GENE DUPLICATION AND ALTERNATIVE SPLICING PLAY A ROLE IN MODULATING THE FUNCTIONS OF THE *ZNF286A* TRANSCRIPTION FACTOR

BY

DEREK CAETANO-ANOLLÉS

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Cell and Developmental Biology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

Professor Lisa J. Stubbs, Chair, Director of Research
Associate Professor Craig A. Mizzen
Associate Professor Stephanie Ceman
Professor Sandra Rodriguez-Zas

ABSTRACT

Neurogenesis, and the processes through which neural stem cells and progenitor cells differentiate into neurons, occurs most actively during embryonic development, although neural differentiation continues at lower levels in certain brain regions well into adulthood. A vast regulatory network involving many known and conserved transcription factors regulates these functions. We have identified a novel zinc finger transcription factor (TF), ZNF286A, which is conserved in all eutherians and marsupials and provide evidence that this novel TF plays a role in regulation of mammalian neurogenesis.

ZNF286 occurs as a unique gene in most species. However, we demonstrate evidence that a gene duplication event in very recent primate history created a human-specific duplicate of *ZNF286A*, called *ZNF286B*. *ZNF286B* arose as part of a larger duplication in human chromosome 17, approximately 600,000 kb in length, that also includes many surrounding genes. Concomitant with (or shortly after) duplication, a processed and incomplete *FOXO3B* pseudo-gene was inserted into the *ZNF286B* genomic sequence and a DNA segment, encompassing a coding exon and regulatory sequences present in the ancestral *ZNF286A* gene, was deleted. As a result, *ZNF286B* encodes a protein with significant structural and expression differences relative to the ancestral gene. Most strikingly, the exon deleted in *ZNF286B* codes for the chromatin-interacting KRAB-domains that are present in the *ZNF286A* gene; in this respect the new human paralog resembles a natural KRAB-less alternative isoform that we demonstrate to be expressed naturally from the parental *ZNF286A* gene.

Using ChIP and siRNA knockdown, we show that ZNF286A protein binds to DNA at or near genes involved in the networks controlling the differentiation of neurons and the formation of axons during neurogenesis, and that both ZNF286A and ZNF286B directly regulate expression of many of those same genes. The pattern of DNA binding closely parallels binding of well-known neuronal differentiation factor, REST, in the same cell lines; siRNA results suggest that ZNF286 proteins act antagonistically to REST during development. We show that the mouse gene, *Zfp286*, is expressed at high levels in the developing nervous system and that both mouse and human genes and proteins are up-regulated transiently over the course of neurogenic differentiation *in vitro,* consistent

with the predicted biological role. We hypothesize that the duplication event that gave rise to ZNF286B allowed for independent regulation of the KRAB-less isoform of the ZNF286 protein, permitting this ancient mammalian gene to take on novel functions in the adult human brain.

Keywords: KRAB-ZNF, recently evolved paralogs, neurogenesis, transcription factors

T o all the ones who pushed me up and
H elped advance me through the ranks,
A nd those who helped me fill the gaps I
N eeded when my mind drew blanks –
K now that I value all your aid,
S o this is my way to say

ACKNOWLEDGEMENTS

Words cannot describe how thankful I am to the many people who have contributed to this thesis project coming to life. It is only natural, then, that I devote the following pages to describing in words how thankful I am to each of them.

First and foremost, I am extremely grateful to my advisor Dr. Lisa Stubbs for her enthusiastic support and guidance throughout the course of this project, and during the completion of my degree. This dissertation would not have been possible without her knowledge, and it would certainly be a confusing mess if not for her patient analysis of my data, or her numerous edits to countless drafts of this thesis and our manuscripts (even though she has a strange habit of typing way too many spaces between words than is necessary). For allowing me to explore different avenues of research, for backing my ideas when they were worthwhile, and for helping me grow as a scientist, thank you.

I would also like to thank the members of my thesis committee for their insightful comments and encouragement, and also for their many questions during my various meetings and exams that incentivized me to better examine the issues being examined. Dr. Craig Mizzen provided a bounty of honest feedback as the chair of both my Qualifying and Preliminary exams. Dr. Sandra Rodriguez-Zas provided her suggestions and continually pushed for me to succeed. Dr. Stephanie Ceman deserves special acknowledgement for her generosity in sharing her lab notes and protocols with me on more than one occasion. I have been fortunate to have such a wonderful and supportive committee.

Many thanks are due to my exceptionally talented friend and colleague Jacqueline Brinkman, who demonstrated great fortitude as my undergraduate assistant as she scaled a mountain of failures to produce the samples and results that have made their way into this project. A heap of gratitude is due to Huimin Zhang, who provided her input into almost every single protocol and technique used in this thesis. Thanks to Dr. Saurabh Sinha and Dr. Majid Kazemian, who provided computational analysis of our ChIP binding motifs in order to confirm a shared ZNF286/REST binding site. Thanks to Dr. Katja Nowick, who laid the groundwork for this project through her post-doctoral work prior to my joining the lab; her notes made a great deal of the early work for this project

easier. Additional thanks to Dr. Elbert Branscomb and Younguk-Calvin Sun for helpful suggestions and comments on the submitted manuscript.

In addition, many thanks are due to my fellow friends and labmates, the (current and former) graduate students and post-doc of the Stubbs Lab who provided input, samples, and suggestions that did not directly contribute to results in this thesis, but contributed to my growth in the lab, and as a person – former graduate students Dr. Chase Bolt and Dr. Annie Weisner (who I watched struggle with their own theses prior to working on mine), along with current graduate students Li-Hsin Chang, Chih-Ying Chen, Soumya Negi, Chris Seward, and Joseph Troy (who will hopefully do the same, in time), and current post-doc Dr. Michael Saul (who watches the curious affairs of graduate students from afar). Additional help locating samples was provided by Veena Chatti, with assistance and troubleshooting on various experiments provided over the years by labmates Erin Borchardt, Bob Chen, Nadya Kholina, Jin Li, Seeon Lee, Dina Leiding, and Xiaochen Lu. In essence, everyone in the Stubbs Lab is better than me at something, and it has been my pleasure to learn from each of them.

Thank you to the CDB departmental staff – Delynn Carter, Shannon Croft, Laura Marie Martin, and Elaine Rodgers – who silently oversaw all the processes that built this thesis, but rarely get recognized for their critical contributions.

And of course, I am very grateful of the support and encouragement I received from my family while pursuing my degree. I received support and encouragement from them both before and after I pursued my degree as well, but the support and encouragement I received *during* the degree-pursuing process is what I was referring to, here. Although, I suppose I am grateful to my family irrespective of the time frame that is selected relative to my graduate work. Regardless, they deserve *much more* than these few sentences in terms of acknowledgement. They have also made me aware that they agree with that sentiment.

Finally, I would like to thank the National Institutes of Health, the Cell and Molecular Biology Institutional Research Training Grant, and the Simons Foundation for providing their generous grant support over the course of my thesis. I want them to know that I made an honest attempt to put their money to good use.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

A bit of context

In 1665, Englishman and natural philosopher Robert Hooke published his *Micrographia*, a book containing thorough descriptions of dozens of plants, animals, and other natural objects as observed under what we would today consider to be a very rudimentary compound microscope. Along with these descriptions were included dozens of very detailed illustrations of his microscopic observations. As described in Hooke's preface to this work, the microscope was an instrument that would open up whole new worlds of exploration, and *Micrographia* served as a demonstration of the power that such technology held (Hooke, 1665).

Among the items Hooke observed under his microscope were very thin slices of cork, which he famously observed as being composed of an arrangement of small pores:

> *"I took a good clear piece of Cork, and with a Pen-knife ſharpen'd as keen as a Razor, I cut a piece of it off, and thereby left the ſurface of it exceeding ſmooth, then examining it very diligently with a Microſcope, me thought I could perceive it to appear a little porous; but I could not ſo plainly diſtinguiſh them, as to be ſure that they were pores… theſe pores, or cells, were not very deep, but conſiſted of a great many little Boxes... I no ſooner diſcern'd theſe (which were indeed the firſt microſcopical pores I ever ſaw, and perhaps, that were ever ſeen, for I had not met with any Writer or Perſon, that had made any mention of them before this) but me thought I had with the diſcovery of them, preſently hinted to me the true and intelligible reaſon of all the Phenomena of Cork."* (Obſerv. XVIII)

This seemingly innocuous observation, made while trying to determine the cause of the material's springiness and swelling nature (the "Phenomena of Cork" as he put it), was in fact a huge milestone in the field of biology, as it represented the very first discovery and naming of the *cell* – the smallest functional unit of life and capable of independent replication. Indeed, Hooke's observations served as an academic launching pad, and in the subsequent 350 years of study our knowledge of the internal structures of these cells and how they function together has grown immeasurably, and our current understanding is that the first life to emerge on the planet originally developed as single-celled organisms – a concept that would have been completely unknown to Hooke.

We now know that within each eukaryotic cell is contained an organism's entire genome. This information is arranged as a long series of nucleotides separated into multiple DNA molecules, with each ensemble of discrete DNA molecules separated into chromosomes. DNA sequences are defined into genes, which can be transcribed into RNA, which act as the template that ribosomes use to translate these sequences into proteins. These proteins then go on to perform the various functions necessary for cellular life. The interactions between these three types of molecules – DNA, RNA, and protein – are what the entire field of molecular biology was developed to study.

This thesis attempts to elucidate just one small aspect of these interactions regarding a small but critical family of proteins, the functions of these proteins in regulating the genes that control neuronal development, and their fascinating heritage through the evolution of mammals and marsupials. The human brain is unique in the animal kingdom (as it is what enables you, dear reader, to understand the words in this chapter), and so it is a matter of keen interest that we study the aspects that make the human brain and its development different from those of other animals.

This thesis represents only a small piece of a much larger puzzle that biologists are trying to put together everyday. While this topic is certainly specialized, representing a fine look at only a handful of genes singled out from a genome of thousands, it is important to remember that an entirely novel field of biology sprung forth following a lone Englishman asking himself why it was that cork floats. So it is that we come to understand that the most innocuous of studies can yield significant conclusions – perhaps even conclusions worth of popping a cork over.

The incredible human brain

There is a huge spectrum of intelligence across the animal kingdom, and humans are widely considered to be the kings in that arena. This self-coronated designation is interesting, but not as significant as the question of what properties of the brain exist beneath that crown, making humans more intelligent than other animals in the first place? Addressing these questions is difficult, and a huge literature of scientific and philosophical thought exists tackling precisely this issue.

Brain size alone does not seem to, by itself, explain intelligence. Many reports have been made of dogs, apes and corvids demonstrating remarkable cognition despite having much smaller brains, (Kaminski et al., 2004; Emery & Clayton, 2004). Therefore, the relationship between brains and intelligence appears to be qualitative in nature, as opposed to quantitative, and so the mechanisms underlying the brain and its development should be where we direct our focus.

The human brain occupies only 2% of human body mass, and yet surprisingly it consumes about 20% of total metabolism, making it one of the most metabolically 'expensive' organs in the body. The heart, the kidneys, the liver and the gastrointestinal tract are also relatively 'expensive' organs, and along with the brain consume about 70% of the basal metabolism of the human body (Aiello and Wheeler, 1995). More remarkable than the human brain's energy consumption is the fact that it is significantly larger than what would be expected from a primate of our mass (***Figure 1.1***).

In fact, human brain mass is disproportionately large when compared to any other mammal when accounting for body mass (***Figure 1.2***). Blue whales have enormous brain masses, but compared to their body masses it is clear that they are expending far less metabolic resources on their brain than what would be expected of an organism of their size.

Encephalization is often brought up as a metric for approximating animal intelligence, expressed by an '*encephalization quotient*', or *EQ* (Roth & Dicke, 2005). *EQ* is calculated as $EQ=E_a/E_e$ (with E_a indicating actual brain size, and E_e indicating expected brain size based on the proportions of a same-taxon animal 'standard'). When observing mammalian encephalization (***Figure 1.3***), the species with the highest level of encephalization is very clearly humans, with brains roughly 7-8 times larger than what is expected for a mammal of human mass, even when compared to most of our hominid ancestors! However, measuring encephalization has its limits, and some "higher cognition" species can find themselves with lower *EQ* scores than species generally accepted as having "lower cognition"; for example, gorillas have *EQ* scores just slightly higher than squirrels, despite countless examples of higher order thought processes and the capacity to express thought through language. As such, *EQ* is not the optimal

predictor for intelligence, but may serve as only one characteristic that explains humans' increased intelligence.

The increase in human brain size is balanced by a reduction in gut size (***Figure 1.1***), which has been compensated for with an increase in food quality (ie. from cooked food, which provides higher nutritional content that is more easily digestible than raw food). Cooked food is also more easily acquired when increased intelligence is applied (Aiello, 1997; Aiello et al., 2001). This connection is supported by a significantly positive correlation between primate diet quality and brain size (Fish & Lockwood, 2003), which is facilitated by the fact that primates with higher cognitive abilities are more capable of recognizing and processing high-quality foods than those with lower cognitive abilities (Aiello et al., 2001). Several hypotheses have attempted to explain the increases in relative brain size that have occurred during mammalian evolution, and while they may differ in their proposed mechanisms, they generally agree that the increase in brain size is primarily due to an increase in total cell number (Fish et al., 2008), with increases in total number of cortical neurons have been used to explain increases in intelligence (Roth & Dicke, 2005).

The most energy-demanding period during the lifetime of the human brain occurs during prenatal development and early childhood growth, and this is also the period when the brain develops the most neuronal cells through *neurogenesis* (Martin, 1996; Blakemore, 2012; Götz & Huttner, 2005). As such, the regulation of the systems responsible for these developmental frameworks is critical to our understanding of what makes us human (Andersen, 2003).

Neurogenesis, or the birth of neurons

Neurogenesis describes the broad range of processes involved in generating neurons from neural stem cells and their derivative progenitor cells. Several regions of the adult brain have been shown to undergo neurogenesis, such as the subventricular zone lining the inside of the lateral ventricles, the interneurons of the striatum, as well as parts of the dentate gyrus of the hippocampus (Ernst et al., 2014; Bergmann et al., 2012; Dayer et al., 2003; Roberts et al., 2014). However, the majority of neurogenesis that occurs does

not take place in adults, but rather during development when neural stem cells give rise to all the neurons of the mammalian central nervous system (Götz & Huttner, 2005). Once the appropriate signal is recognized by a self-renewing *neural stem cell*, it will differentiate into a much more specialized *neural progenitor cell*. From there, the progenitor cell is capable of differentiating into *mature neuronal cells* (ie. neurons, astrocytes, or oligodenrocytes), where their cell fate is locked into that state until their cell death.

Many proteins are involved in neurogenesis, acting within a network of enzymatic pathways that play specific roles in developing the structures and forms involved in a functioning brain and central nervous system (CNS). The three largest and most important pathways for these systems are the *Notch*, *Shh*, and *Wnt* signaling pathways, which have been shown to be hugely important in the process of neurogenesis and also in the regulation of self-renewal in embryonic stem cells.

The *Notch* signaling pathway promotes proliferative signaling during neurogenesis neural differentiation of precursor cells (Artavanis-Tsakonas et al., 1999). Meanwhile, the *sonic hedgehog* (SHH) pathway plays many critical roles in the development and patterning of the vertebrate CNS, including the induction of the floor plate and neural tube, which are precursor structures to the mature CNS (Placzek, 1995; Litingtung & Chiang, 2000). During later stages of development, the *Wnt* signaling pathway has been shown to control the migration of neuroblasts and other embryonic processes included in controlling body axis patterning, cell fate specification, cell proliferation, and cell migration (Pawlowski et al., 2002).

Changes in the ways that these systems are regulated can lead to drastic changes in the ways that the brain and CNS are formed. For this reason, the evolution of transcription factors (and primarily zinc finger transcription factors) plays a critical role in regulating these changes between species.

Handing it over to zinc finger transcription factors

In order to control their unique morphological and functional fates, cells regulate both the transcription of DNA into mRNA, and the translation of mRNA into proteins. A

broad range of *transcription factor* proteins regulates the former (Latchman, 1997; Blackwood & Kadonaga, 1998). These interactions are regulated by transcription factors that bind to DNA elements in order to promote, enhance, or silence gene expression. Promoter elements (like the TATA box) are always located at the 5'-end of the gene they regulate, while enhancer and silencer elements can be located very far away from their targets, utilizing DNA-looping to bridge the gaps (Su et al., 1990). Many of these regulators operate by either sitting at or near transcription start sites and directly regulating pre-initiation complex (PIC) binding specificity, or by recruiting other proteins to modify the chromatin structure of DNA, thereby making it easier or more difficult for the PIC to operate at those sites. These proteins can be broadly categorized into three groups:

(i) *General transcription factors* that bind to specific promoter sites to activate transcription (like the previously-described TFIID).
(ii) *Sequence-specific transcription factors* that interact with DNA-regulatory elements to carefully up- or down-regulate gene transcription.
(iii) *Non-sequence-specific co-factors* that mediate the previously described interactions.

Most sequence-specific transcription factors are composed of three protein structural domains: a conserved DNA-binding domain, a regulatory domain that regulates transcription, and a connective domain that links the two (Latchman, 1997). In general, sequence-specific transcription factors are classified by their DNA-binding motifs, which are conserved throughout evolution. On the other hand, regulatory domains are much more diverse, and can activate or repress transcription through protein-protein interactions. These interactions either occur *directly* with transcriptional components (such as the PIC or RNA polymerases), or *indirectly* with co-factors that modulate transcriptional proteins or regulate transcription at the chromatin-level (Zawal et al., 1996; Maldonado et al., 1999; Spitz & Furlong, 2012). Of these varied families of metazoan transcription factor domains, the most common domain is by far the C2H2 zinc finger motif.

Zinc fingers (sometimes abbreviated as *ZNF*s) are a large family of protein structural motifs that coordinate at least one zinc ion, used for stabilizing the fold structure of the domain. Depending on their structure, these zinc fingers are capable of binding to DNA, RNA, proteins, and/or small molecules. As such, they have proven themselves to be highly useful in serving the regulatory needs of the cell. One motif, the *C2H2 zinc finger motif,* contains conserved cysteine- and histidine-pairs in their amino acid sequence that coordinate a single zinc ion (Miller et al., 1985) and stabilize a "finger" structure that is required for proper folding and DNA binding (Stubbs et al., 2011; Frankel et al., 1987; Pavletich & Pabo, 1991). There is variation between the amino acids that compose individual fingers, however C2H2 motifs are recognizable by the highly conserved, repetitive amino acid pattern – TGEKP(Y/F) – that often links consecutive fingers to one another (***Figure 1.4***).

The most common family of transcription factors to use the C2H2 motif are the "*Krüppel*-type" (KZNF) family of transcription factors, named after their discovery in the regulatory Krüppel-protein of *Drosophila melanogaster* (Ollo & Maniatis, 1987). These proteins are identifiable as having at least 3, and as many as 40, C2H2 finger domains in tandem (Iuchi, 2001; Huntley et al., 2006). Each individual finger recognizes 3 base pairs of DNA at its α-helix motif, such that the array of tandem fingers winds around the major groove of the DNA molecule (***Figure 1.5***) in a highly sequence-specific manner (Pavletich & Pabo, 1991). In addition, there are four specific amino acids that are responsible for DNA-binding, called *fingerprints*. These fingerprints can be compared across species to determine the conservation of a zinc finger domain's binding (Stubbs et al., 2011).

KAP-1's regulation by chromatin modification

The most common method of regulation utilized by sequence-specific transcription factors is to interfere with the structure of chromatin. In eukaryotes, DNA is wrapped around histone octamers in repeating motifs known as *nucleosomes*, colloquially referred to as a "beads-on-a-string" structure, and further packaged into several levels of higher order chromatin structure (Horn & Peterson, 2002; Voss & Hager, 2014). These

higher-order chromatin fibers play a critical role in transcriptional regulation, as their structures can either be modified to be more open and accessible to other proteins (euchromatin) or more closed and inaccessible (heterochromatin). In addition, histones can aid in regulation by undergoing a variety of post-translational modifications that alter the way that they interact with DNA and proteins. These modifications typically occur on the long nucleosomal tails of the H3 and H4 histone proteins; the most common methods include acetylation, phosphorylation, methylation, deimination, ADP-ribosylation, ubiquitylation, sumoylation, and the removal of these aforementioned modifications (Bannister & Kouzarides, 2011). These modifications not only regulate chromatin through their presence, but they also recruit ATP-dependant chromatin remodeling complexes.

Krüppel-type transcription factors typically possess a regulatory domain that is attached to the finger domain by a "tether" region of connective amino acids (***Figure 1.6***). There are many families of effector domains for *Krüppel*-type zinc fingers, such as SCAN, ZAD, BTB/POZ, and Poly-C2H2 (Collins & Sanders, 2000; Jauch et al., 2003; Geyer et al., 2003), but the most common regulatory domain in *Krüppel*-type zinc fingers is the *Krüppel*-Associated Box (KRAB) (Stubbs et al., 2011). Each effector domain functions differently (Collins et al., 2001), and KRABs have been shown to function as repression domains that recruit the *Krüppel*-Associated Protein 1 (KAP-1) cofactor in order to recruit methyltransferases and histone deacetylases (Friedman et al., 1996; Peng et al., 2000; Zeng et al., 2008), which have long been known to modify the structure of chromatin and repress transcription (Allfrey et al. 1964). This strongly implies that KRAB-containing *Krüppel*-type proteins are involved in transcriptional repression (***Figure 1.6***).

The mechanism of KRAB begins with the recruitment of KAP-1 – also known as TRIM28, TF1β, TIF1β, and KRIP-1 (Friedman et al., 1996; Moosmann et al., 1996; Kim et al., 1996) – which has several domains that interact with different cofactors. An HP1-interacting domain binds to the heterochromatin protein, HP1, which anchors heterchromatin subunits, although its specific mechanisms are not known (Singh & Georgatos, 2002). A plant homeodomain finger conjugates SUMO proteins after activation, and these proteins sumoylate a nearby bromodomain on KAP-1 (Hay, 2005;

Ivanov et al., 2007). This SUMOylation of KAP-1 post-translationally modifies it to interact with SETDB methyltransferase, which adds methyl-groups to cysteines on nearby nucleosomes (Schultz et al., 2002). In addition, the SUMOylated bromodomain also recruits a NuRD-complex containing a histone deacetylase, which removes acetyl-groups from lysines on histone H3 (Lai & Wade, 2011). Together, these cofactors tighten chromatin structure and silence genes that are bound in the heterochromatin.

In humans, there are over 400 KRAB-containing, *Krüppel*-type zinc finger transcription factors, each with their own specific DNA-binding sequences and regulatory targets. This work investigates the binding and function of one pair of transcription factors within this large family – ZNF286A and ZNF286B.

The ZNF286 family of transcription factors

In eutherian mammals and marsupials, there exists one *ZNF286* gene that codes for a KRAB-containing *Krüppel*-type zinc finger transcription factor possessing 10 tandem C2H2 zinc fingers separated by a spacer region. *ZNF286* is an especially deeply conserved member of this dynamic gene family, being one of only 37 human KRAB-ZNF genes with a clear and highly conserved 1:1 ortholog in marsupials (Liu & Stormo, 2008). This unusual level of conservation suggests that the gene has acquired an essential mammalian regulatory function.

However, the human version of this gene (referred to as *ZNF286A*) has duplicated to produce a second gene called *ZNF286B*, which is found in humans but is not detected in any other primate species (Nowick et al., 2011). Despite its very recent advent, *ZNF286B* differs from the parental gene in several significant ways. In particular, despite the sequences coding for the zinc finger domains of *ZNF286A* and *ZNF286B* being nearly identical, *ZNF286B* does not include any KRAB-encoding exons, and instead includes novel exons that are not included in the parental gene. In lieu of a KRAB-coding region, *ZNF286B* contains a unique sequence of similar length, identified as a *FOXO3B* pseudogene, inserted into the gene sequence after the duplication event that brought about *ZNF286B*. While previous studies have shown that deletion of KRAB in proteins results in a loss of repression function (Li et al., 2008), this has not been demonstrated for

ZNF286B specifically. Indeed, experimental data demonstrates that this protein includes other types of regulatory functions alongside ZNF286A in neural tissues.

This duplication behavior is entirely commonplace for KRAB-ZNF, and in vertebrates (and especially in mammals) the KRAB-ZNF family has grown through repeated rounds of tandem segmental duplication to include hundreds of lineage-specific genes (Chung et al., 2002, 2007). Differences in KRAB-ZNF copy number are even seen between closely related species, such as humans and chimpanzees (Huntley, 2006; Hamilton, 2006; Nowick et al., 2011). New duplicates can diverge significantly over relatively short evolutionary times, changing in expression pattern, splicing, and/or coding sequence in ways that predict distinct functions (Hamilton, 2006; Nowick, 2011).

Past studies into the history of *Krüppel*-type zinc fingers have annotated the repertoire of human KZNF genes, and carried out a preliminary comparison with genes in other species (Huntley et al., 2006). Most KZNF genes are deeply conserved across evolution, however the KRAB-ZNF is substantially over-represented among genes that have undergone repeated rounds of recent segmental duplications (Nowick et al., 2010), as is the case with *ZNF286A* and *ZNF286B*. In this case, a "recent duplication" is defined as having occurred within the past 35–40 million years. These segmental duplications have produced an evolutionary sandbox for the evolution of proteins with primate-specific functions, and novel lineage-specific genes across many other species (Nowick et al., 2010; Huntley et al., 2006; Bellefroid et al., 1993; Looman et al., 2002). This is because duplications create raw genetic material that can evolve beyond the selective pressures of a deeply conserved gene that may be too critical to mutate (Bailey & Eichler, 2006). Interestingly, these duplication events appear to have significantly accelerated in the last 12 million years within the genome of the African ancestor of the great ape (Marques-Bonet et al., 2009). This evolutionary diversity, and the evidence that KZNF genes play an important role in novel gene expression differences between the human and chimpanzee brain (Nowick et al., 2009), indicate an important role for recently evolved KRAB-ZNF duplicates like ZNF286A and ZNF286B (Nowick et al., 2011).

