JJ-C, Guan B, Jinawath N, Markowski P, et al. 2009. Functional Analysis of 11q13.5 Amplicon Identifies Rsf-1 (HBXAP) as a Gene Involved in Paclitaxel Resistance in Ovarian Cancer. *Cancer Res.* 69(4):1407–15

Choi JW, Herr DR, Noguchi K, Yung YC, Lee C-W, et al. 2010. LPA Receptors: Subtypes and Biological Actions. *Annu. Rev. Pharmacol. Toxicol.* 50(1):157–86

Ciani L, Salinas PC. 2005. Signalling in neural development: WNTS in the vertebrate nervous system: from patterning to neuronal connectivity. *Nat. Rev. Neurosci.* 6(5):351–62

Clapier CR, Cairns BR. 2009a. The Biology of Chromatin Remodeling Complexes. *Annu. Rev. Biochem.* 78(1):273–304

Clapier CR, Cairns BR. 2009b. The Biology of Chromatin Remodeling Complexes. *Annu. Rev. Biochem.* 78(1):273–304

Clapier CR, Cairns BR. 2012. Regulation of ISWI involves inhibitory modules antagonized by nucleosomal epitopes. *Nature*. 492(7428):280–84

Clapier CR, Nightingale KP, Becker PB. 2002. A critical epitope for substrate recognition by the nucleosome remodeling ATPase ISWI. *Nucleic Acids Res*. 30(3):649–55

Close P, East P, Dirac-Svejstrup AB, Hartmann H, Heron M, et al. 2012. DBIRD complex integrates alternative mRNA splicing with RNA polymerase II transcript elongation. *Nature*. 484(7394):386–89

Cohen M, Kicheva A, Ribeiro A, Blassberg R, Page KM, et al. 2015. Ptch1 and Gli regulate Shh signalling dynamics via multiple mechanisms. *Nat. Commun.* 6:6709

Collins N, Poot RA, Kukimoto I, García-Jiménez C, Dellaire G, Varga-Weisz PD. 2002. An ACF1–ISWI chromatin-remodeling complex is required for DNA replication through heterochromatin. *Nat. Genet.* 32(4):627–32

Contos JJ, Fukushima N, Weiner JA, Kaushal D, Chun J. 2000. Requirement for the lpA1 lysophosphatidic acid receptor gene in normal suckling behavior. *Proc. Natl. Acad. Sci.* 97(24):13384–13389

Copp AJ, Greene NDE, Murdoch JN. 2003. The genetic basis of mammalian neurulation. *Nat. Rev. Genet.* 4(10):784–93

Corona DFV, Tamkun JW. 2004. Multiple roles for ISWI in transcription, chromosome organization and DNA replication. *Biochim. Biophys. Acta BBA - Gene Struct. Expr.* 1677(1–3):113–19

Cotton LM, O'Bryan MK, Hinton BT. 2008. Cellular Signaling by Fibroblast Growth Factors (FGFs) and Their Receptors (FGFRs) in Male Reproduction. *Endocr. Rev.* 29(2):193–216

Cutter AR, Hayes JJ. 2015. A brief review of nucleosome structure. *FEBS Lett.* 589(20PartA):2914–22

Dawe CE, Kooistra MK, Fairbridge NA, Pisio AC, McDermid HE. 2011. Role of chromatin remodeling gene Cecr2 in neurulation and inner ear development. *Dev. Dyn.* 240(2):372–83

de Campos-Nebel M, Larripa I, González-Cid M. 2010. Topoisomerase II-Mediated DNA Damage Is Differently Repaired during the Cell Cycle by Non-Homologous End Joining and Homologous Recombination. *PLoS ONE*. 5(9):e12541

de la Serna IL, Ohkawa Y, Imbalzano AN. 2006. Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers. *Nat. Rev. Genet.* 7(6):461–73

Debril M-B, Gelman L, Fayard E, Annicotte J-S, Rocchi S, Auwerx J. 2004. Transcription Factors and Nuclear Receptors Interact with the SWI/SNF Complex through the BAF60c Subunit. *J. Biol. Chem.* 279(16):16677–86

Deindl S, Hwang WL, Hota SK, Blosser TR, Prasad P, et al. 2013. ISWI Remodelers Slide Nucleosomes with Coordinated Multi-Base-Pair Entry Steps and Single-Base-Pair Exit Steps. *Cell*. 152(3):442–52

Denslow SA, Wade PA. 2007. The human Mi-2/NuRD complex and gene regulation. *Oncogene*. 26(37):5433–38

Dhaka A, Murray AN, Mathur J, Earley TJ, Petrus MJ, Patapoutian A. 2007. TRPM8 Is Required for Cold Sensation in Mice. *Neuron*. 54(3):371–78

Dhalluin C, Carlson JE, Zeng L, He C, Aggarwal AK, Zhou M-M. 1999. Structure and ligand of a histone acetyltransferase bromodomain. *Nature*. 399(6735):491–96

Dirscherl SS, Krebs JE. 2004. Functional diversity of ISWI complexes. *Biochem. Cell Biol.* 82(4):482–89

Doerks T, Copley R, Bork P. 2001. DDT–a novel domain in different transcription and chromosome remodeling factors. *Trends Biochem. Sci.* 26(3):145–146

Dong J, Gao Z, Liu S, Li G, Yang Z, et al. 2013. SLIDE, the Protein Interacting Domain of ISWI Remodelers, Binds DDT-domain Proteins of Different Subfamilies in Chromatin Remodeling Complexes: Protein Interactions in the DDT-ISWI Complexes. *J. Integr. Plant Biol.* n/a-n/a

Dowdle JA, Mehta M, Kass EM, Vuong BQ, Inagaki A, et al. 2013. Mouse BAZ1A (ACF1) Is Dispensable for Double-Strand Break Repair but Is Essential for Averting Improper Gene Expression during Spermatogenesis. *PLoS Genet.* 9(11):e1003945

Dykhuizen EC, Hargreaves DC, Miller EL, Cui K, Korshunov A, et al. 2013. BAF complexes facilitate decatenation of DNA by topoisomerase IIα. *Nature*. 497(7451):624–27

Eberharter A, Ferrari S, Langst G, Straub T, Imhof A, et al. 2001. Acf1, the largest subunit of CHRAC, regulates ISWI-induced nucleosome remodelling The EMBO Journal Vol. 20 No. 14 pp. 3781±3788, 2001. *EMBO J.* 20(14):3781–88

Eberharter A, Vetter I, Ferreira R, Becker PB. 2004. ACF1 improves the effectiveness of nucleosome mobilization by ISWI through PHD–histone. *EMBO J.* 23(20):4029–4039

Echelard Y, Epstein DJ, St-Jacques B, Shen L, Mohler J, et al. 1993. Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. Echelard Y1, Epstein DJ, St-Jacques B, Shen L, Mohler J, McMahon JA, McMahon AP. *Cell*. 75(7):1417–30

Eckey M, Kuphal S, Straub T, Rummele P, Kremmer E, et al. 2012. Nucleosome Remodeler
SNF2L Suppresses Cell Proliferation and Migration and Attenuates Wnt Signaling. *Mol.
Cell. Biol.* 32(13):2359–71

Elfring LK, Deuring R, McCallum CM, Peterson CL, Tamkun JW. 1994. Identification and
characterization of Drosophila relatives of the yeast transcriptional activator SNF2/SWI2.
Mol. Cell. Biol. 14(4):2225–2234

Elliott MR, Zheng S, Park D, Woodson RI, Reardon MA, et al. 2010. Unexpected requirement
for ELMO1 in clearance of apoptotic germ cells in vivo. *Nature.* 467(7313):333–37

Emelyanov AV, Vershilova E, Ignatyeva MA, Pokrovsky DK, Lu X, et al. 2012. Identification
and characterization of ToRC, a novel ISWI-containing ATP-dependent chromatin
assembly complex. *Genes Dev.* 26(6):603–14

Erdel F, Rippe K. 2011. Chromatin remodelling in mammalian cells by ISWI-type complexes -
where, when and why?: ISWI chromatin remodellers in mammalian cells. *FEBS J.*
278(19):3608–18

Erdel F, Schubert T, Marth C, Langst G, Rippe K. 2010. Human ISWI chromatin-remodeling
complexes sample nucleosomes via transient binding reactions and become immobilized
at active sites. *Proc. Natl. Acad. Sci.* 107(46):19873–78

Etheridge SL, Ray S, Li S, Hamblet NS, Lijam N, et al. 2008. Murine Dishevelled 3 Functions in
Redundant Pathways with Dishevelled 1 and 2 in Normal Cardiac Outflow Tract,
Cochlea, and Neural Tube Development. *PLoS Genet.* 4(11):e1000259

Euskirchen GM, Auerbach RK, Davidov E, Gianoulis TA, Zhong G, et al. 2011. Diverse Roles
and Interactions of the SWI/SNF Chromatin Remodeling Complex Revealed Using
Global Approaches. *PLoS Genet.* 7(3):e1002008

Ewing AK, Attner M, Chakravarti D. 2007. Novel Regulatory Role for Human Acf1 in

Transcriptional Repression of Vitamin D3 Receptor-Regulated Genes. *Mol. Endocrinol.* 21(8):1791–1806

Fairbridge NA. 2013. University of Alberta

Fairbridge NA, Dawe CE, Niri FH, Kooistra MK, King-Jones K, McDermid HE. 2010. Cecr2

mutations causing exencephaly trigger misregulation of mesenchymal/ectodermal

transcription factors. *Birt. Defects Res. A. Clin. Mol. Teratol.* 88(8):619–25

Falbo KB, Shen X. 2006. Chromatin remodeling in DNA replication. *J. Cell. Biochem.* 97(4):684–89

Falk S, Wurdak H, Ittner LM, Ille F, Sumara G, et al. 2008. Brain Area-Specific Effect of TGF-β

Signaling on Wnt-Dependent Neural Stem Cell Expansion. *Cell Stem Cell*. 2(5):472–83

Fanourgakis G, Lesche M, Akpinar M, Dahl A, Jessberger R. 2016. Chromatoid Body Protein

TDRD6 Supports Long 3' UTR Triggered Nonsense Mediated mRNA Decay. *PLOS Genet.* 12(5):e1005857

Fauvarque M, Laurenti p, Boivin A, Bloyer S, Griffin-Shea R, et al. 2001. Dominant modifiers

of the polyhomeotic extra-sex-combs phenotype induced by marked P element insertional

mutagenesis in Drosophila. *Genet. Res.* 78(2):137–48

Felsenfeld G, Groudine M. 2003. Controlling the double helix. *Nature*. 421(6921):444–48

Feng L, Fong K-W, Wang J, Wang W, Chen J. 2013. RIF1 Counteracts BRCA1-mediated End

Resection during DNA Repair. *J. Biol. Chem.* 288(16):11135–43

Flaus A. 2006. Identification of multiple distinct Snf2 subfamilies with conserved structural

motifs. *Nucleic Acids Res.* 34(10):2887–2905

Footz TK, Brinkman-Mills P, Banting GS, Maier SA, Riazi MA, et al. 2001. Analysis of the cat eye syndrome critical region in humans and the region of conserved synteny in mice: a search for candidate genes at or near the human chromosome 22 pericentromere. *Genome Res.* 11(6):1053–1070

Forcales SV, Albini S, Giordani L, Malecova B, Cignolo L, et al. 2012. Signal-dependent incorporation of MyoD–BAF60c into Brg1-based SWI/SNF chromatin-remodelling complex. *EMBO J.* 31(2):301–16

Fu J, Jiang J, Li J, Wang S, Shi G, et al. 2009. Deleted in Breast Cancer 1, a Novel Androgen Receptor (AR) Coactivator That Promotes AR DNA-binding Activity. *J. Biol. Chem.* 284(11):6832–40

Fyodorov DV, Blower MD, Karpen GH, Kadonaga JT. 2003. Acf1 confers unique activities to ACF/CHRAC and promotes the formation rather than disruption of chromatin in vivo. *GENES & DEVELOPMENT.* 18:170–83

Fyodorov DV, Kadonaga JT. 2002. Binding of Acf1 to DNA Involves a WAC Motif and Is Important for ACF-Mediated Chromatin Assembly. *Mol. Cell. Biol.* 22(18):6344–53

Gangaraju VK, Bartholomew B. 2007. Mechanisms of ATP dependent chromatin remodeling. *Mutat. Res. Mol. Mech. Mutagen.* 618(1–2):3–17

Garapaty S, Xu C-F, Trojer P, Mahajan MA, Neubert TA, Samuels HH. 2009. Identification and Characterization of a Novel Nuclear Protein Complex Involved in Nuclear Hormone Receptor-mediated Gene Regulation. *J. Biol. Chem.* 284(12):7542–52

Gelbart ME, Bachman N, Delrow J, Boeke JD, Tsukiyama T. 2005. Genome-wide identification of Isw2 chromatin-remodeling targets by localization of a catalytically inactive mutant. *Genes Dev.* 19(8):942–954

Goldmark JP, Fazzio TG, Estep PW, Church GM, Tsukiyama T. 2000. The Isw2 chromatin
    remodeling complex represses early meiotic genes upon recruitment by Ume6p. *Cell*.
    103(3):423–433

Goller T, Vauti F, Ramasamy S, Arnold H-H. 2008. Transcriptional Regulator BPTF/FAC1 Is
    Essential for Trophoblast Differentiation during Early Mouse Development. *Mol. Cell.*
    Biol. 28(22):6819–27

Gonzales-Cope M, Sidoli S, Bhanu NV, Won K-J, Garcia BA. 2016. Histone H4 acetylation and
    the epigenetic reader Brd4 are critical regulators of pluripotency in embryonic stem cells.
    BMC Genomics. 17(1):

Goudarzi A, Zhang D, Huang H, Barral S, Kwon OK, et al. 2016. Dynamic Competing Histone
    H4 K5K8 Acetylation and Butyrylation Are Hallmarks of Highly Active Gene Promoters.
    Mol. Cell. 62(2):169–80

Gray JD, Kholmanskikh S, Castaldo BS, Hansler A, Chung H, et al. 2013. LRP6 exerts non-
    canonical effects on Wnt signaling during neural tube closure. *Hum. Mol. Genet.*
    22(21):4267–81

Greene NDE, Copp AJ. 2014. Neural Tube Defects. *Annu. Rev. Neurosci.* 37(1):221–42

Groth A, Rocha W, Verreault A, Almouzni G. 2007. Chromatin Challenges during DNA
    Replication and Repair. *Cell*. 128(4):721–33

Gu L, Frommel SC, Oakes CC, Simon R, Grupp K, et al. 2014. BAZ2A (TIP5) is involved in
    epigenetic alterations in prostate cancer and its overexpression predicts disease
    recurrence. *Nat. Genet.* 47(1):22–30

Guelman S, Kozuka K, Mao Y, Pham V, Solloway MJ, et al. 2009. The Double-Histone-Acetyltransferase Complex ATAC Is Essential for Mammalian Development. *Mol. Cell. Biol.* 29(5):1176–88

Guetg C, Lienemann P, Sirri V, Grummt I, Hernandez-Verdun D, et al. 2010. The NoRC complex mediates the heterochromatin formation and stability of silent rRNA genes and centromeric repeats. *EMBO J.* 29(13):2135–46

Gumienny TL, Brugnera E, Tosello-Trampont A-C, Kinchen JM, Haney LB, et al. 2001. CED-12/ELMO, a novel member of the CrkII/Dock180/Rac pathway, is required for phagocytosis and cell migration. *Cell*. 107(1):27–41

Guschin D, Geiman TM, Kikyo N, Tremethick DJ, Wolffe AP, Wade PA. 2000. Multiple ISWI ATPase Complexes from Xenopus laevis: FUNCTIONAL CONSERVATION OF AN ACF/CHRAC HOMOLOG. *J. Biol. Chem.* 275(45):35248–55

Hamaguchi M, Meth JL, von Klitzing C, Wei W, Esposito D, et al. 2002. DBC2, a candidate for a tumor suppressor gene involved in breast cancer. *Proc. Natl. Acad. Sci.* 99(21):13647–13652

