2010

Functional characterization of a zinc finger protein AEBP2

Hana Kim
Louisiana State University and Agricultural and Mechanical College, hk-one-violin@hanmail.net

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_dissertations/3850

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
FUNCTIONAL CHARACTERIZATION OF A ZINC FINGER PROTEIN
AEBP2

A Dissertation
Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in
The Department of Biological Sciences

by
Hana Kim
B.M., Louisiana State University, 2006
December 2010
ACKNOWLEDGEMENTS

I would like to thank Dr. Joomyeong Kim for being a great mentor and teaching me so much about science. I have learned that it takes patience and persistence to test the hypotheses you believe in. I thank Muhammad Ekram, and Dr. Michelle Thiaville for their ongoing support and suggestions, and previous lab members, Dr. Jennifer Huang, Dr. Chris Faulk, Dr. Jungha Choo and Keunsoo Kang, for their encouragement and help throughout my studies. I thank Dr. Jeongdo Kim and Dr. Sungryul Yu for teaching me many experimental techniques, and Dr. Jungnam Lee and Dr. Kyudong Han for their kind words and advice.

I thank my committee members, Dr. Mark Batzer, Dr. David Donze, Dr. Patrick DiMario, and Dr. Shisheng Li for their generous support and time. I also thank Dr. Kurt Svoboda for allowing me to audit his developmental biology course where I learned about neural crest cells.

Last but not least, I thank my family: my husband, Dr. Thomas Kang, for reading my lengthy papers and giving me suggestions, and my parents for raising me and always teaching me to appreciate knowledge and education.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>v</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>vi</td>
</tr>
<tr>
<td>CHAPTER ONE: BACKGROUND</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER TWO:</td>
<td>20</td>
</tr>
<tr>
<td>AEBP2 AS A POTENTIAL TARGETING PROTEIN FOR POLYCOMB REPRESSION COMPLEX PRC2</td>
<td></td>
</tr>
<tr>
<td>CHAPTER THREE:</td>
<td>52</td>
</tr>
<tr>
<td>AEBP2 AS AN EPIGENIC REGULATOR OF NERUAL CREST CELL DEVELOPMENT</td>
<td></td>
</tr>
<tr>
<td>CHAPTER FOUR: SUMMARY AND FUTURE DIRECTIONS</td>
<td>80</td>
</tr>
<tr>
<td>APPENDIX: LETTER OF PERMISSION</td>
<td>88</td>
</tr>
<tr>
<td>VITA</td>
<td>89</td>
</tr>
</tbody>
</table>
LIST OF TABLES

3.1 Genotype distribution of the mice from the breeding of the Aebp2 .................  57

3.2 Genotype distribution of the embryos from the breeding of F(+/−) x M(+/−) ...  58
# LIST OF FIGURES

1.1 Global ChIP-sequencing .......................................................... 7

1.2 An overview of histone modification marks on the *Aebp2* gene in embryonic stem cells ............................................. 8

2.1 Global protein sequence alignment of AEBP2 and JING .................. 27

2.2 Exon structure and alternative splicing of the mouse *Aebp2* gene .......... 29

2.3 DNA-binding motifs of AEBP2 .................................................. 35

2.4 ChIP cloning scheme and experimental strategies .......................... 35

2.5 Enriched motifs within the AEBP2 and SUZ12 ChIP sequences ........... 39

2.6 Co-occupancy test of the gene loci identified through AEBP2 and SUZ12 ChIPs .......................................................... 44

3.1 Generation of an *Aebp2*-knockin allele .................................... 56

3.2 Spatial and temporal expression patterns of *Aebp2* ......................... 59

3.3 Phenotypes of the *Aebp2* heterozygotes .................................... 61

3.4 *Aebp2* mutation effects on the expression levels of the NCC-associated genes ......................................................... 63

3.5 *In vivo* binding of AEBP2 and EZH2 and the methylation level of H3K27me3 on the NCC-associate genes .......................... 65

3.6 *Aebp2* mutation effects on the PRC2-mediated regulation of the NCC-associated genes .................................................. 66

4.1 A model describing the effect of *Aebp2* heterozygosity on the PRC2 targeting and the subsequent effect on its downstream genes ............. 83
ABSTRACT

AEBP2 is a zinc finger protein that has been shown to interact with the mammalian Polycomb Repression Complex 2 (PRC2). I characterized this unknown protein and tested its potential targeting roles for the PRC2. AEBP2 is an evolutionarily well-conserved gene that is found in animals ranging from flying insects to mammals. The transcription of mammalian AEBP2 is driven by two alternative promoters and produces multiple transcripts that give rise to at least two isoforms of the protein. These isoforms show developmental stage-specific expression patterns: the larger adult-specific form (52 kDa) and the smaller embryo-specific form (31 kDa). The AEBP2 protein binds to a DNA-binding motif with an unusual bipartite structure, CTT(N)15-23cagGCC with lower-case base pairs being less critical. A large fraction of AEBP2's target loci also map closely to the known target loci of the PRC2. In fact, many of these loci are co-occupied by the two proteins, AEBP2 and SUZ12. This suggests that AEBP2 is most likely a targeting protein for the mammalian PRC2 complex. To investigate the in vivo roles of this protein, a mutant mouse line with disrupted Aebp2 transcription has been generated. Breeding experiments demonstrated embryonic lethality in the Aebp2-mutant homozygotes, but survival of the heterozygotes to adulthood with fertility. In developing mouse embryos, Aebp2 is expressed mainly within cells of neural crest origin, such as the dorsal root ganglia, and facial cartilages and bones. In addition, many heterozygotes display a set of phenotypes, including enlarged colon and hypopigmentation, similar to those observed in human patients with Hirschsprung’s disease and Waardenburg syndrome. These phenotypes are caused by the absence of the neural crest-derived ganglia in hindguts and melanocytes. Additional analyses further confirmed changes in
the expression and methylation levels of H3K27me3 on the genes involved in the
development of the neural crest cells in the *Aebp2* heterozygotes. Overall, these results
suggest that *Aebp2* may regulate the development of the neural crest cells through the
PRC2-mediated epigenetic mechanism.
CHAPTER ONE:

BACKGROUND
The Evolution of Genetic Studies: Classical Genetics, Genomics, and Reverse Genetics

Gregor J. Mendel introduced the concept of “factors,” as units of information that are passed on to the next generation. He claimed that these factors, which exist in pairs, give rise to phenotypic traits. In addition, he discovered that these factors segregate independently (independent assortment) during reproduction (Mendel, 1865). His “factors” are now known as genes, and the scientific community often refers to Mendel’s law of inheritance to explain how the phenotype of offspring arises from the parental genotype.

In classical genetic studies, a mutant phenotype is observed and the genes or alleles responsible for that phenotype are determined by genetically mapping them through genetic crosses. Many genes important in the development of organisms have been identified through these genetic studies. We now know that the human genome contains approximately 25,000 genes (Stein, 2004). It would be challenging to characterize all of these genes solely by classical genetic studies, since not all genes correlate with obvious phenotypes, and because there are resolution limits to genetic mapping by genetic crosses. Thus, great effort has been made to sequence the genomes of human, flies, mouse, and yeast. With recent advances in DNA sequencing technologies, hundreds of organisms have been mapped and sequenced. The glut of data generated from whole-genome sequencing has given us new approaches in studying genetics, and a new field called “genomics” has arisen as a powerful way of analyzing genes and regulatory regions in a genome. A gene’s presence in a multitude of organisms can demonstrate that it is highly conserved rather than specific to a particular lineage of animals. In addition, site directed mutagenesis, recombination techniques, and RNA interference can be used to remove or silence a particular gene of interest or to alter their conserved regulatory regions. Some genetic alterations lead to the discovery of a new mutant phenotype and reveal the function of that
particular gene. This genetic approach, in which we select a particular gene and then unearth its phenotype, is called “reverse genetics” (Takahashi et al., 1994). We applied a reverse genetic approach to examine the gene function of Aebp2 (adipocyte enhancer binding protein 2): first by utilizing the whole genome sequence for evolutionary conservation studies, and second by generating Aebp2 mutant mice lines. These studies unraveled the evolutionary conservation of Aebp2; exons in the gene that may encode critical protein domains; and potential function of this protein.

**Epigenetics**

After fertilization, the zygote divides continuously to develop into a full body. In this sense, we are derived from a single cell, and presumably each and every cell in our body (except our germ cells and some immune cells) contains the same genome. However, our liver cells are clearly different from our brain cells. The human body is composed of at least 200 different cell types. How then do these different cell types arise from a single genome? In 1950, Conrad Waddington coined the term “epigenetics” to explain from an organism’s developmental perspective that different cells arise due to the influence of different environments (van Speybroeck, 2002). Recently, this definition of epigenetics has expanded due to the accumulation of various epigenomes observed among different cells and individuals (Feinberg, 2010). Studies suggest that different cell types arise from tissue-specific expression of certain genes in a spatial and temporal manner and that gene regulation occurs at multiple cellular levels. In addition, through mitosis, these cell-specific patterns of gene expression can be passed onto the next generation of daughter cells. This type of genetic inheritance, which is independent of the DNA sequence, is a component of the modern definition of “epigenetics” (Haig, 2004). Epigenetics can be used to understand the different phenotypic outcomes observed between two
individuals with the same genotypes, for example, monozygotic twins. Despite identical genetic makeup, twins may demonstrate variations in height, weight, personalities, and even cancer susceptibilities (Petronis et al., 2006; Fraga et al., 2005). Thus, we suspect that epigenetic inheritance patterns are susceptible to changes in environmental conditions. In addition, epigenetic marks are reversible; unlike our DNA sequence, which is relatively fixed (Handel et al., 2010). Another perception that broadened Waddington’s definition of “epigenetics” is the theory of epigenetic inheritance through meiosis. Some epigenetic marks can be passed on to the next generation through our germ cells. This phenomenon is called “transgenerational epigenetic inheritance” (Youngson and Whitelaw, 2008). For example, studies in mice show that the mother’s nutritional intake of methyl donor supplement throughout pregnancy could give rise to variable coat colors in their offspring and also in their offspring’s next generation (Cropley et al., 2006; Morgan and Whitelaw, 2008). In addition, from human case studies, two families with colorectal cancer showed abnormal epigenetic marks in a tumor suppressor gene (MLH1), which was inherited from their mother’s germ cells (Hitchins et al., 2007). Although the field of transgenerational epigenetic inheritance is relatively new, these studies have brought new insight on the importance of environmental exposure, e.g. diet, and its effect on our developing offspring. The plasticity of epigenetic inheritance is of vital practical importance. Understanding the mechanisms of environmental effects on epigenetic marks may lead to new approaches to disease prevention, cancer treatment, and regenerative medicine.

**Epigenetic Mechanisms: DNA Methylation and Histone Modifications**

Four epigenetic mechanisms for gene regulation have been suggested so far: DNA methylation, histone modification, RNA interference, and chromatin structure. Of these mechanisms, DNA methylation is the most studied and biochemically well understood epigenetic
mechanism (Feinberg, 2010). DNA methylation occurs by the attachment of a methyl group (CH\textsubscript{3}) to a cytosine base that immediately follows a guanine base (CpG). Some of the most important enzymes responsible for DNA methylation are DNA methyltransferase1 (DNMT1) and DNA methyltransferase3 (DNMT3). DNMT3a and DNMT3b are important for de novo methylation (Okano et al., 1999) and DNMT1 is important for DNA methylation maintenance by methylating hemi-methylated DNA (Sharif et al., 2007). Several theories have been suggested to explain the mechanistic role of DNA methylation. Many agree that DNA methylation can block certain transcription factors from binding to regulatory regions. Some suggest that DNA methylation may have an important role in chromatin structure (Weber and Schubeler, 2007).

Although the molecular functions of DNA methylation are in the process of being discovered, the importance of DNA methylation for the survival of an organism is well known. For example, in eukaryotic organisms, many of the repetitive elements in the genome are silenced by DNA methylation, and loss of these methyl marks lead to ectopic expression of nearby genes and instability of transposable elements (Bird, 2002; Cedar and Bergman, 2009). Notably, mutations in the DNMT3b gene are often found in human patients with ICF (immunodeficiency, centromeric instability, facial anomalies) syndrome (Wijmenga et al., 1998). DNA methylation is also important for proper genomic imprinting in mammals (Li et al., 1993). Genomic imprinting is a process in which one specified parental allele is expressed. Many imprinted genes maintain their imprinted status through DNA methylation. For example, Peg3 is a paternally expressed imprinted gene (Kuroiwa et al., 1996; Relaix et al., 1998). The promoter region of the maternal allele of Peg3 is heavily methylated and repressed while the promoter region of the paternal allele is unmethylated and expressed. This methylation pattern has to be maintained throughout cell divisions to maintain the proper expression level of this gene in
somatic cells. In humans, hypermethylation in the regulatory region of the paternal allele of *PEG3* has been observed in patients with ovarian cancer (Feng et al., 2008).

Histone modification is another epigenetic mechanism that is beginning to assert its importance in the epigenetic field. Thomas Jenuwein, David Allis, and Bryan Turner first announced the “histone code” hypothesis in 2001. They claimed that histone tails are modified and that these modifications indicate the active or repressed state of the associated gene (Turner 2000; Jenuwein and Allis, 2001). There are debates over the details of this theory, but recent studies seem to indicate that it is in part true. Innovative molecular techniques such as chromatin immunoprecipitation (ChIP) assay followed by DNA sequencing (ChIP-seq) have allowed us to determine the different histone modification patterns for many genes. An overview of modern ChIP-seq is demonstrated in Figure 1.1.

Various histone modification marks have been discovered, however the two most studied histone marks are histone acetylation and histone methylation. Histone acetylation is generally associated with gene expression while histone deacetylation is generally associated with gene repression. The transcriptional regulatory role of histone methylation is more complex than histone acetylation; i.e. transcriptional regulatory outcome of histone methylation is dependent on the location of the lysine that is methylated. H3K4me3 (histone H3 lysine 4 trimethylation) is associated with active genes while H3K9me3 and H3K27me3 are associated with repressed genes.