Aside from having a gene duplicate, *ZNF286A* also transcribes an alternative transcript that is similar to *ZNF286B* in that it lacks a KRAB domain (***Figure 1.7***), but is conserved in both humans and mice. This aligns with previous studies that show that

conserved human/mice alternative transcripts are enriched in regulatory genes and are frequently expressed in brain (Yeo et al., 2005). Both ZNF286A isoforms, along with ZNF286B, share identical or near-identical DNA binding motifs. For this reason we hypothesize that KRAB-less ZNF286A and ZNF286B act as competitors to ZNF286A and fine-tune the regulation of target genes by varying concentrations of each protein in different tissues at different times. During preliminary tests, we found that *ZNF286A* expresses in tissues at different levels than *ZNF286B*, suggesting that the human-specific duplicate has evolved to fit a functional niche that is not occupied by the full-functioning ancestral gene, however a similar niche may be occupied by the KRAB-less ZNF286A isoform. Alternative splicing acts as a means of creating protein diversity (Keren et al., 2010; Nilsen & Graveley, 2010), so this suggests that it is possible for the function of KRAB-less ZNF286A has become part of the "official" non-alternative proteome in the form of the ZNF286B genetic duplicate. Our data show that *ZNF286A* and *ZNF286B* are of interest because they regulate genes involved in neurogenesis, and bind to regions that are enriched for binding sites of a well-known zinc finger, the RE-1 silencing transcription factor (REST) discussed in the following section, and most likely play complementary roles to REST in transcriptional regulation.

Repressor-element-1 silencing transcription factor (REST)

REST, also known as the neural-restrictive silencer factor (NRSF), is a *Krüppel*-type zinc finger transcription factor that binds to an element known as the NRSE (neuron-restrictive silencer element, also known as RE-1) for the critical role of repressing neuronal gene expression in non-neuronal cells (Schoenherr & Anderson, 1995; Schoenherr et al., 1996). Furthermore, *REST* is required for proper timing of neuronal gene expression in embryonic stem cells by repressing genes that control for the terminal differentiation of cell lines (Chen et al., 1998). It is also directly regulated by the canonical Wnt pathway, which controls the fates of neural progenitor cells (Nishihara et al., 2003). As with many regulatory cells in these pathways, REST has been implicated in different forms of cancer (Bronson et al., 2010; Coulson et al., 2000; Ly et al., 2010).

REST contains eight unique tandem zinc fingers (distinct from ZNF286), and expresses multiple isoforms (Spencera et al., 2006; Raj et al., 2010). It has been shown that in non-neuronal adult tissues, REST is ubiquitously expressed (Scholl et al., 1996), however REST decreases after neuronal differentiation (Nishimura et al., 1996; Singh, 2011; Xiaohua et al., 2004). REST is critical for neurogenesis, and mice that lack the protein actually die during fetal development (Chen et al., 1998).

There are four alternatively spliced isoforms of REST – a full length isoform (122 kDa), a truncated isoform that retains only the first four zinc fingers and localizes to the cytoplasm (35 kDa), a truncated isoform known as REST4 that retains the first five zinc fingers and nuclear localization signal (37 kDa), and an isoform with a selective deletion of the fifth zinc finger (119 kDa) (Palm et al., 1999; Coulson et al., 2000; Shimojo et al., 2001). The full-length REST and REST4 are the most abundantly expressed isoforms.

Full-length REST is localized to the nucleus, but due to the truncation of the nuclear localization signal that directs the protein to the nucleus, some REST isoforms behave differently (Shimojo et al., 2001), with REST4 playing the most active role in regulating full-length REST's function. REST and REST4 are expressed at different levels in different cell types (Spencer et al., 2006; Uchida et al, 2010), even within the same family of derived cell lines – as an example in neuroblastoma cell lines, glial SH-EP cell express full length REST while neuronal SH-SY5Y expresses REST4, with the mixed SK-N-SH cell type expressing both isoforms (Hamelink et al., 2004; Raj et al., 2011). With these two main versions of the same protein existing at varying levels in conjunction with each other, previous studies have shown that REST4 acts to inhibit REST function by competing with REST over NRSE sites (Shimojo et al., 1999; Tabuchi et al., 2002). However, REST4 itself does not actually repress transcription function, or does so very weakly (Magin et al., 2002; Lee et al., 2000). Therefore, REST4 does not act as a transcriptional repressor, but blocks the repressor activity of REST. Under certain circumstances, REST is actively sequestered outside of the cytoplasm by other proteins, which also prevents it from binding to the NRSE altogether (Zuccato et al., 2003; Lunyak & Rosenfeld, 2005).

Repression by REST begins after the protein's zinc finger domain binds to a highly specific and highly conserved DNA sequence known as the neuron-restrictive silencer

element (NRSE) (***Figure 1.8***). An ATP-dependent chromatin-remodeling helicase enzyme called SMARCA4 stabilizes this interaction (Ooi et al., 2006). The canonical NRSE binding element is a ~21 base DNA sequence divided into two "half-sites" by a variant-size "spacer" region. REST can bind to either or both half-sites, depending on the specific locus and the size of the spacer, while maintaining its function (Johnson et al., 2007). This allows REST to assume a less stressful configuration when binding to the DNA molecule (Pavletich & Pabo, 1991; Lee et al., 1989), while simultaneously leaving part of the NRSE exposed for potential binding by another transcription factor. As will be discussed in detail in the following chapters, we hypothesize that ZNF286 may be binding at or near to these exposed half-sites.

After binding to the NRSE, a Sim3 Interaction Domain located at the N-terminus of REST interacts with the Sim3 complex, which recruits several subunits (SAP18, SAP30, RbAP46) including histone deacetylases (HDAC1/2) (Roopra, 2000). REST also has repression domains on both termini of the protein that bind to histone deacetylation complexes for genetic regulation (Roopra et al., 2001). A CoREST-interaction domain located at the C-terminus of REST interacts with a CoREST cofactor, which recruits a variety of proteins including HDAC1/2, the chromatin remodeling SMARCA4, H3K4 demethylase LSD1 and H3K9 methylases G9a (Andrés et al., 1999).

Together, or separately, the co-repressors recruited by Sim3 or CoREST remove several gene-activating chromatin modifications and inhibit RNA polymerases, thus silencing specific genes as a result of REST's presence at the NRSE binding site.

Experimental overview

To uncover the functions of the *ZNF286* genes, we mapped gene and protein expression in human cell lines and tissues, and designed custom siRNA molecules to specifically knock down the expression of both paralogs. We also used custom antibodies designed to unique portions of each protein to carry out chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) in human neuroblastoma cells. Putative regulatory targets, defined as differentially expressed genes (DEGs) that are flanked by or contain ChIP-enriched binding sites, implicate ZNF286A in the negative

regulation of cell proliferation and positive regulation of neuronal differentiation; expression of the genes and proteins during neuronal differentiation in culture support this functional role. Surprisingly, *ZNF286B* knockdowns yielded enrichment for many of the same genes and very similar functional categories, suggesting that (at least in neuroblastoma cells) the KRAB(+) and KRAB(-) versions of the ZNF286 protein do not function as an opposed repressor-competitor pair. Possibly due to the deletion and insertion following *ZNF286B* duplication, the two paralogs are expressed differently: most notably, *ZNF286A* is expressed at its highest levels in adult brain, while *ZNF286B* is highest in fetal neural tissues.

A closer look at the highest affinity ZNF286 binding regions revealed a high degree of overlap with both the known recognition motif and ENCODE-measured binding sites for REST in neuroblastoma cells. Using electrophoretic mobility assays (EMSA), we confirmed that both human ZNF286 proteins indeed recognize the known REST binding motif. The predicted functions of the ZNF286 proteins – inhibition of proliferation and stimulation of neuronal differentiation – are opposite to the known functions of REST (Schoenherr & Anderson, 1995). This fact, combined with the observation that ZNF286 proteins rise as REST falls in expression during differentiation of neuroblastoma cells, support the hypothesis that the ZNF286 proteins evolved as competitive modulators of REST activities in the brain. Given their similar predicted functions, we postulate that duplication and divergence of *ZNF286B* has allowed separate regulation of this deeply conserved protein, adding a novel layer of REST modulation in fetal and adult brain, as well as other human tissues.

Significance

From an *evolutionary* perspective, *ZNF286B* is a perfect example of recent KRAB-ZNF segmental duplication, and provides an opportunity to observe the various ways in which functional novelty arises in a species (Lynch & Conery, 2000), a classical example of this being the modification of duplicates in Antarctic icefish that have led to the introduction of antifreeze proteins (Logsdon & Doolittle, 1997; Chang & Duda, 2012). Few examples are as recently duplicated as *ZNF286B*, however, as alignment data

reveals that *ZNF286B* is a human-specific gene, especially when taking into account the high identity of the inserted *FOXO3B* pseudogene sequence with the human *FOXO3* gene. From a *biological* perspective, ZNF286A and B are proteins that seem to work in conjunction with REST, which implicates them in several neural pathways. For example, knockdown of *REST* leads to down-regulation of *SMARCA2*, a gene that codes for proteins used in the SWI/SNF chromatin remodeling pathway, the down-regulation of which is a phenotype of schizophrenia and leads to abnormal axonogenesis (Loe-Mie et al., 2010). Our preliminary data suggest that knockdown of *ZNF286A/B* up-regulates *SMARCA2* and other related candidate genes that are conversely regulated by REST.

The following chapters will elucidate the functions and regulatory activities of both human *ZNF286A* isoforms, *ZNF286B* and mouse *Zfp286*, and investigate the predicted interactions between the ZNF286 proteins and REST, such that their potential role in the pathways of neurite outgrowth and maintenance might be clarified.

References Cited

Aiello LC. **1997**. Brains and guts in human evolution: The Expensive Tissue Hypothesis. *Braz J Genet* 20(1):141-148.

Aiello LC, Bates N, Joffe T. **2001**. In Defense of the Expensive Tissue Hypothesis: Ontogeny, Maternal Care and Organ Size. In *Evolutionary Anatomy of the Primate Cerebral Cortex* (Gibons KR and Falk D, eds), pp 57-78. Cambridge University Press: Cambridge.

Aiello LC, Wheeler P. **1995**. The Expensive-Tissue Hypothesis: The brain and the digestive system in human and primate evolution. *Curr Anthropol* 36(2):199-221.

Allfrey VG, Faulkner R, Mirsky AE. **1964**. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc Natl Acad Sci USA* 51(5):786–94.

Andersen SL. **2003**. Trajectories of brain development: point of vulnerability or window of opportunity? *Neurosci Biobehav Rev* 27(1):3-18.

Andrés ME, Burger C, Peral-Rubio MJ, Battaglioli E, Anderson ME, Grimes J, Dallman J, Ballas N, Mandel G. **1999**. CoREST: a functional corepressor required for regulation of neural-specific gene expression. *Proc Natl Acad Sci USA* 96(17):9873-8.

Artavanis-Tsakonas S, Rand MD, Lake RJ. **1999**. Notch Signaling: Cell Fate Control and Signal Integration in Development. *Science* 284:(5415)770–776.

Aschoff J, Günther B, Kramer K. **1971**. Energiehaushalt und Temperaturregulation. München: Urban & Schwarzenberg.

Bailey JA, Eichler EE. **2006**. Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat Rev Genet* 7:552-564.

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

Bellefroid EJ, Marine JC, Ried T, Lecocq PJ, Rivière M, Amemiya C, Poncelet DA, Coulie PG, de Jong P, Szpirer C. **1993**. Clustered organization of homologous KRAB zinc-finger genes with enhanced expression in human T lymphoid cells. *Embo J* 12:1363–1374.

Bergmann O, Liebel J, Bernard S, Alkass K, Yeung MSY, Steier P, Kutschera W, Johnson L, Landen M, Druid H, Spalding KL, Frisen J. **2012**. The age of olfactory bulb neurons in humans. *Neuron* 74:634–639.

Blackwood M, Kadonaga JT. **1998**. Going the Distance: A Current View of Enhancer Action. *Science* 281:60-63.

Blakemore SJ. **2012**. Imaging Brain Development: The Adolescent Brain. *NeuroImage* 61:397-406.

Bronson MW, Hillenmeyer S, Park RW, Brodsky AS. **2010**. Estrogen coordinates translation and transcription, revealing a role for NRSF in human breast cancer cells. *Mol Endocrin* 24(6):1120-1135.

Chen ZF, Paquette AJ, Anderson DJ. **1998**. NRSF/REST is required in vivo for repression of multiple neuronal target genes during embryogenesis. *Nat Genet* 2:136-42.

Chung, HR, Lohr U, Jackle H. **2007**. Lineage-specific expansion of the zinc finger associated domain ZAD. *Mol Biol Evol* 24:1934-1943.

Chung, HR, Schafer U, Jackle H, Bohm S. **2002**. Genomic expansion and clustering of ZAD-containing C2H2 zinc-finger genes in Drosophila. *EMBO Rep* 3:1158-1162.

Collins T, Sander TL. **2000**. The Superfamily of SCAN Domain Containing Zinc Finger Transcription Factors. In: *Madame Curie Bioscience Database*. Austin (TX): Landes Bioscience.

Collins T, Stone JR, Williams AJ. **2001**. All in the Family: the BTB/POZ, KRAB, and SCAN Domains. *Mol Cell Biol* 21(11):3609-3615

Coulson JM, Edgson JL, Woll PJ, Quinn JP. **2000**. A splice variant of the neuron-restrictive silencer factor repressor is expressed in small cell lung cancer: a potential role in derepression of neuroendocrine genes and a useful clinical marker. *Cancer Res* 60(7):1840-1844.

Dayer AG, Ford AA, Cleaver KM, Yassaee M, Cameron HA. **2003**. Short-term and long-term survival of new neurons in the rat dentate gyrus. *J Comp Neurol* 460(4):563–572.

Emery NJ, Clayton NS. **2004**. The mentality of crows: Convergent evolution of intelligence in corvids and apes. *Science* 306:1903–1907.

Ernst A, Alkass K, Bernard S, Salehpour M, Perl S, Tisdale J, Possnert G, Druid H, Fisen J. **2014**. Neurogenesis in the striatum of the adult human brain. *Cell* 156(5):1072–1083.

Fish JL, Dehay C, Kennedy H, Huttner WB. **2008**. Making bigger brains – the evolution of neural-progenitor-cell division. *J Cell Sci* 121:2783-2793.

Fish JL, Lockwood CA. **2003**. Dietary constraints on encephalization in primates. *Am J Phys Anthropol* 120:171–181.

Frankel AD, Berg JM, Pabo CO. **1987**. Metal-dependent folding of a single zinc finger from transcription factor IIIA. *Proc Natl Acad Sci USA* 84:4841–4845.

Friedman JR, Fredericks WJ, Jensen DE, Speicher DW, Huang XP, Neilson EG, Rauscher FJ 3rd. **1996**. KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes Dev* 10(16):2067-78.

Geyer R, Wee S, Anderson S, Yates J, Wolf DA. **2003**. BTB/POZ domain proteins are putative substrate adaptors for cullin 3 ubiquitin ligases. *Mol Cell* 12(3):783-90.

Götz1 M, Huttner WB. **2005**. The cell biology of neurogenesis. *Nat Rev Mol Cell Bio* 6:777-788.

Hamelink C, Hahm SH, Huang H, Eiden LE. **2004**. A restrictive element 1 (RE-1) in the VIP gene modulates transcription in neuronal and non-neuronal cells in collaboration with an upstream tissue specific element. *J Neurochem* 88:1091–1101.

Hamilton AT, Huntley S, Tran-Gyamfi M, Baggott DM, Gordon L, Stubbs L. **2006**. Evolutionary expansion and divergence in the ZNF91 subfamily of primate-specific zinc finger genes. *Genome Res* 16:584-594.

Haug H. **1987**. Brain sizes, surfaces, and neuronal sizes of the cortex cerebri: A stereological investigation of man and his variability and a comparison with some mammals (primates, whales, marsupials, insectivores, and one elephant). *Am J Anat* 180(2):126–142.

Hay RT. **2005**. SUMO: a history of modification. *Mol Cell* 18(1):1-12.

Hooke R. **1665**. Micrographia: Or Some Physiological Descriptions of Minute Bodies Made by Magnifying Glasses, with Observations and Inquiries Thereupon. London: J. Martyn and J. Allestry, Printers to the Royal Society. (First Edition).

Horn PJ, Peterson CL. **2002**. Chromatin Higher Order Folding--Wrapping up Transcription. *Science* 297(5588):1824-1827.

Huntley S, Baggott DM, Hamilton AT, Tran-Gyamfi M, Yang S, Kim J, Gordon L, Branscomb E, Stubbs L. **2006**. A comprehensive catalog of human KRAB-associated zinc finger genes: Insights into the evolutionary history of a large family of transcriptional repressors. *Genome Res* 16(5):669–677.

Iuchi S. **2001**. Three classes of C2H2 zinc finger proteins. *Cell Mol Life Sci* 58:625-635.

Ivanov AV, Peng H, Yurchenko V, Yap KL, Negorev DG, Schultz DC, Psulkowski E, Fredericks WJ, White DE, Maul GG, Sadofsky MJ, Zhou M, Rauscher FJ 3rd. **2007**. PHD Domain-Mediated E3 Ligase Activity Directs Intramolecular Sumoylation of an Adjacent Bromodomain which is Required for Gene Silencing. *Mol Cell* 28(5):823–837.

Jauch R, Bourenkov GP, Chung HR, Urlaub H, Reidt U, Jäckle H, Wahl MC. **2003**. The zinc finger-associated domain of the Drosophila transcription factor grauzone is a novel zinc-coordinating protein-protein interaction module. *Structure* 11(11):1393-402.

Johnson DS, Mortazavi A, Myers RM, Wold B. **2007**. Genome-wide mapping of in vivo protein-DNA interactions. *Science* 316(5830):1497-502.

Kaminski J, Call J, Fischer J. **2004**. Word learning in a domestic dog: Evidence for "fast" mapping. *Science* 304:1682–1683.

Keren H, Lev-Maor G, Ast G. **2010**. Alternative splicing and evolution: diversification, exon definition and function. *Nat Rev Genet* 11:345-355.

Kim SS, Chen YM, O'Leary E, Witzgall R, Vidal M, Bonventre JV. **1996**. A novel member of the RING finger family, KRIP-1, associates with the KRAB-A transcriptional repressor domain of zinc finger proteins. *Proc Natl Acad Sci USA* 93:15299-304.

Knight RD, Shimeld SM. **2001**. Identification of conserved C2H2 zinc-finger gene families in the Bilateria. *Genome Biol* 2(5):0016.1-0016.8.

Lai AY, Wade PA. **2011**. Cancer biology and NuRD: a multifaceted chromatin remodelling complex. *Nat Rev Cancer* 11:588-596.

Latchman DS. **1997**. Transcription factors: An overview. *Int J Biochem Cell Biol* 29(12):1305–12.

Lee JH, Chai YG, Hersh LB. **2000**. Expression Patterns of Mouse Repressor Element-1 Silencing Transcription Factor 4 (REST4) and its Possible Function in Neuroblastoma. *J Mol Neuro* 15:205-214.

Lee MS, Gippert GP, Soman KV, Case DA, Wright PE. **1989**. Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science* 245:635–637.

Li Y, Yang D, Bai Y, Mo X, Huang W, Yuan W, Yin Z, Deng Y, Murashko O, Wang Y, Fan X, Zhu C, Ocorr C, Bodmer R, Wu X. **2008**. ZNF418, a novel human KRAB/C2H2 zinc finger protein, suppresses MAPK signaling pathway. *Mol Cell Biochem* 310(1-2):141-51.

Liu J, Stormo GD. **2008**. Context-dependent DNA recognition code for C2H2 zinc-finger transcription factors. *Bioinformatics* 24:1850-1857.

Litingtung Y, Chiang C. **2000**. Control of SHH activity and signaling in the neural tube. *Dev Dynam* 219(2):143–54.

Looman C, Abrink M, Mark C, Hellman L. **2002**. KRAB zinc finger proteins: An analysis of the molecular mechanisms governing their increase in numbers and complexity during evolution. *Mol Biol Evol* 19:2118–2130.

Lunyak VV, Rosenfeld MG. **2005**. No Rest for REST: REST/NRSF Regulation of Neurogenesis. *Cell* 121(4):499-501.

Lv H, Pan G, Zheng G, Wu X, Ren H, Liu Y, Wen J. **2010**. Expression and functions of the repressor element 1 (RE-1)-silencing transcription factor (REST) in breast cancer. *J Cell Biochem* 110(4):968-974.

Lynch M, Conery JS. **2000**. The evolutionary fate and consequences of duplicate genes. *Science* 290(5494):1151-5.

Magin A, Lietz M, Cibelli G, Thiel G. **2002**. RE-1 silencing transcription factor-4 (REST4) is neither a transcriptional repressor nor a de-repressor. *Neurochem Int* 40(3): 195–202.

Maldonado E, Hampsey M, Reinberg D. **1999**. Repression: targeting the heart of the matter. *Cell* 99:455.

Marques-Bonet T, Ryder OA, Eichler EE. **2009**. Sequencing primate genomes: what have we learned? *Annu Rev Genomics Hum Genet* 10:355-386.

Martin, RD. **1996**. Scaling of the mammalian brain: the maternal energy hypothesis. *News Physiol Sci* 11:149–156.

Massiah MA, Matts JA, Short KM, Simmons BN, Singireddy S, Yi Z, Cox TC. **2007**. Solution structure of the MID1 B-box2 CHC(D/C)C(2)H(2) zinc-binding domain: insights into an evolutionarily conserved RING fold. *J Mol Biol* 369(1):1-10.

Miller J, McLachlan AD, Klug A. **1985**. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J* 4:1609–1614.

Moosmann P, Georgiev O, Le Douarin B, Bourquin JP, Schaffner W. **1996**. Transcriptional repression by RING finger protein TIF1 beta that interacts with the KRAB repressor domain of KOX1. *Nucleic Acids Res* 24:4859-4867.

Nilsen TW and Graveley BR. **2010**. Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463:457-463.

Nishihara S, Tsuda L, Ogura T. **2003**. The canonical Wnt pathway directly regulates NRSF/REST expression in chick spinal cord. *Biochem Biophys Res Commun* 311(1):55-63.

Nishimura E, Sasaki K, Maruyama K, Tsukada T, Yamaguchi K. **1996**. Decrease in neuron-restrictive silencer factor (NRSF) mRNA levels during differentiation of cultured neuroblastoma cells. *Neurosci Lett* 211(2):101–104.

Nowick K, Fields C, Gernat T, Caetano-Anollés D, Kholina N, Stubbs L. **2011**. Gain, Loss and Divergence in Primate Zinc-Finger Genes: A Rich Resource for Evolution of Gene Regulatory Differences between Species. *PLOS ONE* 6(6):e21553.

Nowick K, Gernat T, Almaas E, Stubbs L. **2009**. Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain. *Proc Natl Acad Sci USA* 106:22358–22363.

Nowick K, Hamilton AT, Zhang H, Stubbs L. **2010**. Rapid sequence and expression divergence suggests selection for novel function in primate-specific KRAB-ZNF genes. *Mol Biol Evol* 27(11):2606-17

Ooi L, Belyaev ND, Miyake K, Wood IC, Buckley NJ. **2006**. BRG1 chromatin remodeling activity is required for efficient chromatin binding by the transcriptional repressor rest and facilitates rest-mediated repression. *J Biol Chem*. 281:38974–38980.

Ooi L, Wood IC. **2007**. Chromatin crosstalk in development and disease: lessons from REST. *Nat Rev Genet* 8:544-554.

Ollo R, Maniatis T. **1987**. Drosophila Kruppel gene product produced in a baculovirus expression system is a nuclear phosphoprotein that binds to DNA. *Proc Natl Acad Sci USA* 84:5700–5704.

Palm K, Metsis M, Timmusk T. **1999.** Neuron-specific splicing of zinc finger transcription factor REST/NRSF/XBR is frequent in neuroblastomas and conserved in human, mouse and rat. *Brain Res Mol Brain Res* 72(1):30-39.

Pavletich NP, Pabo CO. **1991**. Zinc finger-DNA recognition: crystal structure of a Zif268- DNA complex at 2.1 A. *Science* 252:809–817.

Pawlowski JE, Ertel JR, Allen MP, Xu M, Butler C, Wilson EM, Wierman ME. **2002**. Liganded androgen receptor interaction with beta-catenin: nuclear co-localization and modulation of transcriptional activity in neuronal cells. *J Biol Chem* 277(23):20702-10.

Peng H, Begg GE, Harper SL, Friedman JR, Speicher DW, Rauscher FJ 3rd. **2000**. Biochemical analysis of the Kruppel-associated box (KRAB) transcriptional repression domain. *J Bio Chem* 275(24):18000-18010.

Placzek M. **1995**. The role of the notochord and floor plate in inductive interactions. *Curr Opin Genet Dev* 5(4):499–506.

Raj B, O'Hanlon D, Vessey JP, Pan Q, Ray D, Buckley NJ, Miller FD, Blencowe BJ. **2011**. Cross-Regulation between an Alternative Splicing Activator and a Transcription Repressor Controls Neurogenesis. *Mol Cell* 43:843–850.

Roberts MD, Toedebusch RG, Wells KD, Company JM, Brown JD, Cruthirds CL, Heese AJ, Zhu C, Rottinghaus GE, Childs TE, Booth FW. **2014**. Nucleus accumbens neuronal maturation differences in young rats bred for low versus high voluntary running behaviour. *J Physiol* 592(10):2119–35.

Roopra A, Huang Y, Dingledine R. **2001**. Neurological disease: listening to gene silencers. *Mol Interventions* 1(4):219-228.

Roopra A, Sharling L, Wood IC, Briggs T, Bachfischer U, Paquette AJ, Buckley NJ. **2000**. Transcriptional repression by neuron-restrictive silencer factor is mediated via the Sin3-histone deacetylase complex. *Mol Cell Biol* 20(6):2147-2157.

Roth G, Dicke U. **2005**. Evolution of the brain and intelligence. *Trends Cogn Sci* 9(5):250-257.

Schoenherr CJ, Anderson DJ. **1995**. The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science*. 267(5202):1360-63.

Schoenherr CJ, Paquette AJ, Anderson DJ. **1996**. Identification of potential target genes for the neuron-restrictive silencer factor *Proc Natl Acad Sci USA*. 93(18):9881-86.

Scholl T, Stevens MB, Mahanta S, Strominger JL. **1996**. A zinc finger protein that represses transcription of the human MHC class II gene, DPA. *J Immunol* 4:1448-57.

Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher FJ 3rd. **2002**. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev* 16(8):919–32.

Shimojo M, Paquette AJ, Anderson DJ, Hersh LB. **1999**. Protein Kinase A Regulates Cholinergic Gene Expression in PC12 Cells: REST4 Silences the Silencing Activity of Neuron-Restrictive Silencer Factor/REST. *Mol Cell Bio* 19(10):6788–6795.

Shimojo M, Lee JH, Hersh LB. **2001**. Role of zinc finger domains of the transcription factor neuron-restrictive silencer factor/repressor element-1 silencing transcription factor in DNA binding and nuclear localization *J Biol Chem* 276(16):13121-6.

Singh A, Rokes C, Gireud M, Fletcher S, Baumgartner J, Fuller G, Stewart J, Zage P, Gopalakrishnan V. **2011.** Retinoic acid induces REST degradation and neuronal differentiation by modulating the expression of SCFβ-TRCP in neuroblastoma cells. *Cancer* 117(22):5189–5202.

Singh PB, Georgatos SD. **2002**. HP1: Facts, open questions, and speculation. *J Struct Biol* 140(1-3):10-16.

Snyder WS. **1975**. International Commission on Radiological Protection (ICRP) report on the Task Force on Reference Man. Oxford: Oxford University Press.

Spencer EM, Chandler KE, Haddley K, Howard MR, Hughes D, Belyaev ND, Coulson JM, Stewart JP, Buckley NJ, Kipar A, Walker MC, Quinn JP. **2006**. Regulation and role of REST and REST4 variants in modulation of gene expression in in vivo and in vitro in epilepsy models. *Neurobiol Dis* 24(1):41–52.

Spitz F, Furlong EEM. **2012**. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* 13:613-626.

Stahl WR. **1965**. Organ weights in primates and other mammals. *Science* 150:1039-42.

Stubbs L, Sun Y, Caetano-Anollés D. **2011**. Function and Evolution of C2H2 Zinc Finger Arrays. In TR Hughes (Ed.), *A Handbook of Transcription Factors*, (Chapter 4). Berlin. Springer Science+Business Media.

Su W, Porter S, Kustu S, Exchols H. **1990**. DNA-looping and enhancer activity: Association between DNA-bound NtrC activator and RNA polymerase at the bacterial *glnA* promoter. *Proc Natl Acad Sci USA* 87:5504-5508.

Tabuchi A, Yamada T, Sasagawa S, Naruse Y, Mori N, Tsuda M. **2002**. REST4-Mediated Modulation of REST/NRSF-Silencing Function during BDNF Gene Promoter Activation. *Biochem and Biophys Res Commun* 290:415–420.

Uchida S, Hara K, Kobayashi A, Funato H, Hobara T, Otsuki K, Yamagata H, McEwen BS, Watanabe Y. **2010**. Early Life Stress Enhances Behavioral Vulnerability to Stress through the Activation of REST4-Mediated Gene Transcription in the Medial Prefrontal Cortex of Rodents. *J Neurosci* 30(45):15007-18.

Van Dongen PAM. **1998**. Brain size in vertebrates. In R. Nieuwenhuys et al. (Eds.), *The Central Nervous System of Vertebrates Vol. 3*, pp. 2099–2134. Springer Berlin.

Voss TC, Hager GL. **2014**. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat Rev Genet* 15:69-81.

Xiaohua S, Kameoka S, Lentz S, Majumder S. **2004**. Activation of REST/NRSF Target Genes in Neural Stem Cells Is Sufficient To Cause Neuronal Differentiation. *Mol Cell Biol* 24(18):8018–25.

Yeo GW, Van Nostrand E, Holste D, Poggio T, Burge CB. **2005**. Identification and analysis of alternative splicing events conserved in human and mouse. *Proc Natl Acad Sci USA* 102:2850–2855.

Zawel L, Kumar KP, Reinberg D. **1995**. Recycling of the general transcription factors during RNA polymerase II transcription. *Genes Dev* 9:1479.

Zeng L, Yap KL, Ivanov AV, Wang X, Mujtaba S, Plotnikova O, Rauscher FJ 3rd, Zhou M. **2008**. Structural insights into human KAP1 PHD finger-bromodomain and its role in gene silencing. *Nat Struct Mol Biol* 15(6):626-33.

Zuccato C, Tartari M, Crotti A, Goffredo D, Valenza M, Conti L, Cataudella T, Leavitt BR, Hayden MR, Timmusk T, Rigamonti D, Cattaneo E. **2003**. Huntingtin interacts with REST/NRSF to modulate the transcription of NRSE-controlled neuronal genes. *Nat Genet* 35:76–83.

FIGURES

Figure 1.1. Observed and expected masses for "expensive" human organs. The energy requirements of an organ are not proportional to the mass of that organ. The brain, the heart, the kidneys, the liver and the gastrointestinal tract collectively constitute only 7% of total body mass, but account for almost 70% of the total basal metabolic rate of the body. One way for humans to maintain their relatively larger brains without suffering an increase in basal metabolic rates would be to reduce the size of another expensive tissue in the body; this would offset the metabolic cost of the much larger organ at the expense of the other organ's mass. This explanation is known as the *'Expensive Tissue Hypothesis'* (Aiello and Wheeler, 1995; Aiello, 1997) and was tested by comparing the observed mass of the aforementioned "expensive" organs with the expected mass for each organ in average primates, which were then scaled to human body mass. The observed masses for each human organ (in an average 65 kg human male) were taken from Synder (1975), with the expected organ masses for an average primate having been calculated from the least-squares regression equations for primates (Stahl, 1965). As can be seen, there is an obvious tradeoff between the observed mass of the human brain in relation to its expected mass, at the expense of a decrease in the observed mass of human gastro-intestinal tract. Data was calculated based on a "standard" 65 kg human male with a BMR of 90.6 W (Aschoff et al., 1971); figure modified from Aiello and Wheeler (1995).

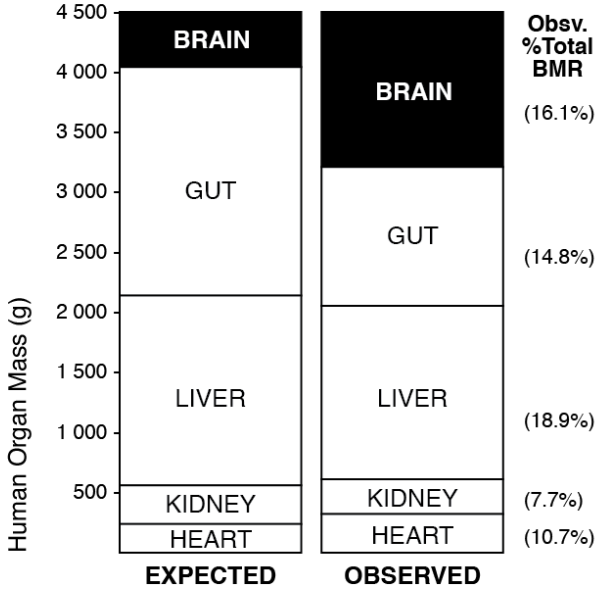


Figure 1.2. Mammalian brain mass and body mass comparison. The brain mass (g) and body mass (kg) of 20 mammals are plotted from smallest body mass (rodents) to largest body mass (whales). In all vertebrates, brain size can be found to scale to body size to the 0.6–0.8 power, increasing with "negative allometry" (meaning that the values are smaller than those predicted through simple "isometric scaling" that is governed simply by the square-cube law). The data in this plot are plotted in log-log coordinates, with a regression line drawn through the data points. The slope of the regression line represents the exponent of the power function. Perpendicular lines can be drawn through the regression line (as has been done with chimpanzee, human, and dusky dolphin), with longer distances above (or below) the line representing greater positive (or negative) brain size deviations from what would be expected based on average mammalian *brain:body* ratios. As can be seen, humans and dolphins possess much larger brains than would be expected for their bodies, while hippos and whales possess much smaller brains compared to their large masses. Figure modified from Roth and Dicke (2005); data from Van Dongen (1998).

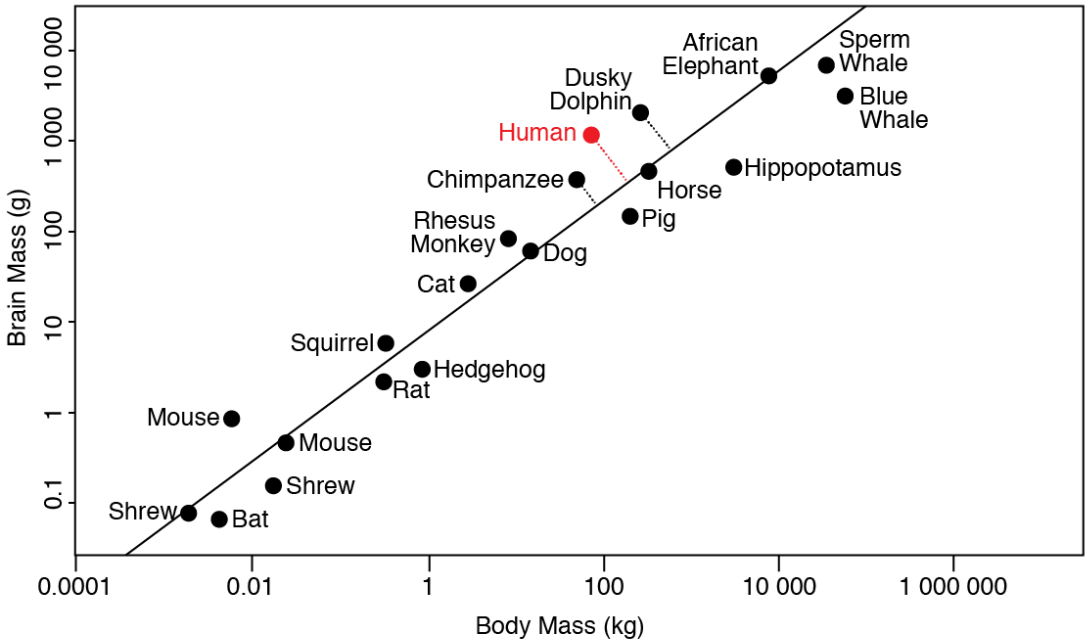


Figure 1.3. Mammalian encephalization quotients. The following diagram plots the average encephalization quotients for 38 mammals, arranged in ascending order of magnitude. Depicted in the plot are ten non-primate mammals (by genus), nineteen monkeys and great apes (by genus), and nine hominids including humans (by species). Encephalization quotient *EQ* is calculated as $EQ=E_a/E_e$ with E_a indicating actual brain size and E_e indicating expected brain size based on the proportions of a same-taxon animal 'standard'. In this case, the cat is used as a standard and has an *EQ* of 1.0 for the purposes of this comparison. Humans, considered to be the most intelligent animal, possesses the highest *EQ* of 7.4-7.8, indicating that the human brain is 7-8 times larger than what is expected of an animal its size. This is substantially greater than the *EQ* of most of humans' closest ancestors – extinct members of the genus homo – that have *EQ* values only half that of humans. However, measuring encephalization has its limits, as can clearly be seen by the fact that chimpanzees and gorillas have lower *EQ* values than squirrel monkeys and capuchin monkeys, despite the fact that Great Apes are in general much more intelligent than New World monkeys. As such, *EQ* is not the optimal predictor for intelligence, but may serve as only one characteristic that explains humans' increased intelligence. Mammalian data assembled from Haug (1987) and Roth & Dicke (2005); hominid data from Aiello & Wheeler (1995).

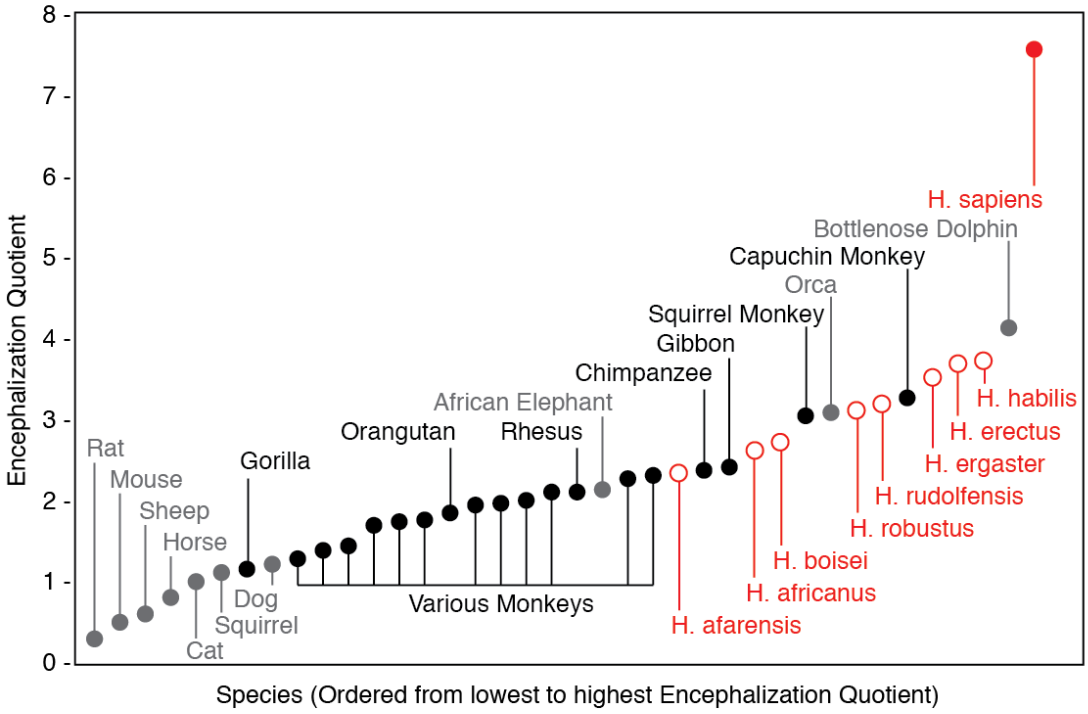


Figure 1.4. Structure of tandem Krüppel-type zinc fingers displaying the C2H2 motif. Individual zinc ions interact with paired cysteine and histidine residues, stabilizing two β-sheets and one α-helix (as indicated by the areas within the dotted lines) into the recognizable "finger" structure shown here, with the latter containing the DNA-binding interface as indicated in the figure at the *–1*, *+2*, *+3*, and *+6* positions relative to the helix. The structure of a finger sequence motif is represented, with *X* denoting an amino acid residue of any type with the subscript representing the number (ie. $X_{2\text{-}4}$ represents a chain of between 2 and 4 non-specified amino acid residues). The consensus sequence – *TGEKP(Y/F)* – is a highly conserved "H/C link" region between consecutive fingers.

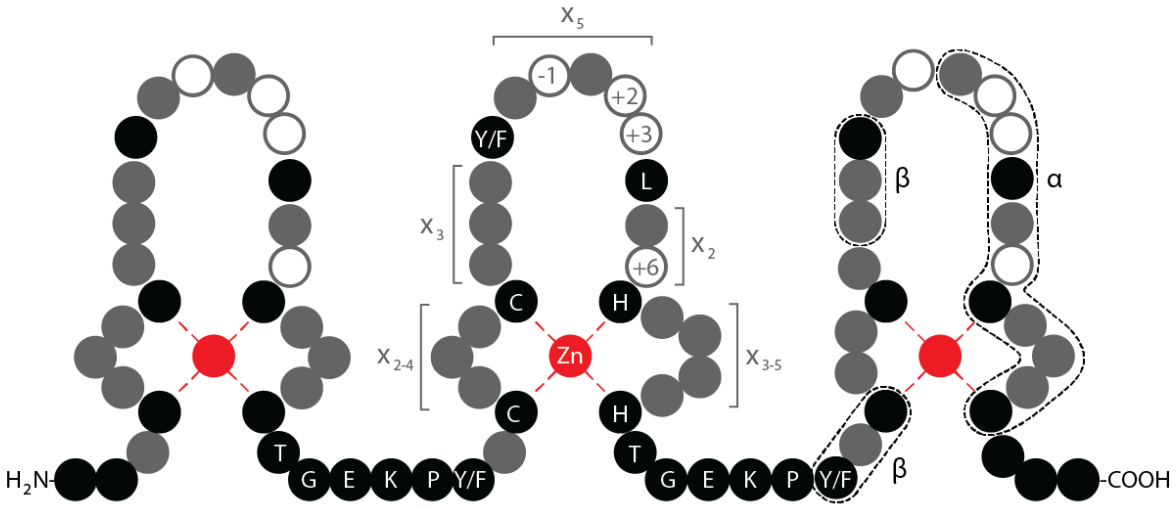


Figure 1.5. DNA binding properties of tandem C2H2 zinc finger domains. The α-helices of KZNF motifs contain amino acid residues that bind to DNA nucleotides at the *-1*, *2*, *3*, and *6* sites (as described in *Figure 1.4*), represented here as white dots on the helices of a set of tandem zinc fingers. The relationship between fingers and nucleotides is not one-to-one, as the amino acid at the *+2* position will interact with the nucleotide complementary to the neighboring finger's *+6* binding site. In this fashion, fingers wind around the major groove of the DNA molecule.

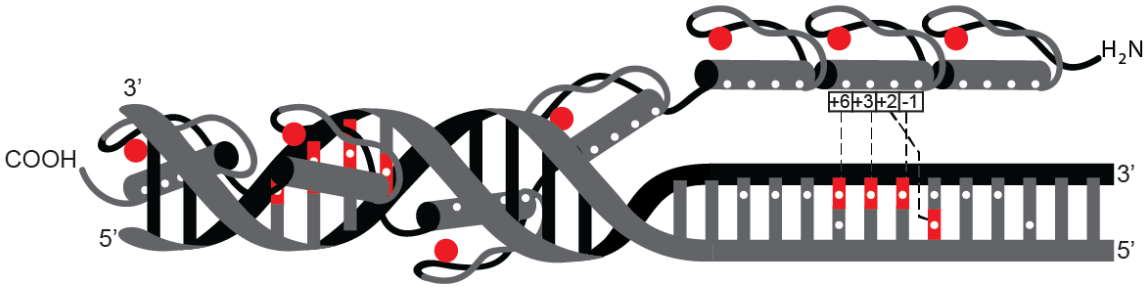


Figure 1.6. Mechanism of KRAB-domain repression. *Krüppel*-type zinc fingers that contain a KRAB-domain regulate gene expression by recruiting the KRAB-Associated Protein (KAP-1), which has several domains that interact with different co-factors. KAP-1 interacts with KRAB through an RBCC (RING Finger-B Box-Coiled Coil) binding domain, which is a common zinc-finger motif found in over 1500 proteins (Massiah et al., 2007). SUMO (Small Ubiquitin-like Modifier) is a 17 kDa ubiquitin-like molecule capable of post-translationally modifying many proteins. Unlike ubiquitin, SUMO does not mark proteins for degradation, and follows an enzymatic cascade to ubiquitin (Hay, 2005) in the KAP1 pathway. SAE1/2 (SUMO-Activating Enzyme Subunits 1 and Subunit 2) act as an E1 SUMO-activating enzyme, after which the E2 SUMO-conjugating enzyme UCB9 (Ubiquitin Carrier Protein 9) attaches SUMO to the E3 intramolecular-ligase PHD (Plant Homeodomain) finger domain of KAP-1 (Ivanov et al., 2007). Once properly recruited to KAP-1, sumoylation may begin on specific lysines on the adjacent Bromodomain. The modified complex is recognized by Histone-Lysine N-Methyltransferase SETDB1 and the NuRD (Nucleosome Remodeling and Histone Deacetylation) Complex. SETDB1 is a histone-H3 lysine-9 specific methyltransferase (Schultz et al., 2002), and the NuRD Complex is a histone deacetylase containing sets of both CHD3/4 (Chromodomain Helicase DNA-Binding Proteins 3 and 4) and HDAC1/2 (Histone Deacetylase 1 and 2) (Lai & Wade, 2011). These interactions are aided by HP1 (Heterochromatin Protein 1), which recognizes an HP1-Interacting Domain on KAP-1 and anchors chromatin subunits. Together, these cofactors tighten chromatin structure and silence genes that are bound in the heterochromatin. Model adapted from Ivanov et al., 2007.

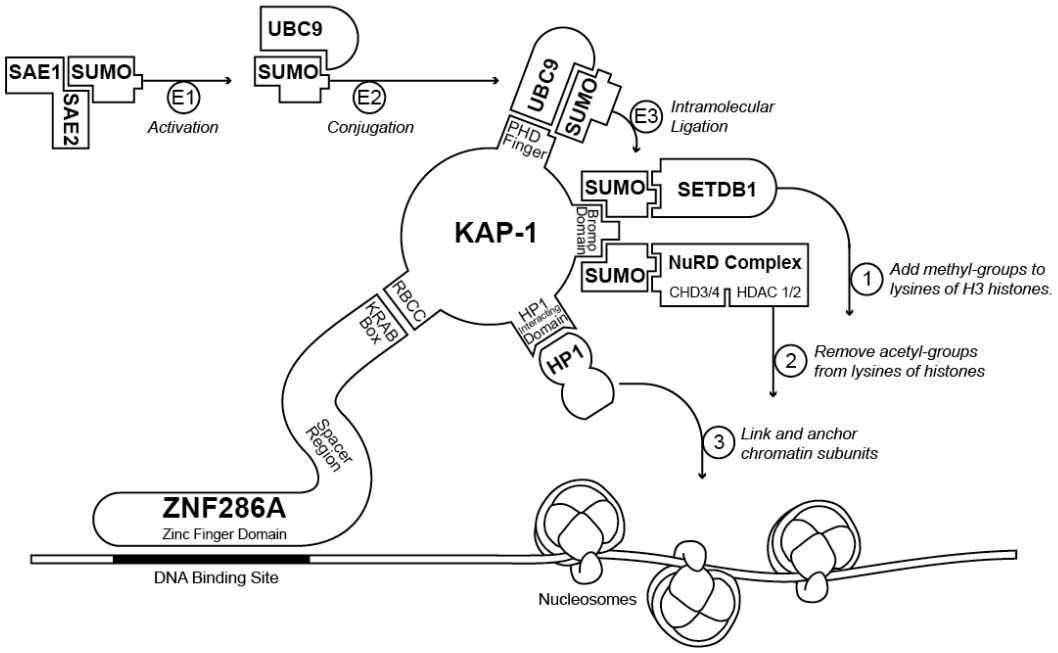


Figure 1.7. Structure of ZNF286 and REST. **(A)** The human ZNF286A protein is a KRAB-containing *Küppel*-type zinc finger transcription factor. It contains ten tandem C2H2 zinc finger within a DNA-binding domain. A spacer/tether region connects this domain to a KRAB-Box, which contains both KRAB-A and KRAB-B protein motifs in ZNF286A. Most mammals also produce a KRAB-less alternate transcript that is expressed alongside the KRAB-containing isoform. Both isoforms are very conserved in mammals. **(B)** Similarly, ZNF286B contains a nearly identical finger domain with a nearly-identical fingerprint. However, instead of a KRAB-domain, ZNF286B contains a FOXO3B pseudo-gene insert in its place. As such, ZNF286B does not have any known direct regulatory function. **(C)** REST is also a *Krüppel*-type zinc finger transcription factor, with seven DNA-binding zinc fingers (and one additional, structural, finger). At both ends of the protein lie cofactor interaction domains – one for Sim3 and one for Co-REST.