Han P, Chang C-P. 2015. Long non-coding RNA and chromatin remodeling. *RNA Biol.* 12(10):1094–98

Han P, Li W, Lin C-H, Yang J, Shang C, et al. 2014. A long noncoding RNA protects the heart from pathological hypertrophy. *Nature*. 514(7520):102–6

Hanai K, Furuhashi H, Yamamoto T, Akasaka K, Hirose S. 2008. RSF Governs Silent Chromatin Formation via Histone H2Av Replacement. *PLoS Genet.* 4(2):e1000011

Hargreaves DC, Crabtree GR. 2011. ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res.* 21(3):396–420

Hartlepp KF, Fernandez-Tornero C, Eberharter A, Grune T, Muller CW, Becker PB. 2005. The
Histone Fold Subunits of Drosophila CHRAC Facilitate Nucleosome Sliding through
Dynamic DNA Interactions. *Mol. Cell. Biol.* 25(22):9886–96

High FA, Epstein JA. 2008. The multifaceted role of Notch in cardiac development and disease.
Nat. Rev. Genet. 9(1):49–61

Hiraike H, Wada-Hiraike O, Nakagawa S, Koyama S, Miyamoto Y, et al. 2010. Identification of
DBC1 as a transcriptional repressor for BRCA1. *Br. J. Cancer*. 102(6):1061–67

Hitoshi S, Ishino Y, Kumar A, Jasmine S, Tanaka KF, et al. 2011. Mammalian Gcm genes induce
Hes5 expression by active DNA demethylation and induce neural stem cells. *Nat.
Neurosci.* 14(8):957–64

Ho L, Crabtree GR. 2010. Chromatin remodelling during development. *Nature*. 463(7280):474–
84

Ho L, Jothi R, Ronan JL, Cui K, Zhao K, Crabtree GR. 2009a. An embryonic stem cell chromatin
remodeling complex, esBAF, is an essential component of the core pluripotency
transcriptional network. *Proc. Natl. Acad. Sci.* 106(13):5187–5191

Ho L, Ronan JL, Wu J, Staahl BT, Chen L, et al. 2009b. An embryonic stem cell chromatin
remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and
pluripotency. *Proc. Natl. Acad. Sci.* 106(13):5181–5186

Holmberg J, Clarke DL, Frisen J. 2000. Regulation of repulsion versus adhesion by different
splice forms of an Eph receptor. *Nature*. 408:203–6

Homanics GE, Maeda N, Traber MG, Kayden HJ, Dehart DB, Sulik KK. 1995. Exencephaly and
hydrocephaly in mice with targeted modification of the Apolipoprotein B (Apob) gene.
Teratology. 51(1):1–10

Hsu C-Y, Chang N-C, Lee MW-Y, Lee K-H, Sun D-S, et al. 2008a. LUZP deficiency affects neural tube closure during brain development. *Biochem. Biophys. Res. Commun.* 376(3):466–71

Hsu C-Y, Chang N-C, Lee MW-Y, Lee K-H, Sun D-S, et al. 2008b. LUZP deficiency affects neural tube closure during brain development. *Biochem. Biophys. Res. Commun.* 376(3):466–71

Huang J, Elicker J, Bowens N, Liu X, Cheng L, et al. 2012. Myocardin regulates BMP10 expression and is required for heart development. *J. Clin. Invest.* 122(10):3678–91

Huet G, Rajakyla EK, Viita T, Skarp K-P, Crivaro M, et al. 2013. Actin-regulated feedback loop based on Phactr4, PP1 and cofilin maintains the actin monomer pool. *J. Cell Sci.* 126(2):497–507

Humbert PO, Verona R, Trimarchi JM, Rogers C, Dandapani S, Lees JA. 2000. E2f3 is critical for normal cellular proliferation. *Genes Dev.* 14(6):690–703

Hwang WL, Deindl S, Harada BT, Zhuang X. 2014. Histone H4 tail mediates allosteric regulation of nucleosome remodelling by linker DNA. *Nature.* 512(7513):213–17

Ito T, Bulger M, Pazin MJ, Kobayashi R, Kadonaga JT. 1997. ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell.* 90(1):145–155

Iwasaki W, Miya Y, Horikoshi N, Osakabe A, Taguchi H, et al. 2013. Contribution of histone N-terminal tails to the structure and stability of nucleosomes. *FEBS Open Bio.* 3(1):363–69

Jensenius JC, Andersen I, Hau J, Crone M, Koch C. 1981. Eggs: Conveniently packeged antibodies. Methods for purification of yolk IgG. *ElsevierNorth-Holl. Biomed. Press.* 46:63–68

Jones MH, Hamana N, Nezu J, Shimane M. 2000. A Novel Family of Bromodomain Genes. *Genomics*. 63(1):40–45

Joshi P, Quach OL, Giguere SSB, Cristea IM. 2013. A functional proteomics perspective of DBC1 as a regulator of transcription. *J. Proteomics Bioinform.*

Juriloff DM, Harris MJ. 2000. Mouse models for neural tube closure defects. *Hum. Mol. Genet.* 9(6):993–1000

Kim J-E, Chen J, Lou Z. 2008. DBC1 is a negative regulator of SIRT1. *Nature*. 451(7178):583–86

Kim JK, Huh S-O, Choi H, Lee K-S, Shin D, et al. 2001. Srg3, a Mouse Homolog of Yeast SWI3, Is Essential for Early Embryogenesis and Involved in Brain Development. *Mol. Cell. Biol.* 21(22):7787–95

Kim JY, Choi SY, Moon Y, Kim HJ, Chin JH, et al. 2012. Different expression patterns of Phactr family members in normal and injured mouse brain. *Neuroscience*. 221:37–46

Kim T-H, Goodman J, Anderson KV, Niswander L. 2007. Phactr4 Regulates Neural Tube and Optic Fissure Closure by Controlling PP1-, Rb-, and E2F1-Regulated Cell-Cycle Progression. *Dev. Cell*. 13(1):87–102

Kim W, Kim J-E. 2013. Deleted in breast cancer 1 (DBC1) deficiency results in apoptosis of breast cancer cells through impaired responses to UV-induced DNA damage. *Cancer Lett.* 333(2):180–86

Kim Y-J, M Wilson III D. 2012. Overview of base excision repair biochemistry. *Curr. Mol. Pharmacol.* 5(1):3–13

Kimple ME, Brill AL, Pasker RL. 2013. Overview of Affinity Tags for Protein Purification: Affinity Tags for Protein Purification. In *Current Protocols in Protein Science*, eds. JE

Coligan, BM Dunn, DW Speicher, PT Wingfield, p. 9.9.1-9.9.23. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Knoll R, Linke WA, Zou P, Miocic S, Kostin S, et al. 2011. Telethonin Deficiency Is Associated With Maladaptation to Biomechanical Stress in the Mammalian Heart. *Circ. Res.* 109(7):758–69

Kooistra MK. 2009. University of Alberta

Kooistra MK, Leduc RYM, Dawe CE, Fairbridge NA, Rasmussen J, et al. 2012. Strain-specific modifier genes of Cecr2-associated exencephaly in mice: genetic analysis and identification of differentially expressed candidate genes. *Physiol. Genomics*. 44(1):35–46

Kowenz-Leutz E, Leutz A. 1999. AC/EBPβ isoform recruits the SWI/SNF complex to activate myeloid genes. *Mol. Cell*. 4(5):735–743

Koyama S, Wada-Hiraike O, Nakagawa S, Tanikawa M, Hiraike H, et al. 2010. Repression of estrogen receptor β function by putative tumor suppressor DBC1. *Biochem. Biophys. Res. Commun.* 392(3):357–62

Krajewski WA. 2013. Comparison of the Isw1a, Isw1b, and Isw2 Nucleosome Disrupting Activities. *Biochemistry (Mosc.)*. 52(40):6940–49

Krebs AR, Demmers J, Karmodiya K, Chang N-C, Chang AC, Tora L. 2010. ATAC and Mediator coactivators form a stable complex and regulate a set of non-coding RNA genes. *EMBO Rep.* 11(7):541–47

Kukimoto I, Elderkin S, Grimaldi M, Oelgeschläger T, Varga-Weisz PD. 2004. The histone-fold protein complex CHRAC-15/17 enhances nucleosome sliding and assembly mediated by ACF. *Mol. Cell*. 13(2):265–277

Kumar R, Cheok CF. 2014. RIF1: A novel regulatory factor for DNA replication and DNA damage response signaling. *DNA Repair*. 15:54–59

Kumar R, Li D-Q, Müller S, Knapp S. 2016. Epigenomic regulation of oncogenesis by chromatin remodeling. *Oncogene*

Kwon SY, Grisan V, Jang B, Herbert J, Badenhorst P. 2016. Genome-Wide Mapping Targets of the Metazoan Chromatin Remodeling Factor NURF Reveals Nucleosome Remodeling at Enhancers, Core Promoters and Gene Insulators. *PLoS Genet*. 12(4):e1005969

Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage. *Nature*. 227:680–85

Lai J-S, Herr W. 1992. Ethidium bromide provides a simple tool for identifying genuine DNA-independent protein associations. *Proc. Natl. Acad. Sci.* 89(15):6958–6962

Lai M-S, Wang C-Y, Yang S-H, Wu C-C, Sun HS, et al. 2016. The expression profiles of fibroblast growth factor 9 and its receptors in developing mice testes. *Organogenesis*. 12(2):61–77

Lalonde M-E, Cheng X, Cote J. 2014. Histone target selection within chromatin: an exemplary case of teamwork. *Genes Dev.* 28(10):1029–41

Lan L, Ui A, Nakajima S, Hatakeyama K, Hoshi M, et al. 2010. The ACF1 Complex Is Required for DNA Double-Strand Break Repair in Human Cells. *Mol. Cell*. 40(6):976–87

Landry J, Sharov AA, Piao Y, Sharova LV, Xiao H, et al. 2008. Essential Role of Chromatin Remodeling Protein Bptf in Early Mouse Embryos and Embryonic Stem Cells. *PLoS Genet.* 4(10):e1000241

Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, et al. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* 22(9):1813–31

Längst G, Bonte EJ, Corona DF, Becker PB. 1999. Nucleosome movement by CHRAC and ISWI without disruption or trans-displacement of the histone octamer. *Cell*. 97(7):843–852

Längst G, Manelyte L. 2015. Chromatin Remodelers: From Function to Dysfunction. *Genes*. 6(2):299–324

Lazzaro MA, Pépin D, Pescador N, Murphy BD, Vanderhyden BC, Picketts DJ. 2006. The Imitation Switch Protein SNF2L Regulates Steroidogenic Acute Regulatory Protein Expression during Terminal Differentiation of Ovarian Granulosa Cells. *Mol. Endocrinol.* 20(10):2406–17

Lazzaro MA, Picketts DJ. 2001. Cloning and characterization of the murine Imitation Switch (ISWI) genes: differential expression patterns suggest distinct developmental roles for Snf2h and Snf2l. *J. Neurochem.* 77(4):1145–1156

Leduc RYM, Singh P, McDermid HE. 2016. Genetic Backgrounds and Modifier Genes of NTD Mouse Models: An Opportunity for Greater Understanding of the Multifactorial Etiology of Neural Tube Defects   Renee Y.M. Leduc, Parmveer Singh, and Heather E. McDermid*

Lee H-S, Park J-H, Kim S-J, Kwon S-J, Kwon J. 2010. A cooperative activation loop among SWI/SNF, γ-H2AX and H3 acetylation for DNA double-strand break repair. *EMBO J.* 29(8):1434–1445

Lee MW-Y, Chang AC, Sun D-S, Hsu C-Y, Chang N-CA. 2001. Restricted expression of LUZP in neural lineage cells: a study in embryonic stem cells. *J. Biomed. Sci.* 8(6):504–511

Lee S-K, Park E-J, Lee H-S, Lee YS, Kwon J. 2012. Genome-wide screen of human bromodomain-containing proteins identifies Cecr2 as a novel DNA damage response protein. *Mol. Cells*. 34(1):85–91

Lefebvre S, Burlet P, Viollet L, Bertrandy S, Huber C, et al. 2002. A novel association of the SMN protein with two major non-ribosomal nucleolar proteins and its implication in spinal muscular atrophy. *Hum. Mol. Genet.* 11(9):1017–1027

LeRoy G. 2000. Purification and Characterization of a Human Factor That Assembles and Remodels Chromatin. *J. Biol. Chem.* 275(20):14787–90

LeRoy G, Orphanides G, S. Lane W, Reinberg D. 1998. Requirement of RSF and FACT for Transcription of Chromatin Templates in Vitro. *Science*. 282(5395):1900–1904

Lessard J, Wu JI, Ranish JA, Wan M, Winslow MM, et al. 2007. An Essential Switch in Subunit Composition of a Chromatin Remodeling Complex during Neural Development. *Neuron*. 55(2):201–15

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25(14):1754–60

Li J, Langst G, Grummt I. 2006. NoRC-dependent nucleosome positioning silences rRNA genes. *EMBO J.* 25(24):5735–41

Li J, Santoro R, Koberna K, Grummt I. 2005. The chromatin remodeling complex NoRC controls replication timing of rRNA genes. *EMBO J.* 24(1):120–127

Li Y, Schulz VP, Deng C, Li G, Shen Y, et al. 2016. Setd1a and NURF mediate chromatin dynamics and gene regulation during erythroid lineage commitment and differentiation. *Nucleic Acids Res.* gkw327

Lichota J, Grasser KD. 2001. Differential Chromatin Association and Nucleosome Binding of the Maize HMGA, HMGB, and SSRP1 Proteins [†]. *Biochemistry (Mosc.)*. 40(26):7860–67

Lin H, Ha K, Lu G, Fang X, Cheng R, et al. 2015. Cdc14A and Cdc14B Redundantly Regulate DNA Double-Strand Break Repair. *Mol. Cell. Biol.* 35(21):3657–68

Liu M, Shi X, Bi Y, Qi L, Guo X, et al. 2014. SHCBP1L, a conserved protein in mammals, is predominantly expressed in male germ cells and maintains spindle stability during meiosis in testis. *Mol. Hum. Reprod.* 20(6):463–75

Liu T, Ortiz JA, Taing L, Meyer CA, Lee B, et al. 2011. Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol*. 12(8):R83

Lourenço D, Brauner R, Rybczyńska M, Nihoul-Fékété C, McElreavey K, Bashamboo A. 2011. Loss-of-function mutation in GATA4 causes anomalies of human testicular development. *Proc. Natl. Acad. Sci.* 108(4):1597–1602

Lu X, Meng X, Morris CA, Keating MT. 1998. A Novel Human Gene, WSTF, Is Deleted in Williams Syndrome Xiaojun Lu,* Xun Meng,† Colleen A. Morris,‡ and Mark T. Keating. *Genomics*. 54:241–49

Luger K, Dechassa ML, Tremethick DJ. 2012. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat. Rev. Mol. Cell Biol.* 13(7):436–47

Ma W, Noble WS, Bailey TL. 2014. Motif-based analysis of large nucleotide data sets using MEME-ChIP. *Nat. Protoc.* 9(6):1428–50

Machanick P, Bailey TL. 2011. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics*. 27(12):1696–97

Magistri M, Faghihi MA, St Laurent G, Wahlestedt C. 2012. Regulation of chromatin structure by long noncoding RNAs: focus on natural antisense transcripts. *Trends Genet.* 28(8):389–96

Maier V, Jolicoeur C, Rayburn H, Takegahara N, Kumanogoh A, et al. 2011. Semaphorin 4C and 4G are ligands of Plexin-B2 required in cerebellar development. *Mol. Cell. Neurosci.* 46(2):419–31

Malik S, Roeder RG. 2005. Dynamic regulation of pol II transcription by the mammalian Mediator complex. *Trends Biochem. Sci.* 30(5):256–63

Maltby VE, Martin BJE, Schulze JM, Johnson I, Hentrich T, et al. 2012. Histone H3 Lysine 36 Methylation Targets the Isw1b Remodeling Complex to Chromatin. *Mol. Cell. Biol.* 32(17):3479–85

Marfella CGA, Imbalzano AN. 2007. The Chd Family of Chromatin Remodelers. *Mutat. Res. Mol. Mech. Mutagen.* 618(1–2):30–40

Margueron R, Reinberg D. 2010. Chromatin structure and the inheritance of epigenetic information. *Nat. Rev. Genet.* 11(4):285–96

Marteijn JA, Lans H, Vermeulen W, Hoeijmakers JHJ. 2014. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat. Rev. Mol. Cell Biol.* 15(7):465–81

Martens MA, Wilson SJ, Reutens DC. 2008. Research Review: Williams syndrome: a critical review of the cognitive, behavioral, and neuroanatomical phenotype. *J. Child Psychol. Psychiatry.* 49(6):576–608

Massa V, Savery D, Ybot-Gonzalez P, Ferraro E, Rongvaux A, et al. 2009. Apoptosis is not required for mammalian neural tube closure. *Proc. Natl. Acad. Sci.* 106(20):8233–38

Mayer C, Neubert M, Grummt I. 2008. The structure of NoRC-associated RNA is crucial for targeting the chromatin remodelling complex NoRC to the nucleolus. *EMBO Rep.* 9(8):774–80

Mayer C, Schmitz K-M, Li J, Grummt I, Santoro R. 2006. Intergenic Transcripts Regulate the Epigenetic State of rRNA Genes. *Mol. Cell.* 22(3):351–61

McConnell AD, Gelbart ME, Tsukiyama T. 2004. Histone Fold Protein Dls1p Is Required for Isw2-Dependent Chromatin Remodeling In Vivo. *Mol. Cell. Biol.* 24(7):2605–13

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, et al. 2010. GREAT improves

functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* 28(5):495–501

Mellacheruvu D, Wright Z, Couzens AL, Lambert J-P, St-Denis NA, et al. 2013. The CRAPome:

a contaminant repository for affinity purification–mass spectrometry data. *Nat. Methods*.