Accumulation of the ChIP-seq (and also ChIP-array) data of these modified histones has provided a superior view of the histone marks in different tissues. For example, an overview of some of the histone marks on the *Aebp2* gene in stem cells can be viewed in Figure 1.2.
**Figure 1.1. Global ChIP-sequencing.** Proteins bound to DNA are crosslinked by formaldehyde. The red boxes represent genes and the blue and yellow symbols represent proteins (e.g. modified histone marks). Sonication fragments the DNA into shorter fragments, and primary antibodies are added to select for DNA that is bound by a protein of interest (in this figure the primary antibody selects for the protein in blue). Agarose beads attach to the primary antibody and precipitate the antibody-protein-DNA complex (agarose beads are indicated in grey). The precipitated complex is decrosslinked and DNA is isolated. These DNA fragments are further amplified by PCR, sequenced, and mapped onto the genome. Based on the enrichment of the DNA fragment, the protein-DNA interaction can be estimated (in the figure above, the blue protein preferably binds the left side of the first gene than in between the two genes).
Figure 1.2. An overview of histone modification marks on the *Aebp2* gene in embryonic stem cells. The UCSC genome browser (http://genome.ucsc.edu/) is utilized to view multiple histone modification marks on a specific loci in a certain cell type. In this example, the *Aebp2* gene in mouse is shown on top in purple (multiple lines indicate alternative transcripts). The arrows in the gene indicate the direction of transcription. The peaks indicated below the gene represent the enrichment of the modified histone marks on the corresponding region. The *Aebp2* gene shows enrichment of histone methylation in H3K4 (in green) and H3K36me3 (in blue), and scarce levels of H3K27me3 (in red) and H3K9me3 (in brown).
These studies suggest that the modified histones each exhibit distinct patterns of enrichment that varies depending on the tissue type, the associated gene, and the preferable enrichment site in the regulatory regions of the gene (Bernstein et al., 2006; Barski et al., 2007). Thus, obtaining a global map of these modified histone marks in various tissues should help us understand the epigenetic signatures in different cell types.

**Epigenetic Modifier: Polycomb Group Proteins**

As much as there has been emphasis to gain a global overview of various histone marks, much effort has also been made to find the key modifiers of these histones. Polycomb group proteins (PcG) are of great interest due to their potential role as epigenetic modifiers. These genes were first identified in *Drosophila* that contained extra sex combs in their legs (Lewis, 1978), hence the name, “polycomb.” Mammalian studies of the polycomb complex showed that they are important in regulating developmental genes such as Hox genes (van der Lugt et al., 1994; Core et al., 1997). The protein complex is composed of two complexes: polycomb repressive complex 1 and 2 (PRC1 and PRC2, respectively). PRC2 contains a histone methyltransferase, Ezh2, which tri-methylates histone H3 lysine 27 (H3K27me3), while PRC1 contains a ubiquitin ligase, RING1, that ubiquitinates histone H2A lysine 119 (H2AK119ub1) (Cao and Zhang, 2004; de Napoles et al., 2004). Due to their enzymatic activities of modifying histones, PcG proteins are considered epigenetic modifiers. The role of the polycomb complex in cell lineage determination has been controversial. Many PcG proteins bind to developmental regulator genes in stem cells. As a result, it was assumed that PcG proteins would play an important role in maintaining the pluripotent state of stem cells (Bernstein et al., 2006; Valk-Lingbeek et al., 2004). However, recent studies have shown that the PcG proteins could be dispensable for stem cell maintenance (Kerppola, 2009; Chamberlain et al., 2008). Now, PcG
proteins are thought to play a regulatory role in cell lineage determination by binding to developmental genes and preparing them to differentiate according to their environmental signals (Rando and Chang, 2009). Many of the PcG target gene expressions are also altered in cancer cells (Baylin, 2009). Hence, much interest in understanding the role of PcG proteins also exists in the field of cancer biology.

Despite the global role of PcG proteins in regulating developmental genes, how PcG proteins are targeted to specific DNA sequences is unclear; especially since none of the core mammalian PcG proteins bind to DNA. Some studies have suggested that noncoding RNA could regulate PcG target specificity (Rinn et al., 2007; Zhao et al., 2008), however, further studies are required to demonstrate if the experimental outcome is a direct or indirect effect.

In addition to regulatory RNAs, transcription factors that target the PcG proteins to DNA likely exist. In flies, the transcription factor Pho is known as the PRC2 targeting protein. As a result, the mammalian homolog YY1 has been suggested as the mammalian targeting protein for PRC2 (Atchison et al., 2003). However, the direct interaction between YY1 and the PRC2 complex has never been demonstrated. Unfortunately, the transcription factors that associate with the PcG proteins have not yet been discovered in mammals. Thus, indentifying and characterizing transcription factors that associate with the PcG proteins would allow better understanding of the functional role of PcG proteins and their target specificity.

Recent biochemical protein purification studies of the PRC2 complex showed that AEBP2 interacts with PRC2 (Cao and Zhang, 2004; Peng et al., 2009; Shen et al., 2009; Li et al., 2010; Pasini et al., 2010). AEBP2 was first identified as a zinc finger protein, which bound to the enhancer of adipose P2 gene and had transcriptional repressor activity (He et al., 1999). Our studies suggest that AEBP2 could be a potential transcription factor responsible for targeting
PRC2 complex to DNA in mammals. Experimental data that supports our theory are illustrated in chapters two and three.

**Transcription Factors**

In bacteria, 98% of the genome encodes for proteins, while in eukaryotic organisms like humans, protein-coding genes comprise only 1.5% of the genome. As we evolved to become a multicellular organism, proper gene regulation has become critical for our development. Our genes contain many regulatory elements, such as promoters, enhancers and insulators, which regulate gene transcription. These regulatory elements are often bound by transcription factors. RNA polymerase II, which is required for transcription, binds to promoters near transcriptional. However, RNA polymerase alone cannot bind to naked DNA, hence basal transcription factors are required for the RNA polymerase to bind DNA (Buratowski et al., 1989; Sota et al., 1989). Different genes have different transcription factor binding sites, and their associated transcription factors are expressed in a highly restricted manner. Most transcription factors bind DNA and activate or suppress transcription of target genes (Struhl, 1995). In addition, they can also modulate the activity of gene expression by interacting with other factors (Brivanlou and Darnell, 2002). Transcription factors are likely among the main components for determining proper cell differentiation and lineage commitment. The significant roles of transcription factors in lineage determination can be appreciated from the recent discovery that four transcription factors, Oct4, Sox2, c-Myc and Klf4, can induce somatic cells to be reprogrammed back into pluripotent stem cells (Takahashi and Yamanaka, 2006). This area of research is very promising for future medicine, because it may be possible to generate personalized cells or tissues such as insulin secreting cells (islets) or nerve cells from patient’s own somatic cells to treat their chronic illnesses.
According to our studies, Aebp2 is an important transcription factor that could facilitate cell lineage commitment through PRC2. Thus, we examined the DNA binding properties of Aebp2 as well as its expression patterns. Our results revealed that Aebp2 indeed binds DNA and that high expression levels of Aebp2 are found in the neural crest cells during embryonic development. This suggests that Aebp2 is an important transcription factor required for neural crest cell development.

**Neural Crest Cells**

Neural crest cells were discovered by the embryologist Wilhelm His. From an evolutionary perspective, neural crest cells exist among land vertebrates and are thought to have evolved in the vertebrate lineage to help the animals adopt a predatory lifestyle (Gans, 1983; Northcutt, 2005). Neural crest cells are generated between the epidermis and the neural plate of the developing embryo (Vogt, 1925). This cell population arises from the ectoderm, and contains the capability to be multipotent and to migrate throughout the developing body (Le Douarain and Kalcheim, 1999). Neural crest cells give rise to many different cell types such as melanocytes, adrenal glands, cranial facial bone, and cells in the peripheral nervous and enteric nervous systems (Le Douarain and Kalcheim, 1999). Due to their significant functions in vertebrate animals, neural crest cells are often referred to as the ‘fourth germ layer’ (Hall, 2000).

Numerous studies have suggested stem-cell properties of this cell lineage (Strobl-Mazzulla et al., 2010). In addition, some suggest that a small population of neural crest cells contain neural crest stem cells (NCSC) (Strobl-Mazzulla et al., 2010, Shakhova and Sommer, 2010; Stemple and Anderson, 1992). Depending on the distribution of the neural crest cells, they are divided into five types: cranial, cardiac, trunk, vagal and sacral. Fluorescent labeling studies show that these four neural crest cell types have distinct migration pathways and preferential cell derivatives.
For example, cranial neural crest cells give rise to the facial cartilage and bone structures, cardiac neural crest cells give rise to the septum between the aorta and pulmonary artery, trunk neural crest cells give rise to the dorsal root ganglion, and vagal and sacral neural crest cells give rise to the enteric nervous system. Improper migration or differentiation of neural crest cells can lead to various genetic disorders, such as Hirschsprung’s disease. Hirschsprung’s disease is characterized by missing ganglion cells in the intestine, which result in aperistalsis in the colon. This is a genetic disorder caused by improper development of vagal neural crest cells in the colon (McCallion et al., 2003; Amiel et al., 2008; Tam and Garcia-Barcelo, 2009).

Several transcription factors have been determined to be important in proper neural crest cell development and migration. Since neural crest cells are a delicate cell population (having to properly maintain a stem cell-like state before lineage commitment, migrate, and differentiate into the proper cell type when expected) it is important to understand how neural crest cell development is regulated by various transcription factors. Transcription factors such as SOX10, BMP4, and MITF are evolutionarily well-conserved among vertebrates and are thought to be important factors for neural crest cell development. However, studies suggest that the regulatory role of these transcription factors may vary among animals. For example, variable results were observed amongst different organisms when Bmp4 was overexpressed in the primary stage of generating neural crest cells. Overexpression of Bmp4 in frogs induced the epidermis (Wilson et al., 1997), while overexpression of Bmp4 in amniotes, such as mice and chickens, had no effect (Streit et al., 1998). Despite these variations, other experimental observations suggest that the fundamental mechanism in regulating neural crest cell development should be conserved (LaBonne and Bronner-Fraser, 1999). Cells extracted from the Hensen’s node of chick embryos were able to induce neural tissue in frogs (Kintner and Dodd, 1991). Also, mutations in the
SOX10 gene result in aganglionic colon phenotypes in both humans and mice (Pingault et al., 1998; Southard-Smith et al., 1998). This suggests that neural crest cells from different species share fundamental transcription factors that fulfill common roles. Consequently, the variable outcome of Bmp4 may have been due to the availability of different protein interaction partners. In addition, there could be some differences in the epigenetic marks between the two organisms. The epigenetic studies in neural crest cells are almost “untouched” (Nelms and Labosky, 2010). Therefore, further studies to reveal other transcription factors and epigenetic modifiers that could participate in the development of neural crest cells are necessary.

The following studies in Aebp2 mutant mice suggest that Aebp2 is an important transcription factor for proper neural crest cell development. Since we suspect that Aebp2 may play a role in epigenetic modifications, our findings should contribute to the understanding of how neural crest cells commit to their cell fate through epigenetic modification.

References


Mendel, G. (1865) *Experiments in Plant Hybridization.* Electronic Scholarly Publishing


CHAPTER TWO:

AEBP2 AS A POTENTIAL TARGETING PROTEIN FOR POLYCOMB REPRESSION COMPLEX PRC2*

*Reprinted by permission of Nucleic Acid Research
**Introduction**

AEBP2 is a Gli-type zinc finger protein, which was originally identified due to its *in vitro* binding capability to the promoter region of adipose P2 (aP2) gene encoding a fatty acid-binding protein (He et al., 1999). This initial study revealed that this protein contains three zinc finger units and a novel basic domain, and also that this protein may function as a repressor based on co-transfection reporter assays. Soon afterwards, the homologous protein, called JING (meaning ‘still’), was also isolated from *Drosophila* (Liu and Montell, 2001). According to the results from several studies, JING is involved in border cell migration (Liu and Montell, 2001) and development of the central nervous system (Sedaghat et al., 2002). Genetic studies further suggested that *jing* may interact with the fly Polycomb Group (PcG) protein complexes (Culi et al., 2006; McClure and Schubiger, 2008). The potential role of AEBP2 as a component of the PcG complexes has been further strengthened by another series of studies using the mammalian cell line system (Cao et al., 2002). Human AEBP2 has been co-purified with the mammalian PcG Repression Complex 2 (PRC2), and the subsequent study revealed that the AEBP2 protein can interact with the three core components of PRC2, including EED, SUZ12 and RbAp48, and that the interaction of AEBP2 with these proteins enhances the catalytic activity of the histone methylation activity of the PRC2 complex (Cao and Zhang, 2004).

Although the core proteins for PRC2 have been identified, the mechanism by which PRC2 is targeted to numerous genomic loci is currently unknown (Kohler and Villar, 2008). This lack of knowledge is mainly due to the facts that the identified core proteins do not have DNA-binding capability, and that DNA-binding proteins have never been consistently co-purified with the PRC2 (Cao and Zhang, 2004; Ringrose and Paro, 2004 Muller and Kassis, 2006; Schwartz and Pirrotta, 2008). In *Drosophila*, however, another Gli-type zinc finger gene, called *pho*
(Pleiohomeotic), has been shown to be a targeting protein for its PcG complexes (Brown et al., 1998). Recent studies further confirmed the presence of two Pho-containing complexes, INO80 and PhoRC, and PhoRC is now regarded as a new member of PcG complexes based on its repression role through another PcG protein, SFMBT [Sex comb on middle leg-related gene with four mbt domains; (Klymenko et al., 2006)]. Along with the other data from several studies, this evidence has long suggested that YY1 (Yin Yang 1), the mammalian homologue of *pho*, might be a targeting protein for the mammalian PRC2. Nevertheless, this possibility has not been formally demonstrated so far. In that regard, it is intriguing to point out that AEBP2 has both DNA-binding capability and PcG connection. Thus, it has been hypothesized that AEBP2 might be a targeting protein for the mammalian PRC2. However, very little is known about the general aspects of AEBP2, in particular its DNA-binding motif and downstream genes.