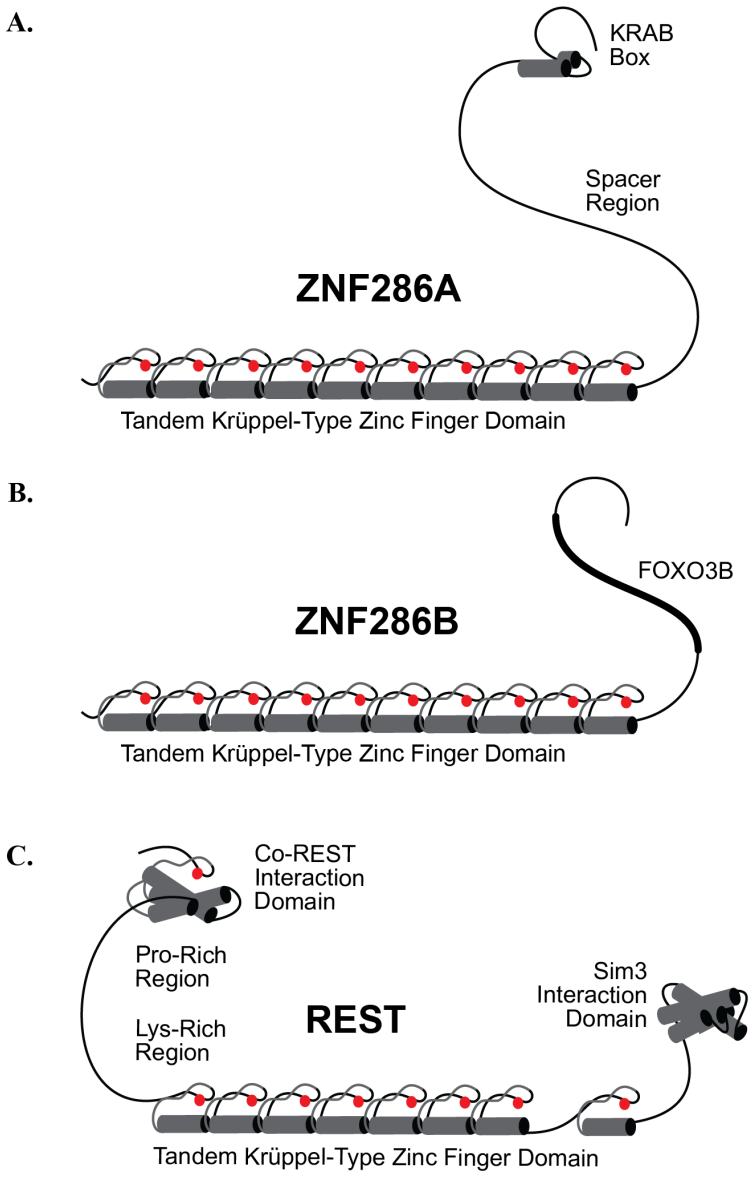
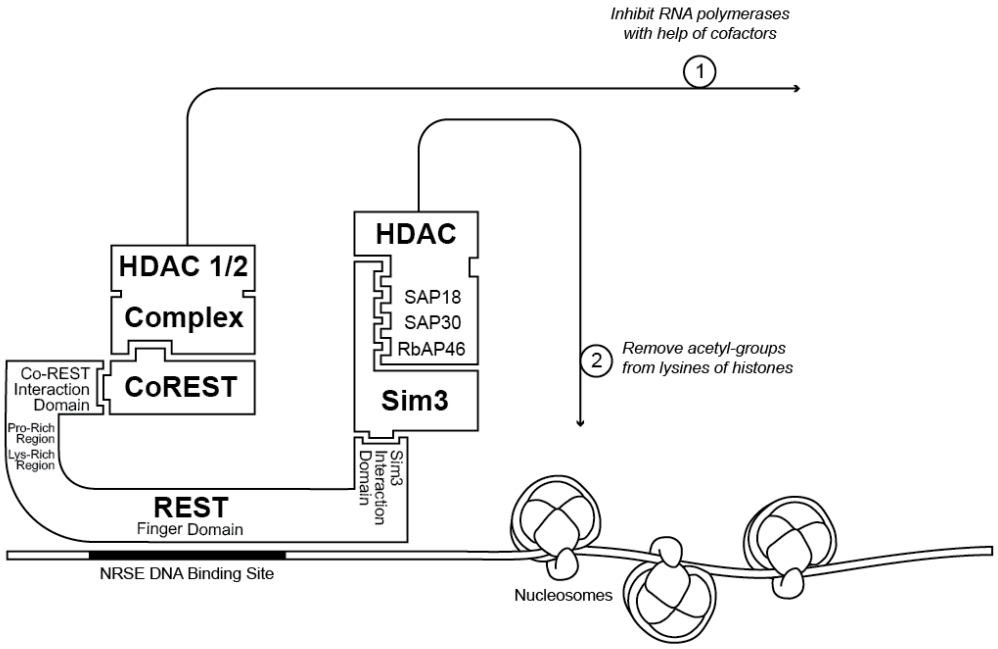


Figure 1.8. Mechanism of REST. Repression by REST begins after the protein's zinc finger domain binds to a highly specific and highly conserved DNA sequence known as the neuron-restrictive silencer element (NRSE). The interaction between REST and the NRSE is stabilized by an ATP-dependent chromatin-remodeling helicase enzyme called SMARCA4. SMARCA4 contains a bromodomain that recognizes acetylated H4K8 and repositions nucleosomes, thereby allowing REST to form more stable DNA interactions – simply increasing H4K8 acetylation leads to increases in REST recruitment (Ooi et al., 2006). A Sim3 Interaction Domain located at the N-terminus of REST interacts with the Sim3 complex, which recruits several subunits (SAP18, SAP30, RbAP46) including histone deacetylases (HDAC1/2) (Roopra, 2000). A CoREST-interaction domain located at the C-terminus of REST interacts with a CoREST cofactor, which recruits a variety of proteins including HDAC1/2, the chromatin remodeling SMARCA4, H3K4 demethylase LSD1 and H3K9 methylases G9a (Andrés et al., 1999). Together, or separately, the corepressors recruited by SIM3 or CoREST remove several gene-activating chromatin modifications and inhibit RNA polymerases, thus silencing specific genes.



CHAPTER 2

MAMMALIAN ZNF286A AND HUMAN-SPECIFIC DUPLICATE ZNF286B BIND RE-1 SITES AND OPPOSE REST ACTIVITY IN NEURAL CELLS

Mammalian ZNF286A and human-specific duplicate ZNF286B bind RE-1 sites and oppose NRSF/REST activity in neural cells

Derek Caetano-Anollés[1,2], Abdol Majid Kazemian[3,4], Huimin Zhang[1,2], Jacqueline A. Brinkman[2], Saurabh Sinha[2,3], and Lisa J. Stubbs[1,2,*]

[1] Department of Cell & Developmental Biology, University of Illinois. Urbana, IL
[2] Carl R. Woese Institute for Genomic Biology, University of Illinois. Urbana, IL
[3] Department of Computer Science, University of Illinois. Urbana, IL
[4] Current Address: National Heart Lung and Blood Institute, National Institutes of Health. Bethesda, MA
[*] Corresponding author: *ljstubbs@illinois.edu*

At the time of submission, this manuscript is under revision. Data and arguments may differ in the published version.

Abstract

ZNF286A is one of only a few members of the large and diverse KRAB-containing zinc finger gene family represented by a unique ortholog in marsupials and most eutherian species. However, this ancient gene gave rise to a single duplicate, called *ZNF286B*, in very recent primate history. Here, we show that *ZNF286B* arose as part of a larger duplication event including many surrounding genes; concomitantly or shortly after duplication, a processed *FOXO3* pseudogene was inserted into *ZNF286B* and DNA encompassing the KRAB-encoding exons and regulatory sequences was deleted. As a result, *ZNF286B* encodes a human-specific protein with structural differences and a distinct expression pattern; most notably, while both genes are most highly expressed in human brain, *ZNF286A* is highest in adults while *ZNF286B* is expressed primarily during fetal development. Chromatin immunoprecipitation (ChIP-seq) and *in vitro* binding assays revealed that ZNF286A and ZNF286B proteins bind preferentially to known motifs for the neuronal regulatory factor, REST/NRSF in neuroblastoma cells. Furthermore, gene knockdown experiments suggested that despite their structural differences, the human ZNF286 proteins act cooperatively to inhibit cell proliferation. Intriguingly, REST is known to *activate* proliferation in neuroblastoma and neural progenitor cells; consistently, we show that as REST protein levels fall, ZNF286 levels rise during neuron differentiation *in vitro*. We hypothesize that the ZNF286A evolved in mammals as a negative modulator of REST activities, and that the ZNF286B duplication allowed for independent transcriptional control and enhancement of this function in the developing human brain.

Introduction

ZNF268A is a member of the large and evolutionarily dynamic KRAB-ZNF family of genes, encoding proteins in which a tandem array of multiple *Krüppel*-type zinc fingers (ZNFs) is typically tethered by a short unstructured sequence to one or more N-terminal *Krüppel*-associated box (KRAB) chromatin interacting domains. The ZNF arrays in these proteins are thought to function primarily in site-specific DNA binding, whereas KRAB (particularly the most common version of this domain, KRAB-A) is predicted to function in transcriptional repression through interactions with a ubiquitous co-repressor called KAP1 (also known as TRIM28, TF1β, TIF1β, and KRIP-1; Friedman et al., 1996; Moosmann et al., 1996; Kim et al., 1996). In vertebrates and especially in mammals, the KRAB-ZNF family has expanded through repeated rounds of tandem segmental duplication to include hundreds of lineage-specific genes (Chung et al., 2002). Differences in KRAB-ZNF copy numbers are seen even between closely related species, such as humans and chimpanzees (Huntley et al., 2006; Hamilton et al., 2006; Nowick et al., 2011). New duplicates can diverge significantly over relatively short evolutionary times, changing in expression pattern, splicing, and/or coding sequence in ways that predict distinct functions (Hamilton et al., 2006; Nowick and Stubbs, 2010; Nowick et al., 2011).

ZNF286A is an especially deeply conserved member of this dynamic gene family, being one of only 37 human KRAB-ZNF genes with a clear 1:1 ortholog in marsupials (Liu et al., 2014). This unusual level of conservation suggests that the gene has acquired an essential mammalian regulatory function. However, *ZNF286A* has given rise to a single gene duplicate in very recent primate history; this duplicate gene, called *ZNF286B*, is found in humans but not in the rhesus macaque, orangutan or chimpanzee genomes (Nowick et al., 2011). Despite its very recent advent, *ZNF286B* differs from the parental gene in several significant respects. In particular, although the ZNF arrays in *ZNF286A* and *ZNF286B* are nearly identical, *ZNF286B* does not include KRAB-encoding exons, and instead includes novel exons that are not present in the parental gene. These novel features suggest that the duplicates may have evolved distinct functions, with *ZNF286A*

acting primarily as a KAP1-binding repressor and *ZNF286B* serving a distinct, perhaps dominant-negative competitive regulatory role.

To uncover the functions of the *ZNF286* genes, we mapped gene and protein expression in human tissues and designed custom siRNA molecules to knock down expression of each paralog specifically. Probably due to deletions and insertions that accompanied *ZNF286B* duplication, the paralogs are differently expressed; most notably, both genes are expressed at highest levels in brain, but *ZNF286A* is expressed most highly in adults whereas *ZNF286B* was is dominant in fetal brain. Surprisingly despite their structural differences, *ZNF286A* and *ZNF286B* knockdowns yielded enrichment for many of the same up- or down-regulated genes, suggesting cooperative rather than competitive functions. We also used a custom ZNF286A antibody to carry out chromatin immunoprecipitation (ChIP-seq), revealing significant overlap between ZNF286 binding regions and the known recognition motif for the important neuronal regulator, RE1-silencing transcription factor (REST), also called NRSF. ChIP-seq with a REST antibody confirmed that ZNF286A and REST binding regions overlap extensively in neuroblastoma cells; we also demonstrate that ZNF286 proteins bind the REST-binding RE-1 motif *in vitro,* as well as an alternative REST motif (Satoh et al., 2013; Rockowitz et al., 2014). However, the dominant function predicted for the ZNF286 proteins – inhibition of cell proliferation – is opposite to the known functions of REST in neural and neuroblastoma cells (Guardavaccaro et. al., 2008; Negrini et al., 2013). Together our data support the hypothesis that the ZNF286 proteins evolved in mammals as competitive modulators of REST activities, and that duplication and divergence of *ZNF286B* provided a novel, independently regulated layer of REST modulation in the developing human brain.

Materials and Methods

QPCR of cell line RNA: RNA was collected by homogenizing tissue samples or cell pellets in TRIzol Reagent (*Invitrogen* Cat. 15596). Samples were purified and DNase-treated using the standard Qiagen protocol with RNeasy columns (*Qiagen* Cat. 74104), from which cDNA was synthesized using using M-MuLV Reverse Transcriptase (*New*

England Biolabs Cat. M0253S). QPCR was performed on each sample using sequence-specific primers (***Supplemental Table S2.1A***) in Power SYBR Green PCR Master Mix (*ThermoFisher* Cat. 4367659). PCR reactions proceeded using the *Applied Biosystems* 7900HT Fast Real-Time PCR System.

Cell line culture and neuroblast differentiation: SH-EP cells acquired from Dr. Martin Reick (Reick et al., 2001) were cultured and maintained in DMEM containing 10% fetal bovine serum (FBS). Neuro2A cells (*ATCC*® CCL-131) were grown in DMEM containing 10% FBS prior to differentiation, and induced to differentiate at 25-40% confluence by replacing growth media with media containing 2% FBS and 20 μM retinoic acid, replacing the media every 24 hours, with cells reaching full differentiation after 72 hours (Tremblay et al., 2010). SH-SY5Y cells (*ATCC*® CCL-131) were grown in 1:1 DMEM/F12 media containing 10% FBS, and were induced to differentiate at 50-60% confluence by replacing growth media with DMEM media containing 15% FBS and 10 μM retinoic acid. Media was replaced every 2-3 days. On the fifth day of RA-treatment, cells were washed three times with DMEM, and 50 ng/mL BDNF was added to the media (without serum) to maintain division. SH-SY5Y cells are fully differentiated after 6-10 days (Encinas et al., 2002). SK-N-SH (*ATCC*® HTB-11) cells were grown in 1X MEM media containing 10% FBS, and were induced to differentiate at 50-60% confluence by replacing growth media with DMEM (without serum) containing 20 μM retinoic acid. Media was replaced every 3 days. SK-N-SH cells are fully differentiated after 6-10 days (Preis et al., 1988). Genomic DNA to test human variation of ZNF286B was collected from a variety of human lymphoblast and fibroblast cell lines (***Supplemental Table S2.1B***) as supplied by the Coriell Institute, and grown in RPMI 1640 containing 15% FBS.

Knockdown using siRNA: Custom siRNA oligonucleotides were designed to target *ZNF286A* sequence 5'-CACCTACCATTCAGTGCTTAT-3' (*Qiagen.* Cat. SI04135600) and 5'-TAGCAGTACGAACATTGTGAA-3' (*Qiagen* Cat. SI03227448). Targets for *ZNF286B* were 5'-TACCATCAATTAAGTTTCATT-3' (*Qiagen* Cat. SI05457697) and 5'-TAACTCTTACCTATTACAGAA-3' (*Qiagen* Cat. SI05457704). In addition, the

AllStars scrambed siRNA with Alexa Fluor 488 modification (*Qiagen* Cat. 1027284) was used as a negative control (called "siScr" in this paper). SH-EP cells were grown under normal growth conditions, and transfection proceeded using 5 nM concentrations of siRNA in HiPerFect Transfection Reagent (*Qiagen* Cat. 301704), following the standard protocol. Cells were incubated under normal growth conditions with Opti-MEM I Reduced Serum Medium (*Gibco* Cat. 31985) for 32 hours, then immediately collected for RNA and protein collection.

Western blots: Nuclear extract was isolated from freshly pelleted mammalian cell lines using the *Active Motif* Nuclear Extract Kit (Cat. 40010), freezing aliquots of the extracted proteins at -80 °C until they were ready for use. Detection of specific proteins was accomplished by Western blot using 10% SDS-PAGE gels and between 1-2 μg of nuclear extracted protein was used per lane as an input. Primary rabbit antibodies used for detection were Rb-α-KAP1 (*Abcam* Cat. ab10483, 1:1000), Rb-a-Lamin-B1 (*Abcam* Cat. ab16048, 1:2000), Rb-α-RCOR1 (*Abgent* AP17018A, 1:300), Rb-α-REST (*Santa Cruz* Cat. Sc-25398, 1:500), Rb-α-TBP (*Santa Cruz* Cat. sc-273, 1:2000), Rb-α-ZNF286A (*Abgent* RB21720, 1:200, custom generated with epitope NGKEPLKLERKAPK), Rb-α-ZNF286B (*Abgent* RB21665, 1:200, generated with epitope PLHPAPAREEIKST), and Rb-α-ZNF431 (*Abgent* RB10670, 1:200, epitope CSVDEYKVHKEGYNE). Primary mouse antibodies were Ms-α-KAP1 (*Abcam* Cat. ab22553, 1:1000) and Ms-α-REST (*Abcam* Cat. ab52849, 1:600). Secondary antibodies used were Bovine-α-rabbit IgG-HRP (*Santa Cruz* Cat. sc-2379, 1:3000) and Bovine-α-mouse IgG-HRP (*Santa Cruz* Cat. sc-2375, 1:3000). *ImageJ* (v1.49) was used to measure western blot band density, using the standard tools to calculate the area of optical spectrum peaks generated for each band, with densitometry values normalized against the control antibody bands of each lane (Tan and Ng, 2008; Gassmann et al., 2009).

ChIP-seq: Cells were cross-linked by adding 1% formaldehyde to cell growth media and incubated for 10 minutes at room temperature. The cross-linking reaction was quenched with 0.125M glycine for 5 minutes at room temperature before washing cells with PBS. The cell suspension was washed using 1X PBS with *Roche* cOmplete™ Protease

Inhibitor Cocktail (*Roche* Cat. 11873580001) and cells collected by centrifugation at 4°C. Washed cell pellets were re-suspended with 2ml per ~10^7 cells in lysis buffer (5% 1M Tris-HCl pH8.1, 0.4% 0.5M EDTA pH8.0, 0.5% NP-40, 10% glycerol) with protease inhibitor and incubated on ice for 30 minutes. Nuclei were collected and re-suspended in 300 µl SDS-lysis buffer (1% SDS, 10mM EDTA, 50mM Tris pH8.1) with protease inhibitor, rotated for 20 mins at 4°C, and sonicated using a *Covaris* M220 Focused-Ultrasonicator (*Covaris* Cat. 500295). The sonicated mixture was centrifuged at 16,000 rcf. Each sonicated sample was diluted 10-fold with cold ChIP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2mM EDTA, 16.7 mM Tris-HCl pH8.1, 167mM NaCl). A total of 75 µl Dynabeads® magnetic beads (*Life Technologies* Cat. 10009D) were added to the 1 mL samples, and incubated for one or more hours 4°C. The beads were magnetically separated, and the supernatant was removed for IP. Samples were prepared using 10-20 µg of IP-antibody or a non-specific Rb-α-IgG antibody (*Santa Cruz* Cat. sc-2027) and incubated for 18 hours at 4°C with rotation. To each solution, 100 µl of magnetic particles were added, incubating for 1 hour at 4°C with rotation before magnetically separating the beads. Samples were sequentially washed at 4°C using Low Salt Buffer (*EMD Millipore* Cat. 20-154), High Salt Buffer (*EMD Millipore* Cat. 20-155), LiCL Buffer (*EMD Millipore* Cat. 20-156), and TE Buffer (*EMD Millipore* Cat. 20-157). Magnetic pellets were then re-suspended in 250 µl ChiP Elution buffer (1% SDS, 0.1M $NaHCO_3$), incubated for 15 minutes and then magnetically separated, transferring the supernatant containing the chromatin complexes. To reverse-crosslink the DNA from the proteins, 20µl of 5M NaCl was added to each eluate and incubated at 65°C overnight, followed by a one hour incubation at 45°C with Proteinase K (*New England Biolabs* Cat. P8107S) and a 30 minute incubation at 37°C with RNaseA (*Qiagen* Cat. 19101). DNA was recovered by phenol-chloroform extraction and isopropanol precipitation. DNA was quantified by Qubit® 2.0 Fluorometer (*Life Technologies* Cat. Q32866), and subsequently used for whole genome amplification using GenomePlex® Complete Whole Genome Amplification (*Sigma-Aldrich* Cat. WGA2). Amplified samples were sequenced using the *Illumina* HiSeq 2500 System. ChIP experiments were conducted in duplicate for each antibody, and duplicates were examined for similarity before being combined for final peak analysis. A total 38,104,016 of 100 bp single-end mapped reads

were analyzed for the ZNF286A ChIPseq experiment and 36,257,606 reads were generated from the sonicated input DNA for comparison. REST ChIP-seq was pooled to generate 8,851,563 reads and compared to a total of 14,894,269 reads from the input control from that experiment.

Computational analysis of microarray and ChIP-seq data: Total RNA from siRNA knockdown experiments were labeled and hybridized to Affymetrix HuGene_1.0-st-v1 microarrays (Affymetrix). Preliminary analysis, including quantitation, RMA normalization and gene annotation, was carried out using the Affymetrix Expression Console software. Differentially expressed gene lists, determined using a multi-tTest in the Bioconductor R package (Gentleman et al., 2004), were corrected for multiple testing using Benjamani-Hochberg (BH) correction. Genes with a BH-corrected *p* value of ≤0.05 and an absolute fold change of ≥1.5 were collected (***Supplemental Table S2.2***) and analyzed as described in the text. Up- or down-regulated differentially expressed genes (DEGs) were examined for functional enrichment using the DAVID program functional clustering settings (Huang et al. 2009a). The top listed category from each cluster with the lowest false discovery rate was used to describe the clusters in Table 2, with only those clusters with an enrichment score of 1.5 reported. ChIPseq data were mapped to the human genome (hg19) using Bowtie2 (Langmead and Salzburg, 2012) allowing 1 mismatch, but otherwise using default settings. Mapped ChIP-seq reads were analyzed using the MACS 1.4.2 package (Zhang et al., 2008) to generate peaks in comparison to the matched input control. REST ChIP-seq was analyzed using default settings; for ZNF286A ChIP-seq we set the –bw at 400 bp to accommodate the slightly longer input fragments used in those experiments, with other settings left as default. Genes were assigned to peak regions using GREAT version 3.0 (McLean et al., 2010) with "Two nearest genes" association rules setting.

Motif identification: To identify enriched motifs, we used a number of different programs, all pointing to a similar set of motifs. Representative examples of the enriched motifs scored most highly by this suite of programs are summarized in Figure 4A, displaying the motifs and *p*-values derived from MEMEchip, from the MEME suite

(Bailey et al., 2009). We used sequence from 200 bp regions surrounding peak summits as estimated by MACS as input for each program. MEME input giving rise to motifs displayed in ***Figure 2.4A*** was selected from the fdr=0 peak sets as described in the text (***Supplemental Table S2.3****).*

EMSA: Sense and antisense 40-mer oligonucleotides were designed (***Supplemental Table S2.1***) and annealed together to form double-stranded probes. Biotinylated duplexes were annealed together using 1X biotinylated sense strands for every 5X unbiotinylated antisense strand, to minimize un-annealed biotinylated products from being detected on the membrane. EMSA was performed using the LightShift™ Chemiluminescent EMSA Kit (*ThermoFisher* Cat. 20148), following the manufacturer's protocol. Samples were prepared using 1 µL of biotinylated probe (1:400 diluted from 10 µM stock solution), 5 µg SH-EP nuclear extract to shift samples, along with 0.5 µL of select antibodies at optimized dilutions to super-shift samples. Super-shift antibodies included Rb-α-YY1 (1:25; *Santa Cruz* Cat. sc-1703), Rb-α-ZNF286A (1:250; *Abgent* Cat. RB21720), Rb-α-ZNF286B (1:250; *Abgent* Cat. RB21665), and Rb-α-REST (1:15; *Santa Cruz* Cat. Sc-25398). If competition was being tested, 1 µL of unbiotinylated probe (10 µM stock solution) was added to the sample. Each sample contained 1X Binding Buffer, 50 ng/µL Poly(dI•dC), 2.5% Glycerol, 0.05% NP-40, and 5mM $MgCl_2$. Ultrapure water was added to a total sample volume of 10 µL. Samples were allowed to sit for 5 minutes before being run through a pre-run, 8% native gel, touch-transferred to a nylon membrane for 1 hour, and cross-linked to the membrane using a 254 nm UV light source for 5 minutes. Biotin-labeled DNA was detected by chemiluminescence using 1:3000 Stabilized Streptavidin-HRP Conjugate (*ThermoFisher* Cat. 89880D).

Co-IP: SH-EP nuclear extract was isolated using the *Active Motif* Nuclear Extract Kit (Cat. 40010) for use in co-IP reactions. For samples that were cross-linked prior to co-IP, cells were incubated for 10 minutes at room temperature using 0.5% formaldehyde and quenched using 0.125M glycine for 5 minutes at room temperature (*sensu* Kim et al., 2009) prior to nuclear extract isolation. Co-IPs proceeded using 14 µg of nuclear extract using the Pierce™ Co-Immunoprecipitation Kit (*ThermoScientific* Cat. 26149) and 10 µg

of affinity-purified antibodies, coupled to AminoLink™ Plus Coupling Resin (*ThermoScientific* Cat. 20501). Coupling-antibodies used in this study include Rb-α-YY1 (*Santa Cruz* Cat. sc-1703), Rb-α-KAP1 (*Abcam* Cat. ab10483), Rb-α-ZNF286A (*Abgent* Cat. RB21720), Rb-α-ZNF286B (*Abgent* Cat. RB21665), Rb-α-REST (*Santa Cruz* Cat. Sc-25398), Rb-α-RCOR1 (*Abgent* AP17018A), Ms-α-KAP1 (*Abcam* Cat. ab22553), and Ms-α-REST (*Abcam* Cat. ab52849). Co-IP products were eluted by boiling resin for 5 minutes and running an SDS-page gel for protein detection. For samples using rabbit antibodies for immunoprecipitation, mouse antibodies were used for detection, to avoid cross-detection of the resin-antibodies. Likewise, rabbit antibodies were used to detect samples immunoprecipitated by mouse antibodies.

Results

The ZNF286B duplicon includes flanking genes and KRAB-sequence deletion

The *ZNF286A* gene is ancient, existing as a unique locus in most eutherian mammals and in the genome of the marsupial *Monodelphus domestica* (Liu et al., 2014), while *ZNF286B* is detected only in the human genome (Nowick et al., 2011). To gain a more concise view of gene structure and duplicon history, we compared the two human paralogs and surrounding genomic sequences (***Figure 2.1***). Alignment between the two genomic regions showed that *ZNF286B* is embedded within a larger segmental duplication located about 2 million base pairs (Mbp) away from *ZNF286A* on human chromosome 17 (chr17). The duplication spans approximately 600 kb and includes 12 neighboring genes that are inverted in telomeric-to-centromeric order relative to genes surrounding *ZNF286A.* Contained as they are within this multigenic duplicon, the two ZNF286 paralogs share a very similar overall genomic environment.

In addition, the two genes encode very similar proteins, including tandem arrays of 10 C-terminal ZNF motifs; except for zinc fingers 1 and 8 (numbered from N- to C-terminal positions (***Table 2.1)*** the two proteins have identical amino acid sequences at DNA-contacting positions (positions -1, 2, 3 and 6 relative to the alpha helical regions in each finger; Choo and Klug, 1994; Wuttke et al., 1997). Surprisingly however, the N-terminal regions of ZNF286A and ZNF286B predicted proteins differ significantly. First, the genomic region encompassing both KRAB-A and KRAB-B-encoding exons of *ZNF286A* has been deleted in the *ZNF286B* gene (***Figure 2.2A***). Furthermore, a retroposed pseudogene of the chr6 gene *FOXO3* was inserted into the duplicated gene, positioned directly downstream of the deletion. In fact, *ZNF286B* co-opted its novel third exon, encoding 30 amino acids that are not present in ZNF286A, from this *FOXO3B* pseudogene. This 90 bp open reading frame (ORF) is not a coding sequence in *FOXO3* but rather, overlaps with 5'-untranslated sequences. Furthermore, *ZNF286B* utilizes a novel splice acceptor site that adds a short segment of amino acid sequence derived from the intron of *ZNF286A* and is only included in the ZNF286B protein (***Figure 2.2A).***

Interestingly, the deletion that removed KRAB-encoding exons also interrupted a predicted enhancer element that is present in a *ZNF286A* intron, marked by modified

histone H3 (H3K4Me1 and H3K27Ac) (***Figure 2.2B***). However, a novel region marked by modified histones and a DNase-sensitive site were brought in by the *FOXO3B* pseudogene insertion (***Figure 2.2C***). Therefore, despite the fact that *ZNF286B* arose as a very recent copy of *ZNF286A*, rearrangements in the novel human duplicate have altered the predicted protein significantly, and have likely altered regulatory element composition of the novel locus as well.