10(8):730–36

Mellor J, Morillon A. 2004. ISWI complexes in Saccharomyces cerevisiae. *Biochim. Biophys.*

Acta BBA - Gene Struct. Expr. 1677(1–3):100–112

Meno C, Shimono A, Saijoh Y, Yashiro K, Mochida K, et al. 1998. lefty-1 is required for left-

right determination as a regulator of lefty-2 and nodal. *Cell*. 94(3):287–297

Merte J, Jensen D, Wright K, Sarsfield S, Wang Y, et al. 2010. Sec24b selectively sorts Vangl2

to regulate planar cell polarity during neural tube closure. *Nat. Cell Biol.* 12(1):41–46

Mi H, Poudel S, Muruganujan A, Casagrande JT, Thomas PD. 2016. PANTHER version 10:

expanded protein families and functions, and analysis tools. *Nucleic Acids Res.*

44(D1):D336–42

Migliorini D, Denchi EL, Danovi D, Jochemsen A, Capillo M, et al. 2002. Mdm4 (Mdmx)

Regulates p53-Induced Growth Arrest and Neuronal Cell Death during Early Embryonic

Mouse Development. *Mol. Cell. Biol.* 22(15):5527–38

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, et al. 2007. Genome-wide maps of

chromatin state in pluripotent and lineage-committed cells. *Nature*. 448(7153):553–60

Mizuguchi G, Tsukiyama T, Wisniewski J, Wu C. 1997. Role of nucleosome remodeling factor

NURF in transcriptional activation of chromatin. *Mol. Cell*. 1(1):141–150

Moreau J-L, Lee M, Mahachi N, Vary J, Mellor J, et al. 2003. Regulated displacement of TBP from the PHO8 promoter in vivo requires Cbf1 and the Isw1 chromatin remodeling complex. *Mol. Cell*. 11(6):1609–1620

Morgan MAJ, Magnusdottir E, Kuo TC, Tunyaplin C, Harper J, et al. 2009. Blimp-1/Prdm1 Alternative Promoter Usage during Mouse Development and Plasma Cell Differentiation. *Mol. Cell. Biol.* 29(21):5813–27

Morris SA, Baek S, Sung M-H, John S, Wiench M, et al. 2013. Overlapping chromatin-remodeling systems collaborate genome wide at dynamic chromatin transitions. *Nat. Struct. Mol. Biol.* 21(1):73–81

Morrison AJ, Highland J, Krogan NJ, Arbel-Eden A, Greenblatt JF, et al. 2004. INO80 and γ-H2AX interaction links ATP-dependent chromatin remodeling to DNA damage repair. *Cell*. 119(6):767–775

Muller S, Filippakopoulos P, Knapp S. 2011. Bromodomains as therapeutic targets. *Expert Rev. Mol. Med.* 13(e29):

Murawska M, Brehm A. 2011. CHD chromatin remodelers and the transcription cycle. *Transcription*. 2(6):244–53

Murdoch JN, Copp AJ. 2010. The relationship between sonic Hedgehog signaling, cilia, and neural tube defects. *Birt. Defects Res. A. Clin. Mol. Teratol.* 88(8):633–52

Nagy Z, Riss A, Fujiyama S, Krebs A, Orpinell M, et al. 2010. The metazoan ATAC and SAGA coactivator HAT complexes regulate different sets of inducible target genes. *Cell. Mol. Life Sci.* 67(4):611–28

Nakanishi S, Prasad R, Wilson SH, Smerdon M. 2007. Different structural states in

   oligonucleosomes are required for early versus late steps of base excision repair. *Nucleic*

   Acids Res. 35(13):4313–21

Narlikar GJ, Sundaramoorthy R, Owen-Hughes T. 2013. Mechanisms and Functions of ATP-

   Dependent Chromatin-Remodeling Enzymes. *Cell*. 154(3):490–503

Nguyen EB, Westmuckett AD, Moore KL. 2014. SPACA7 Is a Novel Male Germ Cell-Specific

   Protein Localized to the Sperm Acrosome That Is Involved in Fertilization in Mice. *Biol.*

   Reprod. 90(1):16–16

Nguyen TN, Goodrich JA. 2006. Protein-protein interaction assays: eliminating false positive

   interactions. *Nat. Methods*. 3(2):135–39

Nichols DH, Pauley S, Jahan I, Beisel KW, Millen KJ, Fritzsch B. 2008. Lmx1a is required for

   segregation of sensory epithelia and normal ear histogenesis and morphogenesis. *Cell*

   Tissue Res. 334(3):339–58

Noguchi C, Rapp JB, Skorobogatko YV, Bailey LD, Noguchi E. 2012. Swi1 Associates with

   Chromatin through the DDT Domain and Recruits Swi3 to Preserve Genomic Integrity.

   PLoS ONE. 7(8):e43988

Okabe I, Bailey LC, Attree O, Srinivasan S, Perkel JM, et al. 1992. Cloning of human and bovine

   homologs of SNF2/SW12: a global activator of transcription in yeast. *Nucleic Acids Res.*

   20(17):4649–55

Pan Y, Bai CB, Joyner AL, Wang B. 2006. Sonic hedgehog Signaling Regulates Gli2

   Transcriptional Activity by Suppressing Its Processing and Degradation. *Mol. Cell. Biol.*

   26(9):3365–77

Pan Y, Wang C, Wang B. 2009. Phosphorylation of Gli2 by protein kinase A is required for Gli2 processing and degradation and the Sonic Hedgehog-regulated mouse development. *Dev. Biol.* 326(1):177–89

Park CY, Pierce SA, von Drehle M, Ivey KN, Morgan JA, et al. 2010. skNAC, a Smyd1-interacting transcription factor, is involved in cardiac development and skeletal muscle growth and regeneration. *Proc. Natl. Acad. Sci.* 107(48):20750–55

Park JH, Lee SW, Yang SW, Yoo HM, Park JM, et al. 2014. Modification of DBC1 by SUMO2/3 is crucial for p53-mediated apoptosis in response to DNA damage. *Nat. Commun.* 5:5483

Parnell TJ, Schlichter A, Wilson BG, Cairns BR. 2015. The chromatin remodelers RSC and ISW1 display functional and chromatin-based promoter antagonism. *Elife*. 4:e06073

Percipalle P, Fomproix N, Cavellán E, Voit R, Reimer G, et al. 2006. The chromatin remodelling complex WSTF–SNF2h interacts with nuclear myosin 1 and has a role in RNA polymerase I transcription. *EMBO Rep.*

Pessina F, Lowndes NF. 2014. The RSF1 Histone-Remodelling Factor Facilitates DNA Double-Strand Break Repair by Recruiting Centromeric and Fanconi Anaemia Proteins. *PLoS Biol.* 12(5):e1001856

Peterson CL, Herskowitz I. 1992. Characterization of the yeast SWI1, SWI2, and SWI3 genes, which encode a global activator of transcription. *Cell*. 68(3):573–83

Piatti P, Zeilner A, Lusser A. 2011. ATP-Dependent Chromatin Remodeling Factors and Their Roles in Affecting Nucleosome Fiber Composition. *Int. J. Mol. Sci.* 12(12):6544–65

Pillai AM, Thaxton C, Pribisko AL, Cheng J-G, Dupree JL, Bhat MA. 2009. Spatiotemporal ablation of myelinating glia-specific *neurofascin* (Nfasc [NF155]) in mice reveals gradual

loss of paranodal axoglial junctions and concomitant disorganization of axonal domains. *J. Neurosci. Res.* 87(8):1773–93

Pinson KI, Brennan J, Monkley S, Avery BJ, Skarnes WC. 2000. An LDL-receptor-related protein mediates Wnt signalling in mice NATURE |VOL 407 | 28 SEPTEMBER 2000 |www.nature.com. *Nature*. 407:535–38

Poot RA, Bozhenok L, van den Berg DLC, Hawkes N, Varga-Weisz PD. 2005. Chromatin Remodeling by WSTF-ISWI at the Replication Site Opening a Window of Opportunity for Epigenetic Inheritance? Cell Cycle 4:4, 543-546; April 2005]; ©. *Cell Cycle*. 4(4):543–46

Poot RA, Bozhenok L, van den Berg DLC, Steffensen S, Ferreira F, et al. 2004. The Williams syndrome transcription factor interacts with PCNA to target chromatin remodelling by ISWI to replication foci. *Nat. Cell Biol.* 6(12):1236–44

Poot RA, Dellaire G, B.Hulsmann B, A.Grimaldi M, F.V.Corona D, et al. 2000. HuCHRAC, a human ISWI chromatin remodelling complex contains hACF1 and two novel histone-fold proteins. *EMBO J.* 19(13):3377–87

Prost S, Bellamy COC, Cunningham DS, Harrison DJ. 1998. Altered DNA repair and dysregulation of p53 in IRF-1 null hepatocytes. *FASEB J.* 12(2):181–188

Qin B, Minter-Dykhouse K, Yu J, Zhang J, Liu T, et al. 2015. DBC1 Functions as a Tumor Suppressor by Regulating p53 Stability. *Cell Rep.* 10(8):1324–34

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 26(6):841–42

Ramirez-Carrozzi VR, Braas D, Bhatt DM, Cheng CS, Hong C, et al. 2009. A Unifying Model for the Selective Regulation of Inducible Transcription by CpG Islands and Nucleosome Remodeling. *Cell*. 138(1):114–28

Rando OJ, Chang HY. 2009. Genome-Wide Views of Chromatin Structure. *Annu. Rev. Biochem.* 78(1):245–71

Rantakari P, Strauss L, Kiviranta R, Lagerbohm H, Paviala J, et al. 2008. Placenta Defects and Embryonic Lethality Resulting from Disruption of Mouse Hydroxysteroid (17-β) Dehydrogenase 2 Gene. *Mol. Endocrinol.* 22(3):665–75

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, et al. 2011. Integrative genomics viewer. *Nat. Biotechnol.* 29(1):24–26

Roux KJ, Kim DI, Burke B. 2013. BioID: A Screen for Protein-Protein Interactions: BioID Screen for Protein Interactions. In *Current Protocols in Protein Science*, eds. JE Coligan, BM Dunn, DW Speicher, PT Wingfield, p. 19.23.1-19.23.14. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Ruthenburg AJ, Li H, Milne TA, Dewell S, McGinty RK, et al. 2011. Recognition of a Mononucleosomal Histone Modification Pattern by BPTF via Multivalent Interactions. *Cell*. 145(5):692–706

Sadeghifar F, Bohm S, Vintermist A, Ostlund Farrants A-K. 2015. The B-WICH chromatin-remodelling complex regulates RNA polymerase III transcription by promoting Max-dependent c-Myc binding. *Nucleic Acids Res.* 43(9):4477–90

Saha A, Wittmeyer J, Cairns BR. 2006. Chromatin remodelling: the industrial revolution of DNA around histones. *Nat. Rev. Mol. Cell Biol.* 7(6):437–47

Sakurabashi A, Wada-Hiraike O, Hirano M, Fu H, Isono W, et al. 2015. CCAR2 negatively
regulates nuclear receptor LXRα by competing with SIRT1 deacetylase. *J. Steroid
Biochem. Mol. Biol.* 149:80–88

Sala A, Toto M, Pinello L, Gabriele A, Benedetto VD, et al. 2011. Genome-wide characterization
of chromatin binding and nucleosome spacing activity of the nucleosome remodelling
ATPase ISWI. *EMBO J.* 30(9):1766–77

Sanchez R, Zhou M-M. 2009. The role of human bromodomains in chromatin biology and gene
transcription. *Curr. Opin. Drug Discov. Devel.* 12(5):659

Santoro R, Li J, Grummt I. 2002. The nucleolar remodeling complex NoRC mediates
heterochromatin formation and silencing of ribosomal gene transcription. *Nat. Genet.*
32(3):393–96

Saucedo L, Buffa GN, Rosso M, Guillardoy T, Góngora A, et al. 2015. Fibroblast Growth Factor
Receptors (FGFRs) in Human Sperm: Expression, Functionality and Involvement in
Motility Regulation. *PLOS ONE.* 10(5):e0127297

Schindler K, Schultz RM. 2009. CDC14B Acts Through FZR1 (CDH1) to Prevent Meiotic
Maturation of Mouse Oocytes. *Biol. Reprod.* 80(4):795–803

Schinzel A, Schmid W, Fraccaro M, Tiepolo L, Zuffardi O, et al. 1981. The "cat eye syndrome":
dicentric small marker chromosome probably derived from a no. 22 (tetrasomy 22pter→
q11) associated with a characteristic phenotype. *Hum. Genet.* 57(2):148–158

Schneider-Maunoury S, Topilko P, Seitanidou T, Levi G, Cohen-Tannoudji M, et al. 1993.
Disruption of Krox-20 results in alteration of rhombomeres 3 and 5 in the developing
hindbrain Author links open the overlay panel. Numbers correspond to the affiliation list
which can be exposed by using the show more link. *Cell.* 75(6):1199–1214

Schorle H, Meier P, Buchert M, Jaenisch R, Mitchell PJ. 1996. Transcription factor AP-2 essential for cranial closure and craniofacial development. *Nature*. 16(381):235–38

Schwanbeck R, Xiao H, Wu C. 2004. Spatial Contacts and Nucleosome Step Movements Induced by the NURF Chromatin Remodeling Complex. *J. Biol. Chem.* 279(38):39933–41

Shim J-H. 2005. TAK1, but not TAB1 or TAB2, plays an essential role in multiple signaling pathways in vivo. *Genes Dev.* 19(22):2668–81

Shimada K, Oma Y, Schleker T, Kugou K, Ohta K, et al. 2008. Ino80 Chromatin Remodeling Complex Promotes Recovery of Stalled Replication Forks. *Curr. Biol.* 18(8):566–75

Shin H, Liu T, Duan X, Zhang Y, Liu XS. 2013. Computational methodology for ChIP-seq analysis. *Quant. Biol.* 1(1):54–70

Shin H, Liu T, Manrai AK, Liu XS. 2009. CEAS: cis-regulatory element annotation system. *Bioinformatics*. 25(19):2605–6

Siddle K. 2012. Molecular Basis of Signaling Specificity of Insulin and IGF Receptors: Neglected Corners and Recent Advances. *Front. Endocrinol.* 3:

Stanne T, Narayanan MS, Ridewood S, Ling A, Witmer K, et al. 2015. Identification of the ISWI Chromatin Remodeling Complex of the Early Branching Eukaryote *Trypanosoma brucei*. *J. Biol. Chem.* 290(45):26954–67

Stopka T, Skoultchi AI. 2003. The ISWI ATPase Snf2h is required for early mouse development. *Proc. Natl. Acad. Sci.* 100(24):14097–14102

Stottmann RW, Berrong M, Matta K, Choi M, Klingensmith J. 2006. The BMP antagonist Noggin promotes cranial and spinal neurulation by distinct mechanisms. *Dev. Biol.* 295(2):647–63

Strohner R, Nemeth A, Jansa P, Hofmann-Rohrer U, Santoro R, et al. 2001. NoRC—a novel

    member of mammalian ISWI-containing chromatin remodeling machines. *EMBO J.*

    20(17):4892–4900

Sun D-S, Chang AC, Jenkins NA, Gilbert DJ, Copeland NG, Chang N-CA. 1996. Identification,

    molecular characterization, and chromosomal localization of the cDNA encoding a novel

    leucine zipper motif-containing protein. *Genomics*. 36(1):54–62

Taatjes DJ. 2010. The human Mediator complex: a versatile, genome-wide regulator of

    transcription. *Trends Biochem. Sci.* 35(6):315–22

Tallant C, Valentini E, Fedorov O, Overvoorde L, Ferguson FM, et al. 2015. Molecular Basis of

    Histone Tail Recognition by Human TIP5 PHD Finger and Bromodomain of the

    Chromatin Remodeling Complex NoRC. *Structure*. 23(1):80–92

Tanaka T, Hosokawa M, Vagin VV, Reuter M, Hayashi E, et al. 2011. Tudor domain containing

    7 (Tdrd7) is essential for dynamic ribonucleoprotein (RNP) remodeling of chromatoid

    bodies during spermatogenesis. *Proc. Natl. Acad. Sci.* 108(26):10579–84

Tanikawa M, Wada-Hiraike O, Yoshizawa-Sugata N, Shirane A, Hirano M, et al. 2013. Role of

    multifunctional transcription factor TFII-I and putative tumour suppressor DBC1 in cell

    cycle and DNA double strand damage repair. *Br. J. Cancer*. 109(12):3042–48

Tate P, Lee M, Tweedie S, Skarnes WC, Bickmore WA. 1998. Capturing novel mouse genes

    encoding chromosomal and other nuclear proteins. *J. Cell Sci.* 111(17):2575–2585

Tatulian SA. 2015. Structural Dynamics of Insulin Receptor and Transmembrane Signaling.

    Biochemistry (Mosc.). 54(36):5523–32

Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ. 2007. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* 14(11):1025–40

Thompson PJ, Norton KA, Niri FH, Dawe CE, McDermid HE. 2012a. CECR2 Is Involved in Spermatogenesis and Forms a Complex with SNF2H in the Testis. *J. Mol. Biol.* 415(5):793–806

Thompson PJ, Norton KA, Niri FH, Dawe CE, McDermid HE. 2012b. CECR2 Is Involved in Spermatogenesis and Forms a Complex with SNF2H in the Testis. *J. Mol. Biol.* 415(5):793–806

Toiber D, Erdel F, Bouazoune K, Silberman DM, Zhong L, et al. 2013. SIRT6 Recruits SNF2H to DNA Break Sites, Preventing Genomic Instability through Chromatin Remodeling. *Mol. Cell*. 51(4):454–68

Toto M, D'Angelo G, Corona DFV. 2014. Regulation of ISWI chromatin remodelling activity. *Chromosoma*. 123(1–2):91–102

Tsukiyama T, Wu C. 1995. Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell*. 83(6):1011–20

Tumurbaatar I, Cizmecioglu O, Hoffmann I, Grummt I, Voit R. 2011. Human Cdc14B Promotes Progression through Mitosis by Dephosphorylating Cdc25 and Regulating Cdk1/Cyclin B Activity. *PLoS ONE*. 6(2):e14711

Turkington RC, Longley DB, Allen WL, Stevenson L, McLaughlin K, et al. 2014. Fibroblast growth factor receptor 4 (FGFR4): a targetable regulator of drug resistance in colorectal cancer. *Cell Death Dis.* 5(2):e1046

Valouev A, Johnson DS, Sundquist A, Medina C, Anton E, et al. 2008. Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nat. Methods*. 5(9):829–34

van Amerongen R, Nusse R. 2009. Towards an integrated view of Wnt signaling in development. *Development*. 136(19):3205–14

Vanolst L. 2005. Toutatis, a TIP5-related protein, positively regulates Pannier function during Drosophila neural development. *Development*. 132(19):4327–38

Varga-Weisz PD. 2010. Insights into how chromatin remodeling factors find their target in the nucleus. *Proc. Natl. Acad. Sci.* 107(46):19611–12

Varga-Weisz PD, Wilm M, Bonte E, Dumas K, Mann M, Becker PB. 1997. Chromatin-remodelling factor CHRACcontains theATPases ISWI and topoisomerase II. *NATURE*. 388:598–602

Vary JC, Gangaraju VK, Qin J, Landel CC, Kooperberg C, et al. 2003. Yeast Isw1p Forms Two Separable Complexes In Vivo. *Mol. Cell. Biol.* 23(1):80–91

Vasicova P, Stradalova V, Halada P, Hasek J, Malcova I. 2013. Nuclear Import of Chromatin Remodeler Isw1 Is Mediated by Atypical Bipartite cNLS and Classical Import Pathway: Nuclear Import of Isw1. *Traffic*. 14(2):176–93

Vasileva A, Tiedau D, Firooznia A, Müller-Reichert T, Jessberger R. 2009. Tdrd6 Is Required for Spermiogenesis, Chromatoid Body Architecture, and Regulation of miRNA Expression. *Curr. Biol.* 19(8):630–39

Vicent GP, Nacht AS, Font-Mateu J, Castellano G, Gaveglia L, et al. 2011. Four enzymes cooperate to displace histone H1 during the first minute of hormonal gene activation. *Genes Dev.* 25(8):845–62

Vidi P-A, Liu J, Salles D, Jayaraman S, Dorfman G, et al. 2014. NuMA promotes homologous recombination repair by regulating the accumulation of the ISWI ATPase SNF2h at DNA breaks. *Nucleic Acids Res.* 42(10):6365–79

Vincent JA, Kwong TJ, Tsukiyama T. 2008. ATP-dependent chromatin remodeling shapes the DNA replication landscape. *Nat. Struct. Mol. Biol.* 15(5):477–84

Vintermist A, Böhm S, Sadeghifar F, Louvet E, Mansén A, et al. 2011. The Chromatin Remodelling Complex B-WICH Changes the Chromatin Structure and Recruits Histone Acetyl-Transferases to Active rRNA Genes. *PLoS ONE*. 6(4):e19184

Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, et al. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*. 457(7231):854–58

Vogel C, Bashton M, Kerrison ND, Chothia C, Teichmann SA. 2004. Structure, function and evolution of multidomain proteins. *Curr. Opin. Struct. Biol.* 14(2):208–16

Vogel-Ciernia A, Wood MA. 2014. Neuron-specific chromatin remodeling: A missing link in epigenetic mechanisms underlying synaptic plasticity, memory, and intellectual disability disorders. *Neuropharmacology*. 80:18–27

Wang J, Rao S, Chu J, Shen X, Levasseur DN, et al. 2006. A protein interaction network for pluripotency of embryonic stem cells. *Nature*. 444(7117):364–68

Wang L, Zhao Z, Meyer MB, Saha S, Yu M, et al. 2014. CARM1 Methylates Chromatin Remodeling Factor BAF155 to Enhance Tumor Progression and Metastasis. *Cancer Cell*. 25(1):21–36

Wei Z, Peddibhotla S, Lin H, Fang X, Li M, et al. 2011. Early-Onset Aging and Defective DNA Damage Response in Cdc14b-Deficient Mice. *Mol. Cell. Biol.* 31(7):1470–77

Weinstein M, Xu X, Ohyama K, Deng C-X. 1998. FGFR-3 and FGFR-4 function cooperatively to direct alveogenesis in the murine lung. *Development*. 125(18):3615–3623

Whitehouse I, Rando OJ, Delrow J, Tsukiyama T. 2007. Chromatin remodelling at promoters suppresses antisense transcription. *Nature*. 450(7172):1031–35

Wiechens N, Singh V, Gkikopoulos T, Schofield P, Rocha S, Owen-Hughes T. 2016. The Chromatin Remodelling Enzymes SNF2H and SNF2L Position Nucleosomes adjacent to CTCF and Other Transcription Factors. *PLOS Genet.* 12(3):e1005940

Wilson BG, Roberts CWM. 2011. SWI/SNF nucleosome remodellers and cancer. *Nat. Rev. Cancer*. 11(7):481–92

Wit JN d., Kessel T va. 1996. Effects of ionic strength on the solubility of whey protein products. A colloid chemical approach. *Food Hydrocoll.* 10(2):143–49

Won K-J, Agarwal S, Shen L, Shoemaker R, Ren B, Wang W. 2009. An Integrated Approach to Identifying Cis-Regulatory Modules in the Human Genome. *PLoS ONE*. 4(5):e5501

Woodcock CL, Ghosh RP. 2010. Chromatin Higher-order Structure and Dynamics. *Cold Spring Harb. Perspect. Biol.* 2(5):a000596–a000596

Wu JI. 2012. Diverse functions of ATP-dependent chromatin remodeling complexes in development and cancer. *Acta Biochim. Biophys. Sin.* 44(1):54–69

Wu JI, Lessard J, Olave IA, Qiu Z, Ghosh A, et al. 2007. Regulation of Dendritic Development by Neuron-Specific Chromatin Remodeling Complexes. *Neuron*. 56(1):94–108

Wurdak H, Ittner LM, Lang KS, Leveen P, Suter U, et al. 2005. Inactivation of TGFbeta signaling in neural crest stem cells leads to multiple defects reminiscent of DiGeorge syndrome. *Genes Dev.* 19(5):530–35

Xiao A, Li H, Shechter D, Ahn SH, Fabrizio LA, et al. 2009. WSTF regulates the H2A.X DNA damage response via a novel tyrosine kinase activity. *Nature*. 457(7225):57–62

Xiao H, Sandaltzopoulos R, Wang H-M, Hamiche A, Ranallo R, et al. 2001. Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. *Mol. Cell*. 8(3):531–543

Xu D, Muniandy P, Leo E, Yin J, Thangavel S, et al. 2010. Rif1 provides a new DNA-binding interface for the Bloom syndrome complex to maintain normal replication. *EMBO J.* 29(18):3140–3155

Yadon AN, Van de Mark D, Basom R, Delrow J, Whitehouse I, Tsukiyama T. 2010. Chromatin Remodeling around Nucleosome-Free Regions Leads to Repression of Noncoding RNA Transcription. *Mol. Cell. Biol.* 30(21):5110–22

Yang JG, Madrid TS, Sevastopoulos E, Narlikar GJ. 2006. The chromatin-remodeling enzyme ACF is an ATP-dependent DNA length sensor that regulates nucleosome spacing. *Nat. Struct. Mol. Biol.* 13(12):1078–83

Ybot-Gonzalez P, Cogram P, Gerrelli D, Copp AJ. 2002. Sonic hedgehog and the molecular regulation of mouse neural tube closure. *Development*. 129(10):2507–2517

Ye Y, Xiao Y, Wang W, Wang Q, Yearsley K, et al. 2009. Inhibition of Expression of the Chromatin Remodeling Gene, SNF2L, Selectively Leads to DNA Damage, Growth Inhibition, and Cancer Cell Death. *Mol. Cancer Res.* 7(12):1984–99

Yen K, Vinayachandran V, Batta K, Koerber RT, Pugh BF. 2012. Genome-wide Nucleosome Specificity and Directionality of Chromatin Remodelers. *Cell*. 149(7):1461–73

Yip DJ, Corcoran CP, Alvarez-Saavedra M, DeMaria A, Rennick S, et al. 2012a. Snf2l Regulates Foxg1-Dependent Progenitor Cell Expansion in the Developing Brain. *Dev. Cell*. 22(4):871–78

Yip DJ, Corcoran CP, Alvarez-Saavedra M, DeMaria A, Rennick S, et al. 2012b. Snf2l Regulates Foxg1-Dependent Progenitor Cell Expansion in the Developing Brain. *Dev. Cell*. 22(4):871–78

Yu H, Smallwood PM, Wang Y, Vidaltamayo R, Reed R, Nathans J. 2010. Frizzled 1 and frizzled 2 genes function in palate, ventricular septum and neural tube closure: general implications for tissue fusion processes. *Development*. 137(21):3707–17

Yung YC, Stoddard NC, Chun J. 2014. LPA receptor signaling: pharmacology, physiology, and pathophysiology. *J. Lipid Res.* 55(7):1192–1214

Zhang J. 2012. *Protein-Protein Interactions in Salt Solutions*. INTECH Open Access Publisher

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9(9):R137

Zhang Z, Witham S, Alexov E. 2011. On the role of electrostatics in protein–protein interactions. *Phys. Biol.* 8(3):35001

Zhao Q, Behringer R, Crombrugghe B. 1996. Prenatal folic acid treatment suppresses acrania and meroanencephaly in mice mutant for the Cart1 homeobox gene. *Nat. Genet.* 13(3):275–83

Zhao W, Kruse J-P, Tang Y, Jung SY, Qin J, Gu W. 2008. Negative regulation of the deacetylase SIRT1 by DBC1. *Nature*. 451(7178):587–90

Zhongyi S, Rantakari P, Lamminen T, Toppari J, Poutanen M. 2007. Transgenic Male Mice Expressing Human Hydroxysteroid Dehydrogenase 2 Indicate a Role for the Enzyme Independent of Its Action on Sex Steroids. *Endocrinology*. 148(8):3827–36

Zhou CJ, Wang Y-Z, Yamagami T, Zhao T, Song L, Wang K. 2009. Generation of Lrp6
conditional gene-targeting mouse line for modeling and dissecting multiple birth
defects/congenital anomalies. *Dev. Dyn.* NA-NA

Zhou P, He A, Pu WT. 2012. Regulation of GATA4 Transcriptional Activity in Cardiovascular
Development and Disease. In *Current Topics in Developmental Biology*, Vol. 100, pp.
143–69. Elsevier

Zhou Y, Grummt I. 2005. The PHD Finger/Bromodomain of NoRC Interacts with Acetylated
Histone H4K16 and Is Sufficient for rDNA Silencing. *Curr. Biol.* 15(15):1434–38

Zohn IE, Anderson KV, Niswander L. 2007. The Hectd1 ubiquitin ligase is required for
development of the head mesenchyme and neural tube closure. *Dev. Biol.* 306(1):208–21

Zou D, Erickson C, Kim E-H, Jin D, Fritzsch B, Xu P-X. 2008. Eya1 gene dosage critically
affects the development of sensory epithelia in the mammalian inner ear. *Hum. Mol.
Genet.* 17(21):3340–56

Appendices

Appendix A: List of primers used in this research

Oligonucleotides sequences used in this project		
Name	Sequence 5'-3'	Use
IngeniousLox1	TTAGAATAGGTGAGGGAGGAG	*Cecr2$^{tm1.1Hemc}$* genotyping
Ingenious SDL2	GTAGCGCCTATTTGTAATGGTCA	
LoxCECR2_DEL3R	AATGGTGGCGAAATCAACTC	
Mmu Intron7 F4	CCCCATTTATTTGCTTGAGCTG	*Cecr2^{Gt45Bic}* genotyping
Mmu Intron7 R4	CACGAACAATGGAAGGAATGA	
pGT1R4	ACGCCATACAGTCCTCTTCACATC	
SRY FOR	GAGAGCATGGAGGGGCAT	
SRY REV	CCACTCCTCTGTGACACT	
WholeCECR2_F	CTA***GAATTC***ATGTGCCCGGAGGAAGGC	*Cecr2* fragment 1-9 cloning
CECR2SalI9R	CTA***GTCGAC***CTAGTGGTGGTGGT GGTGGTGTTCCACTGCCTTGACCTTTT	
CECR2Ecoi17F	CTA ***GAA TTC*** ATG ACA TGT CCA CAG GGC TTT TC	*Cecr2* fragment 17-19 cloning
CECR2Sall19R	CTA ***GTC GAC*** CTA GTG GTG GTG GTG GTG GTG GCT CTG ATC CAG AGG AAG CGT	
ChIP Adapater1	AATGATACGGCGACCACCGAGATCTA CACTCTTTCCCTACACGACGCTCTTCCGATCT	Building library
ChIP Adapater2	GATCGGAAGAGCACACGTCTGAACTCCAGT CACCGATGTATCTCGTATGCCGTCTTCTGCTTG	Building library
ChIP primer1	AATGATACGGCGACCACCGAGATCTACAC	ChIP Primer Mix
ChIP primer2	CAAGCAGAAGACGGCATACGAGAT	

All primers were obtained from Integrated DNA Technologies.