As part of the effort of characterizing this largely unknown gene and to test the above possibility, we have conducted a series of comparative genomics and DNA-binding motif studies in the current study. According to the results derived from this study, AEBP2 is an evolutionarily well-conserved protein that is found in all the animals ranging from flying insects to placental mammals. The exon structure of mouse *Aebp2* indicates the presence of alternative splicing involving both 5′- and 3′-end exons, and subsequently two major forms of AEBP2 with different protein sizes, 52 and 31 kDa. A series of ChIP cloning experiments using anti-AEBP2 and -SUZ12 antibodies also identified many *in vivo* target loci that are bound by these two proteins. Subsequent gel shift assays using the sequences obtained from these target loci revealed one potential DNA-binding motif for AEBP2, CTT(N)15-23cagGCC. Also, individual ChIP experiments further demonstrated that a subset of these identified loci are indeed occupied by
both the AEBP2 and SUZ12 proteins. These results are consistent with the initial prediction that
AEBP2 may be a targeting protein for the mammalian PRC2 complex.

Material and Method

Global Protein Sequence Alignment

AEBP2- and JING-related sequences were collected from NCBI, UniProt, EMBL and UCSC. ClustalW was used to create protein alignments, and the final outcome was produced and edited using CLC free workbench version 4.0.3 (CLC bio A/S, Denmark). The protein alignment was set using the following parameters: gap opening penalty = 10, gap extension penalty = 0.1.

AEBP2 Isoform Confirmation through RT–PCR and Western Blot

Total RNAs were isolated from several tissues of a 3-month-old male mouse using the Trizol RNA isolation kit (Invitrogen). These RNAs were reverse-transcribed using the RT-PCR kit (Invitrogen SuperScript system, Invitrogen). For the 5'-side splicing, the following primer sets were used: mAebp2-a1, 5'-CGGCCAGCGCTACACCCCAAGAACT-3'; mAebp2-a2, 5'-GGGGAGCCGCTGAGCCGAGCATGGACT-3'; and mAebp2-b, 5'-GAAGCATGCCTGGCACTGGTC-3'. For the 3'-splicing, we used the following primer sets: mAb7-F, 5'-GATACTGCCTTGCTGTTGGACC-3'; mAbU2-R, 5'-TCCATGCCATGTGGACTGCAG-3', and mAbU3-R, 5'-CTCCACTTCCACCTACAAGGA-3'. The PCR with the mAb-a1 and mAb-b primer set were performed at an annealing temperature of 61°C for 35 cycles. The remaining primer sets were amplified at an annealing temperature of 60°C for 30 cycles. For the detection of the AEBP protein, we prepared crude tissue extracts from the brain and testis of 1-month-old mouse, and also a 14-day-old embryo using the T-PER Tissue Protein Extraction Reagent kit (Cat. 78510, Thermo Scientific). Each extract (10 µg) was
separated on 10% SDS–PAGE, transferred on to a PVDF membrane (Hybond-P, Amersham), and incubated with the anti-AEBP2 polyclonal antibody (Cat. 11232-2-AP, Proteintech Group).

**AEBP2 GST Fusion Protein Production**

Three different GST fusion proteins were produced through cloning different part of the mouse AEBP2 protein (GenBank accession no. NM_178803). To construct these fusion constructs, we have amplified the coding region of AEBP2 using the following primer set: Acidic-mAb-F (5’-ATGGCCGCGCGCTCGCCGACATG-3’) as a forward primer and mAebp2-b (5’-ATTGCAAAATGTCTCACTGTTTGCT-3’) as a reverse primer for the fusion construct I, mAebp2-a (5’-ATGGCATAAGACACACACACATTTCCAG-3’) as a forward primer and mAebp2-b (5’-ATTGCAAAATGTCTCACTGTTTGCT-3’) as a reverse primer for the fusion construct II, and mAebp2-a (5’-ATGGCATAAGACACACACACACATTTCCAG-3’) as a forward primer and xAbp2 (5’-CTGAAAGTGTGTGGAATCATGC-3’) as a reverse primer for the fusion construct III. These products were first subcloned into the pCR4 TOPO vector (Invitrogen) for sequencing verification, and later subcloned into the BamHI and HindIII restriction enzyme sites of the pGEX-4T-1 vector (Amersham Biosciences). The clones were transformed into the BL21 (DE3) competent cells (Strategen). The transformed cells were grown in LB media in 37°C to an absorbance value of 0.56 at 600 nm. The cells were further induced with 0.4 mM IPTG for 4 hours. The cell pellets were first sonicated, and stored in –80°C for later use in gel shift assays.

**Electro Mobility Shift Assay (EMSA)**

EMSAs were performed as suggested by Promega with alterations in the DNA-binding buffer condition. For most of our EMSAs, we mainly used the NTEN buffer (100 mM NaCl, 1 mM EDTA, 20 mM Tris–Cl pH 6.0, 0.5% NP40). Each reaction contained 10 µg of GST fusion protein along with a given duplex probe (0.007 pmol per reaction), which were labeled with [γ-32
ChIP Cloning and Individual ChIP Assays with Anti-AEBP2 and Anti-SUZ12 Antibodies

We performed ChIP cloning experiments using two polyclonal antibodies: AEBP2 (Cat. 11232-2-AP, Proteintech Group) and SUZ12 (Cat. ab12201, Abcam). The brain tissues of a 1-month-old mouse were used as a starting material, and the detailed protocol for our ChIP cloning is available from our previous study (Huang et al., 2006). DNA products eluted by anti-AEBP2 and SUZ12 ChIP were individually subcloned into the pZErO-2 vector (Invitrogen, Carlsbad, CA). About 200–300 clones were selected and subsequently sequenced using the ABI3130XL automated DNA sequencer (Applied Biosystem).

For individual ChIP experiment, the AEBP2 antibody (20 μl) was added into each fraction (500 μl) of cross-linked and sonicated mouse brain tissue. One mouse brain (1 g) was usually divided into 10 fractions. We followed the protocol of ChIP assay provided by the Upstate company (Upstate Biotech.). The immunoprecipitated DNA was dissolved in 40 μl of TE, and 1 μl of this eluted DNA was used as template DNA for one PCR-based ChIP assay. PCR conditions are as follows: 95°C for 5 min, 40 repetitions of the following cycle of 90°C for 30 s, 60°C for 30 s, 72°C for 30 s, and final extension at 70°C for 15 min. Each reaction included a pair of 25 ng of oligonucleotide primers. The information regarding the sequence of each primer set is available in our website (http://jookimlab.lsu.edu/?q=node/36).

Motif Analysis of ChIP-Cloned Sequences

Both 5’- and 3’-end regions, 4 bp in length, corresponding to the recognition sites for restriction enzymes were removed from each ChIP sequence for motif analyses. Any regions containing repeat elements were also removed from each ChIP sequence using RepeatMasker.
Our motif analyses used a total of 126 and 71 sequences that were obtained from AEBP2 and SUZ12 ChIP-cloned sequences, respectively. A standalone MEME (4.0.0) was used to derive overrepresented motifs among each set of the ChIP-cloned sequences. We set parameters to ensure that at least a half of the sequences contain potential motifs in each set of ChIP sequences using a ‘-minsites’ parameter. We used the following parameters: -minsites (a half number of the sequences), -minw 5, -nmotifs 3, -revcomp, -dna. MAST (4.0.0) was also used for the motif occurrence test with default setting.

Results

Identification of AEBP2-Related Sequences from Vertebrates and Flying Insects

The protein sequence of mouse AEBP2 (GenBank accession no. NP_001005605, 496 amino acids long) was used to identify its related sequences from all available genome sequences. This search identified six AEBP2-related sequences from flying insects, including flies, mosquitoes, honeybees, beetles, and wasp. These insect sequences were previously identified as JING. The same search also identified 22 related sequences from vertebrates, ranging from urochordates to placental mammals: one sea urchin, five fish, one lizard, one chicken, 14 mammal sequences. The average sizes of the predicted ORFs (Open Reading Frames) for these AEBP2-related sequences are as follow: 1744 amino acids for insects’ JING, 450 amino acids for fish's AEBP2, and 500 amino acids for mammalian AEBP2. The sizes of the AEBP2 sequences identified from lizard, frogs, sea urchins, cannot be determined due to the incompleteness of their genome sequences. All these AEBP2-related sequences are available at

http://jookimlab.lsu.edu/?q=node/36.
The amino-acid sequences of 20 full-length AEBP2-related sequences were used for global sequence alignment (Figure 2.1).

**Figure 2.1. Global protein sequence alignment of AEBP2 and JING.** AEBP2 and JING protein sequences of different organisms were aligned using the ClustalW program. Different amino acids are represented in different colors and shades. The conservation level of each position is indicated in the graph below the alignment. Six conserved domains are indicated with different colors and patterns. The mouse AEBP2 protein was used as a reference to indicate the position of each conserved domain. The zinc finger and basic domains are the most conserved and show sequence conservation from flying insects to mammals. The zoom-in version of this alignment is available as Supplementary Data 5 or the following website (http://jookimlab.lsu.edu/?q=node/81).

As shown in Figure 2.1, the mammalian AEBP2 sequences can be divided into six protein domains: acidic, neutral, serine-rich, zinc finger, basic and lysine-rich domains. Among these six domains, two domains (zinc finger and basic) show the highest levels of amino-acid sequence conservation among all different lineages. The two domains of the mammalian AEBP2 (a.a. 256–496 in mouse AEBP2 Figure 2.2A) show an average of 38% amino acid sequence identity with those of the insect JING (Figure 2.2A). The serine-rich domain (a.a. 202–255) also shows high levels of sequence conservation: 83% amino acid sequence identity between the mammal
and fish lineage. In contrast, the two domains located in the N-terminal portion of AEBP2 are lineage-specific. Although these two domains show an average 90% amino acid sequence identity between the mammalian species, these domains do not show any obvious similarity to the respective regions of the AEBP2 sequences derived from insects and fish. The acidic domain of mammalian AEBP2 is mainly characterized by arrays of glutamic and aspartic acid residues, whereas the neutral domain is characterized by arrays of glycines and serines. The AEBP2 sequences of the fish lineage also have similar acidic and neutral domains, displaying 38% amino acid sequence identity within the fish lineage. These two domains are localized within one large exon in both mammals and fish (Figure 2.2B), and the sequences of this exon in both lineages exhibit tandem repeat structure with high CpG densities. As such, many insertions/deletions are detected between the two closely related species of both mammals and fish (data not shown). On the other hand, the lysine-rich domain located in the C-terminus of mouse AEBP2 is found only in mammals, but is well conserved within mammalian species (a.a. 497–511).

Overall, mammalian AEBP2 is comprised of six protein domains: four domains are lineage-specific whereas two domains (zinc finger and basic) appear to be well conserved throughout all the lineages.

**Exon Structure and Isoforms of Mammalian AEBP2**

Inspection of all the available cDNA and EST (Expressed Sequence Tag) sequences derived from the mouse *Aebp2* gene revealed that mouse *Aebp2* is comprised of 11 individual exons that spread over a 60-kb genomic interval in mouse chromosome 6 (Figure 2.2B). A single exon (exon 1b) encodes both the acidic and neutral domains of mouse AEBP2 (a.a. 1–201).
Figure 2.2. Exon structure and alternative splicing of the mouse Aebp2 gene. (A) The same colors and patterns as Figure 1 were used to represent different protein domains. The percent identities were calculated through comparing the amino-acid sequences of AEBP2 from individual organisms vs. the mouse. The N terminus of the insects’ JING (dotted-lines) was omitted in this comparison due to the lack of any detectible sequence similarity. The AEBP2 of the fish lineage also has similar acidic and neutral domains, but these do not show any similarity to those of mammalian AEBP2 (indicated with a green box). (B) Isoforms and stage-specific expression of the mouse Aebp2. A total of 11 exons have been found in the mouse Aebp2 so far. Three START codons are indicated: two within Exon 1b and the third one in Exon 2. Three STOP codons are also indicated within exons 8, 9a and 9b. Alternative splicing of these exons could result in at least six different isoforms, including the two major forms detected in this study, Isoform 1 (52 kDa) and Isoform 2 (31 kDa). (C) Alternative splicing confirmed through RT–PCR. This analysis used total RNA from individual tissues, the different amounts of which were normalized to an internal control gene, p53. The combination of exons 1a and 2 is highly expressed in a 14-day-old embryo, which belongs to Isoform 2. (D) Western blot of the mouse AEBP2. Isoform 1 (52 kDa) is dominant in the most tissues examined, while Isoform 2 (31 kDa) was detected only in a 14-day-embryo.
A) % Identities

<table>
<thead>
<tr>
<th></th>
<th>Acidic</th>
<th>Neutral</th>
<th>Zinc finger</th>
<th>Basic</th>
</tr>
</thead>
<tbody>
<tr>
<td>human</td>
<td>(93%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>chicken</td>
<td>(87%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fish</td>
<td>(83%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>insect</td>
<td>(38%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B) Mouse exon structure

Isoform 1 (52 kDa)

Isoform 2 (31 kDa)