Despite these and other internal rearrangements, the overall sequence similarity across the two gene duplicates is 98.6% for regions flanking the *FOXO3B* insertion site, including several short insertion-deletion events (indels). Similarity between the *FOXO3B* insertion and the parental *FOXO3* gene is similar although slightly higher, at 99% identity confirming a very recent duplication event (not shown). In a previous study we used PCR to demonstrate that this juxtaposition of ZNF286 and FOXO3 sequences can be detected only in the human genome (Nowick et al., 2011), and a BLAST search for a similar fusion event in the updated chimpanzee, orangutan and other primate genomes found matches only to separate unique *ZNF286A* and *FOXO3* genes (not shown).

These results prompted the question of whether the current version of the *ZNF286B* gene has been fixed in the human population. To test this hypothesis we used the same *FOXO3*- and *ZNF286B*-based PCR primers to screen DNA from a panel of lymphoblastoid cells from humans of diverse American, African, Asian, and European origins. A PCR product of the correct length and sequence was detected in all of the 46 DNA samples we tested (***Supplemental Figure S2.1***). These data support the hominin (and for extant species, the human) specificity of the *ZNF286B* locus, and demonstrates the fixation of this duplication in the human population, suggesting that local rearrangements not only altered protein-coding capacity but also potentially altered the expression patterns of otherwise very similar genes.

ZNF286 paralogs are differentially expressed in human tissues

To examine expression of the two human genes, we compared steady-state RNA levels of *ZNF286A* and *ZNF286B* transcripts in quantitative RT-PCR (qRT-PCR) experiments in a panel of RNA from human tissues. We used primer pairs designed

against sequences that are uniquely included in transcripts arising from each gene (locations illustrated in ***Figure 2.2A***). These data confirmed that ZNF286A and ZNF286B are expressed with different tissue-specific patterns. Notably, of the tissues we tested, ZNF286A displayed highest levels of expression in adult brain with relatively lower levels in fetal brain (***Figure 2.3A***), whereas *ZNF286B* displayed highest levels of expression in fetal brain with substantially reduced expression in adults (***Figure 2.3B***). These data therefore indicate that the two human duplicates have indeed evolved distinct patterns of tissue-specific expression.

ZNF286A and ZNF286B regulate genes involved in cell proliferation

Developing paralog-specific reagents

Our next goal was to identify the functions associated with the two paralogous human genes. Our initial hypothesis was that the KRAB-containing ZNF286A should function primarily as a repressor, whereas the KRAB-less ZNF286B protein would function distinctly, possibly as a "dominant negative" binding competitor to the parental protein. This hypothesis would predict that "knockdown" of ZNF286A and ZNF286B would produce very different, and potentially opposite, effects on expression of direct and secondary target genes. To test this hypothesis we looked for cell lines in which both *ZNF286A* and *ZNF286B* were expressed at relatively robust levels so that results of functional assays could be directly compared. Among the human cell lines we tested, the S-type SH-EP neuroblastoma cell line (Reddy et al., 1991) was one of the few that expressed relatively robust levels of both *ZNF286A* and *ZNF286B,* and we selected this cell line for further analysis.

To test gene function, we designed custom siRNAs targeting the diverged transcribed sequences in each paralog. After siRNA treatment for 32 hours, *ZNF286A* RNA was reduced by over 70% of levels in cells treated with a "scrambled" control, whereas *ZNF286B* transcripts levels were not affected significantly. Likewise, *ZNF286B* RNA levels were similarly reduced by over 70% when targeted by custom siRNA, whereas *ZNF286A* transcripts were not affected (***Fig 2.3C***). These results indicated that the reagents were acting relatively efficiently and specifically.

We also developed custom rabbit polyclonal antibodies from unique protein regions (epitope locations presented in ***Figure 2.2***, and sequences detailed in Methods). Both antibodies detected proteins of expected size (~60 kD) in SH-EP cells (***Figure 2.3D***). In proteins prepared from nuclear extracts of the siRNA-treated and control cells, the proteins were also specifically decreased, confirming specificity of both antibodies as well as the efficient knockdown of the proteins in the siRNA experiments (***Figure 2.3C***).

siRNA knockdown reveals overlapping collections of differentially expressed genes

We analyzed global gene expression after siRNA knockdown in the SH-EP cell line on microarrays using two siRNAs from each gene that gave the highest knockdown levels (***Figure 2.3***). These experiments yielded a total of 261 significant (at least 2x fold change; $p \leq 0.05$ after correction for multiple testing) differentially expressed genes (DEGs) after knockdown of *ZNF286A*, including 209 up-regulated and 52 down-regulated unique genes. Knockdown of *ZNF286B* yielded similar ratios of up- and down-regulated DEGs: 300 and 78 respectively (***Supplemental Table S2.2***). Surprisingly given our original hypothesis, DEG lists from the two knockdown experiments were very similar, with significant up- or down-regulation of many of the same genes. In fact, 233 out of 261 DEGs obtained after *ZNF286A* knockdown (89.2%) were also found on the *ZNF286B* DEG list and all of these genes were differentially expressed in the same direction (Table S2). These data indicated that when they are present in the same cell types, the two human ZNF286 paralogs serve to regulate the same genes, and in the same fashion. The two human genes do not appear to be redundant in function, since the knockdown of one gene mimics the effects of knocking down the other. Rather, these results suggested that *ZNF286A* and *ZNF286B* might in fact act cooperatively and dependently when expressed together in human cells.

Not surprisingly therefore, DEGs in the two experiments were enriched in very similar functional categories. Up-regulated DEGs, corresponding to genes that are repressed by *ZNF286* expression, were very highly enriched in functions generally related to mitosis, DNA replication, chromatin structure, and hormone response, whereas down-regulated DEGs (activated by ZNF286 expression) were commonly enriched in functions related to cellular signaling, and regulation of steroid biosynthesis (***Table 2.2***).

Identification of ZNF286A binding sites and target genes

We carried out ChIP with the ZNF286A antibody in SH-EP cells, identifying 1,502 ChIP peaks (False Discovery Rate (FDR)<5% and Enrichment Factor (EF)>5) and their nearest genes (***Supplemental Table S2.3)***. Unfortunately the custom ZNF286B antibody did not work well for ChIP. However, given the similarity between the fingerprints of the two human genes and especially the very similar effects of their knockdowns, we identified potential direct target genes by overlapping the list of genes nearest ZNF286A ChIP peaks with DEGs from both the ZNF286A and ZNF286B SH-EP knockdown experiments (adjusted p-value ≤ 0.05 and absolute value of fold-change greater than 1.5). We found 165 peak regions that either flanked or contained DEGs, including 50 down-regulated and 74 up-regulated genes (***Suppl Table S2.4***). Thus despite the predominance of down-regulated genes after siRNA knockdown, the genes nearest ZNF286A peaks could be either activated or repressed.

Of note, predicted repressed targets of ZNF286 proteins (siRNA up-regulated), were dominated by essential mitotic regulators (including *AURKB, NCS1, CLSPN, NEK6, CDC25B, CDCA8, E2F7, CKS2, GSG2, SMARCA2)*. On the other hand, positively regulated targets (siRNA down-regulated DEGs located near ChIP peaks) included genes involved in neural differentiation and development (*ID4, SMAD6, INA, TLE3, NEUROD6, SEMA3E, GPRC5B)* or adult neuron function (*GLRA3, CADM1, SYT12, SCG3)*. Most peaks were distal (>10kb) to the transcription start sites (TSS) of the nearest DEGs. However, a small number of DEG TSS (32 genes) was located within 10kb of ZNF286A peaks. These included 21 up-regulated (e.g. *CDC25B, CENPJ, SMARCA2)* and 11 down-regulated genes (e.g. *GLRA3, INA, ID4, SMAD6)*.

We validated a selection of predicted target genes in repeat siRNA knockdown experiments followed by quantitative PCR to compare experimental samples to the appropriate controls. These data confirmed gene expression changes and ChIP peak enrichment for several genes and peaks (***Supplemental Table S2.5***). Together these data suggested that ZNF286A and ZNF286B proteins are involved in negative regulation of cell proliferation and positive regulation of neuron differentiation and function in neuroblastoma cells.

Analysis of ZNF286A binding motifs reveals high enrichment for RE-1 motifs

We examined ZNF286A bound regions identified from the ChIP-seq for evidence of a potential DNA recognition motif, and potentially, of evidence for TF binding sites that might correspond to interacting TF proteins. We used the MEME suite (Bailey et al., 2009; Machanick and Bailey, 2011) to derive motifs from 200 bp regions centered at 826 FDR0 peak summits. The analysis yielded strong enrichment for two sequence motifs. The first motif which was present in 199 peaks (E-value 6.7 e-698), showed a remarkable level of similarity to the consensus motif for the well known neuronal ZNF TF, REST; this protein is also referred to as NRSF, and the binding site called RE-1 or NRSE (***Figure 2.4A***). This motif is centrally located in the ZNF286A peak summit regions and detected as high-scoring matches in a significant fraction of the ZNF286A FDR0 peaks (199 peaks, or 24%). The RE-1 motif is organized as two adjacent "half sites" with variable spacing, which can also bind REST in isolation (Johnson et al., 2007) and a number of ZNF286A peaks that did not include the full RE-1 motifs contained one of these half sites (91 peaks; not shown).

In addition to RE-1 a second, distinct G-rich motif was, detected in 46 peaks (E-value 6.6e-15), was a (***Fig 2.4B***); this motif is very similar to a G-rich motif identified as an alternative binding site for REST in neuronal cells (Johnson et al., 2008; Satoh et al., 2013). Other than these two motifs, no other significant enrichments were identified. Thus, REST motifs are highly enriched within ZNF286A binding regions in SH-EP cells.

To examine whether ZNF286A peaks also overlap with actual REST binding, we examined overlap between ZNF286A peaks and REST binding sites as determined in multiple cell types in the ENCODE project, including SH-SY-5Y which is related to SH-EP (Biedler et al., 1973). Indeed, many of the ZNF286A peak regions coincide with high-scoring NRSF/REST ENCODE binding sites (***Supplemental Table 2.2***). However since SH-EP was not included in ENCODE ChIP experiments, we carried out ChIP with the REST antibody in SH-EP chromatin to permit direct comparison. This analysis confirmed that more than half of the ZNF286A binding peaks – a total of 899 of the 1,502 FDR<5, EF>5 sites (59.9%), and 607 of the 876 peaks detected at FDR0 (69.3%) – can also be occupied by REST in SH-EP cells (***Supplemental Table 2.3***). The ZNF286A peaks that did not overlap with REST binding in SH-EP chromatin were still highly enriched in the

RE-1 binding site (with an E-value score of 5.3 e-57) and the G-rich motif as the second best match (E-value 6.2e-12), thus showing no clear distinction in sequence between the overlapping and non-overlapping peaks.

ZNF286 proteins bind RE-1 and an alternative REST motif

To confirm ZNF286 binding, we designed 40-mer oligonucleotide duplexes from two ZNF286A binding regions. The first region was identified to be centered on a strong consensus RE-1 motif (chr20:36255472-36255692); this region is also occupied by the REST protein in ENCODE ChIP assays and in our SH-EP ChIP (***Supplemental Table S2.3***). We will refer to this 40-mer as "Oligo1" in the following discussion (***Figure 2.4D***). A second 40-mer was designed from a region including the alternative "G-rich" motif (chr20:36201383-36201602); this sequence will be referred to as "Oligo2" (***Figure 2.4E*).** We also generated three unlabeled "competitor" 40-mer duplexes that overlap each labeled Oligo sequence in 20 bp steps (downstream competitor [-20], upstream competitor [+20], and a direct [0] competitor that completely overlaps the labeled oligonucleotide (***Figure 2.4D***; ***Figure 2.4E***). We incubated the biotin-labeled 40-mer with protein extract from SH-EP cells to conduct electrophorectic mobility assays (EMSA) with SH-EP protein extracts; we included competition experiments for the ZNF286A and ZNF286B super-shift experiments, in order to more precisely determine the locations of protein binding within the sequence.

SH-EP protein incubation caused a clear shift in labeled oligonucleotide mobility in the case of both Oligo1 and Oligo2 sequences, indicating that the 40-mers are bound by SH-EP proteins (***Figure 2.4F***). We then completed "super-shift" assays with several different antibodies. For example, as a negative control, we used an antibody to a ubiquitous zinc-finger protein, YY1, which recognizes a distinct motif and did not shift Oligo1 or Oligo2 sequences (***Figure 2.4F*** lanes 3 and 9). As a positive control, we added REST antibody to the mixture to complete a "super-shift", and as expected saw an additional upward shift in complex size for both labeled Oligos (***Figure 2.4F*** lanes 4 and 10).

Furthermore, both ZNF286A and ZNF286B antibodies generated a clear super-shift of both Oligo1 and Oligo2 when incubated with the same mixtures (***Figure 2.4F***

lanes 5, 6, 11, and 12**)**. Upon pre-incubation of the protein mixture with the unlabeled competitor sequences overlapping Oligo1, ZNF286A super-shift was most significantly reduced with upstream and direct competition, indicating a potential binding preference for sequences surrounding the 5'-most REST half site. Competition experiments confirmed binding within the center portion of Oligo2, which includes the match to the G-rich motif (***Figure 2.4G***). Competition experiments with the ZNF286B antibody, on the other hand, suggested a preference for the 3'-most RE-1 half-site and possibly a preference also for the 5' side of the G-rich motif (***Figure 2.4H)***.

ZNF286A binds to KAP1 but not REST in SH-EP cells

Together these data show that ZNF286A protein can occupy binding sites that overlap with the known REST canonical and non-canonical motifs, and that both ZNF286A and ZNF286B proteins can both bind to those motif sequences *in vitro*. One possible explanation for this finding is that the ZNF286 proteins interact with REST, either binding indirectly to the RE-1 sites through REST interaction or binding directly, but adjacent to the REST protein at bound sites. In either case one would expect that the two proteins might be detected as interacting intimately in SH-EP cells. A third possibility would be that the ZNF286 proteins compete with REST for binding and exclude it from engagement with the binding sites; in this case interaction would not be expected.

To examine these possibilities, we used co-immunoprecipitation (co-IP) in SH-EP protein extracts. We carried out reciprocal co-IP experiments, first pulling down REST with the anti-REST antibody, and then staining eluted proteins on a western blot to query the possible presence of ZNF286A, while also carrying out IP with the ZNF286A antibody and staining co-IPed proteins with an antibody for REST (***Figure 2.5A***). In repeated attempts to demonstrate interaction between the proteins, we could find no evidence of direct binding between REST and ZNF286A. This included attempts to trap each protein in immuoprecipitations with each other antibody after cross-linking, which would trap the two sequences on the same DNA fragments or more solidly, through covalently linked protein:protein bonds. Since interactions were also not detected after cross-linking (not shown), we conclude that ZNF286A and REST proteins do not interact

intimately in SH-EP cells. This was surprising considering how closely ZNF286A and REST would be juxtaposed if located on the same binding site at the same time; the co-IPs were repeated using cross-linked SH-EP chromatin in case the binding between the two proteins was particularly weak and disturbed by the experimental manipulations. However, no such binding was observed, further validating the previous co-IP results (not shown).

We also used co-IP experiments to determine whether ZNF286A is able to recruit the repressive co-factor KAP1, as it contains KRAB-A and KRAB-B domains. However the KRAB-A domain of ZNF286A has an unusual structure suggesting that it might not recruit the co-repressor. In particular, it is significantly shorter than the canonical KRAB-A sequence due to a 5 amino acid deletion near the C-terminal KAP1 interaction region, despite the fact that it includes all amino acids known to be critical to KAP1 binding (indicated by the asterisks in ***Table 2.3***). To test this function, we carried out co-IP experiments using nuclear extracts with SH-EP proteins. In addition to the ZNF286A and ZNF286B antibodies, we also included two controls: (1) an antibody designed against another KRAB-ZNF protein with a canonical KRAB domain, ZNF431, as a positive control, and (2) the YY1 antibody as a negative control.

These results indicated clearly that despite its unusual conserved structure, the ZNF286A KRAB-A domain efficiently recruits KAP1 (***Figure 2.5B***). In contrast, as expected, ZNF286B does not recruit the co-repressor. The available literature about KAP1 function would predict that through KAP1 binding, ZNF286A should possess histone deacetylase (HDAC) recruiting activities similarly to REST, which is also known to recruit HDAC to repress target genes (Huang et al., 1999; Ballas and Mandel, 2005).

Opposite expression of ZNF286 proteins and REST in differentiating cells

These data suggested that ZNF286 proteins might compete with REST binding at RE-1 sites. REST is well known to be involved in neuronal differentiation, and was originally described as a negative regulator of neuronally-expressed genes in neuroblasts and non-neuronal cells (Schoenherr and Anderson, 1995). More recently REST has been demonstrated to play a wider role in stem cell development, acting in neurons to stimulate cell proliferation thus preventing neuronal commitment (Jørgensen et al., 2009).

To ask how the ZNF286 proteins are expressed during neuron differentiation, we examined two additional neuroblastoma lines, SH-SY-5Y and SK-N-SH, both of which undergo neuronal differentiation in culture when treated with retinoic acid (RA) (Sidell et al., 1983; Encinas et al., 2000). In fact, although SH-EP does not respond to RA treatment, SH-EP and SH-SY-5Y were both originally derived as subclones of the SK-N-SH line (Biedler et al., 1973). REST transcripts are produced throughout differentiation in these cell models; however, the REST protein is rapidly degraded after synthesis in the differentiating cells, and therefore protein levels decline once the cells are committed to neuronal lineages (Nishimura et al., 1996; Chen et al., 1998).

We treated SH-SY-5Y cells with RA and examined expression of ZNF286A and ZNF286B proteins over the course of 8 days of differentiation. ZNF286A was dramatically increased by the end of the differentiation period (approximately 6.3 fold according to densitometry***, Figure 2.5D; Supplemental Table S2.4***); ZNF286B was also up-regulated but more modestly, to approximately 2.9 fold. REST, on the other hand, was expressed at highest levels in the undifferentiated cells with protein expression dropping dramatically as differentiation nears completion (by day 8 of differentiation; a reduction to a ratio of 0.13 relative to day 0, or a change of -7.69 fold) consistent with the published reports. RA treatment of the parental SK-N-SH cell line yielded very similar results (not shown). The up-regulation of ZNF286A and B proteins, on the one hand, and down-regulation of REST on the other, is consistent with the idea that these proteins exert opposite functional effects on neuron development.

Discussion

ZNF286A is an unusual member of the mammalian KRAB-ZNF gene family, having arisen in marsupial lineages and having been highly conserved throughout most of mammalian history. Unlike most other members of the mammalian KRAB-ZNF gene family, *ZNF286A* was maintained as a unique gene, suggesting an essential function and important interactions that ruled out profligate expansion. The retroposition of the *FOXO3B* pseudogene into the *ZNF286B* sequence, which may or may not have occurred concomitantly with the duplication, is arguably a hominin-specific event; as detailed here this insertion event changed protein structure and also altered regulatory elements within the gene that may explain the novel pattern of *ZNF286B* expression. Therefore, regardless of the precise timing of duplication, the modern version of *ZNF286B* is very likely to be human-specific; our data also indicate that this novel TF gene has been stably fixed in the human population.

Fortunately, these structural innovations provided us with the means to examine and target *ZNF286A* and *ZNF286B* transcripts and proteins specifically. We began with the hypothesis, based on the widely accepted notions of KRAB-KAP1 function (Schultz et al., 2001; Urrutia, 2003), KRAB-containing ZNF286A and KRAB-less ZNF286B should have distinct regulatory functions. Specifically, despite containing an unusual and truncated KRAB domain, the ZNF286A protein does interact with KAP1 in SH-EP neuroblastoma cells and thus, theoretically, should function as a potent transcriptional repressor. On the other hand, ZNF286B – with very similar zinc fingers but no KRAB domain – should be able to bind to ZNF286A binding sites but not to exert a repressive function.

We were thus surprised to find that siRNA knockdowns of *ZNF286A* and *ZNF286B* led to differential expression of very similar sets of genes, a result suggesting cooperative rather than competitive function. In both cases, most DEGs were up-regulated and very highly enriched in functions related to DNA replication, chromosome structure, cell-cycle checkpoint activities and cytokinesis. The results point clearly to the hypothesis that human *ZNF286A* and *ZNF286B* commonly work to negatively regulate cell cycle progression in SH-EP cells.

Intriguingly - especially given the high levels of overlap we discovered between ZNF286A and REST binding regions and motifs - those same functions, and many of the same genes, are known to be *activated* by REST in neuroblastoma and other types of neural cells (Su et al., 2004; Tomasoni et al., 2011; Pozzi et al., 2013). However, co-immunoprecipitation experiments provided no suggestion that REST and ZNF286A or ZNF286B interact, either directly (as tested in native protein extracts), or indirectly through adjacent DNA binding (as tested after trapping the proteins to cross-linked chromatin).

Although other explanations are possible, together these data suggest that genomic binding sites may be occupied *either* by REST *or* by ZNF286 proteins in a particular cell. The results are most consistent with a model in which ZNF286 proteins compete with REST by excluding or inhibiting stable REST binding at occupied RE-1 or G-rich sites. Such a competitive model would be consistent with opposing functions, such as the opposite effects of ZNF286 proteins and REST on genes promoting the progression of mitosis. In further agreement with this either/or competitive model, we found that while REST decreases during neuroblastoma cell differentiation *in vitro,* ZNF286 proteins rise in the same cells.

Given these findings, what can we predict about the functions of and interactions between human ZNF286A and ZNF286B? Certainly, we found no evidence of competitive function and rather -with siRNA knockdown of one gene virtually equivalent to knockdown of the other, and with EMSA suggesting "side-by-side" occupancy at adjacent REST motifs (as shown in ***Figure 2.4***) - our data are indicative of a positive, and likely cooperative function. If this model is correct, KRAB-less ZNF286B could "borrow" KAP1 interaction and regulatory function through cooperative binding with ZNF286A.

However, present evidence does not support the idea that KAP1 binding translates simply into repression of nearby genes. On the one hand, the majority of all DEGs including most DEGs closest to ZNF286A peaks (21 of 32 genes) were up-regulated in siRNA experiments; this is consistent with repression by ZNF286A. Furthermore, several of these have important mitotic functions and their release from repression could explain the overwhelming signal of cell-cycle up-regulation in the overall DEG set. For example,

up-regulated DEG *CDC25B* is a well-known positive regulator of cell-cycle progression with a special role in nervous system development (Agius et al., 2015). However, another 11 DEGs near ZNF286A peaks are down-regulated, and these also include some intriguing candidates. For example, down-regulated DEG, *INA,* with both ZNF286A and REST binding sites located 1876 bp downstream of its TSS, encodes the alpha-internexin component of intermediate neurofilaments with a role in neurite outgrowth; through REST binding to this intronic enancer, *INA* is known to be negatively regulated by REST (Ching and Liem, 2009). Therefore this down-regulated DEG is also likely to have a real regulatory interaction with ZNF286A.

This leaves us with a puzzle regarding the regulatory functions of ZNF286A in living cells. KRAB repressive functions have are well documented, although mostly in studies involving over-expressed, engineered proteins *in vitro* (e.g. Moosmann et al., 1996; Okumura et al., 1997; Peng et al., 2000). But genome-wide binding sites have been determined for only a handful of KRAB-ZNF proteins, and in even fewer cases have binding sites and target gene expression been compared. In the first-published study of this type, the authors concluded that ZNF263, a KAP1- binding KRAB-ZNF protein, can serve either to activate or repress neighboring target genes (Frietze et al., 2010). It is possible then, that like ZNF263 and also like REST (Negrini et al. 2013; Rockowitz et al., 2014; Baldelli and Meldolesi, 2015), the ZNF286 proteins function neither as classic repressive or activating TFs but modulate chromatin environment in a more complex, context-dependent mode. It is further possible to imagine a complex relationship between ZNF286 and REST proteins, dependent upon the balance of isoforms, paralogs and interacting proteins expressed in particular types of cells.

These data highlight ZNF286 proteins as a novel, mammalian-specific competitors of REST. Given suggestions of cooperative ZNF286A:ZNF286B binding and functions developed here, we speculate that cooperation between KRAB- and KRAB-less ZNF286 isoforms – the latter of which is generated in some cellular contexts by human, mouse and other mammalian genomes through alternative splicing – may be a natural property of this ancient protein. Furthermore we speculate that ZNF286B has evolved in recent primate history to provide an independently transcribed and regulated KRAB-less partner for this cooperation. In this model, ZNF286B has evolved to

supplement the ancestral protein's functions in modulating REST's control of cell cycle exit and neuron progenitor cell differentiation in the developing human brain.

References Cited

Agius E, Bel-Vialar S, Bonnet F, Pituello F. **2015**. Cell cycle and cell fate in the developing nervous system: the role of CDC25B phosphatase. *Cell Tissue Res* 359(1):201-13.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. **2009**. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37:W202-8.

Bailey TL, Elkan C. **1994**. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology.* AAAI Press, Menlo Park, California. pp. 28-36.

Ballas N, Mandel G. **2005** The many faces of REST oversee epigenetic programming of neuronal genes. *Curr Opin Neurobiol* 15:500–506.

Best SA, Hutt KJ, Fu NY, Vaillant F, Liew SH, Hartley L, Scott CL, Lindeman GJ, Visvader JE. **2014**. Dual roles for Id4 in the regulation of estrogen signaling in the mammary gland and ovary. *Development* 141(16):3159-64.

Biedler JL, Helson L, Spengler BA. **1973**. Morphology and growth, tumorigenicity, and cytogenetics of human neuroblastoma cells in continuous culture. *Cancer Res* 33:2643–2652.

Chen ZF, Paquette AJ, Anderson DJ. **1998**. NRSF/REST is required in vivo for repression of multiple neuronal target genes during embryogenesis. *Nat Genet* 2:136-42.

Ching GY, Liem RK. **2009**. RE1 silencing transcription factor is involved in regulating neuron-specific expression of alpha-internexin and neurofilament genes. *J Neurochem* 109(6):1610-23.

Choo Y, Klug A. **1994**. Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage. *Proc Natl Acad Sci USA* 91:11163-11167.

Chung HR, Schafer U, Jackle H, Bohm S. **2002**. Genomic expansion and clustering of ZAD-containing C2H2 zinc-finger genes in Drosophila. *EMBO Rep* 3:1158-1162.