Appendix B: CECR2 gene fragment1 and fragment2 nucleotide and amino acid sequences used for raising rabbit polyclonal antibody. Two fragments of mouse CECR2 protein were selected and used in the production of rabbit polyclonal antibodies to mouse CECR2 protein. Each *Cecr2* gene fragment was amplified using forward and reverse primers (in blue) containing EcoRI and SalI restriction sites (in bold blue) and a sequence coding for an N or C-terminal Hexahisitidine (His6)-tag (in green).

```
Fragment 1: exons 1-9

Primers used:
Left: WholeCECR2_F (B3P80)
5'-CTAGAATTCATGTGCCCGGAGGAAGGC-3'
GAATTC= EcoR1

Right: CECR2SalI9R
5'-CTAGTCGACCTAGTGGTGGTGGTGGTGGTGTTCCACTGCCTTGACCTTTT-3'
GTCGAC=SalI

CTAGAATTCATGTGCCCGGAGGAAGGCGGCGCGGCCGGGCTGGGCGAGCTCCGCTCCTGGTGGGAAGTCCCGGCCATC
GCGCACTTCTGCTCGCTCTTTCGCACGGCGTTCCGCCTGCCCGACTTCGAGATCGAGGAGTTAGAGGCAGCTCTTCAC
AGAGATGACGTGGAGTTTATCAGTGATCTGATTGCCTGCTTGCTTCAAGGCTGCTATCAGCGAAGAGATATCACGCCT
CAGACATTCCACAGCTACTTGGAAGATATCATTAACTACCGTTGGGAGCTTGAAGAAGGAAAGCCCAATCCTCTGAGA
GAAGCTAGTTTCCAAGACCTTCCTCTGCGTACTCGAGTAGAGATCCTGCATCGCCTCTGTGATTACCGGCTTGATGCA
GATGATGTCTTTGATCTTCTCAAGGGTCTGGATGCAGACAGTCTCAGAGTTGAACCTCTGGGTGAAGACAATTCAGGT
GCACTCTACTGGTACTTCTATGGGACACGAATGTACAAAGAAGACCCAGTACAAGGAAGGTCCAATGGAGAACTCTCC
TTGTGCAGGGAAAGTGAAAGGCAAAAAAATGTCTCAAATGTTCCTGGAAAAACAGGCAAAAGAAGAGGAAGGCCTCCA
AAGCGAAAAAAGCTACAGGAGGAGATTATATCCAGTGAAAAGCAGGAAGAAAACTCCTTGACATCTGACCTACAGACA
AGAAATGGATCCCGAGGGCCAGGCCAAGGTACTTGGTGGCTCCTGTGTCAAACAGAAGAAGAATGGAGACAGGTCACT
GAGAGCTTCCGAGAGCGGACCTCTCTCCGAGAGCGGCAACTCTACAAGCTCCTTAGCGAGGACTTCCTGCCTGAAATT
TGCAACATGATTGCCCAGAAGGGAAAACGTCCACAGCGCACAAAGCCAGAGTTGCAACATAGGTTTATGTCTGACCAC
CTGTCCATCAAATCCATCAAGCTAGAGGAAACTCCCATGCTCACCAAGATAGAAAAACAGAAGCGCAGGGAGGAGGAG
GAGGAGCGACAGCTCCTTCTTGCCGTGCAGAAGAAGGAGCAAGAGCAGATGTTAAAGGAGGAGAGAAAGCGGGAAATG
GAGGAAAAGGTCAAGGCAGTGGAACACCACCACCACCACCACTAGGTCGACTAG
```

```
Polypeptide ~45 kDa
5'→3'
```

Met C P E E G G A A G L G E L R S W W E V P A I A H F C S L F R T A F R L P D F E I E E L E A A L H R D D V E F I S D L I A C L L Q G C Y Q R R D I T P Q T F H S Y L E D I I N Y R W E L E E G K P N P L R E A S F Q D L P L R T R V E I L H R L C D Y R L D A D D V F D L L K G L D A D S L R V E P L G E D N S G A L Y W Y F Y G T R **Met** Y K E D P V Q G R S N G E L S L C R E S E R Q K N V S N V P G K T G K R R G R P P K R K K L Q E E I I S S E K Q E E N S L T S D L Q T R N G S R G P G Q G T W W L L C Q T E E E W R Q V T E S F R E R T S L R E R Q L Y K L L S E D F L P E I C N **Met** I A Q K G K R P Q R T K P E L Q H R F **Met** S D H L S I K S I K L E E T P **Met** L T K I E K Q K R R E E E E E R Q L L L A V Q K K E Q E Q **Met** L K E E R K R E **Met** E E K V K A V E H H H H H H **Stop**

Fragment 2: exons 17-19

Primers used:
Left: CECR2Ecoi17F
5'-CTA**GAATTC**ATGACATGTCCACAGGGCTTTTC-3'

Right: CECR2Sall19R
5'-CTA**GTCGAC**CTAGTGGTGGTGGTGGTGGTGGCTCTGATCCAGAGGAAGCGT-3'

CTA**GAATTC**ATGACATGTCCACAGGGCTTTTCTGATTGGCAGAGATCTCTCCCTTCCCAGAGAAGCCCAAGTGGACCC
CCAGGAAGTCACCCTCCACGCTCCCTCTTTTCAGAGAAGAATGTCCTGTCTAGTCTGCAAGGTTGTGAAACATTAAAT
ACTGCCTTAACTTCTCCAACTCAAATGGATGTAGTGACTGCTAAAGTTGTTCCTCCTGATGGCACAATTCTGGTCCA
GAAGAAGAAAAGATGGATGAATCTGTGGAGAGGCCAGAGAGTCCCAAAGAATTTCTAGATCTCGACAACCATAATGCA
GCTACCAAGCGGCAGAACTCACTGTCAACCAGTGACTATCTTTATGGAACTCCACCACCATCTCTGAGTTCAGGAATG
ACTTTTGGTTCATCTGCTTTCCCACCCCACAGTGTGATGCTGCAGACAGGGTCTCCATACACTCCTCAGCGATCAGCC
AGTCATTTTCAGCCCAGGGCATACCCATCCCCTGTGCCTGCCCATCCACCTCCCCATCCAGTAGCCACCCAACCCAAT
GGCCTCTCTCCAGAGGATTCCCTCTACTGCTGTCAGGAAGAGGGCCTAGGTCACTTTCAGGCTTCGATGATGGAGCAG
ACTGGTACCGGAAGCGGATTAAGAGGCTCATTTCAAGAAGTACACAGACCACCAGGATTGCAGATGCATCCAGTGCAG
TCACAGTCCTTATTCCCAAAGACCCCGGCACCTGCAGCATCACCAGAACAGTTGCCACCGCATAAGACCCCA<u>ACG**CTT**</u>
<u>**CCTC**</u>TGGATCAGAGCCACCACCACCACCACCACTAG**GTCGAC**TAG

Polypeptide ~30 KDa
5'→3'

Met T C P Q G F S D W Q R S L P S Q R S P S G P P G S H P P R S L F S E K N V L S S L Q
G C E T L N T A L T S P T Q **Met** D V V T A K V V P P D G H N S G P E E E K **Met** D E S V
E R P E S P K E F L D L D N H N A A T K R Q N S L S T S D Y L Y G T P P P S L S S G **Met**
T F G S S A F P P H S V **Met** L Q T G S P Y T P Q R S A S H F Q P R A Y P S P V P A H P P
P H P V A T Q P N G L S P E D S L Y C C Q E E G L G H F Q A S **Met** **Met** E Q T G T G S G
L R G S F Q E V H R P P G L Q **Met** H P V Q S Q S L F P K T P A P A A S P E Q L P P H K T
P T L P L D Q S H H H H H H **Stop**

Appendix C: Conditions used for large scale IP reactions for mass spectrometry analysis.
The table includes the conditions for 9 experiments for ES cells, 5 experiments for adult FVB/N testis and an experiment for neural stem cells. List of identified proteins are shown in appendices D-O. CECR2-containing complexes were isolated from ES cells lysate. Co-immunoprecipitated proteins were separated by SDS-PAGE gel and stained with Coomassie blue. The gel were cut and submitted to liquid chromatography (LC-MS/MS) mass spectrometry analysis. Proteins also identified in the IgG control were considered as non-specific interactions and removed from the list, as were known background contaminants.

Experi ment	Tissue	IP condition	Treatments	Gel pieces submitted (kDa)	Appendix
#1	TT2	NaCl:420mM, Agarose beads, O/N IP, 3X5min wash, Ag-Antibody first	No CTRL	130, 170	Table 3-1
#2	TT2	NaCl:420mM, Agarose beads, O/N IP, 3X5min wash, Ag-Antibody first	No CTRL EtBr RNase	75, 130, 200, 250	App-D
#3	TT2	NaCl:420mM, Agarose beads, O/N IP, Antigen-Antibody first	IgG as CTRL EtBr RNase	100-200	App-E
#4	TT2	NaCl:420mM, Agarose beads, O/N IP, 3X5min wash, Ag-Antibody first	IgG as CTRL	100-200	App-F
#5	TT2	NaCl:420mM, 2hrs IP, Agarose beads	IgG as CTRL	130, 150	App-G
#6	TT2	NaCl:420mM Dynabeads, O/N IP, Ag-beads first 3Xwash in NaCl:150mM	IgG as CTRL	Whole gel: 12 gel pieces	Table 3-2
#7	TT2	NaCl:420mM, Dynabeads, O/N IP, 4Xwash in NaCl:150mM, Ag-beads first	IgG as CTRL	Whole gel: 3 gel pieces	App-H
#8	TT2	NaCl:150mM, Dynabeads, O/N IP, 3Xwash in NaCl:150mM, Ag-beads first	IgG as CTRL	Whole gel: 3 gel pieces	App-I
#9	TT2	NaCl:420mM, Dynabeads, 2.5hr IP, Cross-linked Ab-dynabeads	IgG as CTRL (Chemistry facility)	Whole gel: 3 gel pieces (excluding Ig band)	App-J

#1	Testis	NaCl:420mM, Agarose beads, O/N IP, Antigen-Antibody first	EtBr RNase IgG as CTRL	Whole gel: 3 gel pieces	Table 3-2
#2	Testis	NaCl:420mM, Nuclear extract, Agarose beads, 2hrs IP, Antigen-Antibody first	IgG as CTRL	130, 170	App-K
#3	Testis	NaCl:150mM, Dynabeads, O/N IP, 3Xresuspension in [150], Ag-beads first	Del/Del as CTRL	Whole gel: 3 gel pieces	App-L
#4	Testis	NaCl:420mM, Dynabeads, O/N IP, Antigen-beads first, 5Xwash in NaCl:420mM	Del/Del as CTRL	Whole gel: 3 gel pieces	App-M
#5	Testis	NaCl:420mM, Dynabeads, 3hrs IP Cross-linked Ab-dynabeads	IgG as CTRL Ingel digestion	Whole gel: 10 gel pieces	App-N
#1	Neurosphere	NaCl:150mM, Dynabeads, O/N IP, 3Xwash in NaCl:150mM, Ag-beads first	Del/Del as CTRL	Whole gel: 3 gel pieces	App-O

Appendix D Experiment # 2 using ES cells. Proteins identified using LC-MS/MS. Only gel pieces corresponding to 130 kDa and 200 kDa were analyzed. In this experiment IgG control was used. See Appendix C for the test condition. The confirmed components of the CECR2 complex are highlighted.

Accession (UniProt)	Gene name	Score[1]	Coverage%[2]	# Peptides[3]
Q6PR54	*Rif1*	50.10	7.32	11
Q91ZW3	*Smarca5 (Snf2h))*	9.45	3.71	3
Q01320	*Top2a*	5.65	1.64	2

[1] Score: (Sequest Score) The sum of the ion scores of all peptides that were identified.
[2] Coverage: The percentage of the protein sequence covered by identified peptides.
[3] # Peptides: The total number of distinct peptide sequences identified in the protein group

Appendix E: Experiment # 3 using ES cells. Proteins identified using LC-MS/MS. Only gel pieces corresponding to 100-200 kDa were analyzed. ES cell lysate was treated with EtBr and RNase A in this experiment to remove the interactions mediated by DNA and RNA. See the Appendix C for the test condition. The confirmed components of the CECR2 complex are highlighted.

Accession (UniProt)	Gene name	Score[1]	# Peptides[2]
Q8VDP4	*Ccar2*	105.35	16
Q64523	*Hist2h2ac*	68.65	1
Q7TT37	*Ikbkap*	36.48	7
Q64522	*Hist2h2ab*	28.84	2
Q8BMB0	*Emsy*	14.95	4
O09106	*Hdac1*	14.10	3
Q569Z6	*Thrap3*	12.71	2
P28656	*Nap1l1*	12.38	3
Q7JJ13	*Brd2*	12.29	4
Q91ZW3	*Smarca5(SNF2H)*	11.58	3
Q5SVQ0	*Kat7*	10.85	2
Q0P678	*Zc3h18*	10.84	3
P29341	*Pabpc1*	10.08	2
Q9Z204	*Hnrnpc*	9.98	2
Q8R4U6	*Top1mt*	9.04	2
Q3UMU9	*Hdgfrp2*	8.66	3
Q60596	*Xrcc1*	8.36	2
Q8R2M2	*Dnttip2*	7.42	3
P60122	*Ruvbl1*	7.14	3
Q3UKC1	*Tax1bp1*	7.10	2
P62996	*Tra2b*	6.55	2
Q9ESX5	*Dkc1*	6.10	2
P25206	*Mcm3*	5.97	2
Q9JHJ0	*Tmod3*	5.52	2
P68040	*Rack1*	5.35	2
O35286	*Dhx15*	5.34	2
Q8K310	*Matr3*	5.10	2
A2AQ19	*Rtf1*	4.60	2

[1] Score: (Sequest Score) The sum of the ion scores of all peptides that were identified.
[2] # Peptides: The total number of distinct peptide sequences identified in the protein group

Appendix F: Experiment # 4 using ES cells. Proteins identified using LC-MS/MS. Only gel pieces corresponding to 130 and 200 kDa were analyzed. ES cell lysate was treated with EtBr and RNase A in this experiment to remove the interactions mediated by DNA and RNA. See the Appendix C for the test condition. The confirmed components of the CECR2 complex are highlighted.

Accession (UniProt)	Gene name	Score[1]	# Peptides[2]
P09405	*Ncl*	43.54	3
Q7TT37	*Ikbkap*	22.54	3
Q8VDP4	*Ccar2*	17.43	4
P23116	*Eif3a*	6.66	2
Q9DBD5	*Pelp1*	6.05	2
Q7TPV4	*Mybbp1a*	4.84	2

[1] Score: (Sequest Score) The sum of the ion scores of all peptides that were identified.
[2] # Peptides: The total number of distinct peptide sequences identified in the protein group

Appendix G: Experiment # 5 using ES cells. Proteins identified using LC-MS/MS. Only gel pieces corresponding to 130 and 150 kDa were analyzed. In this experiment antibody and antigen reaction was incubated for 2 hours instead of overnight. See Appendix C for the test condition. The confirmed components of the CECR2 complex are highlighted.