C) Western blot

D) Western blot
The following three exons (exons 2–4) encode the serine-rich domain and the three zinc finger units (a.a. 201–345), and the next four exons (exons 5–8) encode the basic domain (a.a. 346–496). This exon structure also involves two sets of alternative splicing: one is between the two 5′-end exons (1a and 1b) and the other is between three individual 3′-end exons (8, 9a and 9b). The first alternative splicing yields the two different forms of AEBP2 cDNAs: Isoform 1 with exon 1b and Isoform 2 with exon 1a (Figure 2.2B). The Isoform 1 cDNA has three potential START codons that are in-frame with the rest of the AEBP2 exons. The first two are located within exon 1b, and the third one is within exon 2. In contrast, exon 1a does not contain any in-frame ATG codon, and thus the Isoform 2 cDNA starting from exon 1a likely uses the third START codon located within exon 2. Since the two potential START codons for Isoform 1 and 2 are separated by 222 amino-acid residues, a large protein size difference is predicted between these two forms. Also, the acidic and neutral domains should be included only in the larger form (Isoform 1), but not in the smaller form (Isoform 2). The second alternative splicing occurs between several 3′-end exons: Exon 8, 9a and 9b. The first form of the 3′-end alternative splicing simply ends at Exon 8 with its STOP codon and 3′-UTR (Figure 2.2B). The second form splices out the 3′-UTR of Exon 8, and join only the coding region of Exon 8 (named Exon 8s) to another downstream exon (exon 9a). Exon 9a has an additional 14-amino-acid-long coding region with its STOP codon. In fact, this small peptide region from exon 9a corresponds to the lysine-rich domain that is conserved within every mammalian species (Figures 2.1 and 2.2A). The third form of the 3′-end alternative splicing connects exon 8s to another further downstream exon (exon 9b), and this form also has an additional 7-amino-acid-long coding region. The evolutionary conservation of this exon is, however, currently unknown. According to our own survey on ESTs and cDNAs, most cDNAs derived from different tissues start with exon 1b and
end with exon 8 (496 a.a. long), but a subset of cDNAs from early embryonic stages start with exon 1a, and end with either exons 8 or 9a (274 or 288 a.a. long). The other combinations of cDNAs are also likely, but the above two forms are believed to be the major forms for mouse AEBP2 cDNAs. The two sets of alternative splicing predicted from cDNA sequences were tested through RT-PCR-based experiments using total RNAs isolated from several mouse tissues and cell lines (Figure 2.2C). As shown in Figure 2.2C (second and third row), the exon combination of 1b and 2 (1b + 2) was detected throughout all the tissues tested. The exon combination of 1a and 2 (1a + 2) was similarly detected in all the samples except the Neuro2A cell line. The overall expression levels of the 1b + 2 combination were higher than those of the 1a + 2 combination. However, this trend was reversed in the embryonic tissues: the expression of the 1a + 2 combination was much higher than that of the 1b + 2 combination. This is also consistent with the fact that all of the available EST clones containing the 1a+2 combination were derived from early embryonic tissues. The detection of these two alternative first exons further suggests the presence of two alternative promoters for mouse Aebp2: one may be responsible for the ubiquitous expression whereas the other one for the embryo-specific expression. We also performed another set of RT-PCR analyses to confirm the presence of the 3′-end alternative splicing (Figure 2.2C, fourth and fifth row). Since exon 9a is still part of the 3′-UTR of exon 8, we avoided testing the second exon combination (7 + 8s + 9a). We mainly analyzed two different combinations of the 3′-end alternative splicing (7 + 8 and 7 + 8s + 9b). Both exon combinations were detected in most of the tissues examined, but the expression levels of the 7 + 8 combination appear to be higher than those of the 7 + 8s + 9b combination. Overall, RT-PCR analyses indeed confirmed the presence of two alternative splicing patterns, and detected somewhat stage and tissue specificity of these splicing patterns.
We further tested the presence of multiple isoforms of the AEBP2 protein with western blot experiments using a polyclonal antibody raised against the human AEBP2 protein (Figure 2.2D). This analysis detected two main forms of AEBP2 (52 and 31 kDa, respectively). The larger form (52 kDa) was detected in the HeLa nuclear extracts as well as in the several tissues of the mouse, including brain and testis. Detection of another band in the brain sample was likely caused by non-specific binding to other unknown proteins. The 52-kDa protein appears to correspond to the largest ORF predicted from Isoform 1 cDNAs based on its similar size and ubiquitous expression in most tissues. In the brain of a 14-day-old embryo, however, the same analysis detected only the smaller form (31 kDa). The 31-kDa protein likely corresponds to the ORF derived from Isoform 2 cDNAs based on its smaller size and limited expression in embryonic stages. Due to the limited separation capability of SDS-PAGE gel electrophoresis, however, it is currently unknown whether these two isoforms also have different C-terminal endings, as predicted from the 3’-end alternative splicing. Nevertheless, the above analysis confirmed the existence of two major forms of AEBP2 in vivo: the adult-specific larger form (52 kDa) and the embryo-specific smaller form (31 kDa).

**DNA-Binding Motifs of AEBP2**

To characterize DNA-binding motifs for AEBP2, we made three GST-fusion constructs containing different isoforms of mouse AEBP2: Construct I (a.a. 1–496) corresponding to the 52-kDa larger form, Construct II (a.a. 223–496), corresponding to the 31-kDa smaller form, and Construct III (a.a. 223–348), corresponding to a truncated version lacking the basic domain. All of these GST-fusion proteins were successfully expressed in bacteria. However, only the two GST-fusion proteins from Constructs II and III exhibited some levels of DNA-binding activity. The reason for the inactivity of the GST-fusion protein I is currently unknown. We also tested
the binding capability of the GST-fusion protein II to the sequences of several individual DNA fragments, which have been derived from ChIP cloning experiments designed to identify the *in vivo* target loci of the AEBP2. This will be described in detail in the following section. Among these short sequences, one sequence named T1 showed consistently high levels of DNA-binding affinity to the GST-fusion protein II. Thus, we have selected and used this particular sequence as a main probe for our DNA-binding motif assays (*Figure 2.3*). Several mutant series of the T1 duplex probes were designed and used for our DNA binding motif studies of AEBP2. First, an internal 28-bp-long region of the T1 sequence (8th to 35th position) was divided into four individual 7-bp-long sections, and the sequence of each section was changed into a 7-bp-long stretch of A's (Probe II-1 through 4). Each of these mutant probes was used as a competitor to the P$^{32}$-labeled T1 probe for gel shift assays (*Figure 2.3A*). A shift band was completely abolished in a self-competition experiment, using a 100 to 1 molar ratio of the P$^{32}$-unlabeled to labeled probes (*Figure 2.3A, Lane 1*). Similarly, the third mutant (lanes 6 and 7) competed and abolished the band, indicating that this region is dispensable for the binding activity.

In contrast, the three remaining mutants did not compete at all (lanes 2, 3, 4, 5, 8, 9), indicating that the three regions covered by these mutants are important for the binding to AEBP2. This initial series of competition experiments demonstrated that the two regions (8th to 21st and 29th to 35th position) of the T1 sequence are critical for the binding to AEBP2. These two regions were further analyzed using a second series of mutants, each of which has a 3-bp-long stretch of A's (*Figure 2.3B*). This series of experiments identified 3 smaller regions showing relatively weak competition (*Figure 2.3B, lanes 6, 8, 9, marked with asterisk*), indicating that these three small regions (CTT, CAG, GCC) are likely critical for the binding to AEBP2.
Figure 2.3. DNA-binding motifs of AEBP2. Gel shift assays of the GST-AEBP II fusion proteins. (A) Competition assay with the first series of mutants (II-1, II-2, II-3, II-4). Each contains an internal 7-bp-long region substituted with a stretch of As. (B) Competition assay with the second series of mutants. Each contains a 3-bp-region substituted with a stretch of As. Mutants 6, 8 and 9 did not compete well (as indicated with asterisk), and thus these three small regions are thought to be the most critical for the AEBP2 binding. (C) The T1 probe was also competed with other short sequences derived from the ChIP cloning experiment using the AEBP2 antibody. The actual sequences for these ChIP fragments are shown on the right column. (D) Competition assay with different probes: the AE-1 probe from the original sequence that was used to identify the AEBP2 (1), the AB-13 probe from an independent ChIP-derived sequence with high affinity to AEBP2, the Zipped probe with the II-3 region removed and the Wide probe with the II-3 region duplicated.
We also performed another independent series of gel shift assays using the DNA sequences derived from the 20 shortest DNA fragments derived from the ChIP cloning experiments using the AEBP2 and SUZ12 antibodies (Figure 2.4). This survey identified four individual sequences showing high levels of binding affinity to AEBP2 (Figure 2.3C, S8, S10, AB-13 and S17). Further inspection of the four sequences, averaging 50 bp in length, revealed that all of these sequences share two small motifs with the T1 sequences, CTT and GCC. However, the distance between these two motifs is somewhat variable among these potential binding sites of AEBP2, ranging from 15 to 23 bp in length. Thus, the effect of spacing between motifs was tested using two mutants of the T1 sequences (Figure 2.3D, lanes 6–9). The spacing region of the T1 sequence was either deleted (Zipped) or duplicated (Wide). The Zipped probe did not compete but the Wide probe competed very well, demonstrating that the spacing between the two motifs requires some minimum distances, but that the size can be variable without any major effect on the binding affinity to AEBP2. We also tested the binding affinity of the original sequence that was used to identify the AEBP2 protein ((He et al., 1999); AE-1, lanes 2 and 3). This sequence competed at some levels, but the binding affinity was much lower than the other sequences used for this study. We repeated the above experiments using GST-fusion protein III, which lacks the basic domain. The results did not show any difference from those of GST-fusion protein II (data not shown). This suggests that the three Gli-type zinc finger motifs shared by both fusion proteins are mainly responsible for the DNA-binding activity of AEBP2. In sum, the above series of experiments identified a DNA-binding motif for AEBP2, which displays an unusual bipartite motif structure, CTT(N)15-23cagGCC with the lowercase bases being less critical for binding.
ChIP Cloning of \textit{in vivo} Target Loci Bound by AEBP2 and SUZ12 Proteins

A series of ChIP cloning experiments were performed to identify \textit{in vivo} target loci bound by AEBP2 (Figure 2.4). We previously developed a modified version of ChIP cloning method, which can be used to directly clone very short DNA fragments without PCR amplification (Huang et al., 2006). This method performs restriction enzyme digestion with 4-bp cutters, such as Sau3AI or Tsp509, directly on the DNA while it is still cross-linked to a target protein as a chromatin complex. This enzyme digestion usually generates much shorter DNA fragments that are compatible with subcloning. Using two polyclonal antibodies against AEBP2 and SUZ12, we have generated two individual libraries containing ChIP-derived DNA fragments. We have sequenced a subset of these two libraries, 250 and 165 clones for AEBP2 and SUZ12, respectively. The average length of the inserts from each library was about 140 bp in length. Individual sequences are available in our website (http://jookimlab.lsu.edu/?q=node/36) and also the associated information can be viewed using a custom track view of the UCSC genome web browser (http://genome.ucsc.edu/cgi-bin/hgTracks?db=mm9&hgt.customText=http://jookimlab.lsu.edu/sites/default/files/Aebp2_bed.txt). This method performs restriction enzyme digestion with 4-bp cutters, such as Sau3AI or Tsp509, directly on the DNA while it is still cross-linked to a target protein as a chromatin complex. This enzyme digestion usually generates much shorter DNA fragments that are compatible with subcloning. Using two polyclonal antibodies against AEBP2 and SUZ12, we have generated two individual libraries containing ChIP-derived DNA fragments. We have sequenced a subset of these two libraries, 250 and 165 clones for AEBP2 and SUZ12, respectively.
**Figure 2.4. ChIP cloning scheme and experimental strategies.** Our modified ChIP cloning method has one additional step compared to other existing protocols: the restriction enzyme digestion step (step 4) right before the elution step. This allows immediate cloning of shorter ChIP DNA fragments without PCR amplification. Also, this shortening of ChIP fragments further trims other unnecessary long regions from either side of each ChIP fragment while preserving the actual binding site for a given DNA-binding protein. This further facilitates accurate prediction of DNA-binding motifs. The isolated ChIP fragments were subsequently used for the following three experiments. First, we used the sequences derived from the shortest ChIP fragments as probes for our gel shift assays of AEBP2. Second, we used the sequences from the ChIP cloning to identify enriched DNA motifs. Third, we also identified in vivo target loci for both AEBP2 and SUZ12. These confirmed loci were later tested for the co-occupancy by AEBP2 and SUZ12.

The average length of the inserts from each library was about 140 bp in length. Individual sequences are available in our website (http://jookimlab.lsu.edu/?q=node/36) and also the associated information can be viewed using a custom track view of the UCSC genome web browser (http://genome.ucsc.edu/cgi-bin/hgTracks?db=mm9&hgt.customText=http://jookimlab.lsu.edu/sites/default/files/Aebp2_bed).
Initial inspection of the sequences from these two libraries derived the following conclusions. First, both libraries contain fractions of repeat sequences, 124/250 for the AEBP2 set and 94/165 for the SUZ12 set. Second, the remaining non-repeat sequences of both sets mapped closely to gene regions of the mouse genome. The list of the genes associated with each set was compared with the list of the mouse genes that are known to be bound by the PRC2 (Boyer et al., 2006). This comparison confirmed that 53 out of the 126 sequences of the AEBP2 set (42%) were derived from the known PcG target loci while 18 of the 71 sequences of the SUZ12 set (25%) came from the PcG target loci. It has been shown that only a small fraction of mammalian genes are controlled by the PRC2 (less than 5% of the entire gene set of mammals) (Boyer et al., 2006). Thus, the observed high levels of enrichments of the PcG downstream genes among the ChIP cloning sets of AEBP2 (42%) and SUZ12 (25%) strongly suggest that both proteins, AEBP2 and SUZ12, are likely involved in the targeting of the PRC2 complex. Some of the notable PcG loci that were identified through our ChIP cloning trials include: Grm8, Abcc3, and Phkb from the AEBP2 set and Pax1, Acvrip1 and A20Rik from the SUZ12 set. The summary of this comparison is available on our website (http://jookimlab.lsu.edu/?q=node/36).

Third, analysis of the non-repeat sequences by the MEME program (http://meme.sdsc.edu/meme/intro.html) revealed the presence of several DNA motifs that were overrepresented within each set of ChIP sequences. The two most significant motifs from each set are shown in sequence logo format (Figure 2.5). The first motif of the AEBP2 set is 22 bp long, and shared by 77 out of the 126 individual sequences. Interestingly, this motif contains several small regions showing sequence similarity to the two 3-bp-long critical regions, which were shown to be critical for the binding to AEBP2 by our previous gel shift assays, CTT and GCC (Figure 2.3).
Figure 2.5. Enriched motifs within the AEBP2 and SUZ12 ChIP sequences. Motifs were predicted with each set of AEBP2 and SUZ12-ChIP sequences using the MEME program (http://meme.sdsc.edu/meme4/cgi-bin/meme.cgi). (A) The two most significant motifs identified from each set are shown in the sequence logo format. (B) The most significant motif from the AEBP2 set was shown with the two small motifs, which have been also independently identified through gel shift assays. This most significant motif was also aligned with 20 individual sequences.