Encinas M, Iglesias M, Liu Y, Wang H, Muhaisen A, Ceña V, Gallego C, Comella JX. **2000**. Sequential Treatment of SH-SY5Y Cells with Retinoic Acid and Brain-Derived Neurotrophic Factor Gives Rise to Fully Differentiated, Neurotrophic Factor-Dependent, Human Neuron-Like Cells. *J Neurochem* 75: 991–1003.

ENCODE Project Consortium. **2012**. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57-74.

Friedman JR, Fredericks WJ, Jensen DE, Speicher DW, Huang X, Neilson EG, Rauscher FJ 3rd. **1996**. KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes Dev* 10(16):2067-78.

Frietze S, Lan X, Jin VX, Farnham PJ. **2010**. Genomic targets of the KRAB and SCAN domain-containing zinc-finger protein 263. *J Biol Chem* 285(2):1393-1403.

Gassmann M, Grenacher B, Rohde B, Vogel J. **2009**. Quantifying western blots: pitfalls of densitometry. *Electrophoresis* 30:1845-1855.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang J. **2004**. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5(10):R80

Guardavaccaro D, Frescas D, Dorrello NV, Peschiaroli A, Multani AS, Cardozo T, Lasorella A, Iavarone A, Chang S, Hernando E, Pagano M. **2008**. Control of chromosome stability by the beta-TrCP-REST-Mad2 axis. *Nature* 452(7185):365-9.

Hamelink C, Hahm SH, Huang H, Eiden LE. **2004**. A restrictive element 1 (RE-1) in the VIP gene modulates transcription in neuronal and non-neuronal cells in collaboration with an upstream tissue specific element. *J Neurochem* 88:1091–1101.

Hamilton AT, Huntley S, Tran-Gyamfi M, Baggott DM, Gordon L, Stubbs L. **2006**. Evolutionary expansion and divergence in the ZNF91 subfamily of primate-specific zinc finger genes. *Genome Res* 16(5):584-94.

Huang DW, Sherman BT, Lempicki RA. **2009a**. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc* 4(1):44-57.

Huang DW, Sherman BT, Lempicki RA. **2009b**. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37(1):1-13.

Huang Y, Myers SJ, Dingledine R. **1999**. Transcriptional repression by REST: recruitment of Sin3A and histone deacetylase to neuronal genes. *Nat Neurosci* 2(10):867-72.

Huntley S, Baggott DM, Hamilton AT, Tran-Gyamfi M, Yang S, Kim J, Gordon L, Branscomb E, Stubbs L. **2006**. A comprehensive catalog of human KRAB-associated zinc finger genes: insights into the evolutionary history of a large family of transcriptional repressors. *Genome Res* 16(5):669-77.

Johnson DS, Mortazavi A, Myers RM, Wold B. **2007**. Genome-wide mapping of in vivo protein-DNA interactions. *Science* 316:1497-1502.

Jones NC, Pevzner PA. **2006**. Comparative genomics reveals unusually long motifs in mammalian genomes. *Bioinformatics* 22(14):e236-42.

Jørgensen HF, Terry A, Beretta C, Pereira CF, Leleu M, Chen ZF, Kelly C, Merkenschlager M, Fisher AG. **2009**. REST selectively represses a subset of RE1-containing neuronal genes in mouse embryonic stem cells. *Development* 136(5):715-21.

Kim SS, Chen YM, O'Leary E, Witzgall R, Vidal M, Bonventre JV. **1996**. A novel member of the RING finger family, KRIP-1, associates with the KRAB-A transcriptional repressor domain of zinc finger proteins. *Proc Natl Acad Sci USA* 93(26):15299-304.

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

Lim FL, Soulez M, Koczan D, Thiesen HJ, Knight JC. **1998**. A KRAB-related domain and a novel transcription repression domain in proteins encoded by SSX genes that are disrupted in human sarcomas. *Oncogene* 17:2013-2018.

Liu H, Chang LH, Sun Y, Lu X, Stubbs L. **2014**. Deep vertebrate roots for mammalian zinc finger transcription factor subfamilies. *Genome Biol Evol* 6(3):510-25.

Livak KJ, Schmittgen TD. **2001**. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods* 25:402–408.

Machanick P, Bailey TL. **2011**. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* 27(12):1696-7.

Margolin JF, Friedman JR, Meyer WKH, Vissing H, Thiesen HJ, Rauscher FJ. **1994**. Kruppel-associated boxes are potent transcriptional repression domains. *Proc Natl Adac Sci USA* 91:4509-4513.

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. **2010**. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 28(5):495-501.

Moosmann P, Georgiev O, Le Douarin B, Bourquin JP, Schaffner W. **1996**. Transcriptional repression by RING finger protein TIF1 beta that interacts with the KRAB repressor domain of KOX1. *Nucleic Acids Res* 24(24):4859-67.

Negrini S, Prada I, D'Alessandro R, Meldolesi J. **2013**. REST: an oncogene or a tumor suppressor? *Trends Cell Biol* 23(6):289-95.

Nishimura E, Sasaki K, Maruyama K, Tsukada T, Yamaguchi K. **1996**. Decrease in neuron-restrictive silencer factor (NRSF) mRNA levels during differentiation of cultured neuroblastoma cells. *Neuroscience Letters* 211(2):101-104.

Nowick K, Fields C, Gernat T, Caetano-Anollés D, Kholina N, Stubbs L. **2011**. Gain, Loss and Divergence in Primate Zinc-Finger Genes: A Rich Resource for Evolution of Gene Regulatory Differences between Species. *PLoS One* 6(6):e21553.

Nowick K, Stubbs L. Lineage-specific transcription factors and the evolution of gene regulatory networks. **2010**. *Brief Funct Genomics* 9(1):65-78.

Okumura K, Sakaguchi G, Naito K, Tamura T, Igarashi H. **1997**. HUB1, a novel Krüppel type zinc finger protein, represses the human T cell leukemia virus type I long terminal repeat-mediated expression. *Nucleic Acids Res* 25(24):5025-32.

Ovcharenko I, Loots GG, Hardison RC, Miller W, Stubbs L. **2004**. zPicture: dynamic alignment and visualization tool for analyzing conservation profiles. *Genome Res* 14(3):472-7.

Peng H, Begg GE, Harper SL, Friedman JR, Speicher DW, Rauscher FJ 3rd. **2000**. Biochemical analysis of the Kruppel-associated box (KRAB) transcriptional repression domain. *J Biol Chem* 275(24):18000-10.

Pozzi D, Lignani G, Ferrea E, Contestabile A, Paonessa F, D'Alessandro R, Lippiello P, Boido D, Fassio A, Meldolesi J, Valtorta F, Benfenati F, Baldelli P. **2013**. REST/NRSF-mediated intrinsic homeostasis protects neuronal networks from hyperexcitability. *EMBO J* (22):2994-3007.

Raj B, O'Hanlon D, Vessey JP, Pan Q, Ray D, Buckley NJ, Miller FD, Blencowe BJ. **2011**. Cross-Regulation between an Alternative Splicing Activator and a Transcription Repressor Controls Neurogenesis. *Mol Cell* 43:843–850.

Raney BJ, Dreszer TR, Barber GP, Clawson H, Fujita PA, Wang T, Nguyen N, Paten B, Zweig AS, Karolchik D, Kent WJ. **2014**. Track Data Hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics* 30(7):1003-5.

Reick M, Garcia J, Dudley C, McKnight SL. **2001**. NPAS2: An Analog of Clock Operative in the Mammalian Forebrain. *Science* 293(5529):506-509.

Reddy UR, Venkatakrishnan G, Roy AK, Chen J, Hardy M, Mavilio F, Rovera G, Pleasure D, Ross AH. **1991**. Characterization of Two Neuroblastoma Cell Lines Expressing Recombinant Nerve Growth Factor Receptors. *J Neurochem* 56(1):67–74.

Rockowitz S, Lien WH, Pedrosa E, Wei G, Lin M, Zhao K, Lachman HM, Fuchs E, Zheng D. **2014**. Comparison of REST Cistromes across Human Cell Types Reveals Common and Context-Specific Functions. *PLOS Comput Biol* 10(6):e1003671.

Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, Kirkup VM, Wong MC, Maddren M, Fang R, Heitner SG, Lee BT, Barber GP, Harte RA, Diekhans M, Long JC, Wilder SP, Zweig AS, Karolchik D, Kuhn RM, Haussler D, Kent WJ. **2012**. ENCODE Data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res* 41(D1):D56-63.

Satoh T, Inagaki T, Liu Z, Watanabe R, Satoh AK. **2013**. GPI biosynthesis is essential for rhodopsin sorting at the trans-Golgi network in Drosophila photoreceptors. *Development* 140(2):385-394.

Schoenherr CJ, Anderson DJ. **1995**. The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science* 267(5202):1360-63.

Schoenherr CJ, Paquette AJ, Anderson DJ. **1996**. Identification of potential target genes for the neuron-restrictive silencer factor. *Proc Natl Acad Sci USA* 93:9881–988693(18):9881–9886.

Schultz DC, Friedman JR, Rauscher FR. **2001**. Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2α subunit of NuRD. *Genes Dev* 15(4): 428-443.

Sharma P, Knowell AE, Chinaranagari S, Komaragiri S, Nagappan P, Patel D, Havrda MC, Chaudhary J. **2013**. Id4 deficiency attenuates prostate development and promotes PIN-like lesions by regulating androgen receptor activity and expression of NKX3.1 and PTEN. *Mol Cancer* 12:67.

Sidell N, Altman A, Haussler MR, Seeger RC. **1983**. Effects of retinoic acid (RA) on the growth and phenotypic expression of several human neuroblastoma cell lines. *Exp Cell Res* 148(1):21-30.

Su X, Kameoka S, Lentz S, Majumder S. **2004**. Activation of REST/NRSF target genes in neural stem cells is sufficient to cause neuronal differentiation. *Mol Cell Biol* 24(18):8018-25.

Tan HY, Ng TW. **2008**. Accurate step wedge calibration for densitometry of electrophoresis gels. *Opt Commun* 281:3013-3017.

Thiel G, Lietz M, Cramer M. **1998**. Biological Activity and Modular Structure of RE-1-silencing Transcription Factor (REST), a Repressor of Neuronal Genes. *J Bio Chem* 273:26891-26899.

Tomasoni R, Negrini S, Fiordaliso S, Klajn A, Tkatch T, Mondino A, Meldolesi J, D'Alessandro R. **2011**. A signaling loop of REST, TSC2 and β-catenin governs proliferation and function of PC12 neural cells. *J Cell Sci* 124(18):3174-86.

Urrutia R. **2003**. KRAB-containing zinc-finger repressor proteins. *Genome Biol* 4(10): 231.

Vissing H, Meyer WK, Aagaard L, Tommerup N, Thiesen HJ. **1995**. Repression of transcriptional activity by heterologous KRAB domains present in zinc finger proteins. *FEBS Lett* 369(2-3):153-7.

Witzgall R, O'Leary E, Leaf A, Önaldi D, Bonventre JV. **1994**. The Kruppel-associated box-A (KRAB-A) domain of zinc finger proteins mediates transcriptional repression. *Proc Natl Acad Sci USA* 91:4514-4518.

Wuttke DS, Foster MP, Case DA, Gottesfeld JM, Wright PE. **1997**. Solution structure of the first three zinc fingers of TFIIIA bound to the cognate DNA sequence: determinants of affinity and sequence specificity. *J Mol Biol* 273:183-206.

Yun K, Mantani A, Garel S, Rubenstein J, Israel MA. **2004**. Id4 regulates neural progenitor proliferation and differentiation in vivo. *Development* 131(21):5441-8.

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

FIGURES

Figure 2.1. Duplicon located at human chromosome 17. A *Blastz* nucleotide alignment dotplot assembled by zPicture (Ovcharenko et al., 2004) of the ~600 kb long duplication that spans chr17:15300000-15900000 and chr17:18300000-18900000 and includes 12 neighboring genes, all of which are inverted in telomeric-to-centromeric order relative to ZNF286A. Dots correspond to identity between the sequences represented on the X- and Y-axes. The coordinates representing the locations of *ZNF286A* and *ZNF286B* are indicated by the space within the dotted lines, along with representations of the corresponding genes at the X- and Y-axis boundaries of the box. The non-duplicated region of *ZNF286B* that differs from *ZNF286A* is colored red, corresponding to a gap in the dotplot.

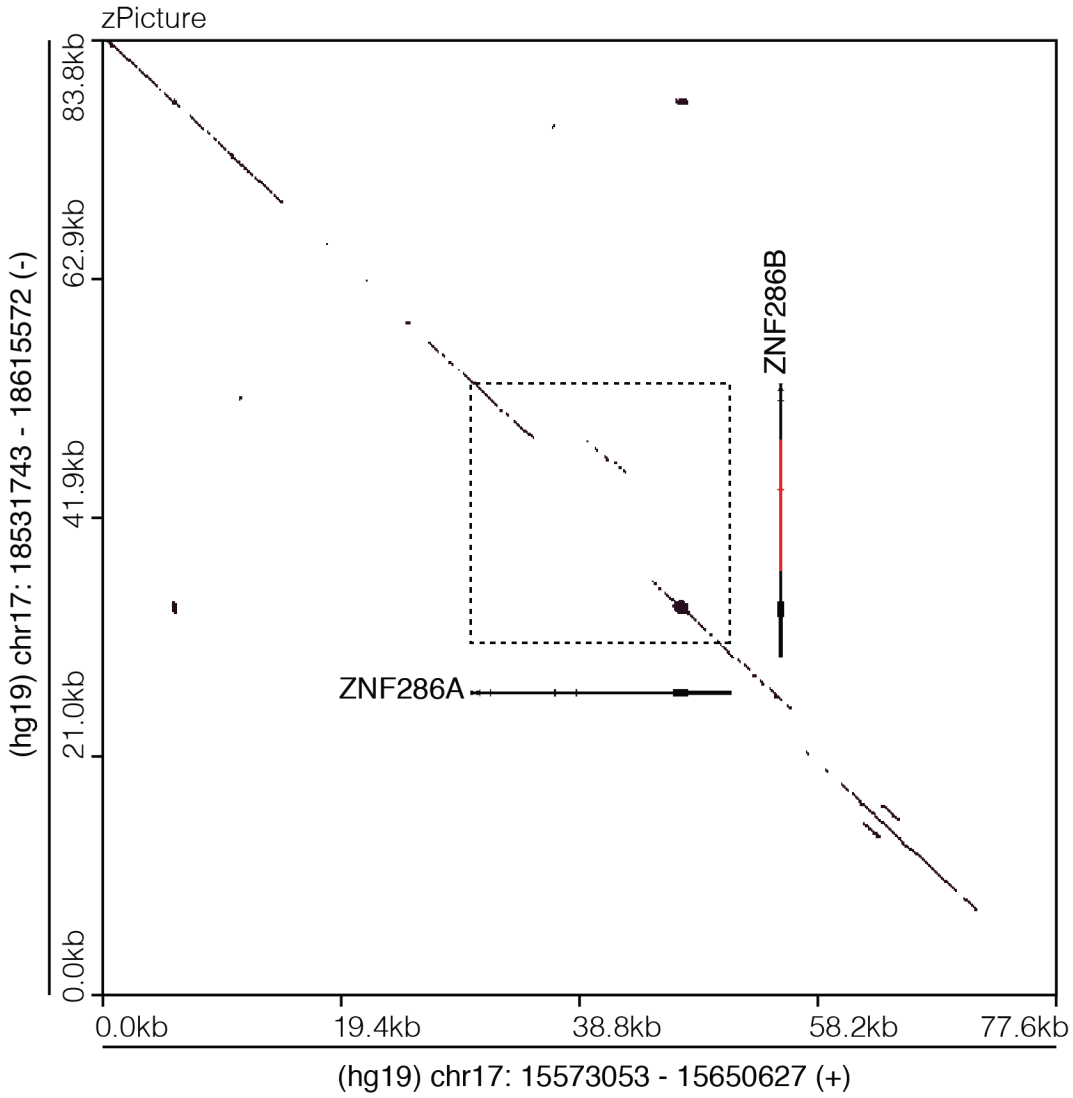


Figure 2.2. Insertion of FOXO3B into ZNF286B after duplication. **(A)** Domains regions for *ZNF286A* and *ZNF286B* are represented with letters above each exon: 5' UTR ("5"), pre-KRAB region ("P"), KRAB-A ("A") and KRAB-B ("B") domains, the tether region ("T"), zinc finger domain ("F"), and the 3' UTR ("3"). In addition, roman numerals below each gene represent the locations of primers, antibody epitopes, and siRNA targets used in this study: (i) *ZNF286A* forward primer; (ii) *ZNF286A* reverse primer; (iii) *Abgent* RB21720; (iv) *Qiagen* SI04135600; (v) *Qiagen* SI03227448; (vi) *Abgent* RB21665; (vii) SI05457704; (viii) *ZNF286B* forward primer; (ix) *ZNF286B* reverse primer; (x) *Qiagen* SI05457697. The *FOXO3B* sequence (colored in red) was inserted into *ZNF286B* at the location of a hAT-Charlie Charlie1a DNA repeat element, splitting it in two and end-capping the insertion. *FOXO3B* contains 98% identity with the originating *FOXO3* gene. The figure shows *ZNF286A*, the hypothetical duplicate deletion product prior to *FOXO3B* insertion, and *ZNF286B* containing the pseudogene insertion in red. **(B)** Representations of both *ZNF286A* and **(C)** *ZNF286B* from the UCSC Genome Browser (ENCODE Project Consortium, 2012; Raney et al., 2014), along with the "*Conservation: Vertebrate Multiz Alignment & Conservation (100 Species)*" track, show that the highest levels of conservation are located in the finger exons and inter-exon region between the promoter and KRAB-A domain. The *FOXO3B* region has high levels of conservation carried over from the more ancient *FOXO3* gene. In addition, histone modification in both genes is represented from the "*ENCODE Regulation Super-track*", showing where H3K4Me1 and H3K27Ac histone modification is suggestive of enhancer activity. These histone modifications have similarly been conserved between *ZNF286A* and *ZNF286B*, but it is important to note that the *FOXO3B* region of *ZNF286B* has histone modifications that may indicate novel regulatory activity following the insertion event. Each histone mark uses colored transparent overlays to display data from a number of cell lines, outlined in the figure key at the bottom of the figure.

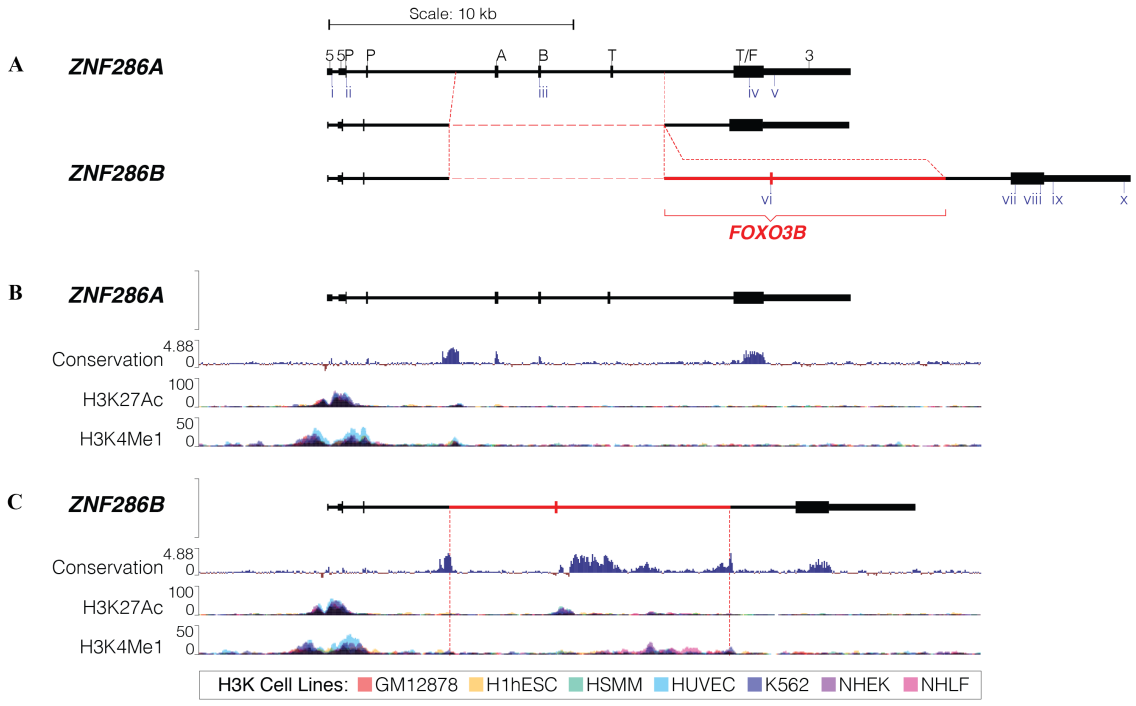


Figure 2.3. Relative expression of ZNF286 transcripts in adult and fetal human tissues. The expression of both **(A)** *ZNF286A* and **(B)** *ZNF286B* transcripts in various human tissues was determined by performing QPCR on each sample using sequence-specific primers for each gene. The $2^{-\Delta Ct}$ values for each sample were calculated in relation to *GAPDH* positive controls (Livak and Schmittgen, 2001) and then normalized against the highest-expressed tissue for each primer set, in order to minimize primer-specific effects when comparing both genes. Samples are ordered from highest-to-lowest *ZNF286A* expression, with fetal/placental samples separated. The relative expression of *ZNF286A* and *ZNF286B* is not identical: *ZNF286A* expresses higher in adult brain than in fetal brain, while *ZNF286B* is expressed significantly higher in fetal brain compared to adult brain. **(C)** Using custom-designed siRNA oligonucleotides, knockdown of *ZNF286A* and *ZNF286B* was conducted in SH-EP neuroblasts over the course of 32 hours. Scrambled siRNA oligonucleotides ("siScr") and transfection reagent with no siRNA ("0si") were used as negative controls. This knockdown was validated using QPCR, with RNA expression independently calculated as a percentage siScr (which is set as 100% for each primer set), revealing 70-80% knockdown of target RNA after 32 hours of treatment, with minimal comparable effect of the paralog gene, confirming that the siRNA knockdown is specific to ZNF286A and ZNF286B. **(D)** Western blots detected protein expression for ZNF286A, ZNF286B, and REST in the knockdown samples, using control α-TBP antibody (*Santa Cruz* Cat. sc-273). Decrease in levels of the appropriate protein was observed for each knockdown, validating knockdown and additionally demonstrating antibody specificity for the custom αZNF286A and αZNF286B antibodies.

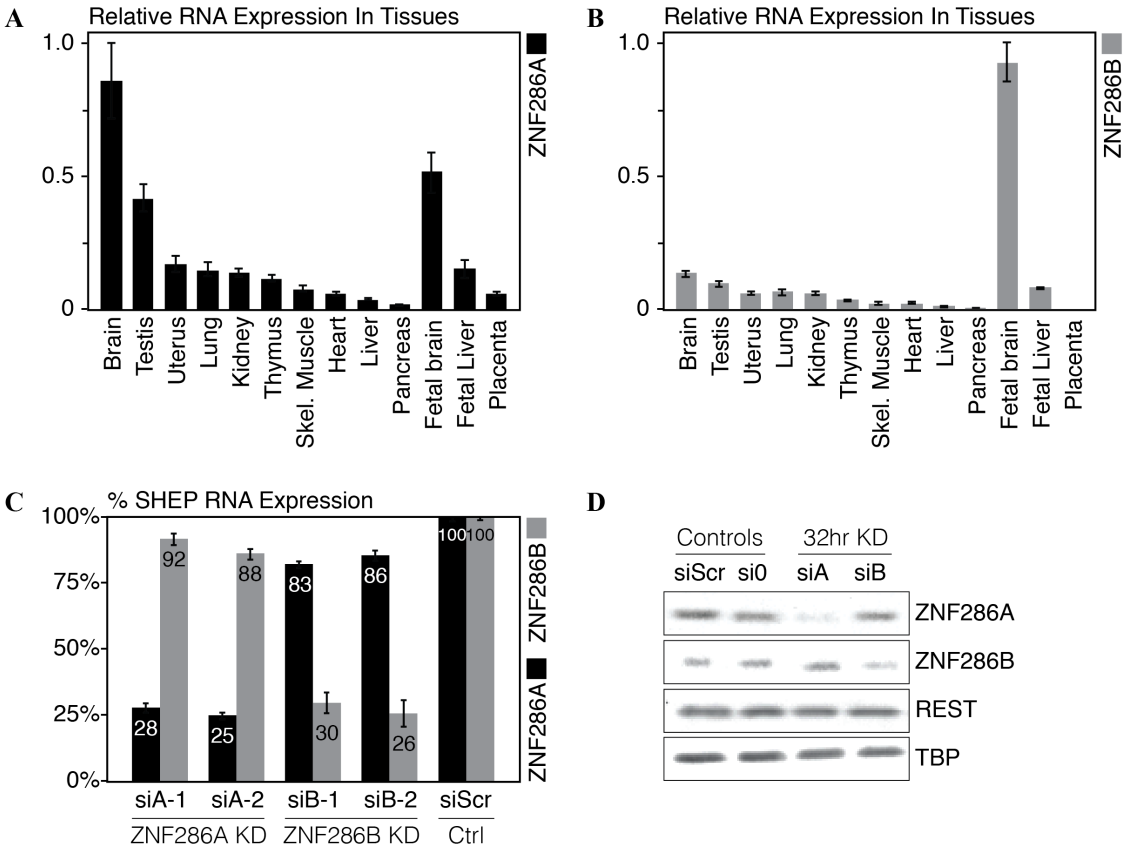


Figure 2.4. ZNF286 binding motif analysis and EMSA. ZNF286 binding motif analysis and EMSA. **(A)** The consensus REST binding motif is split into two "half-sites", both of which REST is capable of binding (Schoenherr et al., 1996; Thiel et al., 1998). **(B)** ChIP-seq was performed on ZNF286A using ChIP-grade polyclonal rabbit antibodies; the highest-ranked binding motif prediction (called "Motif1") was identified using the MEME suite (Bailey and Elkan, 1994) was found to match with the REST consensus binding site. **(C)** A secondary G-rich binding motif was predicted by MEME (called Motif2). **(D)** An example of Motif1 was located in the human genome (chr20:36255472-36255692) and used as the basis for the EMSA probe ("Oligo1") that included a 10 bp buffer on each side of the binding motif. Five unbiotinylated probes were designed, acting as direct binding competitor "0", a direct downstream competitor "-20", a more distant downstream competitor "-30", upstream competitor "+20", and a more distant upstream competitor "+30". **(E)** Similarly, an example of Motif2 was identified in the genome (chr20:36201383-36201602) and used as the template for an EMSA probe ("Oligo2"), with the five competition probes shown. **(F)** EMSA using Oligo1 and Oligo2, after being super-shifted using specific antibodies shows that all three proteins bind to the predicted binding motifs in SH-EP, with YY1 serving as a negative control. **(G)** When the competitor probes were included in the mixture with the biotinylated probe and super-shifted with ZNF286A, the signal strength decreased as predicted for both oligos. **(H)** Likewise, super-shifted ZNF286B signal strength was affected by exposure to the five competition probes.