Accession (UniProt)	Gene name	Score[1]	# Peptides[2]
Q91ZW3	*Smarca5 (Snf2h)*	302.54	18
Q6PGB8	*Smarca1(Snf2l)*	207.37	7
Q6PDG5	*Smarcc2*	72.09	6
Q8R4U7	*Luzp1*	62.64	14
P97496	*Smarcc1*	34.90	6
Q8K4P0	*Wdr33*	26.33	5
Q3TKT4	*Smarca4*	24.45	5
Q7TQH0	*Atxn2l*	16.90	4
Q8BX22	*Sall4*	16.14	6
Q8CFI7	*Polr2b*	12.34	5
Q6NZQ4	*Paxip1*	12.13	3
E9QA25	*Cecr2*	10.58	9
Q8VDP4	*Ccar2*	10.40	2
Q8CGZ0	*Cherp*	10.02	2
Q9CZU3	*Skiv2l2*	9.70	2
Q9DBE9	*Ftsj3*	7.62	3
O35691	*Pnn*	7.60	3
Q9Z207	*Diaph3*	6.35	2
Q8R2M2	*Dnttip2*	6.06	2
Q8VDM6	*Hnrnpul1*	5.78	2
Q8K019	*Bclaf1*	5.65	2
O08810	*Eftud2*	5.56	2
Q8R5K4	*Nol6*	4.44	2

[1] Score: (Sequest Score) The sum of the ion scores of all peptides that were identified.
[2] # Peptides: The total number of distinct peptide sequences identified in the protein group

Appendix H: Experiment # 7 using ES cells. Proteins identified using LC-MS/MS. Co-immunoprecipitated proteins were separated by SDS-PAGE gel. The gel was cut into bands and submitted to liquid chromatography (LC-MS/MS) mass spectrometry analysis. In this experiment IP reactions were washed with less stringent condition. The confirmed components of the CECR2 complex are highlighted.

Accession (UniProt)	Gene name	Score[1]	# Peptides[2]
E9QA25	*Cecr2*	65.80	13
Q8C7S2	*Lima1*	32.95	7
Q5SQB0	*Npm1*	32.47	5
Q641N9	*Kpna2*	31.07	7
Q3UJZ7	*Nop56*	29.50	8
Q3UBZ3	*Capza2*	21.56	4
A0JLV3	*Hist1h2bj*	17.98	3
Q3TS88	*Ppp1r12a*	17.15	4
Q5SW83	*Actr2*	13.98	5
Q3UBB6	*Nolc1*	13.71	3
Q3UAS2	*Capza1*	13.57	3
Q3U0F8	*Smarca5 (Snf2h)*	12.54	4
Q6GQV8	*Specc1l*	12.46	2
B7XGA6	*Alpl*	11.59	4
F6S5I0	*Mprip*	10.49	2
F6Z6F4	*Smarca1 (Snf2l)*	10.46	3
Q3TEU8	*Coro1c*	10.43	2
Q3UJS2	*Fbl*	9.99	4
Q61769	*Mki67*	9.14	3
Q3TBA2	*Arpc1b*	9.13	3
H0YT69	*LUZP1*	8.48	1
Q3UFZ6	*Caprin1*	7.66	2
Q99KC3	*Nop2*	7.37	2
Q3U868	*Parp1*	6.87	3
Q3U931	*G3bp2*	5.78	2
D3YXT4	*Gm5619*	5.54	2
Q3TTX0	*Matr3*	5.52	2
D3YZ09	*Gar1*	4.96	2
Q3TII3	*Eef1a1*	4.87	2
Q9R2A1	*Hsc70t*	4.84	2
H3BLF7	*Gnb1*	4.57	2
H7C4G5	*Ruvbl1*	Rerun*	1

[1] Score: (Sequest Score) The sum of the ion scores of all peptides that were identified.
[2] # Peptides: The total number of distinct peptide sequences identified in the protein group
[3] * Detected in second run against customized database.

Appendix I: Experiment # 8 using ES cells. Proteins identified using LC-MS/MS. Co-immunoprecipitated proteins were separated by SDS-PAGE gel. The gel was cut into bands and submitted to liquid chromatography (LC-MS/MS) mass spectrometry analysis. In this experiment Protein extractions were performed in salt concentration of 150 mM instead of 420 mM (Appendix C). The confirmed components of the CECR2 complex are highlighted.

Accession (UniProt)	Gene name	Score[1]	# Peptides[2]
E9QA25	*Cecr2**	164	42
Q6PR54	*Rif1*	147	37
Q91ZW3	*Smarca5*	142	36
Q01320	*Top2a*	55	17
Q6NZJ6	*Eif4g1*	56	15
Q9JHU4	*Dync1h1*	39	12
G5E870	*Trip12*	33	11
P68040	*Rack1*	25	8
Q6P4T2	*Snrnp200*	24	8
B2RQC6	*Cad*	21	7
Q6NS46	*Pdcd11*	21	7
P17182	*Eno1*	19	6
P51881	*Slc25a5*	16	6
P62827	*Ran*	21	6
P17225	*Ptbp1*	16	5
Q6DFW4	*Nop58*	18.11	5
Q8CGC7	*Eprs*	13.45	5
Q8K3F7	*Tdh*	25.55	5
Q99PV0	*Prpf8*	17.41	5
Q9D6Z1	*Nop56*	19.59	5
Q9JIK5	*Ddx21*	14.66	5
P29341	*Pabpc1*	11.76	4
P49718	*Mcm5*	10.56	4
P54276	*Msh6*	12.85	4
Q61990	*Pcbp2*	10.74	4
Q64012	*Raly*	13.85	4
Q6P5F9	*Xpo1*	20.93	4
Q8K363	*Ddx18*	12.20	4
Q9ERK4	*Cse1l*	12.68	4
Q9ERU9	*Ranbp2*	12.54	4
P70372	*Elavl1*	11.14	3
P70398	*Usp9x*	7.73	3
Q60865	*Caprin1*	9.66	3
Q6P5B0	*Rrp12*	9	3
Q6PGB8	*Smarca1 (Snf2l)*	41.99	3
Q6ZQ08	*Cnot1*	7.61	3
Q8BX17	*Gemin5*	8.3	3
Q8K310	*Matr3*	13.69	3

Q8VEM8	*Slc25a3*	8.88	3
Q99NB9	*Sf3b1*	12.39	3
Q9EPU0	*Upf1*	7.31	3
A2AN08	*Ubr4*	6.67	2
P63087	*Ppp1cc*	7.67	2
Q02257	*Jup*	5.87	2
Q03265	*Atp5a1*	7.61	2
Q1HFZ0	*Nsun2*	5.79	2
Q3UJB9	*Edc4*	7.61	2
Q5F2E7	*Nufip2*	5.61	2
Q60848	*Hells*	5.51	2
Q62167	*Ddx3x*	10.6	2
Q62318	*Trim28*	5.96	2
Q6ZQ38	*Cand1*	5.65	2
Q6ZQ58	*Larp1*	5.32	2
Q7TQH0	*Atxn2l*	4.68	2
Q8BMS1	*Hadha*	8.91	2
Q8C0C7	*Farsa*	8.62	2
Q8CI11	*Gnl3*	7.78	2
Q8K2V6	*Ipo11*	6.66	2
Q8K2Z4	*Ncapd2*	5.53	2
Q91YE6	*Ipo9*	11.9	2
Q922K7	*Nop2*	4.61	2
Q9CWX9	*Ddx47*	8.35	2
Q9DBD5	*Pelp1*	8.89	2
Q9DBE9	*Ftsj3*	8.84	2

[1] Score: (Sequest Score) The sum of the ion scores of all peptides that were identified.
[2] # Peptides: The total number of distinct peptide sequences identified in the protein group

Appendix J: Experiment # 9 using ES cells. Co-immunoprecipitated proteins were separated by SDS-PAGE gel. The gel was cut into 3 pieces and proteins were in-gel digested with trypsin and extracted and submitted to liquid chromatography (LC-MS/MS) mass spectrometry analysis. In this experiment antibody was cross-linked to dynabeads to decrease Ig contamination. The IP reactions were incubated for 2.5 hours instead of overnight (Appendix C). The confirmed components of the CECR2 complex are highlighted.

Accession (UniProt)	Gene name	Score[1]	# Peptides[2]
Q91ZW3	*Smarca5 (Snf2h)*	306	
E9QA25	*Cecr2*	134	
P0CG50	*Ubc*	81	
Q9D6Z1	*Nol5a*	75	
P29341	*Pabpc1 Pabp1*	66	
Q8R4U7	*Luzp1*	60	
Q9JM76	*Arpc3*	53	
Q8K224	*Nat10*	47	
Q6DFW4	*Nol5*	46	
Q8CGP2	*Hist1h2bp*	41	
Q8CCE5	*Top2a*	41	
Q8BWM4	*Arid4a*	41	
Q9CSR3	*Ddx21*	36	
Q0VF22	*Ccdc138*	35	
Q6PGB8	*Smarca1(Snf2l)*	34	
Q9WV32	*Arpc1b*	33	
Q8BVY0	*Rsl1d1*	32	
Q6P9T4	*Hdac5*	28	

[1] Score: (Sequest Score) The sum of the ion scores of all peptides that were identified.
[2] # Peptides: The total number of distinct peptide sequences identified in the protein group

Appendix K: Experiment # 2 using testis. Proteins identified using LC-MS/MS. Only gel pieces corresponding to 130 kDa and 170 kDa were analyzed. In this experiment no IgG control was used. In this experiment IP reactions were incubated for 2 hours instead of overnight. See Appendix C for the test condition. The confirmed components of the CECR2 complex are highlighted.

Accession (UniProt)	Gene name	Score[1]	# Peptides[2]
O35691	*Pnn*	20.82	5
Q99MV1	*Tdrd1*	15.74	3
E9QA25	*Cecr2*	10.58	9
Q91ZW3	*Smarca5 (Snf2h)*	9.87	3
Q5SSW2	*Psme4*	9.30	3
Q8VDP4	*Ccar2*	7.20	3
Q8K310	*Matr3*	5.00	2
Q99PU8	*Dhx30*	4.71	2

[1] Score: (Sequest Score) The sum of the ion scores of all peptides that were identified.
[2] # Peptides: The total number of distinct peptide sequences identified in the protein group

Appendix L: Experiment # 3 using testis. Co-immunoprecipitated proteins were separated by SDS-PAGE gel. The gel was cut into 3 pieces and submitted to liquid chromatography (LC-MS/MS) mass spectrometry analysis. In this experiment, protein extractions were performed in salt concentration of 150 mM instead of 420 mM (Appendix C). As a negative control, IP assay was performed using homozygous $Cecr2^{Tm1.1Hemc}$ mutant. The confirmed components of the CECR2 complex are highlighted.

Accession (UniProt)	Gene name	Score[1]	# Peptides[2]
Wild type			
Q91ZW3	*Smarca5*	127.07	30
Q8K310	*Matr3*	67.39	9
Q9Z2V5	*Hdac6*	41.26	10
P61407	*Tdrd6*	37.28	11
Q3UJB9	*Edc4*	33.93	7
Q8VDP4	*Ccar2*	31.25	6
P16381	*D1Pas1*	24.53	2
Q9JIX8	*Acin1*	23.85	5
Q8BMD2	*Dzip1*	20.87	6
Q99PV0	*Prpf8*	17.3	3
Q9Z1X4	*Ilf3*	15.34	4
P29341	*Pabpc1*	15.03	5
Q7M6Z4	*Kif27*	14.98	3
Q9Z2C8	*Ybx2*	14.77	3
P17225	*Ptbp1*	11.83	2
Q497K7	*Tmem247*	11.12	2
Q61584	*Fxr1*	10.27	3
O35286	*Dhx15*	9.77	2
Q7TT37	*Ikbkap*	9.23	3
Q64012	*Raly*	9.14	3
P10126	*Eef1a1*	8.89	2
Q6PDK2	*Kmt2d*	8.85	3
Q99NB9	*Sf3b1*	8.14	3
Q8K4Z5	*Sf3a1*	7.01	2
Q3TTP0	*Shcbp1l*	6.35	2
Q9R0A0	*Pex14*	5.46	2
Q80XI3	*Eif4g3*	5.04	2
Cecr2$^{Tm1.1Hemc}$ **mutant**			
Q9JJ28	*Flii*	49.89	10
P39447	*Tjp1*	42.52	10
Q9EP71	*Rai14*	41.23	10
Q9DBR7	*Ppp1r12a*	35.01	9
Q8K310	*Matr3*	33.78	7
P59242	*Cgn*	33.57	8
P16546	*Sptan1*	28.00	4

P16381	*D1Pas1*	21.56	2
Q9DA79	*Dpep3*	20.68	2
Q9QXS1	*Plec*	19.88	7
Q91WK0	*Lrrfip2*	14.03	2
P10126	*Eef1a1*	13.80	2
Q6PR54	*Rif1*	11.69	2
Q9Z2C8	*Ybx2*	11.18	3
Q3UJB9	*Edc4*	10.77	2
P98086	*C1qa*	8.69	3
P17225	*Ptbp1*	8.42	2
Q9QXL2	*Kif21a*	8.17	2
P61407	*Tdrd6*	8.16	2
P62141	*Ppp1cb*	7.86	2
Q99NB9	*Sf3b1*	7.57	2
Q9DBG3	*Ap2b1*	7.45	2
Q9QZB7	*Actr10*	7.15	2
P14106	*C1qb*	7.05	2
Q9Z2V5	*Hdac6*	6.39	2
Q497K7	*Tmem247*	6.20	2
Q8R1A4	*Dock7*	6.07	2
Q62261	*Sptbn1*	5.56	2
P13020	*Gsn*	5.53	2
Q9D8Y0	*Efhd2*	5.09	2

[1] Score: (Sequest Score) The sum of the ion scores of all peptides that were identified.
[2] # Peptides: The total number of distinct peptide sequences identified in the protein group

Appendix M: Experiment # 4 using testis. Co-immunoprecipitated proteins were separated by SDS-PAGE gel. The gel was cut into 3 pieces and submitted to liquid chromatography (LC-MS/MS) mass spectrometry analysis. As a negative control, IP assay was performed using homozygous *Cecr2*[Tm1.1Hemc] mutant. Proteins identified in both wild type and mutant samples are shown in the list. CECR2 was detected in the mutant. The confirmed components of the CECR2 complex are highlighted.