We performed another gel shift assays using several AEBP2-ChIP fragments, which are shown in Figure 2.5B, and the results confirmed again that these ChIP fragments indeed contain the DNA-binding sites for AEBP2 (http://jookimlab.lsu.edu/?q=node/36). The second motif from the AEBP2 set is also 22 bp long, and displays a somewhat similar C-rich consensus sequence as the first motif. This motif is shared by the 56 individual sequences of the AEBP2 set. Similar analyses also identified two motifs from the SUZ12 set, which are shared by 45 and 35
sequences of the total 71 individual sequences, respectively. It is interesting to note that both motifs contain small regions similar to the GAGA motif, which is a frequent DNA motif in the Polycomb Response Element (PRE) of *Drosophila* (Ringoes and Paro, 2004, Ringoes et al., 2003). However, the two motifs from the SUZ12 set are shorter and also shared by fewer of the individual sequences than the two motifs from the AEBP2 set. This suggests that the sequences of the SUZ12 set are more heterogeneous than those of the AEBP2 set. In sum, many *in vivo* target loci of AEBP2 appear to be derived from the known PcG target regions, and these loci display C-rich sequences with several small motifs, which are reminiscent of the two critical DNA-binding sites of AEBP2.

**Co-occupancy Test with AEBP2 and SUZ12 ChIP Assays**

The identified genomic loci by AEBP2 and SUZ12-ChIP cloning were further analyzed using individual ChIP experiments (*Figure 2.6*). These individual ChIP experiments were performed to measure what fraction of each library contains genuine *in vivo* target loci for each protein. According to the results from four different trials, 18 out of 19 tested loci from the AEBP2 set showed consistent enrichment with the AEBP2 antibody, indicating that about 94% of the AEBP2 set likely contains genuine *in vivo* target loci. A similar test indicated that about 70% of the SUZ12 set (16/23) likely contains *in vivo* target loci. The representative results from these series of ChIP experiments are shown in *Figure 2.6A* and the remaining portion of the results along with other relevant information are also available on our website (http://jookimlab.lsu.edu/?q=node/36).

Since AEBP2 is a potential targeting protein for PRC2, we further tested this possibility through performing co-occupancy tests. If the two proteins, AEBP2 and SUZ12, bind to target loci together as a protein complex, many confirmed target loci from one protein (AEBP2) should
be also positive with another ChIP experiment using the antibody against the second protein (SUZ12), and vice versa. The results from this co-occupancy test are as follows. Out of the 19 AEBP2 loci tested, 15 (79%) were positive with the SUZ12-ChIP experiments. None of the negative loci from the AEBP2-ChIP were positive with the SUZ12-ChIP. On the other hand, 16 out of the 23 confirmed loci of the SUZ12 set (70%) were also positive with the AEBP2-ChIP. We also extended this co-occupancy test to the known target loci of the PcG complex. Out of the six loci tested, three loci turned out to be positive with both AEBP2 and SUZ12-ChIP experiments (Figure 2.6A and B; http://jookimlab.lsu.edu/?q=node/36). We also included YY1 ChIP experiments to test if these loci are bound by YY1. None of the tested loci were positive with the YY1-ChIP, suggesting that YY1 may not be involved in the targeting of the mammalian PRC2 to these loci. The HoxA9 locus was analyzed in further detail by including two additional primer sets. The precise PcG target region within this locus is located 3.75-kb upstream of the gene as demonstrated in the high levels of the DNA enrichment by the SUZ12-ChIP experiments (Figure 2.6B). The two other regions also show some levels of the enrichment with the SUZ12-ChIP, which is consistent with the previous study (Cao et al., 2004). However, the AEBP2-ChIP showed high levels of the enrichment only at the 3.75-kb upstream region, demonstrating high levels of target selectivity by the AEBP2-ChIP experiment. In sum, the co-occupancy tests revealed that an unusually large fraction of in vivo target loci are co-occupied by both AEBP2 and SUZ12 proteins, further supporting the initial idea that AEBP2 is likely a targeting protein for PRC2.
Figure 2.6. Co-occupancy test of the gene loci identified through AEBP2 and SUZ12 ChIPs.

(A) Co-occupancy test of the genes identified from AEBP2 and SUZ12-ChIP sequencing. The left panel indicates the genes derived from each round of ChIP sequencing. Grm8, Abcc3 and Phkb were derived from the AEBP2-ChIP sequencing, while Pax1, Acvrinp1 and A20Rik were from the SUZ12-ChIP sequencing. The three loci (Grm8, Abcc3 and Phkb) were first tested through individual ChIP assays using the AEBP2 antibody, and later using another antibody (SUZ12) for the co-occupancy test. This was also repeated for the SUZ12 set (Middle). A subset of the known Polycomb target loci (Barx1 and Zic1) were also included for the co-occupancy test (bottom). (B) Co-occupancy of AEBP2 and SUZ12 on the HoxA9 locus. The co-occupancy of AEBP2 and SUZ12 was only detected at the 3.75-kb upstream region of the transcription start site of HoxA9. We also performed another independent ChIP using the YY1 antibody to test if YY1 is also involved in the targeting of the PRC2 to these loci.
Discussion

In the current study, we have characterized AEBP2 in terms of evolutionary conservation, genomic structure, DNA binding motifs and potential targeting roles for the Polycomb group repression complex 2 (PRC2). AEBP2 contains two evolutionarily conserved protein domains, the zinc finger and basic domains, and these two domains are also shared by the flying insect protein JING. JING has recently been recognized as a member of the PcG in *Drosophila*. Mammalian AEBP2 is driven by two alternative promoters and produces at least two major forms of the protein, and these isoforms show developmental stage-specific expression patterns: the adult-specific larger form (52 kDa) and the embryo-specific smaller form (31 kDa). The AEBP2 protein binds to a DNA-binding motif with an unusual bipartite structure, CTT(N)15-23cagGCC. A large fraction of AEBP2's target loci also map closely to the known target loci of the mammalian PRC2. We further confirmed that many of these loci are indeed co-occupied by the two proteins AEBP2 and SUZ12. This supports the prediction that AEBP2 is a targeting protein for the mammalian PRC2 complex.

Global alignment of 20 AEBP2 sequences identified two evolutionarily conserved domains, the zinc finger and basic domains, which are located at the C-terminus of the protein (Figure 2.1). These two domains maintain very high levels of sequence conservation throughout all the vertebrates, greater than 80% sequence identity in the 280 amino-acid long region. Similar domains are also found even in the flying insect protein JING, sharing an overall 38% sequence identity with AEBP2 (Figure 2.2A). Although the observed sequence similarity is relatively low, the insects' JING is thought to be a homolog to vertebrates’ AEBP2 based on the following reasons. First, although the zinc finger domains of both genes are comprised of three typical Gli-type finger units that are quite prevalent in eukaryotic genome, the 2nd finger shows
relatively high levels of sequence similarity (63%) between the two groups. Interestingly, the second finger is also three amino acids longer than typical Gli-type finger units (He et al., 1999). Yet, this unique variation is also detected in all the sequences of both AEBP2 and JING proteins (http://jookimlab.lsu.edu/?q=node/36). Second, the basic domain is characterized by a stretch of basic (Lys and Arg) amino-acid residues at the beginning and another stretch of hydrophobic (Leu, Val and Ile) and aromatic (Phe, Tyr and Trp) amino acid residues at the end. According to database search, this unusual domain is only found in the two proteins, AEBP2 and JING. Furthermore, the unique combination of this novel basic domain along with three Gli-type zinc finger units is found again only within these two proteins. Therefore, it is highly likely that vertebrates’ AEBP2 and the flying insects’ JING have been derived from a common ancestor. According to recent genetic studies in flies, the jing locus genetically interacts with several members of PcG members (Culi et al., 2006). This further suggests that AEBP2 and JING still play a similar role, perhaps in the PcG-mediated repression. If this is the case, the two conserved domains likely play the most central roles for this repression mechanism, DNA binding by the zinc finger units and protein–protein interaction by the basic domain.

One of the unexpected features associated with the Aebp2 gene is the presence of several combinations of alternative splicing, which involve the 5′-end two exons and 3′-end three exons (Figure 2.2B–D). In principle, six different types of AEBP2 protein isoforms are possible although we have detected only two major forms through western blot analyses. The expression of these two major forms is very developmental stage-specific: the larger form (52 kDa) is mainly detected in adult tissues, whereas the smaller form (31 kDa) is found only in embryonic tissues. This stage-specific expression is thought to be driven by the two different promoters located upstream of the two alternative first exons. According to our own surveys using the EST
database, a similar 5'-end alternative splicing of AEBP2 is also detected in other mammals, and the expression of each of the two splicing variants can be easily categorized into either embryonic or adult-specific. In the other vertebrates and invertebrates, however, there appears to be only one first exon for AEBP2 and JING, and also the expression pattern of this cDNA form appears to be spatially and temporally ubiquitous. This suggests that the alternative splicing of AEBP2 and stage-specific expression are unique features found only in mammals. Then, what is the major impetus for the sudden implementation of this alternative splicing for the mammalian AEBP2? This could be explained by the actual products of the alternative splicing: a smaller embryonic form with two conserved domains versus a larger adult form with additional lineage-specific domains (Figure 2.2). Given the similarities in domain structure between the smaller form and other vertebrate AEBP2 proteins, the smaller form is likely involved in more fundamental biological processes than the larger form, such as determining the pattern and axis of animal body during early development. On the other hand, the larger form with lineage-specific protein domains is likely involved in cellular processes that are more species-specific, such as determining the lineage and location of different cell types within the adult tissues. It is interesting to note that the smaller form with conserved domains participates in earlier developmental processes than the larger form with additional lineage-specific domains. This could be another case of ontogeny recapitulating phylogeny in animal evolution (Gould, 1977).

Overall, the alternative splicing and subsequent formation of mammalian AEBP2 isoforms represents a case where alternative splicing has driven functional division and adaptation of genes.

According to DNA-binding motif studies (Figure 2.3), AEBP2 binds to a consensus sequence with bipartite structure, CTT(N)15-23cagGCC, and this binding is mainly driven by the
three zinc finger units. This consensus DNA-binding motif has been further substantiated by the independent observation that the in vivo target loci of AEBP2 also show similar motifs (Figure 2.5). However, the AEBP2 binding to a bipartite structure motif was unexpected given the fact that the three zinc finger units are juxtaposed right next to each other. This suggests that the recognition of the two smaller motifs within the bipartite motif, which is separated by a spacing, (N)15-23, may be driven by individual zinc finger units of either one or two proteins. The model of single protein binding posits bending of the DNA because of the predicted close proximity between individual fingers, whereas the model involving binding by two proteins hypothesizes potential dimer formation of the AEBP2 protein. We favor the first model based on the following reasons. First, the truncated version of AEBP2, GST-fusion protein III, lacks any domains that could function in dimerization, and it still showed unchanged binding preference to the bipartite motif (data not shown). Second, the bipartite motif tends to show higher affinity to AEBP2 when the spacing region, (N)15-23, of the bipartite motif is either homopolymeric or polypyrinidine stretch sequences, such as polyA or poly(CT). These types of sequences are known to be common in bending regions of the genome (De’ jardín, 2004). This is further supported by our independent observation that many confirmed target loci of AEBP2 also exhibit polypyrinidine sequence structures (Figure 2.5). Although we cannot rule out the other possibilities of DNA binding driven by the dimer structure of AEBP2, the above results suggest potential binding of AEBP2 to bent DNA.

The co-occupancy test with the ChIP experiment clearly demonstrated that AEBP2 and SUZ12 bind to similar genomic regions (Figure 2.6). Since none of the PRC2 core proteins are known to be DNA-binding proteins, this further implicates that AEBP2 may act as a targeting protein for this complex. According to our recent data from a mouse model disrupting the Aebp2
locus (H. Kim et al., unpublished results), some of the known PcG downstream genes are indeed de-repressed in these mutant mice (http://jookimlab.lsu.edu/?q=node/36), further confirming this possibility. Also, the previously reported activity of AEBP2 as a transcriptional repressor supports this possibility (He et al., 1999). We do not predict, however, all of the identified target loci of AEBP2 to be PcG target loci based on the following reasons. First, we expect that AEBP2 should be also involved in many other cellular processes besides the predicted PcG-targeting role based on its evolutionary age and also various protein isoforms detected in mammals. YY1 is an example of a similar case: its role has diversified tremendously from its original evolutionarily conserved role in the Polycomb-mediated repression since the split of insects and vertebrates (Gordon et al, 2006; Kim, 2007). Second, as demonstrated in flies, PcG targeting is likely mediated through a combination of several DNA-binding proteins along with critical DNA structures which are yet unrevealed (Ringose et al., 2003). A similar conclusion has been drawn from the current study: although many SUZ12 confirmed loci are also bound by AEBP2, the DNA-binding motifs of AEBP2 were not significantly overrepresented in this pool of the genomic sequences (Figure 2.5). This suggests that AEBP2 may be one of several DNA-binding proteins involved in the targeting of the mammalian PRC2. In that regard, characterizing the functional contexts of each of the AEBP2 binding to the identified in vivo target loci will be of great interest in the near future.

References


CHAPTER THREE:

AEBP2 AS AN EPIGENETIC REGULATOR OF NEURAL CREST CELL DEVELOPMENT
Introduction

* Aebp2 is an evolutionarily well conserved *Gli*-type zinc finger gene that is found in species ranging from flying insects to humans (Kim et al. 2009). This gene was initially identified due to its binding capability to the promoter of the adipocyte P2 gene, hence named Adipocyte Enhancer Binding Protein 2 (*Aebp2*) (He et al. 1999). Since then, *Aebp2* has been increasingly recognized as a component of the mammalian Polycomb Repression Complex 2 (PRC2) due to its frequent co-purification with the other components of PRC2 (Cao and Zhang 2004; Peng et al. 2009; Shen et al. 2009; Li et al. 2010; Pasini et al. 2010). According to recent studies, AEBP2 is indeed a DNA-binding protein with its consensus DNA-binding motif being CTT(N)_{15-23}cagGCC. Also, the majority of its genome-wide target sites overlap very well with the known target loci of PRC2, suggesting AEBP2 is a targeting protein for the mammalian PRC2 (Kim et al. 2009). The *in vivo* functions of *Aebp2* are currently unknown, but are likely involved in cell migration based on the following observations. First, *jing*, a *Drosophila* homolog of *Aebp2*, was identified as a gene controlling the border cell migration within eggs (Liu and Montell 2001). Second, the expression of mouse *Aebp2* is mainly detected within cells of neural crest origin (this study), which are notable for their migratory capability during vertebrate development. Thus, the *in vivo* roles of *Aebp2* are most likely associated with the migration and development of neural crest cells.