Figure 2.4 (Continued)

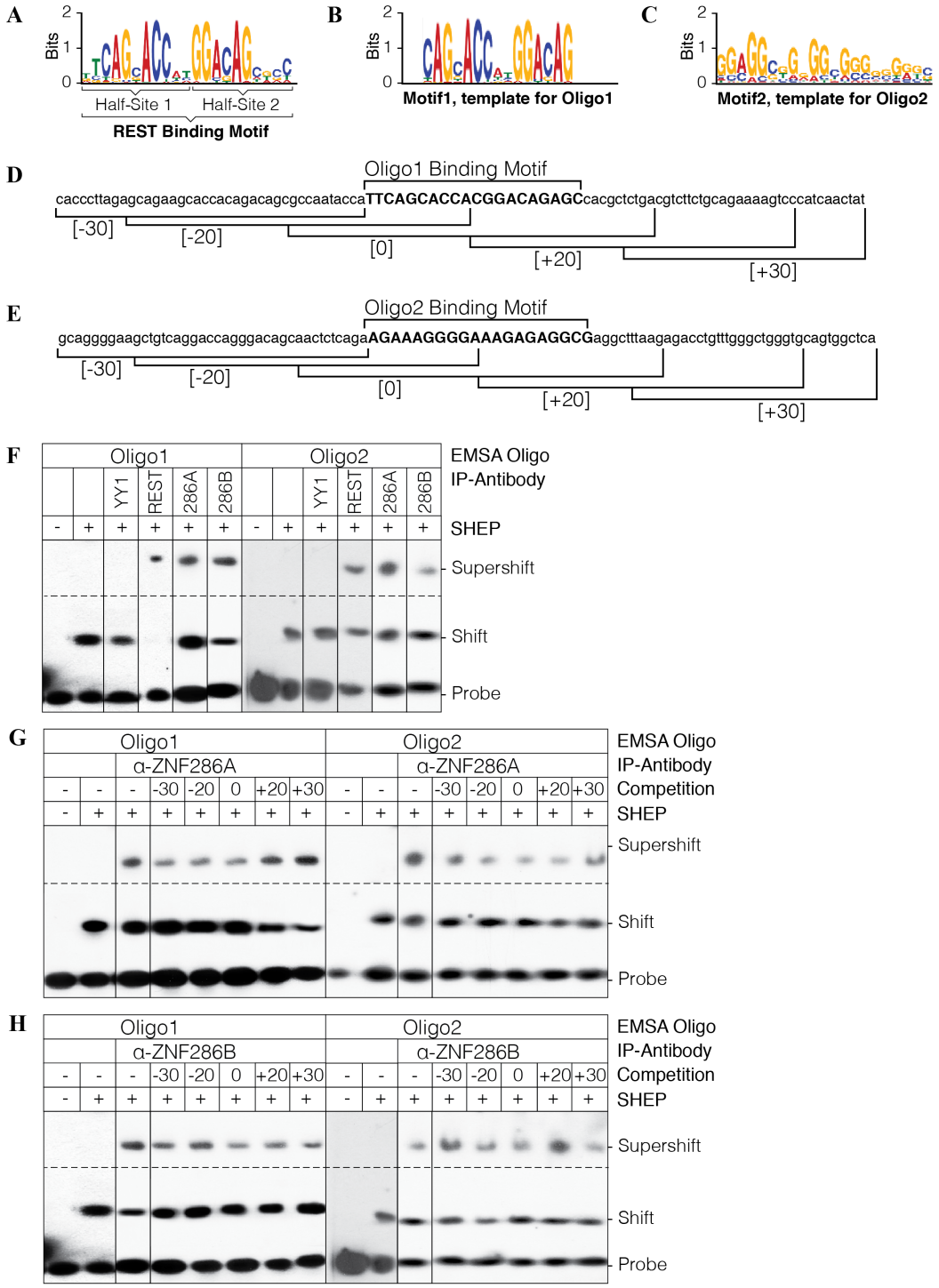
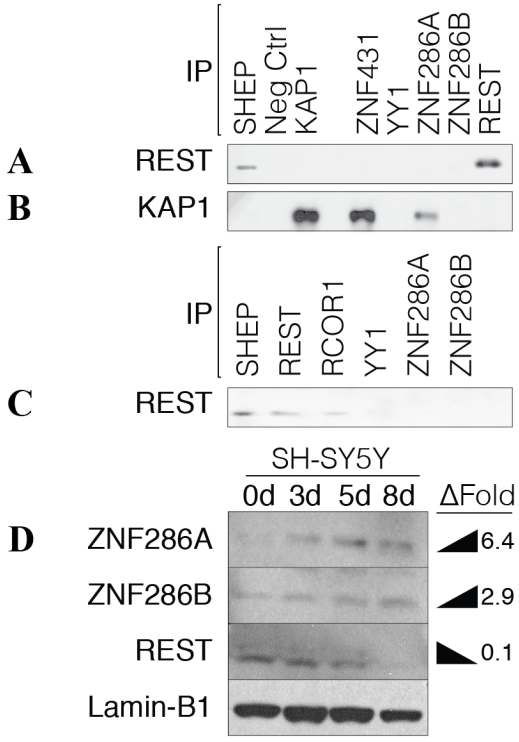


Figure 2.5. Validation of function by co-immunoprecipitation and RA-differentiation. Co-IP with 14 µg of input SH-EP nuclear extract ("*SH-EP*") used for the IP, negative control using *Pierce* coupling resin ("*Neg. Ctrl.*") and IPs with Rb-α-KAP1, Rb-α-ZNF431, Rb-α-YY1, Rb-α-ZNF286A, Rb-α-ZNF286B, and Rb-α-REST. **(A)** REST-pulldown for each IP reaction was detected by Western blotting with Ms-α-REST (1:400) – REST is only detected in IPs for REST, confirming that there are no protein-protein interactions between REST and either ZNF286 protein. **(B)** KAP1-pulldown for each IP reaction was detected by Western blotting with Ms-α-KAP1 (1:1000) – KAP1 is confirmed in IPs for KAP1 and ZNF431 (a consensus-KRAB containing zinc finger transcription factor) as positive controls, while absent in YY1 negative control. KAP1 co-IPs with ZNF286A, confirming that it has a functional KRAB-domain, but is absent from ZNF286B and REST pull-downs as predicted, failing to demonstrate any protein-protein interactions with those proteins and KAP1. **(C)** Positive control co-IP using Ms-α-REST (1:600) to confirm REST-RCOR1 interactions, demonstrating the efficiency of the REST antibody used in this study. Input SH-EP nuclear extract used for the IP ("*SH-EP*"), along with IPs using Rb-α-REST , Rb-α-RCOR1, Rb-α-YY1, Rb-α-ZNF286A, and Rb-α-ZNF286B. The REST-RCOR1 interaction was detected, confirming Co-IP using the Ms-α-REST antibody. **(D)** ZNF286A and ZNF286B (top two panels) increase over the course of RA-differentiation in SH-SY5Y neuroblastoma cells (which are fully differentiated after 7 days of retinoic acid (RA) exposure); densitometry confirmed a rise at 8 days compared to 0 days of RA treatment of 6.3 and 2.9-fold increase in protein levels respectively (***Suppl. Table S4***). In contrast, REST protein (third panel) decreased significantly (7.7 fold) over the same period of differentiation. The total protein amount for each time point was normalized to detected levels of Lamin-B1 (bottom panel).v



TABLES

Table 2.1. Alignments of the four DNA-binding amino acids ("fingerprints") of each of the ten ZNF286 fingers shows that they have very high identity between organisms, as well as between ZNF286A and ZNF286B in humans. Non-identical nucleotides are color-coded in red. The high conservation of ZNF286 fingerprints implies that the DNA-binding sites for these proteins are also conserved, as their binding affinity is fingerprint-specific. As such, the binding motifs for ZNF286A and ZNF286B is predicted to be the same.

Table 2.1. Alignment of DNA-contacting residues in ZNF286A and ZNF286B fingerprints

Finger Number:	*01*	*02*	*03*	*04*	*05*	*06*	*07*	*08*	*09*	*10*
ZNF286A - HUMAN	YHVR	HRNK	ESST	RSHQ	HSAK	HCSK	QSHQ	RSNK	HSAQ	CSSR
ZNF286B - HUMAN	CHVQ	HRNK	ESST	RSHQ	HSAK	HCSK	QSHQ	QSNK	HSAQ	CSSR
Znf286 - CHIMPANZEE	YHVR	HRNK	ESST	RSHQ	HSAK	HCSK	QSHQ	RSNK	HSAQ	CSSR
Znf286 - MACAQUE	YHVR	HRNK	ESST	RSHQ	HSAK	HCSK	QSHQ	RSNK	HSAQ	CSSR
Znf286 - ORANGUTAN	YHVR	HRNK	ESST	RSHQ	HSAK	HCSK	QSHQ	RSNK	HSAQ	CSSR
Znf286 - PANDA	YHVR	HRNK	ESSI	RSHQ	HSAK	HCSK	QSHQ	RSNK	HSAQ	CSSR
Znf286 - CATTLE	YHVR	HRNK	ESSI	RSHQ	HSAK	HCSK	QSHQ	RSNK	HSAQ	CSSR
Zfp286 - MOUSE	YRVR	HRNK	ESSV	RSHQ	HSAK	HCSK	QSHQ	RSNK	HSAQ	SSSR
Znf286 - OPOSSUM	YHVR	HKNK	ESAE	RTHQ	HSAK	HCSK	QSHQ	RSNK	HSAQ	CSSR
	*	* **	**	****	****	****	****	***	****	***

Table 2.2. The Database for Annotation, Visualization and Integrated Discovery (DAVID) is a functional annotation tool used for analyzing gene enrichment, and describes biological patterns found in a user-inputted gene list (Huang et al., 2009a; 2009b). This table summarizes data generated by DAVID, including the functions of genes (GO term clusters) that Illumina microarrays indicate are enriched after *ZNF286A* or *ZNF286B* knockdown. Here, results are shown at enrichment > 2.0 levels, although results were comparable at high, moderate, and low enrichment stringencies (>1, >1.5, and >2 respectively).

Table 2.2. Functionally enriched categories in DEGs detected at ≥2 fold change after ZNF286A and ZNF286B gene knockdown.

siRNA Effect	*ZNF286A*	*ZNF286B*
Up-Regulated	DNA replication	DNA replication
	DNA metabolic process	Cell cycle
	Cell cycle	DNA metabolic process
	Chromosome organization	Chromosome organization
	Steroid biosynthesis	DNA recombination
	ATP binding	Condensed chromosome
	Cell cycle checkpoint	Steroid biosynthesis
	Condensed chromosome	ATP binding
	DNA recombination	Homologous recombination
	DNA helicase activity	Interphase of mitotic cell cycle
	Chromosome segregation	DNA helicase activity
	DNA integrity checkpoint	Double strand break repair
	Response to endogenous stimulus	Microtubule-based process
		Cell cycle checkpoint
		Response to endogenous stimulus
Down-Regulated	Regulation of steroid biosynthesis	Secreted/signal
	Extracellular region	Response to hormone stimulus
	Gonad development	Ureteric bud development
		Negative regulation of cell proliferation

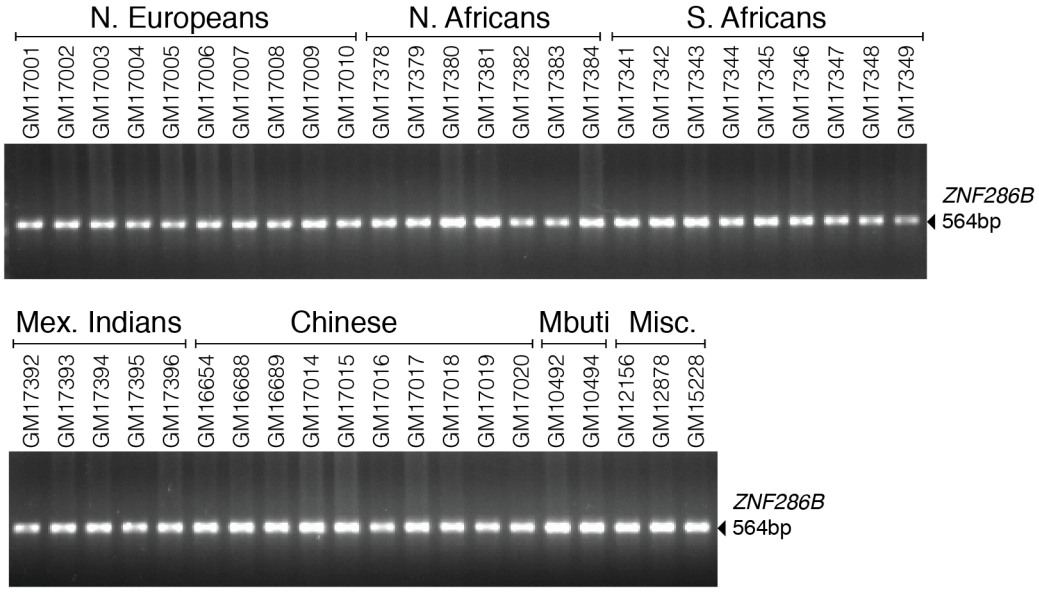
Table 2.3. Comparison table of the amino acids making up the KRAB-A domains among ZNF286 orthologs in human, common chimpanzee, rhesus macaque, Sumatran orangutan, giant panda, cattle, house mouse, and opossum. These domains are further compared to the human KOX1 (also known as ZNF10) domain, with non-identical amino acids colored red. In relation to the consensus KRAB domain present in KOX1 which contains the consensus amino acids at the top of the table, present in >90% of KRAB-A domains (Witzgall et al., 1994), alignments of the ZNF286A KRAB-A domain show that it is somewhat unique among KZNF proteins in that it possesses a four amino acid deletion that is conserved across all mammalian species. Only targeted deletions in the DV and MLE amino acids of the consensus KRAB-A (underlined) have been shown to negatively impact KRAB-A interactions with KAP-1 (Margolin et al., 1994; Lim et al., 1998), so the deletion present in ZNF286 proteins is not predicted to have substantial impact on domain function.

Table 2.3. KRAB-domain alignment of ZNF286 proteins and the canonical domain of KOX1

CONSENSUS KRAB-A:	F DV F EEW L Q LY VMLEN L
KOX1 - HUMAN	MVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSL
ZNF286A - HUMAN	TVTFKDVAMDFTPEEWGKLDPAQ----RDVMLENYRNLVSL
Znf286 - CHIMPANZEE	TVTFKDVAMDFTPEEWGKLDPAQ----RDVMLENYRNLVSL
Znf286 - MACAQUE	TVTFKDVAMDFTPEEWGKLDPAQ----RDVMLENYRNLVSL
Znf286 - ORANGUTAN	KVTFKDVAMDFTPEEWGKLDPAQ----RDVMLENYRNLVSL
Znf286 - PANDA	SVTFKDVAMDFTPEEWGKLDPTQ----REVMLENYRNLVSL
Znf286 - CATTLE	SVTFKDVAMDFTPEEWGRLDPTQ----RDVMLENYRNLVSL
Zfp286 - MOUSE	AMTFKDVAMDFTPEEWGKLDPAH----RAVMLENYRNLVSL
Znf286 - OPOSSUM	-VTFKDVAMDFTPEEWGQLDPAQ----RDVMLENYRNLVSL
	***** ******* ** * ****** *****

SUPPLEMENTARY FIGURES

Supplemental Figure S2.1. Human variation of ZNF286B. *FOXO3B*- and *ZNF286B*-based PCR primers used in Nowick et al. (2011) were used to screen genomic DNA from a panel of lymphoblastoid cells from humans of diverse American, African, Asian, and European origins. All 46 cell lines used are outlined in ***Supplementary Table S2.1B***. The 564 bp product is consistently produced in all samples, demonstrating the fixation of *ZNF286B* among the human population.



SUPPLEMENTARY TABLES

Supplemental Table S2.1. Primer and probe sequences

Gene Target	FWD Primer (5′-to-3′)	RVS Primer (5′-to-3′)
ARAP1	CTCCCACGTGGCTGCCATCG	GCGGTCAGGCTGCTCTGAGC
EGFR	CTGTGCCCAGCAGTGCTCCG	GGCAGACCAGGCAGTCGCTC
GAPDH	*Qiagen* Cat. QT01192646	
GRIA4	TGCTCCATGGGGCCAGGGAA	AGCAGCTGTGTCATTGCCAAGAGT
HOMER1	CGCTGTCTCCTCCGCAAAGCA	CCACCCCCACCCCCAGATCC
KCNH7	TTGTGCGCGTGGCCAGGAAA	TGCAAGCCAGCCAGTGAGCA
MIR9-3	TGTCTGTCCATCCCCTCTG	GCTTTATGACGGCTCTGTGG
NCS1	ACCCGTCCATTGTGCAGGCG	AGCCGGCTGTGTGTGGATGC
NRG1	GGACTCGCTGCTCACCGTGC	TTCCGGCCCGTCTCCAGAGG
NRXN3	TCCACCTCTTCCTCGCCGGG	TAGCGCCAGCGTGTCCTCCT
REST	TGGGGCCTGCTCCCACAGAG	GCAGGCCGTATCTGGGCACC
SNTG1	CGGGAGCAAGACCCCCTCCA	CCGTCGGTCCAGCAGGTCAC
TLE3	GACGCCCCAACTCCAGGCAC	CGAAGGGCGCCGCATAGGAG
ZNF286A	TGTGAGGCCCGGGATGGGAG	TGCCTCAGTCCTACGGCGCT
ZNF286B	GCCATTCAGTGCATATTCAACACCAG	TACACTCATAGGGTTTCTCTCCAGTGTGA

Peak Target	FWD Primer (5′-to-3′)	RVS Primer (5′-to-3′)
ARAP1	ACCCTCGGCCTAGCTCAGGC	GCTGCCCCTTGCCCTCTGTG
EGFR	ACTGCGTCCTCCTGGAAGAGGT	TGCACATCTGAGGAAGTGGGGAGG
GRIA4	ACTGCCTACATTGTCCCCACATCCT	TGTGAGCTAGCCAAGTCAGGCACT
GRM1	CGGCGAGAGCAACCCGGAAG	TTGTTTCCCGCGGTGGGTGG
HOMER1	GAGGGCACGGTGAGGGCAAC	ACACTTTACTGGTCAGACTGCTGCA
KCNH7	TGCAGCTGTGCTTGGGAGACG	TGTCCCTGGTTGGTTGTAGGCAGA
MIR9-3	CCCTGAGGGAAACGCGGAGC	TCACCGGCAAACTCGCTGGG
NCS1	TGCAGCCATTGGGAAGCCGT	GCTGACCAACGCTGGGGGTG
NRG1	TCTTCCGGGAGCTATCCACGGT	GGTTTGTTTTGCGGGGTTACAGGA
NRXN3	TGGGGGTGGACCTCTCCATGG	TCCTGGCAGGGCCAGTGAGT
PCSK1	GCCCACACACGATGAGCCCT	TGGGAATGGCCTGGGTCACCA
SMARCA2	GGCGGGGTGAGAGTAGAAATATCCT	AGCAGCATGGTTTGGAGCAGTCT
SNTG1	GCTTAGGGGCTCCCGTCCCC	CCGGGAGGCGAGGGGGTAATA
TLE3	TGACCAGCCAACCGTCCTGGT	CGGCTGGCCACCCTATGAGGA

EMSA Probe (S/AS)	40-mer Sequence (5′-to-3′)
Oligo1_Probe_S	**/5BiodT/**GCCAATACCATTCAGCACCACGGACAGAGCCACGCTCTGA
Oligo1_Probe_AS	TCAGAGCGTGGCTCTGTCCGTGGTGCTGAATGGTATTGGC
Oligo1_[-20]_S	GCAGAAGCACCACAGACAGCGCCAATACCATTCAGCACCA
Oligo1_[-20]_AS	TGGTGCTGAATGGTATTGGCGCTGTCTGTGGTGCTTCTGC
Oligo1_[0]_S	GCCAATACCATTCAGCACCACGGACAGAGCCACGCTCTGA
Oligo1_[0]_AS	TCAGAGCGTGGCTCTGTCCGTGGTGCTGAATGGTATTGGC
Oligo1_[+20]_S	CGGACAGAGCCACGCTCTGACGTCTTCTGCAGAAAAGTCC
Oligo1_[+20]_AS	GGACTTTTCTGCAGAAGACGTCAGAGCGTGGCTCTGTCCG
Oligo2_Probe_S	**/5BiodT/**AACTCTCAGAAGAAAGGGGAAAGAGAGGCGAGGCTTTAAG
Oligo2_Probe_AS	CTTAAAGCCTCGCCTCTCTTTCCCCTTTCTTCTGAGAGTT
Oligo2_[-20]_S	CTGTCAGGACCAGGGACAGCAACTCTCAGAAGAAAGGGGA
Oligo2_[-20]_AS	TCCCCTTTCTTCTGAGAGTTGCTGTCCCTGGTCCTGACAG
Oligo2_[0]_S	AACTCTCAGAAGAAAGGGGAAAGAGAGGCGAGGCTTTAAG
Oligo2_[0]_AS	CTTAAAGCCTCGCCTCTCTTTCCCCTTTCTTCTGAGAGTT
Oligo2_[+20]_S	AAGAGAGGCGAGGCTTTAAGAGACCTGTTTGGGCTGGGTG
Oligo2_[+20]_AS	CACCCAGCCCAAACAGGTCTCTTAAAGCCTCGCCTCTCTT

B. Supplemental Table S2.1B. Coriell Institute human variation panel cell lines

Population	Coriell Cat.	Sex	Age	Description
N. Europeans	GM17001	Female	12 YR	Northern Europeans
	GM17002	Female	NA	Northern Europeans
	GM17003	Male	39 YR	Northern Europeans
	GM17004	Female	39 YR	Northern Europeans
	GM17005	Female	33 YR	Northern Europeans
	GM17006	Female	40 YR	Northern Europeans
	GM17007	Male	39 YR	Northern Europeans
	GM17008	Male	NA	Northern Europeans
	GM17009	Female	NA	Northern Europeans
	GM17010	Male	45 YR	Northern Europeans
N. Africans	GM17378	Male	4 YR	Africans N. of the Sahara
	GM17379	Female	NA	Africans N. of the Sahara
	GM17380	Female	1 MO	Africans N. of the Sahara
	GM17381	Male	16 YR	Africans N. of the Sahara
	GM17382	Male	6 DA	Africans N. of the Sahara
	GM17383	Male	1 YR	Africans N. of the Sahara
	GM17384	Male	3 YR	Africans N. of the Sahara
S. Africans	GM17341	Female	16 YR	Africans S. of the Sahara
	GM17342	Male	53 YR	Africans S. of the Sahara
	GM17343	Male	16 YR	Africans S. of the Sahara
	GM17344	Female	10 YR	Africans S. of the Sahara
	GM17345	Male	16 YR	Africans S. of the Sahara
	GM17346	Male	17 YR	Africans S. of the Sahara
	GM17347	Male	36 YR	Africans S. of the Sahara
	GM17348	Female	5 YR	Africans S. of the Sahara
	GM17349	Male	NA	Africans S. of the Sahara
Mex. Indians	GM17392	Male	NA	Mexican Indians
	GM17393	Female	NA	Mexican Indians
	GM17394	Female	NA	Mexican Indians
	GM17395	Male	NA	Mexican Indians
	GM17396	Male	NA	Mexican Indians
Chinese	GM16654	Male	51 YR	Chinese
	GM16688	Female	75 YR	Chinese
	GM16689	Male	74 YR	Chinese
	GM17014	Male	4 YR	Chinese
	GM17015	Male	51 YR	Chinese
	GM17016	Male	NA	Chinese
	GM17017	Male	NA	Chinese
	GM17018	Female	NA	Chinese
	GM17019	Female	NA	Chinese
	GM17020	Male	NA	Chinese
Mbuti	GM10492	Male	NA	Mbuti, N.E. Zaire
	GM10494	Male	NA	Mbuti, N.E. Zaire
Misc	GM12156	Female	NA	CEPH/Utah Pedigree 1463
	GM12878	Female	NA	CEPH/Utah Pedigree 1463
	GM15228	NA	NA	DNA Polymorph. Disc. Res.

Supplemental Table S2.2. DEGs following siRNA knockdown of ZNF286A or ZNF286B.