Accession (UniProt)	Gene name	Score[1]	# Peptides[2]
Wild type			
P00342	*Ldhc*	40.30	9
P16858	*Gapdh*	39.59	6
Q9D051	*Pdhb*	34.51	8
P17182	*Eno1*	29.76	9
Q8BMS1	*Hadha*	28.23	6
P48774	*Gstm5*	26.04	4
P54869	*Hmgcs2*	25.18	7
P29341	*Pabpc1*	24.64	8
Q3V132	*Slc25a31*	24.43	8
P61205	*Arf3*	21.86	6
Q8K310	*Matr3*	21.74	5
Q91ZW3	*Smarca5 (Snf2h)*	27.9	9
P20029	*Hspa5*	18.49	2
O70325	*Gpx4*	16.37	6
P16381	*D1Pas1*	15.96	3
P35487	*Pdha2*	14.67	5
Q9DB20	*Atp5o*	14.21	5
P62827	*Ran*	14.07	5
P06151	*Ldha*	11.56	2
Q9JHU4	*Dync1h1*	11.13	4
Q62452	*Ugt1a9*	10.98	2
Q99JY0	*Hadhb*	10.57	4
Q9QZ82	*Cyp11a1*	10.53	2
Q61990	*Pcbp2*	10.38	4
Q8VEM8	*Slc25a3*	10.14	3
P47738	*Aldh2*	10.11	3
P27786	*Cyp17a1*	10.08	2
Q9Z2I9	*Sucla2*	9.87	4
Q61656	*Ddx5*	9.59	3
Q61545	*Ewsr1*	8.69	3
P00405	*Mtco2*	8.59	3
Q6PDK2	*Kmt2d*	8.44	3
Q921F2	*Tardbp*	8.26	3
P51881	*Slc25a5*	7.93	2
Q9D023	*Mpc2*	7.77	2
P02301	*H3f3c*	7.73	2
Q9CQQ7	*Atp5f1*	7.14	2

279

E9QA25	*Cecr2*	*	6
Q6P1B1	*Xpnpep1*	6.91	2
Q9Z2C8	*Ybx2*	6.55	2
P19157	*Gstp1*	6.49	2
P68254	*Ywhaq*	6.45	2
Q9DBG6	*Rpn2*	6.35	2
P45376	*Akr1b1*	6.24	2
Q9CYH2	*Fam213a*	6.19	2
Q03265	*Atp5a1*	6.13	2
P60867	*Rps20*	5.67	2
Q9WUM5	*Suclg1*	5.57	2
P70168	*Kpnb1*	5.51	2
Q9Z2V5	*Hdac6*	5.46	2
P60766	*Cdc42*	5.45	2
Q9CXY6	*Ilf2*	5.31	2
P10649	*Gstm1*	5.23	2
Q8R127	*Sccpdh*	5.20	2
Q9Z204	*Hnrnpc*	5.19	2
Q9JIX8	*Acin1*	4.90	2
Q6PR54	*Rif1*	4.74	2
P61407	*Tdrd6*	4.58	2
Q9ESG2	*Ropn1*	4.54	2

Cecr2[Tm1.1Hemc] **mutant**

E9QA25	*Cecr2*	*[3]	4
Q6PR54	*Rif1*	47.34	11
P16858	*Gapdh*	29.80	7
P29341	*Pabpc1*	28.11	9
Q8K310	*Matr3*	27.29	6
P20029	*Hspa5*	21.68	4
P00342	*Ldhc*	19.27	5
Q9Z2V5	*Hdac6*	18.27	6
P16381	*D1Pas1*	15.86	3
O70133	*Dhx9*	12.65	4
P59242	*Cgn*	11.15	3
Q9R0A0	*Pex14*	10.78	3
Q9Z2C8	*Ybx2*	10.60	3
Q9DBR7	*Ppp1r12a*	9.90	2
Q9D051	*Pdhb*	9.12	3
P51881	*Slc25a5*	8.91	2
P54869	*Hmgcs2*	8.10	3
Q60930	*Vdac2*	7.24	3
Q61656	*Ddx5*	6.93	2
Q02257	*Jup*	6.59	2
P0CG49	*Ubb*	6.54	2
P00405	*Mtco2*	6.32	2
Q99MW1	*Stk31*	5.90	2
Q61990	*Pcbp2*	5.76	2

Q9JMB7	*Piwil1*	5.74	2
Q9D824	*Fip1l1*	5.70	2
Q9D023	*Mpc2*	5.51	2
Q99NB9	*Sf3b1*	5.37	2
Q9QXS1	*Plec*	5.25	2
Q8VEM8	*Slc25a3*	5.13	2
Q9JIX8	*Acin1*	5.12	2
Q9EP71	*Rai14*	4.92	2

1 Score: (Sequest Score) The sum of the ion scores of all peptides that were identified.

2 # Peptides: The total number of distinct peptide sequences identified in the protein group

3 * Detected in second run against customized database.

Appendix N: Experiment # 5 using testis. Co-immunoprecipitated proteins were separated by SDS-PAGE gel. The gel was cut into 10 pieces and proteins were in-gel digested with trypsin and extracted and submitted to liquid chromatography (LC-MS/MS) mass spectrometry analysis. In this experiment antibody was cross-linked to dynabeads to decrease Ig contamination. The IP reactions were incubated for 3 hours instead of overnight (Appendix C). Proteins also identified in the IgG control were considered as non-specific interactions and removed from the list, as were known background contaminants. The confirmed components of the CECR2 complex are highlighted.

Accession (UniProt)	Gene name	Score[1]	# Peptides[2]
E9QA25	*Cecr2*	98.01	21
Q91ZW3	*Smarca5 (Snf2h)*	39.68	11
Q61545	*Ewsr1*	6.88	2

[1] Score: (Sequest Score) The sum of the ion scores of all peptides that were identified.
[2] # Peptides: The total number of distinct peptide sequences identified in the protein group

Appendix O: Experiment # 1 using neurospheres. Co-immunoprecipitated proteins were separated by SDS-PAGE gel. The gel was cut into 3 pieces and submitted to liquid chromatography (LC-MS/MS) mass spectrometry analysis. In this experiment, protein extractions were performed in salt concentration of 150 mM instead of 420 mM (Appendix C). As a negative control, IP assay was performed using homozygous *Cecr2*[Tm1.1Hemc] mutant cells. The confirmed components of the CECR2 complex are highlighted.

Accession (UniProt)	Gene name	Score[1]	# Peptides[2]
Wild type			
Q8C525	*Mb21d2*	103.91	13
Q3UEB3	*Puf60*	46.51	8
Q76KF0	*Sema6d*	27.87	4
Q8VH51	*Rbm39*	22.61	4
Q8K310	*Matr3*	20.13	3
P26369	*U2af2*	13.84	3
O70133	*Dhx9*	11.77	2
B2RY56	*Rbm25*	9.87	2
Q91ZW3	*Smarca5 (Snf2H)*	7.69	2
P39447	*Tjp1*	6.36	2
P51881	*Slc25a5*	6.36	2
P60335	*Pcbp1*	5.45	2
***Cecr2*[Tm1.1Hemc] mutant**			
P39447	*Tjp1*	39.74	10
Q3UEB3	*Puf60*	36.37	9
Q8K310	*Matr3*	27.29	3
Q8VH51	*Rbm39*	22.17	6
Q9Z0U1	*Tjp2*	13.38	2
Q9CVB6	*Arpc2*	12.80	2
P21956	*Mfge8*	11.06	3
Q9JJ28	*Flii*	10.68	2
Q9JIS8	*Slc12a4*	9.68	3
P0CG49	*Ubb*	8.70	3
Q68FD5	*Cltc*	8.51	2
P60335	*Pcbp1*	8.33	3
Q921M3	*Sf3b3*	6.98	2
Q9JHJ0	*Tmod3*	6.90	2
Q7TNC4	*Luc7l2*	6.82	2
P26369	*U2af2*	6.42	2
Q8C525	*Mb21d2*	6.12	2
P51881	*Slc25a5*	5.85	2
P18760	*Cfl1*	5.83	2
Q99JY9	*Actr3*	5.80	2
Q76KF0	*Sema6d*	5.37	2

[1] Score: (Sequest Score) The sum of the ion scores of all peptides that were identified.
[2] # Peptides: The total number of distinct peptide sequences identified in the protein group

Appendix P: Genes associated with "molecular function" GO terms assigned for CECR2 peaks in ES cells. The √ shows the overlap between at least one of the binding targets of CECR2 with SNF2H or LUZP1 or both. "Overlap in three proteins" indicates the presence of the binding site that is shared with the three proteins.

GO Term (ES cells)	Gene	Peak genomic region (distance to TSS)	Overlap with SNF2H	Overlap with LUZP1	Overlap in three protein
Wnt-protein binding (ID:GO:0017147)	*Apcdd1*	chr18:63079363-63079863 (-2368)		√	
	Frzb	chr2:80260812-80261312 (+26720)			
	Fzd1	chr5:4749449-4749949 (+8517), chr5:4581102-4581602 (+176864)			
	Fzd10	chr5:128958545-128959045 (-147924), chr5:128987489-128987989 (-118980), chr5:129031714-129032214 (-74755), chr5:129123719-129124219 (+17250)	√	√	√
	Fzd2	chr11:102503075-102503575 (+37580)			
	Fzd5	chr1:64823532-64824032 (-39458), chr1:64803531-64804031 (-19457)			
	Fzd6	chr15:38775838-38776338 (-61791)			
	Fzd7	chr1:59314982-59315482 (-223759), chr1:59321603-59322103 (-217138), chr1:59361504-59362004 (-177237), chr1:59463082-59463582 (-75659), chr1:59464275-59464775 (-74466)	√	√	√
	Fzd8	chr18:9199032-9199532 (-13572)			
	Fzd9	chr5:135765446-135765946 (-38596), chr5:135748302-135748802 (-21452), chr5:135741213-135741713 (-14363)	√	√	√
	Lrp5	chr19:3606186-3606686 (+80120), chr19:3595550-3596050 (+90756)			
	Lrp6	chr6:134566192-134566692 (-49511), chr6:134524166-134524666 (-7485), chr6:134520334-134520834 (-3653), chr6:134507188-134507688 (+9493), chr6:134493669-134494169 (+23012), chr6:134385101-134385601 (+131580)	√		
	Ptpro	chr6:137229216-137229716 (+28646), chr6:137232446-137232946 (+31876), chr6:137240183-137240683 (+39613), chr6:137285381-137285881 (+84811), chr6:137351794-137352294 (+151224), chr6:137407684-137408184 (+207114), chr6:137430801-137431301 (+230231), chr6:137504888-137505388 (+304318), chr6:137528654-137529154 (+328084), chr6:137580999-137581499 (+380429)	√	√	√
	Ror2	chr13:53301954-53302454 (+79289), chr13:53174028-53174528 (+207215)	√	√	√

	Ryk	chr9:102718416-102718916 (-18581)			
	Sfrp1	chr8:24388159-24388659 (-133565), chr8:24510956-24511456 (-10768), chr8:24649450-24649950 (+127726), chr8:24689566-24690066 (+167842), chr8:24702699-24703199 (+180975), chr8:24714274-24714774 (+192550)	√	√	√
	Tdgf1	chr9:110824358-110824858 (+24054), chr9:110815760-110816260 (+32652)			
	Wif1	chr10:120433477-120433977 (-37289), chr10:120529886-120530386 (+59120), chr10:120533950-120534450 (+63184), chr10:120572584-120573084 (+101818		√	
	Wls	chr3:159500823-159501323 (-1628)			
Wnt-activated receptor activity (ID:GO:0042813)	Frzb	chr2:80260812-80261312 (+26720)			
	Fzd1	chr5:4749449-4749949 (+8517), chr5:4581102-4581602 (+176864)			
	Fzd10	chr5:128958545-128959045 (-147924), chr5:128987489-128987989 (-118980), chr5:129031714-129032214 (-74755), chr5:129123719-129124219 (+17250)	√	√	√
	Fzd2	chr11:102503075-102503575 (+37580)			
	Fzd5	chr1:64823532-64824032 (-39458), chr1:64803531-64804031 (-19457)			
	Fzd6	chr15:38775838-38776338 (-61791)			
	Fzd7	chr1:59314982-59315482 (-223759), chr1:59321603-59322103 (-217138), chr1:59361504-59362004 (-177237), chr1:59463082-59463582 (-75659), chr1:59464275-59464775 (-74466)	√	√	√
	Fzd8	chr18:9199032-9199532 (-13572)			
	Fzd9	chr5:135765446-135765946 (-38596), chr5:135748302-135748802 (-21452), chr5:135741213-135741713 (-14363)	√	√	√
	Lrp5	chr19:3606186-3606686 (+80120), chr19:3595550-3596050 (+90756)			
	Lrp6	chr6:134566192-134566692 (-49511), chr6:134524166-134524666 (-7485), chr6:134520334-134520834 (-3653), chr6:134507188-134507688 (+9493), chr6:134493669-134494169 (+23012), chr6:134385101-134385601 (+131580)	√		
	Ryk	chr9:102718416-102718916 (-18581)			
	Sfrp1	chr8:24388159-24388659 (-133565), chr8:24510956-24511456 (-10768), chr8:24649450-24649950 (+127726), chr8:24689566-24690066 (+167842), chr8:24702699-24703199 (+180975), chr8:24714274-24714774 (+192550)	√	√	

Appendix Q: Genes associated with brain "biological process" GO terms assigned for CECR2 peaks in ES cells. The √ shows the overlap between the binding targets of CECR2 with SNF2H or LUZP1 or both. "Overlap in three proteins" indicates the presence of the binding site that is shared with the three proteins. The shared genomic regions are highlighted.

Gene	Peak genomic region (distance to TSS)	Overlap with SNF2H	Overlap with LUZP1	Overlap in three protein	Mutant abnormality
Aars	chr8:113569940-113570440 (+12388)				
Agtpbp1	chr13:59680817-59681317 (-22417), chr13:59450964-59451464 (+207436), chr13:59111614-59112114 (+546786)				cochlear outer hair cell degeneration
Ascl1	chr10:86759542-86760042 (+196613)				
Atg7	chr6:114355872-114356372 (-237027)				
Atp2b2	chr6:114013108-114013608 (-171825)				
Bcl2	chr1:108635163-108635663 (-24562), chr1:108387501-108388001 (+223100), chr1:108375814-108376314 (+234787)		√		
Cacna1a	chr8:86819427-86819927 (-119586), chr8:86820662-86821162 (-118351), chr8:86868712-86869212 (-70301), chr8:86884303-86884803 (-54710), chr8:86909071-86909571 (-29942), chr8:86915622-86916122 (-23391), chr8:86930368-86930868 (-8645), chr8:86988683-86989183 (+49670), chr8:86995690-86996190 (+56677), chr8:87010870-87011370 (+71857), chr8:87073551-87074051 (+134538), chr8:87114656-87115156 (+175643)		√		infertility
Cdk5	chr5:23926797-23927297 (+2301)				
Cdk5r1	chr11:80410391-80410891 (+120093)		√		
Cdk5r2	chr1:74843614-74844114 (-57644), chr1:74872075-74872575 (-29183)				
Cntn1	chr15:91680117-91680617 (-278252)				
Cyp11a1	chr9:57869912-57870412 (+7342), chr9:57902115-57902615 (+39545)	√			
Dab1	chr4:104134631-104135131 (+94686), chr4:104332368-104332868 (+292423)		√		infertility
Dlc1	chr8:37976198-37976698 (+25239), chr8:37890311-37890811 (+111126), chr8:37641833-37642333 (+359604), chr8:37553859-37554359 (+447578)				open neural tube
Ezh2	chr6:47702301-47702801 (-157522), chr6:47698136-47698636 (-153357), chr6:47600269-47600769 (-55490), chr6:47598831-47599331 (-54052),	√ √ √	√	√	

	chr6:47597485-47597985 (-52706), chr6:47596773-47597273 (-51994)	√			
Foxp2	chr6:15124994-15125494 (-10262)				
Gas1	chr13:60529869-60530369 (-251393), chr13:60416076-60416576 (-137600), chr13:60062681-60063181 (+215795), chr13:60028404-60028904 (+250072)				coloboma
Gbx2	chr1:91854442-91854942 (-24827)				
Gli2	chr1:121263271-121263771 (-313325), chr1:121224795-121225295 (-274849), chr1:121192633-121193133 (-242687), chr1:121144839-121145339 (-194893), chr1:121065380-121065880 (-115434), chr1:121060504-121061004 (-110558), chr1:121039145-121039645 (-89199), chr1:121027929-121028429 (-77983), chr1:120993910-120994410 (-43964), chr1:120834325-120834825 (+115621), chr1:120719709-120720209 (+230237), chr1:120672722-120673222 (+277224), chr1:120651819-120652319 (+298127), chr1:120648485-120648985 (+301461), chr1:120608845-120609345 (+341101)	√ √ √ √	√ √	√	exencephaly, polydactyly
Hes1	chr16:29697284-29697784 (-367892), chr16:30135498-30135998 (+70322), chr16:30192036-30192536 (+126860)		√		exencephaly,increased cochlear inner hair cell number
Hes3	chr4:151670209-151670709 (-4688)				exencephaly
Hoxa1	chr6:52072434-52072934 (+35632)				
Hoxb1	chr11:96241185-96241685 (+14369), chr11:96287143-96287643 (+60327)				
Klhl1	chr14:97752650-97753150 (-834579), chr14:97403408-97403908 (-485337)				
Ldb1	chr19:46095275-46095775 (+24175)	√	√	√	
Lhx1	chr11:84464837-84465337 (-125570), chr11:84440123-84440623 (-100856), chr11:84424059-84424559 (-84792), chr11:84419853-84420353 (-80586), chr11:84360709-84361209 (-21442), chr11:84347541-84348041 (-8274)	√ √	 √		acephaly