The neural crest cell (NCC) is a transient, multipotent cell population that gives rise to many different cell types for vertebrate organs, including those in the enteric nervous system and endocrine system, facial cartilage and bone, and melanocytes. One unique feature associated with NCC is its migration capability from the neural crest to various locations in the developing vertebrate (Sauka-Spengler and Bronner-Fraser 2008a, 2008b). Several signaling pathways are
involved in this migration process, including RET and EDNRB pathways. RET encodes a receptor tyrosine kinase that recognizes GDNF (Glial cell line-Derived Neurotrophic Factor) whereas EDNRB (Endothelin Receptor B) encodes a G protein-coupled receptor that recognizes EDN3 (Endothelin 3). Mutations in these two pathways quite often manifest as human genetic disorders, including Hirschsprung’s disease (HSCR) and Waardenburg syndrome (WS). The disease phenotype of HSCR is obstruction of the gastrointestinal tract, resulting in a pathologically enlarged colon, or ‘megacolon.’ This is caused by the absence of NCC-derived ganglia and subsequent aperistalsis in the colon (McCallion et al. 2003; Amiel et al. 2008; Tam and Garcia-Barcelo 2009). More than half of familial and sporadic cases have been shown to be linked to the RET locus although a small fraction of cases are also linked to the EDNRB pathway. On the other hand, the core disease phenotypes of WS are sensorineuronal hearing loss and pigmentary disturbance, which are usually caused by the absence of NCC-derived melanocytes. WS can be further divided into four subgroups based on the presence of additional disease traits: WS Types 1 through 4 (Tachibana et al. 2003; Baxter et al. 2004; Pingault et al. 2010). For example, WS Type 4 (Waardenburg-Shah syndrome) exhibits a similar megacolon phenotype as seen in HSCR in addition to the two WS core traits. WS Type 4 is often caused by mutational defects in several genes in the EDNRB pathway, including EDNRB, EDN3, and SOX10 (Tachibana et al. 2003; Baxter et al. 2004; Pingault et al. 2010). Similarly, WS Types 1 through 3 are also linked to the genes that play significant roles in the migration and development of NCC, such as PAX3 for WS Type 1 and 3, and MITF and SNAI2 for WS Type 2.

In the following, the in vivo roles of Aebp2 have been investigated using a mutant mouse line disrupting its transcription. Aebp2 is essential for early mouse development based on the lethality observed from Aebp2-mutant homozygotes. During embryogenesis, Aebp2 is expressed
mainly in cells of neural crest origin. Consistently, the heterozygotes display a set of phenotypes that are usually caused by defects in the migration of NCC, suggesting critical roles for \textit{Aebp2} in the migration and development of NCC. The results supporting this conclusion have been presented and discussed in this manuscript.

\textbf{Results}

\textbf{Generation of a Mutant Mouse Line Targeting \textit{Aebp2}}

To characterize the \textit{in vivo} functions of \textit{Aebp2}, we generated a mutant mouse line with one gene trap ES clone (BC0681; http://www.sanger.ac.uk/PostGenomics/genetrap/). After we established this mutant line, we first characterized the insertion position of the gene trap vector (\textbeta-Geo). As shown in Figure 3.1A, the \textbeta-Geo vector has inserted into the 1\textsuperscript{st} intron of \textit{Aebp2}. We identified the 5’- and 3’-side junction regions between the \textbeta-Geo vector and the surrounding genomic regions, which subsequently allowed us to develop a set of three primers that could be used for genotyping the embryos derived from the breeding of this mutant line (Figure 3.1B). We also confirmed that the gene trap vector inserted into only the \textit{Aebp2} gene locus with a series of Southern blot experiments (Figure 3.1C). To test the truncation of Aebp2 transcription by the \textbeta-Geo vector, we performed qRT-PCR assays using total RNA isolated from the brains of one-day-old neonates [wild-type (+/+) and heterozygotes (+/-)] (Figure 3.1D). According to separate qRT-PCR runs measuring the expression levels of two alternative forms, the expression levels of \textit{Aebp2} in the heterozygote were lower (about 30\%) than those detected in the wild-type littermate, confirming the proper truncation of Aebp2 expression by the gene trap vector (\textbeta-Geo). We also confirmed this through western blotting (data not shown).
Breeding Experiments of the Aebp2 Mutant Line

We performed two series of breeding experiments to test potential roles of Aebp2 in normal development and survival of the mouse. First, we performed the following three breeding experiments: male or female heterozygotes with their littermates and an intercrossing between two heterozygotes (Table 1). The results revealed a slight reduction in the litter size for both breeding: 8 for both F(+/-) x M(+/-) and F(+/-) x M(+/-) vs. 9 for the control breeding F(+/-) x
M(+/+). The ratios between the heterozygote and wild type in both breeding were very close to the expected mendelian ratio (1:1). In contrast, the intercrossing between two heterozygotes derived a much smaller litter size (6) than that of the control breeding (9). Also, none of the homozygotes for the Aebp2-mutant allele were found among the offspring derived from 19 litters, confirming the embryonic lethality associated with the Aebp2 locus. To determine the exact time point of this lethality, we performed another series of intercrossing breeding experiments with timed-mating, which allowed us to harvest embryos at two different stages: 10.5 and 14.5 day post coitum (dpc), but we did not obtain any homozygotes among the harvested embryos, suggesting that the lethality likely occurs at least before the organogenesis stage (Table 2). In sum, these breeding experiments confirm an essential role for Aebp2 during early mouse development.

**Table 3.1. Genotype distribution of the mice from the breeding of the Aebp2 knockin mice**

<table>
<thead>
<tr>
<th>Genotype</th>
<th>F(+/−) X M(+/−)</th>
<th>F(+/+) X M(+/−)</th>
<th>F(+/−) X M(+/+)</th>
<th>F(+/+) X M(+/+)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+/−</td>
<td>35 (30°)</td>
<td>58 (50)</td>
<td>43 (44)</td>
<td>72 (72)</td>
</tr>
<tr>
<td>+/−</td>
<td>84 (60)</td>
<td>42 (50)</td>
<td>44 (44)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>−/−</td>
<td>0 (30)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No. of mice</th>
<th>119</th>
<th>100</th>
<th>87</th>
<th>72</th>
</tr>
</thead>
</table>

| Average Litter size | 6 | 8 | 8 | 9 |

°This indicates the expected number of F2 pups based on the Mendelian ratio.
### Spatial and Temporal Expression Patterns of Mouse *Aebp2*

Since the *Aebp2* locus in the mutant line has been targeted by the promoterless gene trap vector (β-Geo), we took advantage of this β-Geo reporter system for analyzing the temporal and spatial expression patterns of mouse *Aebp2*. First, we performed a series of β-Gal staining with whole-mount and cryo-sectioned embryos that had been harvested at various developmental stages (Figure 3.2). In the sectioned 6.5-dpc embryos, the Aebp2 expression was detected at the highest levels in the embryonic ectoderm (Ect) and primitive streak (PS), and at moderate levels in chorion (Ch) and allantois (Al) (Figure 3.2A). In the whole-mount embryos (9.5, 13.5, and 14.5 dpc), the Aebp2 expression was detected in the midbrain, the branchial arches and along the somites (Figure 3.2A). In the sagittal-sectioned 15.5-dpc embryos, the Aebp2 expression was detected in relatively high levels in tissues derived from neural crest cells, including dorsal root ganglia, endocrine organs, facial cartilage and bone, and the surface of intestine, heart, and lung (Figure 3.2B-D). Second, we also surveyed the sectioned tissues derived from 2-month-old adult mice of both genders. The most obvious expression sites include brain and testes (data not shown). These results are consistent with those from previous studies, revealing high levels of expression in early embryonic stages and adult brains (He et al. 1999; Kim et al. 2009). Overall,
it is intriguing that Aebp2 expression was most obvious in tissues derived from the neural crest cell, suggesting significant functional roles for Aebp2 in the development of this cell lineage.

**Figure 3.2. Spatial and temporal expression patterns of Aebp2.** (A) β-Gal staining of whole-mount embryos with different developmental stages. In 6.5-dpc embryos, high levels of Aebp2 is detected in ectoderm (Ect) and primitive streak (PS), modest levels in chorion (Ch) and allantois (Al). In 9.5-, 11.5-, 14.5-dpc embryos, Aebp2 expression is consistently detected in the midbrain section and also along the somites. (B) β-Gal staining of a sagittal-sectioned slide from a 15.5-dpc embryo. DRG (Dorsal Root Ganglion), H (Heart), and L (Liver). (C) A different sagittal section of a 15.5-dpc embryo showing the expression of Aebp2 in thymus and pituitary gland. (D) Zoom-in views of the Aebp2 expression in the thymus, DRG, intestine, and heart of a 15.5-dpc embryo.
Visible Phenotypes of the Aebp2 Heterozygotes

While breeding the Aebp2 mutant line, we observed the following phenotypes from the Aebp2 heterozygotes. First, about one quarter of the Aebp2 heterozygotes tend to show a pot-shaped belly, and seem to have difficulty in discharging feces. Furthermore, when we examined the internal organs of these mice, some of these mice displayed enlarged, green-colored colons (megacolon, Figure 3.3A). This megacolon phenotype is caused by the absence of neural crest-derived ganglia and subsequent aperistalsis in the colon (McCallion et al. 2003; Amiel et al. 2008; Tam and Garcia-Barcelo 2009). Thus, the intestines harvested from the Aebp2 heterozygotes were analyzed using the acetylcholine esterase staining method (Enomoto et al. 1998; Carrasquillo et al. 2002). Out of the 28 Aebp2 heterozygotes examined, 8 mice showed a significantly reduced density of ganglion cells in the section between the anus and cecum as compared to the wild-type littermates (Figure 3.3A). Also, this megacolon phenotype seemed to be more pronounced among the older mice.

Second, although we maintained this mutant strain in the 129/B6-mixed background with the black coat color (a/a), we observed a large fraction of the Aebp2 heterozygotes with white spotting at the tail tip (Figure 3.3B). The length of the white spot area varied among the individual mice of the same litter ranging from 0.2 to 1.5 cm, but the lengths of the white area in the littermates from the intercrossing between the Aebp2 heterozygotes were longer than those from the crossing between the wild type and heterozygotes. Some of the Aebp2 heterozygotes even showed white toes on their hind feet. Third, a large portion of the Aebp2 heterozygotes did not exhibit a brisk acoustic startle response to clapping sounds, suggesting potential hearing defects, although this needs to be further substantiated through more physiologic and pathologic tests. In addition to these three phenotypes, we also occasionally observed overgrown teeth.
among the Aebp2 heterozygotes, but much less frequently than the white spotting on the tail tip (Figure 3.3C). Overall, the three phenotypes observed from the Aebp2 heterozygotes are similar to those observed from Waardenburg syndrome Type 4 (WS4): megacolon, hypopigmentation, and auditory defect.

**Figure 3.3. Phenotypes of the Aebp2 heterozygotes.** (A) Comparison of internal organs between the wild-type (WT) and Aebp2 heterozygotes (upper panel). Some of the Aebp2 heterozygotes display an enlarged green-colored colon (Megacolon), which is easily detectible as compared to the normal-size colon from the wild-type mice. Acetylcholinesterase staining further indicates that the Aebp2 heterozygotes have much less ganglion cells in the intestinal section between the anus and cecum than the wild-type mice. The ganglion cells are shown as brown thin fibers on the surface of the intestines (lower panel). Some of the Aebp2 heterozygotes also display white spotting at the tail tip (B) as well as overgrown teeth (C).
Aebp2 Mutation Effects on the Expression Levels of the Disease Genes of NCC

The HSCR and WS phenotypes observed in the Aebp2 mutant are frequently associated with mutations within a set of about 10 susceptible genes that are involved in the RET and EDNRB signaling pathways (Amiel et al. 2008; Tam and Garcia-Barcelo 2009). Since AEBP2 is a DNA-binding protein with NCC-specific expression, it is likely that Aebp2 may control these susceptible loci as a regulator, and subsequently that de-regulation of some of these genes may be responsible for the observed phenotypes in the Aebp2 heterozygotes. To test this prediction, we measured and compared the expression levels of a set of 10 susceptible genes between the Aebp2 heterozygotes and wild-type littermates (Figure 3.4). Since the gene dosage (or expression levels) of these loci are critical during embryogenesis, this series of qRT-PCR analyses mainly used the total RNA isolated from the two groups of embryos with three different stages, 10.5, 14.5 and 17.5 dpc (Figure 3.4).

For this series of qRT-PCR analyses, we first calculated the expression level of each gene relative to that of an internal control, β-actin, and later compared these relative values derived from the Aebp2 heterozygotes and wild-type littermates. As shown in Figure 3.4, the expression levels of Aebp2 in the heterozygotes decreased 40-50 percent (0.5-0.6 fold) compared to those from the wild-type littermates, confirming the disruption of the Aebp2 transcription. In 10.5-dpc embryos, all of the analyzed genes, with the exception of Mitf, showed relatively high levels of expression based on their Ct values ranging from 21 through 29 (Ct value of β-actin being 19). Most genes were down-regulated in the Aebp2 heterozygotes: the genes with the most significant changes were Sox10 (0.5 fold) and Pax3 (0.5 fold). In contrast, Snai2 showed an up-regulation (2 fold), and this up-regulation appears to be very significant based on its high levels of expression (Ct value 21.3).
In 14.5-dpc embryos, the majority of the genes in the Aebp2 heterozygotes were also down-regulated as seen in the 10.5-dpc embryos. The most significant down-regulation was also observed in Sox10 (0.5 fold). However, the down-regulation of Pax3 became much milder in the 14.5-dpc embryos than in the 10.5-dpc embryos. This was also true for the up-regulation of
Snai2: 1.1 fold in the 14.5-dpc embryos compared to 2.0 fold in the 10.5-dpc embryos. This trend was also detected in the 17.5-dpc embryos: the majority of the genes displayed very marginal differences in their expression levels between the Aebp2 heterozygotes and wild-type littermates (data not shown). Overall, the expression analyses revealed that the majority of the genes involved in the migration and development of NCC are down-regulated during the organogenesis stage (E10.5 to 14.5), and that the expression levels of one gene, Sox10, is significantly affected in the Aebp2 heterozygotes.