Gene	Coordinates				Phenotype
Lhx5	chr5:120770194-120770694 (-111335), chr5:120786212-120786712 (-95317)	√			
Lmx1a	chr1:169588328-169588828 (-31111), chr1:169964873-169965373 (+345434), chr1:170282262-170282762 (+662823)	√			infertility
Lmx1b	chr2:33492367-33492867 (+3414), chr2:33446215-33446715 (+49566), chr2:33409323-33409823 (+86458), chr2:33382356-33382856 (+113425), chr2:33359627-33360127 (+136154)	√			
Lrp6	chr6:134566192-134566692 (-49511), chr6:134524166-134524666 (-7485), chr6:134520334-134520834 (-3653), chr6:134507188-134507688 (+9493), chr6:134493669-134494169 (+23012), chr6:134385101-134385601 (+131580)				exencephaly, polydactyly,coloboma
Mtpn	chr6:35790531-35791031 (-300893)				
Myh10	chr11:68647781-68648281 (+142213)	√			exencephaly
Myo16	chr8:10054114-10054614 (-218208), chr8:10190110-10190610 (-82212), chr8:10325984-10326484 (+53662), chr8:10405045-10405545 (+132723), chr8:10407187-10407687 (+134865), chr8:10427952-10428452 (+155630), chr8:10504659-10505159 (+232337), chr8:10647155-10647655 (+374833), chr8:10649209-10649709 (+376887), chr8:10659181-10659681 (+386859), chr8:10704770-10705270 (+432448), chr8:10709148-10709648 (+436826), chr8:10755751-10756251 (+483429), chr8:10798718-10799218 (+526396), chr8:10846066-10846566 (+573744), chr8:10863386-10863886 (+591064), chr8:10864858-10865358 (+592536), chr8:10898083-10898583 (+625761), chr8:10914370-10914870 (+642048)	√ √ √ √	√ √ √	√	
Naglu	chr11:100900428-100900928 (-30648), chr11:100917414-100917914 (-13662)				abnormal cochlea morphology
Neurod1	chr2:79323762-79324262 (-27104)				
Neurod2	chr11:98174349-98174849 (+16363), chr11:98137152-98137652 (+53560)	√			
Nfix	chr8:87327475-87327975 (-30430), chr8:87324778-87325278 (-27733), chr8:87321626-87322126 (-24581)				
Nr2c2	chr6:91946959-91947459 (-94175)				subfertilit

					y
Otx1	chr11:22128078-22128578 (-226710), chr11:21616275-21616775 (+285093), chr11:21613229-21613729 (+288139)		√		
Otx2	chr14:49608344-49608844 (-321406), chr14:49081143-49081643 (+205795)	√			exencephaly,subfertility
Prox1	chr1:192726628-192727128 (-732319), chr1:192702218-192702718 (-707909), chr1:192694465-192694965 (-700156), chr1:192501051-192501551 (-506742), chr1:192495421-192495921 (-501112), chr1:192461547-192462047 (-467238), chr1:192434077-192434577 (-439768), chr1:192399398-192399898 (-405089), chr1:192308003-192308503 (-313694), chr1:192258896-192259396 (-264587), chr1:192202815-192203315 (-208506), chr1:192189563-192190063 (-195254), chr1:192154585-192155085 (-160276), chr1:192146505-192147005 (-152196), chr1:192132457-192132957 (-138148), chr1:191975798-191976298 (+18511), chr1:191853254-191853754 (+141055), chr1:191759192-191759692 (+235117), chr1:191752445-191752945 (+241864)	√ √ √ √ √ √			
Ptprs	chr17:56497254-56497754 (+118402)		√		exencephaly
Rfx4	chr10:84314890-84315390 (+96333)				hydroencephaly
Rora	chr9:68864151-68864651 (+362808), chr9:69110421-69110921 (+609078)	√			subfertility
Rpgrip1l	chr8:93706988-93707488 (+129923)	√	√	√	exencephaly
Sec24b	chr3:129680019-129680519 (+83556)	√			craniorachischisis, open eyelid,abnormal cochlear inner hair cell morphology,omphalocele

Sema4c	chr1:36664248-36664748 (-49304), chr1:36654110-36654610 (-39166), chr1:36653260-36653760 (-38316)	√ √			exenceph aly
Serpine 2	chr1:79915463-79915963 (-60442), chr1:79904584-79905084 (-49563)				infertility
Sez6	chr11:77715749-77716249 (-28303), chr11:77756264-77756764 (+12212)				
Sez6l	chr5:113315020-113315520 (-309065), chr5:113286497-113286997 (-280542), chr5:113207503-113208003 (-201548), chr5:113144906-113145406 (-138951), chr5:113071303-113071803 (-65348), chr5:113038530-113039030 (-32575), chr5:113028076-113028576 (-22121), chr5:112982416-112982916 (+23539), chr5:112931789-112932289 (+74166)		√ √		
Skor2	chr18:76537301-76537801 (-557593), chr18:76814033-76814533 (-280861), chr18:76872679-76873179 (-222215), chr18:76907979-76908479 (-186915), chr18:76949485-76949985 (-145409), chr18:77116305-77116805 (+21411)	√ √	√		
Sstr2	chr11:113504112-113504612 (+23572)	√	√	√	
Wnt7a	chr6:91365030-91365530 (-3923)				infertility, polydacty ly
Zbtb18	chr1:179232816-179233316 (-141726), chr1:179309175-179309675 (-65367), chr1:179330564-179331064 (-43978), chr1:179514644-179515144 (+140102), chr1:179555330-179555830 (+180788), chr1:179556535-179557035 (+181993)	√			

Appendix R: Heart genes associated with "biological process" GO terms assigned for CECR2 peaks in ES cells. The √ shows the overlap between the binding targets of CECR2 with SNF2H or LUZP1 or both. "Overlap in three proteins" indicates the presence of the binding site that is shared with the three proteins and the shared genomic regions are highlighted.

GO Term (ES cells)	Gene	Peak genomic region (distance to TSS)	Overlap with SNF2H	Overlap with LUZP1	Overlap in three protein
cardiac muscle tissue growth (ID: GO:0055017)	*1190002N15Rik*	chr9:94002360-94002860 (+435890)			
	Agt	chr8:127085765-127086265 (+7591)			
	Ctnnb1	chr9:120637421-120637921 (-204847), chr9:120769704-120770204 (-72564), chr9:120787702-120788202 (-54566), chr9:120788627-120789127 (-53641), chr9:121058660-121059160 (+216392), chr9:121068681-121069181 (+226413)			
heart growth (ID: GO:0060419)	*Foxc2*	chr8:123647749-123648249 (+7928), chr8:123655482-123655982 (+15661), chr8:123664818-123665318 (+24997), chr8:123672919-123673419 (+33098), chr8:123674600-123675100 (+34779), chr8:123693482-123693982 (+53661), chr8:123702840-123703340 (+63019), chr8:123709109-123709609 (+69288), chr8:123744950-123745450 (+105129), chr8:123751301-123751801 (+111480), chr8:123759593-123760093 (+119772), chr8:123798299-123798799 (+158478), chr8:123803421-123803921 (+163600), chr8:123806578-123807078 (+166757), chr8:123809043-123809543 (+169222), chr8:123811838-123812338 (+172017), chr8:123826396-123826896 (+186575), chr8:123833151-123833651 (+193330), chr8:123845216-123845716 (+205395), chr8:123865235-123865735 (+225414), chr8:123908057-123908557 (+268236), chr8:123910281-123910781 (+270460), chr8:123944381-123944881 (+304560), chr8:123953837-123954337 (+314016), chr8:123966361-123966861 (+326540), chr8:123974786-	√ (chr8:123664818-123665318), √ (chr8:123672919-123673419), √ (chr8:123908057-123908557)	√ (chr8:123672919-123673419), √ (chr8:123809043-123809543), √ (chr8:123944381-123944881)	√ (chr8:123672919-123673419)

			√		
		123975286 (+334965), chr8:124017135-124017635 (+377314), chr8:124061160-124061660 (+421339)	√	√	√
	Mesp1	chr7:86897940-86898440 (+40484)			
	Myocd	chr11:65474713-65475213 (-391520), chr11:65440606-65441106 (-357413), chr11:65366078-65366578 (-282885), chr11:65316178-65316678 (-232985), chr11:65238149-65238649 (-154956)	√		
	Nkx2-5	chr17:26994977-26995477 (-16717)			
	Notch1	chr2:26387911-26388411 (-28819)			
	Prkar1a	chr11:109478361-109478861 (-33624), chr11:109502872-109503372 (-9113)			
	Rxra	chr2:27475979-27476479 (-56492), chr2:27483441-27483941 (-49030), chr2:27621348-27621848 (+88877)			
	Smad1	chr8:82107541-82108041 (-184374), chr8:82088414-82088914 (-165247), chr8:81950665-81951165 (-27498), chr8:81944718-81945218 (-21551)		√ √	
	Tenm4	chr7:103727700-103728200 (+367769)			
	Tgfb2	chr1:188445287-188445787 (+84334), chr1:188353542-188354042 (+176079), chr1:188284940-188285440 (+244681), chr1:188283178-188283678 (+246443), chr1:188277850-188278350 (+251771), chr1:188259866-188260366 (+269755), chr1:188177773-188178273 (+351848), chr1:188163401-188163901 (+366220), chr1:188105374-188105874 (+424247), chr1:188094649-188095149 (+434972), chr1:187947736-187948236 (+581885)	√ √ √	√	√
	Tgfbr3	chr5:107615968-107616468 (+102430), chr5:107586898-107587398 (+131500), chr5:107572903-107573403 (+145495)		√	

Appendix S: List of the genes bearing CECR2 peaks within their proximal promoter region (up to 1 kb upstream of TSS) in ES cells.

1190005I06Rik	*B3gnt3*	*Dux*	*Hoxd11*	*Nudt7*	*Sebox*	*Tmem50a*
1500012F01Rik	*BB031773*	*Edar*	*Hoxd3*	*Nufip2*	*Sema4a*	*Tnxb*
1700029F09Rik	*BC017643*	*Emp2*	*Hsd11b1*	*Nynrin*	*Serinc2*	*Trpm8*
1700029M20Rik	*Bdh1*	*Fam126b*	*Hsd17b2*	*Oas3*	*Setd6*	*Unc13a*
1700042O10Rik	*Best2*	*Fam160b2*	*Icmt*	*Oasl1*	*Sh3bgr*	*Usp4*
1700101E01Rik	*Btbd2*	*Fam46c*	*Ifi204*	*Olfml2a*	*Six3*	*Vash2*
1700125H03Rik	*C1qtnf1*	*Fam69b*	*Ifna12*	*Olfr117*	*Slc11a1*	*Vmn1r44*
1810010H24Rik	*C230024C17*	*Fam71a*	*Itih4*	*Olfr1256*	*Slc22a8*	*Vsx1*
2010009K17Rik	*Cacng1*	*Fcer2a*	*Katnal1*	*Olfr1286*	*Slc25a19*	*Xrcc2*
2010015L04Rik	*Caly*	*Fcgbp*	*Kcnip3*	*Olfr1375*	*Slc2a4*	*Zan*
2310010M20Rik	*Capg*	*Ffar1*	*Kif2b*	*Olfr1415*	*Slc36a1*	*Zfp462*
4930404A10Rik	*Casp8*	*Fgd4*	*Klhl38*	*Olfr372*	*Slc38a4*	*Zfp52*
4930430J02Rik	*Ccdc77*	*Fkbp5*	*Krtap17-1*	*Olfr384*	*Slc46a1*	*Zfp764*
4930504O13Rik	*Ccl6*	*Flot2*	*Lag3*	*Olfr392*	*Slc4a9*	*Zfp827*
4930558C23Rik	*Ccr10*	*Foxi1*	*Lemd1*	*Olfr412*	*Slc6a1*	*Zfp846*
4930562C15Rik	*Ccr6*	*Foxm1*	*Leprel2*	*Olfr568*	*Slc6a15*	*Zp1*
4930578C19Rik	*Cd244*	*G630071F17*	*Lgals9*	*Olfr633*	*Slco1a5*	*NR_002848*
4931408C20Rik	*Cd247*	*Galc*	*LOC105242399*	*Olfr658*	*Slfn4*	*NR_027883*
4931429L15Rik	*Cd300ld*	*Gas8*	*Lpl*	*Osbpl10*	*Smok2a*	*NR_029952*
4931431B13Rik	*Cd84*	*Gbx1*	*Lrrc36*	*Otop2*	*Snord37*	*NR_045466*
4933416E03Rik	*Cdh7*	*Gcm2*	*Lyg1*	*Pde2a*	*Snx8*	*NR_073523*
4933436E23Rik	*Ces1e*	*Gfra2*	*Mcat*	*Pde3a*	*Sowaha*	*NR_073524*
5430435G22Rik	*Ces2h*	*Ggt6*	*Mecom*	*Pde4a*	*Sox4*	*NR_103716*
5730590G19Rik	*Chst10*	*Gm10142*	*Mfng*	*Phrf1*	*Sp6*	*NR_104263*
6720456B07Rik	*Cldn7*	*Gm11202*	*Mir124a-1*	*Pisd-ps1*	*Sp7*	*NR_105824*
A330048O09Rik	*Clec2i*	*Gm11544*	*Mir138-1*	*Pisd-ps3*	*Spaca3*	*NR_105883*
A530065N20	*Cmas*	*Gm11545*	*Mir141*	*Pkd1l2*	*Spata19*	*NR_105889*
AA465934	*Cnga3*	*Gm11744*	*Mir1897*	*Plb1*	*Specc1*	*NR_105944*
Actg2	*Cntnap1*	*Gm13154*	*Mir195*	*Plcg2*	*Speer4c*	*NR_110380*
Actn2	*Cntnap5a*	*Gm4850*	*Mir208b*	*Pmp2*	*Speer7-ps1*	*NR_110488*
Adora3	*Cox7a2l*	*Gm4871*	*Mir28b*	*Pon1*	*Spink2*	*NR_121616*
AI314180	*Cox8a*	*Gm4981*	*Mir3078*	*Ppapdc3*	*Spink2*	*NR_126072*
AI429214	*Creb3l1*	*Gm53*	*Mir3107*	*Prickle2*	*Spns3*	*NR_126073*
AI462493	*Cst11*	*Gm6116*	*Mir497*	*Prr15l*	*Ssr2*	*NR_130325*
AI846148	*Ctc1*	*Gm7904*	*Mir762*	*Prss39*	*Stau2*	*NR_130345*
Alox12b	*Ctnnd1*	*Gng7*	*Mnda*	*Prtn3*	*Sulf1*	*NR_130904*
Alx4	*D10Wsu52e*	*Gpm6a*	*Mrln*	*Rab7b*	*Taar1*	*NR_130966p*
Amac1	*Dbx1*	*Gpr133*	*Mrps2*	*Rasef*	*Tal2*	*NR_131021*
Amica1	*Def6*	*Gpr77*	*Mthfsd*	*Rbm8a*	*Tas2r135*	*NR_131113*
Ap1s3	*Def8*	*Gpr84*	*Myadml2*	*Rcvrn*	*Tbx2*	*NR_131114*
Apoa2	*Defb28*	*Gpr97*	*Napsa*	*Rftn2*	*Tcap*	*NR_131129*
Aqp3	*Defb29*	*Gsdma*	*Nat1*	*Rn4.5s*	*Tcf23*	*NR_131138*
Arfgap3	*Dennd5b*	*Gsg1*	*Nceh1*	*Rn5s10*	*Tex24*	*NR_131173*
Arhgef15	*Dffa*	*Gsta2*	*Ndfip2*	*Rn5s17*	*Tfap2a*	*NR_131182*

Armc6	Dgcr6	Hmgxb4	Nenf	Rn5s23	Tgfb3	NR_132336
Atn1	Dnajc11	Hnrnpl	Nfix	Rn5s28	Timp2	NR_132634
Avpr1b	Dpep2	Hnrnpu	Nmur2	Rn5s30	Tmem119	