**In vivo Binding of AEBP2 and PRC2 to the Disease Loci Associated with NCC**

The *in vivo* binding of AEBP2 to the disease loci of HSCR and WS was further tested using Chromatin ImmunoPrecipitation (ChIP) experiments (*Figure 3.5*). Since AEBP2 is a potential targeting protein for PRC2, we also tested the binding of EZH2 and the methylation on Lys27 of Histone 3 (H3K27me3) to these loci, which represent a key component and a functional outcome of PRC2, respectively. For this series of ChIP experiments, we prepared one set of the cross-linked chromatin isolated from the wild-type littermates of the 14.5-dpc stage (*Figure 3.5*). We selected the promoter region of each of these disease loci for this survey. First, the majority of these loci except Zfhx1 were indeed bound by AEBP2 based on the detection of enrichment of the immunoprecipitated DNA by polyclonal AEBP2 antibodies. This was also true for EZH2 and H3K27me3: the majority of the loci except Zfhx1 showed the enrichment of the immunoprecipitated DNA by the EZH2 and H3K27me3 antibodies. These results confirmed the *in vivo* binding of AEBP2 and PRC2 to the disease loci of HSCR and WS. Also, the co-occupancy by AEBP2 and PRC2 further supports that AEBP2 may be a targeting protein for PRC2 (Kim et al. 2009).
Aebp2 Mutation Effects on the PRC2 Involvement in the Disease Genes of NCC

Since AEBP2 is a potential targeting protein for PRC2, we further hypothesized that the changes observed in the expression levels of several genes may be a result of de-regulation of the PRC2-mediated control in the Aebp2 heterozygotes (Figure 3.4). To test this hypothesis, we performed another series of similar ChIP experiments as described above, and compared the levels of the binding of AEBP2, EZH2, and H3K27me3 to these loci between the wild type and Aebp2 heterozygotes (Figure 3.6).
Figure 3.6. Aebp2 mutation effects on the PRC2-mediated regulation of the NCC-associated genes. The levels of AEBP2 and EZH2-binding to the NCC-associated genes were compared between wild-type (blue) and Aebp2 heterozygous (yellow) embryos with qPCR to measure the immunoprecipitated DNA derived from 14.5-dpc embryos (A,B). The methylation levels of H3K27me3 was also compared between the two types of embryos (C). The amount of each precipitated DNA is presented as a relative value (%) to that of the input DNA (y-axis). The values derived from the wild-type and Aebp2 heterozygote embryos are presented together per each gene (x-axis).
In the majority of the tested loci, the enrichment levels of the precipitated DNA by the AEBP2 antibody were lower in the Aebp2 heterozygotes than in the wild-type embryos (Figure 3.6A). This is expected since the protein levels of AEBP2 should be lower in the Aebp2 heterozygotes than in the wild-type embryos. This was also the case for EZH2: the enrichment levels on the majority of the loci were similarly lower in the Aebp2 heterozygotes (Figure 3.6B). Interestingly, however, the methylation levels of H3K27me3 on these loci were higher in the Aebp2 heterozygotes than in the wild-type embryos (Figure 3.6C). The higher levels of the methylation of H3K27me3 were unexpected since the AEBP2 and EZH2 binding to these loci (and thus the functional involvement of PRC2) was found to be less in the Aebp2 heterozygotes than in the wild-type embryos. This may indicate the possibility that the methylation level of H3K27me3 may not be simply proportional to the level of the PRC2 binding to any given locus, which has been observed in the recent studies of Jarid2, another potential targeting protein for PRC2 (Peng et al. 2009; Shen et al. 2009; Li et al. 2010).

Overall, the changes in the methylation levels of H3K27me3 observed in the Aebp2 heterozygotes supports the initial prediction that Aebp2 likely controls the genes associated with the migration and development of NCC through the PRC2-mediated mechanism. Also, since the H3K27me3 mark is regarded as a repression signal, the increased levels of methylation of H3K27me3 is somewhat consistent with the reduced levels of the transcription of the genes associated with NCC in the Aebp2 heterozygotes (Figure 3.4).

Discussion

In this chapter, the in vivo roles of Aebp2 have been investigated using a mutant mouse line disrupting its transcription. Aebp2 is essential for early mouse development based on the lethality observed from Aebp2-mutant homozygotes. Furthermore, the half dosage of Aebp2
appears to be insufficient for the proper migration of some neural crest cells as the Aebp2 heterozygotes display a set of phenotypes very similar to those from HSCR and WS. The majority of the genes involved in the RET and EDNRB signaling pathways appear to be downstream target genes of Aebp2 and PRC2, and also changes in the expression levels and in the methylation levels of H3K27me3 of some of these genes are likely accountable for the phenotypes observed in the Aebp2 heterozygotes. These results suggest that Aebp2 may control these genes through the PRC2-mediated epigenetic mechanism, and also that epigenetic mechanisms are likely involved in the pathogenesis of WS and HSCR.

Genetic breeding experiments revealed embryonic lethality in the Aebp2-mutant homozygotes but survival of the heterozygotes to adulthood with fertility (Table 1). The breeding experiments also estimated the timing of the observed lethality to be before 10.5 dpc (Table 2). The exact timing and cause of this lethality remain to be further investigated, but are likely similar to those observed from the other components of PRC2, such as Ezh2, Eed, and Suz12 (Faust et al. 1995; O’Carroll et al. 2001; Pasini et al. 2004). The null mutants for these genes fail to form the three germ layers after implantation, suggesting essential roles for these genes in the lineage specification of the germ layers. Given the tight interactions between Aebp2 and PRC2 (Schuettengruber and Cavalli 2009), we predict that Aebp2 plays critical roles in establishing the three germ layers, and thus the observed lethality in the Aebp2-null mutants may be an outcome of the failure of formation of the three germ layers during the gastrulation stage. The evolutionary conservation of Aebp2 is noteworthy: its homologues are present in species ranging from flying insects to humans (Kim et al. 2009). Given this evolutionary conservation, Aebp2 is likely involved in the regulation of a large number of genes and pathways, and thus its depletion should be detrimental for the survival of the embryos. Overall, the embryonic lethality
observed from the Aebp2-null mutants suggests an essential role for this PcG gene during early embryogenesis.

According to the results derived from the previous studies, the expression patterns of Aebp2 are considered to be ubiquitous (He et al. 1999; Kim et al. 2009). However, one unique observation from this study is the detection of very high levels of Aebp2 expression in neural crest cells during embryogenesis (Figure 3.2). This unexpected observation appears to be somewhat consistent with Aebp2’s functional connection with PRC2. The migratory NCC is regarded as a multipotent stem cell since it gives rise to so many different cell types in the major organs of adult vertebrates (Stemple and Anderson 1992; Sauka-Spengler and Bronner-Fraser 2008a, 2008b). Stem cells are characterized by two core features, multipotency and self-renewal without differentiation, and these features are usually maintained by epigenetic mechanisms, especially by PRC2 (Boyer et al. 2006; Lee et al. 2006; Chen et al. 2008; Jaenisch and Young 2008). Migratory NCC likely employs PRC2 to maintain these properties during embryonic development. Therefore, Aebp2 expression in NCC may be required to provide these two properties to this stem cell population. If this is the case, the other components of PRC2 should also be highly expressed in NCC, as is Aebp2. This will require further testing in the near future.

Although the homozygotes for the Aebp2-knockin allele are lethal, the heterozygotes are viable, fertile, and display intriguing phenotypes such as enlarged colon and hypopigmentation (Figure 3.3). Since the Aebp2-knockin allele disrupts the transcription of Aebp2, this mutation is regarded as a loss-of-function-type mutation. The phenotypes generated by the Aebp2 mutation are also regarded as dominant traits based on their detection in heterozygotes. Therefore, the dominance of these phenotypes is likely an outcome of haploinsufficiency, meaning the reduced dosage of Aebp2 is responsible for the observed phenotypes. Similar
situations also occur in human patients with Hirschsprung’s disease (HSCR) and Waardenburg Syndrome (WS). In most cases of these disorders, mutational defects are found in the genes involved in the migration process of NCC, RET and EDNRB signaling pathways (Amiel et al. 2008; Tam and Garcia-Barcelo 2009). The disease alleles are also loss-of-function-type mutations, and inherited as autosomal dominant traits. Therefore, haploinsufficiency is also the primary mode for the dominant phenotypes by these disease alleles. Overall, there are many similarities between the Aebp2-knockin allele and the disease alleles of HSCR and WS. In particular, the similar mode of the phenotype dominance, haploinsufficiency, may indicate that the migration process of NCC is very susceptible to changes in the gene dosage of the participating loci. Thus, it is likely that the gene dosage of Aebp2 is very critical for the proper migration and development of NCC.

As a DNA-binding protein, AEBP2 most likely exerts its in vivo roles through its unknown downstream genes. As predicted, ChIP experiments confirmed that AEBP2 indeed binds to the majority of the genes involved in the development and migration of NCC during embryogenesis (Figure 3.5). Expression analyses further confirmed changes in the expression levels of some of these genes by the half dosage of Aebp2 (Figure 3.4). In particular, one gene (Sox10) is consistently down-regulated in the Aebp2 heterozygotes. This is analogous to the reduced gene dosage of SOX10 frequently linked to WS Type 4 in humans. Also, the phenotypes observed in the Aebp2 heterozygotes are seen in human patients with WS Type 4 (Pingault et al. 2010). It is possible that Aebp2 is responsible for the observed phenotypes via Sox10. However, we cannot rule out the possibility that the effects of the Aebp2 mutation might occur more globally and at much earlier stages than described. If this is the case, the observed phenotypes should not be accounted for by the mis-expression of a single gene. This is evidenced by the
observation that other genes involved in the migration of NCC are also affected in the Aebp2 heterozygotes. These possibilities require further investigation. It will be very interesting to determine if the changes in the Sox10 expression are primarily responsible for the phenotypes observed in the Aebp2 heterozygotes.

HSCR and WS demonstrate incomplete penetrance mainly due to their oligogenic nature and other non-genetic factors involved in their pathogenesis (McCallion et al. 2003; Owens et al. 2005; Amiel et al. 2008; Tam and Garcia-Barcelo 2009). Identification of Aebp2 as a potential disease locus for these disorders is an intriguing possibility. If, as our data suggest, Aebp2 exerts its roles through PRC2, it may require optimal concentrations of the cellular enzymes and substrates necessary for histone modification reactions. The outcome of these reactions may vary depending on the nutritional status and environmental conditions of developing embryos, resulting in different levels of histone modification among individuals. This type of inter-individual differences, also known as epigenetic variations, may be a major factor contributing to phenotypic variations (e.g. incomplete penetrance) (Jirtle and Skinner 2007; Bollati and Baccarelli 2010). Unfortunately, epigenetic variations have not been discernible by traditional genetic studies, which rely on genetic variations. We predict that this is the case for both HSCR and WS since the majority of the associated disease genes are modified by PRC2 (Figure 3.5). It is possible that different levels of histone modifications of the disease alleles are accountable for the phenotypic variations (incomplete penetrance) observed for HSCR and WS. In sum, characterizing Aebp2 as an epigenetic regulator may provide a new and exciting direction for the study of HSCR, WS, and other related disorders.


**Materials and methods**

**Generation and Breeding of the Aebp2 Knockin Mutant Mice**

One gene trap clone, BC0681 (strain 129/OlaHsd) from SIGTR (Sanger Institute Gene Trap Resource, http://www.sanger.ac.uk/PostGenomics/genetrap/), was injected into mouse blastocysts to generate chimeric mice. Injection of these cells into C57BL/6 blastocysts was performed at The Darwin Transgenic Mouse Core Facility (Baylor College of Medicine, Houston, TX, USA). The male chimeric mice were bred with female C57BL/6 mice, and the following F1 offspring with agouti coat color was further genotyped to confirm the germline transmission of the Aebp2-knockin allele. This initial genotyping was performed with PCR using a primer set targeting the NeoR coding region of the gene trap vector (pGT2lxr). All the experiments related to mice were performed in accordance with National Institutes of Health guidelines for care and use of animals.

**Southern Blot and Genotyping by PCR**

Genomic DNA was purified from the spleens of the wild-type and Aebp2 heterozygote mice with DNAzol (Invitrogen). Ten µg of these genomic DNA was used for each of EcoRV and SacI digestion reactions, separated on a 0.8% agarose gel, and finally transferred onto Hybond nylon membranes (Amersham) by capillary blotting. Membranes were hybridized with a $^{32}$P-labeled probe corresponding the 1st intron region of Aebp2 (Figure 3.1).

The mice were genotyped by PCR using the following three primers: F1, 5'-ACCAGGGTGTGAAACAGAAGAACTCTG-3'; R1, 5'-AGGTGCTGCACCTCACCTCCCA-3'; R2, 5'-AACGGTAGGATCCCAAGGGCAGTA-3'. The 570-bp product generated by F1 and R1 primers is amplified from the endogenous allele of Aebp2, thus representing the wild-type allele. In contrast, since the R2 primer is derived from the gene trap vector, the 304-bp product by F1
and R2 represents the Aebp2 knockin allele. PCR conditions were 33 cycles at 95°C for 30 seconds, 60°C for 30 seconds, and 72°C for 30 seconds.

The genders of neonatal mice and embryos were determined by PCR using the primer set of the mouse Sry gene under the same PCR conditions described above; mSry-F (5'-GTCCCGTGGTGAGGGCACAAG-3) and mSry-R (5'-GCAGCTCTACTCCAGTCTTTGCCC-3). To prepare genomic DNA from clipped tails or ears, each tissue was incubated overnight at 55°C in the lysis buffer (0.1 M Tris-Cl, 5 mM EDTA, 0.2% SDS, 0.2 M NaCl, pH 8.0, 20 µg/ml Proteinase K). One µl of the lysed extract was first diluted with 30 µl of TE, and one µl of the diluted extract was finally used for each PCR amplification.

**β-Galactosidase Staining**

Pregnant dams with time-mating were sacrificed at various stages during embryonic development. The embryos were fixed overnight in fixing solution (0.2% paraformaldehyde, 0.1 M PIPES buffer pH 6.9, 2 mM MgCl2, 5 mM EGTA). The fixed embryos were then cryo-protected in PBS buffer containing 30% sucrose and 2 mM MgCl2 at 4°C overnight, or until the embryos sank to the bottom. The embryos were further embedded in OCT and frozen at -80°C.

The embedded embryos were sectioned on a crytome (Leica) to 50 micron thickness and placed onto poly-L-lysine coated slides. The sections were further immobilized in the fixing solution for 10 minutes. After rinsing in PBS for 10 minutes, they were placed in detergent rise solution (2 mM MgCl2, 0.01% sodium deoxycholate, 0.02% NP-40 in PBS) for 10 minutes. The sections were then placed at 37°C overnight in the staining solution (2mM MgCl2, 0.01% sodium deoxycholate, 0.02% NP-40, 5mM potassium ferricyanide, 5mM potassium ferrocyanide) containing 1 mg/ml of bromo-chloro-indolyl-galactopyranoside (X-gal). For better contrast, the heart and thymus tissue sections were counterstained with eosin Y.
For whole-mount staining, embryos were fixed in 4% paraformaldehyde for 2 hours and stained overnight at 37°C in the staining solution containing 1 mg/ml of X-gal. Tissue sections and whole-mount embryos were visualized using a dissecting stereo light microscope (Leica MZ75). Images were captured with a digital camera (Model #4.2 Color Mosaic, Diagnostic Instruments Inc.).

Acetylcholinesterase Staining

The intestines from one-month old mice were harvested and fixed in 4% paraformaldehyde for 1 hour at 4°C. After incubation in saturated sodium sulfate overnight at 4°C, the intestines were further incubated for 4 hours in staining buffer (0.2 mM ethopropazine HCl, 4 mM acetylthiocholine iodide, 10 mM glycine, 2 mM cupric sulfate, and 65 mM sodium acetate, pH 5.5). Lastly, the acetylcholinesterase activity was detected by incubating the intestines in 1.25% sodium sulfide, pH 6, for 1.5 minutes.

Quantitative Reverse Transcription PCR and Data Analysis

Total RNA was extracted from tissues using Trizol (Invitrogen). Reverse transcription was performed using the M-MLV kit (Invitrogen). Quantitative real time PCR was performed with the iQ SYBR green supermix (Thermo Scientific) using the iCycler iQ multicolor real-time detection system (Bio-Rad). All qRT-PCRs were carried out for 40 cycles under the standard PCR conditions. We analyzed the results of qRT-PCR based on the threshold (Ct) value. A Δ Ct was first calculated through subtracting the average Ct value of an internal control (β-actin) from the average Ct value of a given target gene. Later, the ΔΔ Ct was calculated through subtracting the Δ Ct value of the target gene in the Aebp2 heterozygote from the Δ Ct value of the same gene in the wild-type littermate. Fold differences were determined by raising 2 to the Δ
Δ Ct powers (Winer et al. 1999). Information regarding individual primer sequences and PCR conditions is available upon request (or see http://jookimlab.lsu.edu/?q=node/36).

**Chromatin ImmunoPrecipitation (ChIP) Experiments**

Chromatin immunoprecipitations were performed according to the protocol provided by Upstate Biotechnology (Upstate Biotech.) with some modification as described previously (Kim et al. 2003). Briefly, mouse embryos at various stages were harvested and homogenized in 10 ml PBS. The samples were treated with formaldehyde to the final concentration of 1% and incubated at 37°C for 10 minutes. Treated samples were sheared by sonication and immunoprecipitated with anti-AEBP2 (Cat. No. 11232-2-AP, ProteinTech Group), EZH2 (Cat. No. ab3748, Abcam), and H3K27me3 (Cat. No. 07-449, Upstate Biotech.) antibodies. Precipitated DNA and protein complexes were reverse cross-linked and purified through phenol/chloroform extraction. Purified DNA was used as template DNA for PCR amplification. PCR reactions were carried out for 40 cycles using standard PCR conditions. The resulting PCR products were run on 1.6% agarose gels containing ethidium bromide. All ChIP assays were performed independently at least three times. The oligonucleotide sequences used for this study are available upon request (or see http://jookimlab.lsu.edu/?q=node/36).

**References**


CHAPTER FOUR:

SUMMARY AND FUTURE DIRECTIONS
Of approximately 25,000 genes in the human genome, 10% encode transcription factors (Babu et al., 2004). Those that are fundamental to organism survival are evolutionarily well conserved, while those that gained their function in adaptation to the environment are species-specific (Edger and Pires, 2009). Genomic studies can reveal the history of a gene and generate new theories based on their evolutionary conservation.

In this dissertation, I have characterized the transcription factor AEBP2, a highly conserved DNA-binding protein that may target the PRC2 complex. Our data revealed that Aebp2 is predominantly expressed in neural crest cells. Thus, Aebp2 may regulate neural crest development through the PcG-mediated machinery.

Chapter two discusses the evolutionary conservation, gene structure and DNA binding properties of Aebp2. Genomic studies showed that the zinc finger and the basic domains are the most evolutionarily conserved domains. In addition, our studies identified two alternative promoters of Aebp2 that can generate two separate protein products: one 31kDa protein and one 51kDa protein. Many genes in our genome have been reported to contain alternative promoters (Kimura et al., 2006). Characterizing these two promoters would be important for understanding the function of Aebp2. Future experiments will characterize these promoters by: performing reporter assays with the two alternative promoters; and generating conditional knockout mice with deletion of each promoter. Most likely, the two Aebp2 isoforms will have different effects on their downstream genes since their interacting protein partners will vary.

For the DNA binding studies of Aebp2, we performed a series of EMSA and ChIP-cloning experiments. The results of these experiments identified the consensus DNA binding motif CTT(N)_{15-23}cagGCC. In addition, ChIP assays demonstrated that Aebp2 co-occupies many
target loci with PRC2. Thus, we propose Aebp2 as a new mammalian targeting protein for PRC2 to specific regulatory regions.

Chapter three addresses the in vivo function of Aebp2. Breeding experiments of Aebp2 knockin mice revealed that Aebp2 is critical for proper embryonic development. Aebp2 homozygous mutants were lethal at an early stage, presumably soon after gastrulation. It has been reported that PcG proteins may be involved in cell lineage determination (Kerpolla, 2009). If this were the case, proper cell differentiation cannot occur in the Aebp2 mutant mice. Since transcription factors are usually expressed in a highly restricted manner, spatially and temporally characterizing the expression sites of the protein is critical to discern its function. The β-gal staining system in the Aebp2 knockin mice allowed us to visualize the expression sites of Aebp2. Although Aebp2 expression was seen throughout embryonic development, high levels of Aebp2 expression were observed in the neural crest cells of E15.5 embryos. Consistently, some adult Aebp2 heterozygotes showed megacolon and white tail tip phenotypes, which are caused typically by defects in neural crest cell development or migration. We also examined the expression of transcription factors, growth factors and receptors known to be critical for proper neural crest development in wild-type and mutant embryos to understand the molecular effect of Aebp2 haploinsufficiency in neural crest development. ChIP assays confirmed that many of these genes were indeed bound by both Aebp2, Ezh2 and contained H3K27me3 in their promoter regions. In addition, when we compared the binding enrichment of Aebp2 and Ezh2 to the promoter regions of the neural crest genes in wild-type and mutant mice, both Aebp2 and Ezh2 binding were reduced in the Aebp2 heterozygous mutant mice. These results support our hypothesis that Aebp2 is responsible for recruiting the PRC2 complex to the genes that regulate neural crest development.
Interestingly, however, the enrichment of H3K27me3 was increased in the Aebp2 heterozygotes compared to the wild-type mice. The elevation of H3K27me3 levels seems to correlate with the reduced expression of the downstream genes we observed in the Aebp2 mutant mice. However, the increase in H3K27me3 in Aebp2 mutant mice was unexpected. A model summarizing our results is in Figure 5.1.

**Figure 5.1.** A model describing the effect of Aebp2 heterozygosity on the PRC2 targeting and the subsequent effect on its downstream genes. **Figure 5.1A** illustrates the binding of PRC2 components to the promoter region of PcG target genes in wild type mice. The histone tails are indicated in a grey protruding away from the nucleosomes (which are indicated in white circles with DNA wrapped around it) and the H3K27me3 are indicated in red. An ideal level of transcription exists when all the components are properly bound. If Aebp2 is downregulated (as in the Aebp2 heterozygous mice), other PRC2 components cannot bind to their targeting region efficiently. As a result, other factors such as Ezh1 or PRC1 may compete against the PRC2 complex and induce ectopic histone methylation on H3K27 and the downstream genes are further repressed (**Figure 5.1B**).
Previous biochemical studies have demonstrated that Ezh2 is the enzyme responsible for trimethylation of the lysine 27 in histone H3 (Cao et al., 2004). If the recruitment of Ezh2 is reduced due to insufficient levels of Aebp2, there should also be reduced levels of H3K27me3. Then why do we observe more histone methylation on H3K27? It has recently been suggested that Jarid2 potentially targets the PRC2 complex. This protein belongs to the family of the Jumonji proteins, which are responsible for histone demethylation. However, the enzymatic domain in the Jarid2 protein is missing, so that it does not exhibit any histone demethylation properties (Takeuchi et al., 2006). According to the four independent experiments of Jarid2, two groups also observed that the enrichment of H3K27me3 increased when the core PRC2 proteins recruitment was reduced (Peng et al., 2009; Shen et al., 2009). These observations are consistent with our data demonstrating increased H3K27me3.

I propose three possibilities for the increased levels of H3K27me3 in the Aebp2 heterozygote mice. First, PRC1 may induce the H3K27me3 machinery when PRC2 cannot fulfill its function, and this may lead to over-methylation of H3K27 in the Aebp2 heterozygous mutant cells (Figure 5.1B). As a result, cell properties that arise may be abnormal. Although the H3K27me3 mark, an outcome of PRC2, is a recruiting signal for PRC1 (Wang et al., 2004; Fischle et al., 2003; Min et al., 2003), some studies have shown that PRC1 can bind to PcG target genes independently of PRC2 (Schoeftner et al., 2006). For instance, in cancer cells the PcG target genes are often methylated in their DNA (Iacobuzio-Donahue, 2009). Lack of critical transcription factors such as Aebp2 could trigger compensatory histone methylation and also DNA methylation. Although Ezh2 is the histone methyltransferase that methylates H3K27, the correlation between PRC2 and H3K27me3 may not be as simple as we imagine. Secondly, Ezh1, an isoform of Ezh2, has also been shown to methylate H3K27 when Ezh2 is not present.
(Shen et al, 2008). Factors responsible for recruiting Ezh1 may be favored when the targeting proteins for PRC2 are missing. This could also lead to compensatory histone methylation on H3K27 (Figure 5.1B). Thirdly, reduced levels of Aebp2 and Ezh2 could cause delay in the removal of H3K27me3 marks when the neural crest genes need to be activated. In this case, improper demethylation of H3K27me3 by trithorax groups (trxG) factors such as Jmjd3 and Utx, might have caused the reduced expression of neural crest genes. Further investigation is required to explain our unexpected experimental results. Nevertheless, it is most likely that Aebp2 is responsible for targeting the PRC2 complex to the genes that are important in neural crest cell development.

In the near future, we plan to test potential roles of PRC2 in neural crest development using Ezh2 conditional knockout mice. We should be able to knockout Ezh2 in the neural crest cell lineage using Wnt1-cre, which is neural crest-specific. If, as we hypothesized, the defects in the neural crest cells observed in Aebp2 heterozygotes are caused by the PRC2 mediated mechanism, similar phenotypes should also be visible in the knockout mice of the PRC2 proteins. It is possible that megacolon phenotypes will arise in these conditional knockout mice. We will also generate conditional knockout mouse lines for Aebp2. These conditional knockout lines may allow us to ascertain when and where Aebp2 is critical for neural crest cell development.

Many diseases are associated with defects in neural crest cells. Because of its multipotent nature, it is an interesting cell lineage to study. Our studies are among the first to merge the field of neural crest cell development and the field of epigenetics. This approach may give us a better understanding of a variety of cell types. I believe the study of the epigenetic properties of neural crest cells deserves as much attention as has been given to the study of the
epigenetic properties of cancer and stem cells. In fact, it has been suggested that adult stem cells might be a derivative of neural crest cells (Slack, 2007). Robert Weinberg, the noted cancer biologist, claimed that metastasizing cancer cells resemble neural crest cells migrating throughout our body. Many of the transcription factors important in cancer metastasis, such as Twist and Snail, are also critical transcription factors for neural crest cell migration (Weinberg, 2007; Yang et al., 2006). Most importantly, neural crest cells, cancer cells, and stem cells all seem to be prone to epigenetic changes. Thus, it would be worthwhile to understand and compare the epigenetic machinery behind these cell types.

This dissertation chronicles the function of an evolutionarily well-conserved gene, Aebp2. It illustrates how we first characterized a gene from an evolutionary perspective and then determined its molecular function in vitro and in vivo. Our studies not only reveal a new function of an uncharacterized gene, but also bring new insight into the field of developmental biology, genetics, and epigenetics.

Reference


Dear Hana Kim,


Thank you for your email dated 22 October 2010 requesting permission to reprint the above material. Our permission is granted without fee to reproduce the material. Use of the above article is restricted to reproduction in a thesis to Louisiana State University available in print/electronic format only, to be used only in the English language. This permission is limited to this particular use and does not allow you to use it elsewhere or in any other format other than specified above. Please include a credit line in your publication citing full details of the Oxford University Press publication which is the source of the material and by permission of Oxford University Press or the sponsoring society if this is a society journal. If the credit line or acknowledgement in our publication indicates that material including any illustrations/figures etc was drawn or modified from an earlier source it will be necessary for you to also clear permission with the original publisher. If this permission has not been obtained, please note that this material cannot be included in your publication/photocopies.

Please do not hesitate to contact me if I can be of any further assistance.

Yours sincerely,

Mubashar Khattak
Permissions Assistant
Oxford University Press
Great Clarendon Street
Oxford
OX2 6DP
Email: mubashar.khattak@oup.com
VITA

Hana Kim is the daughter of Dr. Sungeon Kim and Ms. Gyuwon Hwang. She was born 1982 in Seoul, South Korea. Hana graduated with a Bachelor of Music degree from Louisiana State University in 2006. She began her studies in science as an undergraduate student through the Howard Hughes Medical Institute (HHMI) research program with Dr. Joomyeong Kim at Louisiana State University. Subsequently she began her doctoral research under the guidance of Dr. Kim in the fall of 2006 in the Department of Biological Sciences at Louisiana State University. Hana will graduate with the degree of Doctor of Philosophy in December 2010.