OVERLAP	GENE	EXPT	PROBESET	SI-1	SI-2	CTRL-1	CTRL-2	rawP	BH corrected P-values
	A4GALT	286A	8076497	6.08	6.02	6.73	6.76	0	0
ab	AACS	286B	7959786	8.62	8.65	7.82	7.53	9.06E-11	3.90E-09
ab	AACS	286A	7959786	8.58	8.81	7.77	7.53	2.26E-10	1.28E-08
	AADACL4	286B	7897955	3.99	4.05	4.77	5.11	6.74E-08	2.14E-06
	AARSD1	286B	8015741	7.22	7.27	6.99	6.89	3.75E-08	1.22E-06
	ABCA11P	286B	8098752	5.3	5.32	4.84	4.37	3.07E-03	0.03
	ABCA6	286A	8017964	4.1	4.06	4.6	4.64	0	0
ab	ABCB10	286A	7924956	8.57	8.48	7.94	8.16	8.14E-05	2.02E-03
ab	ABCB10	286B	7924956	8.42	8.39	7.77	7.97	1.82E-07	5.46E-06
ab	ABCB6	286B	8059111	8.21	8.18	6.72	6.38	0	0
ab	ABCB6	286A	8059111	8.12	8.24	6.68	6.32	0	0
	ABCC12	286B	8001235	4.37	4.32	4.64	4.88	7.48E-04	0.01
ab	ABCC3	286A	8008454	7.71	7.85	7.3	7.4	3.16E-07	1.22E-05
ab	ABCC3	286B	8008454	7.92	7.96	7.15	7.46	5.99E-05	1.21E-03
ab	ABCC5	286A	8092418	7.52	7.55	6.99	6.95	0	0
ab	ABCC5	286B	8092418	7.44	7.48	6.79	6.56	1.87E-11	8.56E-10
	ABHD3	286A	8022488	6.25	6.15	5.76	5.54	7.56E-06	2.34E-04
ab	ABR	286A	8010983	8.66	8.98	8.17	7.72	1.71E-03	0.02
ab	ABR	286B	8010983	8.78	8.76	8.08	7.7	3.91E-06	9.85E-05
ab	ABT1	286A	8117522	7.64	7.46	7.14	6.95	1.18E-04	2.81E-03
ab	ABT1	286B	8117522	7.59	7.34	6.99	6.78	4.12E-04	6.80E-03
	ACADS	286B	7959220	5.99	6.06	6.38	6.71	2.26E-03	0.02
ab	ACAT2	286A	8123137	9.72	9.43	7.1	6.45	2.89E-15	2.48E-13
ab	ACAT2	286A	8123137	9.72	9.43	7.1	6.45	2.89E-15	2.48E-13
ab	ACAT2	286B	8123137	9.68	9.61	7.01	6.34	0	0
	ACAT2	286B	8123137	9.68	9.61	7.01	6.34	0	0
	ACBD6	286B	7922656	6.41	6.58	6.15	6.15	9.34E-05	1.81E-03
	ACBD7	286A	7932214	6.16	6.32	5.7	5.18	3.25E-03	0.04
ab	ACD	286A	8002057	8.22	8.27	7.6	7.92	2.60E-03	0.03
ab	ACD	286B	8002057	8.48	8.39	7.56	7.84	3.76E-07	1.09E-05
	ACLY	286B	8015460	11.83	11.88	11.14	10.39	3.47E-03	0.04
	ACLY	286B	8015460	11.83	11.88	11.14	10.39	3.47E-03	0.04
	ACMSD	286A	8045368	4.41	4.39	4.6	4.54	3.66E-06	1.18E-04
	ACOT2	286A	7975602	8.07	8.16	7.76	7.48	5.96E-04	0.01
ab	ACSBG2	286B	8025011	3.86	3.59	4.16	4.31	6.17E-04	9.70E-03
	ACSL1	286B	8103951	8.2	8.08	6.96	5.75	3.52E-03	0.04
	ACSL5	286B	7930498	5.28	5.32	5.93	5.66	3.14E-04	5.36E-03
	ACSS2	286B	8062041	8.6	8.61	7.57	7.4	0	0
	ACSS2	286B	8062041	8.6	8.61	7.57	7.4	0	0
	ACSS2	286B	8062041	8.6	8.61	7.57	7.4	0	0
ab	ACTBL2	286A	8112198	4.47	4.54	6.6	5.71	2.36E-04	5.13E-03
ab	ACTBL2	286B	8112198	4.12	4.29	6.29	5.33	9.82E-04	0.01
	ACTC1	286B	7987315	5.55	5.4	5.1	4.98	3.83E-06	9.69E-05
ab	ACTR5	286A	8062545	7.03	7.13	6.79	6.65	1.80E-05	5.24E-04
ab	ACTR5	286B	8062545	7.31	7.39	6.88	6.94	0	0
	ACYP1	286B	7980265	7.19	7.04	6.46	6.73	8.25E-04	0.01
ab	ADAM20	286A	7979927	4.38	4.71	5.23	5.72	1.59E-03	0.02
ab	ADAM20	286B	7979927	4.51	4.43	5.1	5.6	5.91E-04	9.33E-03
	ADAM3A	286B	8150375	3.03	3.03	3.23	3.26	0	0
ab	ADAMTS1	286A	8069676	5.43	5.64	7.07	6.42	4.72E-04	9.33E-03
ab	ADAMTS1	286B	8069676	5.71	5.7	6.9	6.34	1.02E-03	0.01
ab	ADAMTS6	286A	8112342	4.83	4.63	5.2	5.27	1.37E-06	4.81E-05
ab	ADAMTS6	286B	8112342	4.71	4.51	5.12	4.96	7.71E-04	0.01
ab	ADAMTSL1	286B	8154512	6.44	6.68	7.91	7.49	2.56E-06	6.66E-05
ab	ADAMTSL1	286A	8154512	6.55	6.72	7.74	7.36	7.49E-06	2.32E-04
	ADC	286A	7899851	6.82	6.87	7.07	7.02	4.99E-09	2.44E-07
	ADCY2	286B	8104394	5.24	5.29	5.56	5.69	1.08E-07	3.32E-06
	ADCY3	286B	8050766	9.26	9.33	8.35	7.38	3.34E-03	0.04
ab	ADD1	286B	8093643	9.13	8.92	8.58	8.33	5.36E-04	8.57E-03
ab	ADD1	286A	8093643	9.03	9.01	8.61	8.19	3.24E-03	0.04
	ADD2	286B	8052882	7.65	7.53	6.85	6.77	0	0
	ADHFE1	286A	8146687	4.48	4.55	4.93	5.25	3.81E-04	7.71E-03
ab	ADO	286B	7927767	8.86	8.76	8.65	8.59	1.56E-03	0.02
ab	ADO	286A	7927767	8.62	8.7	8.4	8.5	5.75E-04	0.01
	ADPRHL2	286A	7900087	7.98	7.95	7.89	7.86	1.51E-07	6.12E-06
	ADRM1	286B	8063893	10.59	10.67	10.25	9.92	1.10E-03	0.01
	ADRM1	286B	8063893	10.59	10.67	10.25	9.92	1.10E-03	0.01
	AFARP1	286B	7904084	5.55	5.62	6.11	6.37	1.30E-06	3.54E-05
	AFF2	286B	8170364	5.28	5.28	5.82	6.08	1.57E-07	4.74E-06
ab	AGAP4	286A	7927231	6.83	7.07	7.3	7.45	2.71E-03	0.04
ab	AGAP4	286B	7927231	6.99	7.05	7.28	7.36	2.46E-08	8.23E-07
ab	AGAP5	286A	7927305	6.64	6.81	7.12	7.11	1.79E-06	6.19E-05
ab	AGAP5	286B	7927305	6.7	6.82	7	6.98	2.24E-04	3.98E-03
	AGPAT9	286A	8096116	6.59	6.45	5.95	5.63	2.94E-05	8.16E-04
	AGXT2	286B	8111474	4.56	4.72	5.17	5.58	9.74E-04	0.01
	AHNAK2	286B	7981514	8.32	8.31	7.53	7.02	4.23E-05	8.86E-04
ab	AHSP	286A	7995237	3.64	3.81	4.1	4.2	8.50E-06	2.60E-04

Supplemental Table S2.2. (Continued)

OVERLAP	GENE	EXPT	PROBESET	SI-1	SI-2	CTRL-1	CTRL-2	rawP	BH corrected P-values
ab	AHSP	286B	7995237	3.71	3.78	4.04	4.2	2.47E-05	5.42E-04
	AKAP7	286B	8122045	5.75	5.78	5.34	4.94	1.66E-03	0.02
	AKR1C1	286A	7925918	6.2	5.98	6.58	6.66	7.03E-06	2.18E-04
	AKR1C3	286B	7925929	4.86	5.13	5.36	5.47	4.20E-03	0.04
	ALDH3A2	286B	8005638	8.79	8.81	7.74	6.61	4.09E-03	0.04
	ALDH5A1	286A	8117207	5.6	5.65	5.92	6.13	2.59E-04	5.54E-03
	ALOX5AP	286A	7968344	5.6	5.27	6.09	6.26	8.31E-05	2.06E-03
	ALOXE3	286B	8012326	5.21	5.36	5.73	5.95	2.64E-05	5.76E-04
ab	AMELY	286B	8177061	3.48	3.57	4.15	4.49	5.72E-06	1.41E-04
ab	AMELY	286A	8177061	3.39	3.56	4.21	4.16	1.31E-14	1.06E-12
	AMFR	286A	8001477	9.44	9.46	8.87	8.27	3.22E-03	0.04
	AMHR2	286B	7955797	5.19	5.31	5.57	5.78	5.30E-04	8.50E-03
ab	AMMECR1	286B	8174496	9.66	9.66	9.06	8.5	1.67E-03	0.02
ab	AMMECR1	286A	8174496	9.6	9.78	9.03	8.48	1.19E-03	0.02
ab	ANAPC10	286A	8103005	6.47	6.17	5.64	5.79	4.14E-04	8.28E-03
ab	ANAPC10	286B	8103005	6.6	6.53	5.36	5.61	2.22E-16	1.41E-14
	ANK3	286A	7933772	8.25	8.44	7.29	6.55	1.81E-04	4.07E-03
	ANK3	286B	7933772	8.55	8.64	7.34	6.48	1.04E-04	2.00E-03
	ANKAR	286B	8047006	4.87	4.88	4.64	4.66	0	0
	ANKRD16	286A	7931888	5.06	5.32	5.57	5.58	2.61E-03	0.03
	ANKRD18A	286B	8161270	4.9	5.15	5.41	5.62	3.18E-03	0.03
ab	ANKRD29	286A	8022559	4.74	4.98	5.79	6.43	2.07E-04	4.57E-03
ab	ANKRD29	286B	8022559	4.97	5.05	5.75	6.47	2.37E-03	0.03
	ANKRD36BP1		286B	7922121	5.45	5.21	4.27	4.29	0 0
	ANKRD45	286B	7922382	5.17	5.1	4.5	3.87	2.65E-03	0.03
	ANKRD49	286B	7943231	7.13	7.15	6.65	6.89	2.57E-03	0.03
	ANKRD6	286A	8121095	5.12	5.13	5.56	5.54	0	0
ab	ANO5	286B	7938951	2.82	3.01	3.46	3.34	1.32E-05	3.06E-04
ab	ANO5	286A	7938951	3.23	2.96	3.53	3.54	1.16E-03	0.02
	ANUBL1	286B	7933237	6.32	6.33	5.61	5.58	0	0
	ANXA2P1	286B	8103240	7.3	7.17	6.77	6.31	3.90E-03	0.04
ab	ANXA3	286B	8095986	3.68	3.64	4.11	3.9	1.18E-03	0.01
ab	ANXA3	286A	8095986	4.1	4.15	4.34	4.28	1.91E-07	7.56E-06
	AOC3	286B	8007420	4.91	4.6	5.46	5.83	1.78E-04	3.23E-03
	AP1M1	286B	8026548	9.71	9.77	9.29	9.1	7.50E-08	2.36E-06
ab	APEX2	286A	8167854	7.9	7.93	7.53	7.33	8.41E-07	3.05E-05
ab	APEX2	286B	8167854	8.05	8.02	7.65	7.41	1.51E-05	3.46E-04
	APOBEC3H	286A	8073096	5.41	5.4	6.04	5.93	0	0
	APOLD1	286B	7954055	6.8	6.8	7.26	7.71	2.80E-03	0.03
	APOM	286A	8118209	6.03	6.06	6.33	6.6	2.28E-03	0.03
	APOM	286A	8178043	6.03	6.06	6.33	6.6	2.28E-03	0.03
	APOM	286A	8179291	6.03	6.06	6.33	6.6	2.28E-03	0.03
	AQP7	286A	8160663	6.48	6.45	6.66	6.81	4.10E-04	8.23E-03
	ARAF	286B	8167165	9.06	9.11	8.7	8.64	0	0
	ARFGAP2	286B	7947784	8.43	8.44	7.89	7.31	4.04E-03	0.04
	ARHGAP28	286B	8019964	3.76	3.81	4.36	4.25	0	0
ab	ARHGDIA	286B	8019263	11.4	11.4	10.82	10.77	0	0
ab	ARHGDIA	286A	8019263	11.18	11.5	10.75	10.69	1.13E-04	2.70E-03
ab	ARHGDIA	286A	8019765	11.25	11.57	10.84	10.67	3.38E-04	6.96E-03
ab	ARHGDIA	286B	8019765	11.4	11.42	10.84	10.55	5.26E-07	1.50E-05
	ARHGEF10	286B	8144281	7.13	7.12	6.84	6.56	3.16E-03	0.03
	ARHGEF11	286B	7921179	8.03	7.93	7.78	7.63	2.12E-03	0.02
	ARHGEF11	286B	7921179	8.03	7.93	7.78	7.63	2.12E-03	0.02
	ARHGEF7	286B	7970111	8.32	8.47	7.85	7.68	3.23E-08	1.07E-06
	ARID3B	286A	7984843	5.54	5.49	5.67	5.63	8.31E-05	2.06E-03
	ARL4A	286B	8131573	5.65	5.58	5.88	5.79	8.73E-05	1.70E-03
	ARPC5L	286A	8157828	7.56	7.59	7.13	7.34	1.02E-03	0.01
	ARRDC1	286B	8159692	8.5	8.45	8.22	8.3	3.15E-06	8.06E-05
	ARRDC4	286B	7986350	4.69	4.77	6.46	5.98	1.19E-09	4.59E-08
ab	ARSG	286A	8009443	6.34	6.24	7.07	6.98	0	0
ab	ARSG	286B	8009443	6.33	6.15	6.98	6.83	1.29E-08	4.47E-07
	ASB1	286A	8049657	8.34	8.4	8.02	8.13	6.48E-06	2.02E-04
ab	ASB4	286A	8134376	4.14	4.17	4.43	4.55	2.38E-08	1.07E-06
ab	ASB4	286B	8134376	4.42	4.41	4.62	4.74	1.68E-05	3.83E-04
	ASB5	286B	8103812	3.87	3.86	4.19	4.4	4.63E-05	9.62E-04
	ASB7	286B	7986433	7.78	7.76	7.42	7.24	7.42E-07	2.08E-05
	ASCL3	286B	7946436	4.17	4.36	4.9	5.38	8.08E-04	0.01
ab	ASF1B	286A	8034772	10.19	10.32	8.92	8.46	5.74E-11	3.51E-09
ab	ASF1B	286B	8034772	10.49	10.33	8.96	8.53	1.60E-13	8.58E-12
	ASPA	286B	8003814	2.78	2.84	3.01	3.18	2.21E-03	0.02
ab	ASPHD1	286B	7994675	7.16	7.15	7.53	7.78	6.19E-05	1.25E-03
ab	ASPHD1	286A	7994675	7.47	7.3	7.7	7.89	1.53E-03	0.02
ab	ATAD3A	286A	7896952	9.03	9.08	8.38	8.6	3.79E-07	1.45E-05
ab	ATAD3A	286B	7896952	9.11	9.1	8.49	8.47	0	0
	ATF7IP2	286A	7993167	3.4	3.16	3.65	3.7	1.43E-03	0.02
	ATL1	286B	7974270	6.29	6.24	5.04	4.79	0	0
	ATP6AP1L	286A	8106722	5.98	6.18	6.45	6.57	2.71E-04	5.75E-03

Supplemental Table S2.2. (Continued)

OVERLAP	GENE	EXPT	PROBESET	SI-1	SI-2	CTRL-1	CTRL-2	rawP	BH corrected P-values
ab	ATP6V0C	286A	7992646	11.92	11.93	11.51	11.51	0	0
ab	AURKB	286B	8012403	9.8	9.74	9.12	8.85	1.04E-08	3.64E-07
ab	AURKB	286A	8012403	9.67	9.66	9.04	8.67	1.65E-05	4.84E-04
	AVIL	286A	7964555	5.38	5.45	5.79	5.86	0	0
	B3GNT5	286B	8084206	3.65	3.53	4.09	4.09	8.22E-15	4.84E-13
ab	B4GALNT1	286A	7964484	9.2	9.52	8.25	8.33	6.99E-11	4.25E-09
ab	B4GALNT1	286B	7964484	9.47	9.53	8.4	8.39	0	0
ab	B9D1	286B	8013331	6.86	6.71	7.67	7.44	2.05E-08	6.94E-07
ab	B9D1	286A	8013331	6.9	7.13	7.72	7.66	2.12E-08	9.57E-07
	BACE2	286B	8068671	5.09	5.08	5.77	6.44	2.14E-03	0.02
	BAG1	286B	8160647	6.97	7.13	6.71	6.59	3.30E-05	7.05E-04
	BAG3	286B	7930921	8.95	8.96	8.24	8.45	9.34E-09	3.29E-07
	BAG5	286A	7981439	8.02	7.96	7.52	7.13	8.90E-04	0.01
	BAG5	286A	7981439	8.02	7.96	7.52	7.13	8.90E-04	0.01
	BAK1	286B	8125766	8.28	8.29	7.91	7.79	8.59E-14	4.70E-12
ab	BAP1	286B	8087885	9.31	9.37	9.06	8.95	2.46E-08	8.23E-07
ab	BAP1	286A	8087885	9.25	9.47	8.99	8.92	3.60E-04	7.33E-03
	BASP1	286B	8104601	8.19	8.17	8.57	8.9	7.57E-04	0.01
	BATF2	286B	7949340	6.66	6.73	7.3	7.86	1.90E-03	0.02
ab	BBOX1	286A	7939056	3.2	3.16	3.46	3.64	4.68E-05	1.25E-03
ab	BBOX1	286B	7939056	3.38	3.42	3.61	3.77	3.05E-04	5.23E-03
	BCL10	286B	7917338	7.94	7.95	7.72	7.57	8.10E-05	1.59E-03
ab	BCL11A	286A	8052399	5.59	5.65	6.64	7.13	3.59E-07	1.38E-05
ab	BCL11A	286A	8052399	5.59	5.65	6.64	7.13	3.59E-07	1.38E-05
ab	BCL11A	286A	8052399	5.59	5.65	6.64	7.13	3.59E-07	1.38E-05
ab	BCL11A	286B	8052399	5.81	5.68	6.58	7.14	1.02E-04	1.97E-03
ab	BCL11A	286B	8052399	5.81	5.68	6.58	7.14	1.02E-04	1.97E-03
ab	BCL11A	286B	8052399	5.81	5.68	6.58	7.14	1.02E-04	1.97E-03
ab	BCL2A1	286A	7990818	4.6	4.4	4.97	5.06	3.91E-06	1.26E-04
ab	BCL2A1	286B	7990818	4.4	4.41	4.73	5.07	3.86E-03	0.04
	BCL2L14	286B	7953993	5.17	5.29	5.45	5.43	7.94E-04	0.01
ab	BCMO1	286B	7997401	4.88	4.75	5.37	5.68	2.52E-05	5.52E-04
ab	BCMO1	286A	7997401	4.9	4.86	5.19	5.5	2.57E-03	0.03
ab	BDKRB2	286B	7976560	6.16	6.43	7.85	7.6	3.33E-15	1.99E-13
ab	BDKRB2	286A	7976560	6.75	6.71	8.17	7.94	0	0
	BEND6	286B	8120362	6.67	6.48	5.64	4.88	7.33E-04	0.01
	BEND6	286B	8120362	6.67	6.48	5.64	4.88	7.33E-04	0.01
	BEX4	286B	8169009	6.55	6.56	7.07	7.51	1.05E-03	0.01
	BIRC3	286A	7943413	3.53	3.43	3.95	4	0	0
ab	BIRC5	286A	8010260	8.62	8.56	7.99	8.15	2.32E-09	1.19E-07
ab	BIRC5	286B	8010260	8.71	8.57	7.94	8.18	2.09E-05	4.68E-04
ab	BLM	286A	7986068	8.91	8.92	7.32	5.86	1.48E-03	0.02
ab	BLM	286B	7986068	9.06	8.9	7.27	5.75	1.26E-03	0.01
	BLVRA	286B	8132515	7.94	7.95	7.33	7.09	4.22E-10	1.71E-08
ab	BMP3	286B	8096070	5.36	5.4	6.21	6.24	0	0
	BNC2	286B	8160260	7.72	7.75	7.16	6.59	2.45E-03	0.03
	BOP1	286A	8153678	9	8.82	8.14	8.47	9.79E-04	0.01
	BPESC1	286A	8083025	4.89	5.05	5.46	5.37	2.06E-06	7.02E-05
	BPY2	286B	8176821	4.54	4.54	4.88	5.09	2.56E-05	5.59E-04
	BPY2	286B	8176867	4.54	4.54	4.88	5.09	2.56E-05	5.59E-04
	BPY2	286B	8177449	4.54	4.54	4.88	5.09	2.56E-05	5.59E-04
ab	BRCA1	286A	8015769	8.19	8.26	6.72	5.39	1.06E-03	0.01
ab	BRCA1	286B	8015769	8.36	8.42	6.65	5.22	5.66E-04	8.98E-03
	BRD2	286B	8118580	10.57	10.58	10.27	10.07	4.59E-05	9.55E-04
	BRD2	286B	8179504	10.39	10.4	10.08	9.76	3.60E-03	0.04
ab	BRI3BP	286A	7959777	9.3	9.39	8.49	7.67	2.26E-03	0.03
ab	BRI3BP	286A	7959777	9.3	9.39	8.49	7.67	2.26E-03	0.03
ab	BRI3BP	286B	7959777	9.75	9.69	8.78	8.06	3.59E-04	6.02E-03
ab	BRI3BP	286B	7959777	9.75	9.69	8.78	8.06	3.59E-04	6.02E-03
	BST2	286B	8035304	5	4.87	6.13	6.44	1.55E-15	9.55E-14
	BTBD11	286A	7958352	4.73	4.81	5.12	5.22	1.21E-09	6.34E-08
	BTBD12	286B	7999008	7.57	7.51	7.06	7.16	3.09E-14	1.72E-12
ab	BTBD6	286A	7977340	8.12	8.17	6.92	7.4	4.90E-05	1.30E-03
ab	BTBD6	286B	7977340	8.25	8.07	6.87	7.15	9.35E-12	4.36E-10
ab	BTN2A1	286B	8117485	7.55	7.36	7.13	7.15	1.10E-03	0.01
ab	BTN2A1	286A	8117485	7.46	7.35	7	7.1	1.95E-06	6.68E-05
	BUD13	286A	7951826	6.78	6.67	6.38	6.42	1.52E-08	6.99E-07
	BUD13	286B	7951826	6.83	6.65	6.26	6.28	5.37E-08	1.72E-06
ab	BYSL	286A	8119492	8.69	8.47	8.02	8.01	2.08E-07	8.20E-06
ab	BYSL	286B	8119492	8.59	8.57	8.12	7.87	2.66E-06	6.90E-05
	C10orf129		286A	7929497	4.38	4.47	4.7	4.76	1.12E-07 4.59E-06
	C10orf31	286B	7932019	4.92	4.97	5.18	5.42	3.10E-03	0.03
ab	C10orf55	286A	7934505	4.4	4.29	4.54	4.66	1.78E-03	0.02

Note: This supplementary table only includes an excerpt of 232 of 2,618 entries of the originating data file. The full supplementary file is available as part of the submitted manuscript.

Supplemental Table S2.3. ZNF286A ChIP-seq peak fold enrichment ($2^{-\Delta Ct}$) validation

ChIPExpt	chr	SummitStart-End	FoldEnr.	FDR	siRNADEG	siExpt	BH Adj p-values	FC
ZNF286A	chr1	12663940-12664140	6.3	0.1	DHRS3	286B	0	-3.97
ZNF286A	chr1	12663940-12664140	6.3	0.1	DHRS3	286A	3.64E-09	-3.39
REST	chr1	44434106-44434306	35.41	0	DPH2	286B	2.40E-05	1.57
ZNF286A	chr1	44434159-44434359	10.23	0	DPH2	286B	2.40E-05	1.57
ZNF286A	chr1	65888130-65888330	5.34	0	PDE4B	286A	5.51E-07	-1.65
ZNF286A	chr1	65888130-65888330	5.34	0	PDE4B	286A	5.51E-07	-1.65
ZNF286A	chr1	199149349-199149549	5.32	1.89	NR5A2	286B	2.21E-07	-1.55
REST	chr1	205428152-205428352	36.54	0	CDK18	286A	3.36E-12	-1.77
REST	chr1	205428152-205428352	36.54	0	CDK18	286B	0	-1.74
ZNF286A	chr1	205428171-205428371	11.54	0	CDK18	286A	3.36E-12	-1.77
ZNF286A	chr1	205428171-205428371	11.54	0	CDK18	286B	0	-1.74
ZNF286A	chr1	205457231-205457431	6.92	0.1	LEMD1	286B	3.88E-07	-1.83
ZNF286A	chr1	205457231-205457431	6.92	0.1	LEMD1	286A	0.02	-1.55
ZNF286A	chr10	4715327-4715527	5.19	0.07	KLF6	286A	1.55E-04	2.14
ZNF286A	chr10	4715327-4715527	5.19	0.07	KLF6	286B	9.62E-04	2.19
ZNF286A	chr10	13391966-13392166	5.91	0	SEPHS1	286B	0.04	1.95
ZNF286A	chr10	61297338-61297538	5.24	0.07	SLC16A9	286B	0.01	1.59
ZNF286A	chr10	79088214-79088414	16.75	0	KCNMA1	286B	0	3.15
ZNF286A	chr10	79088214-79088414	16.75	0	KCNMA1	286A	0	3.77
REST	chr10	79088247-79088447	64.1	0	KCNMA1	286B	0	3.15
REST	chr10	79088247-79088447	64.1	0	KCNMA1	286A	0	3.77
REST	chr10	79397258-79397458	16.03	0	KCNMA1	286B	0	3.15
REST	chr10	79397258-79397458	16.03	0	KCNMA1	286A	0	3.77
ZNF286A	chr10	80872497-80872697	5.43	0	PPIF	286A	0	1.50
REST	chr10	81002959-81003159	14.19	0.44	PPIF	286A	0	1.50
ZNF286A	chr10	97151797-97151997	13.28	0	PDLIM1	286A	1.46E-07	-2.26
ZNF286A	chr10	97151797-97151997	13.28	0	PDLIM1	286B	2.36E-03	-1.62
REST	chr10	97151826-97152026	50.9	0	PDLIM1	286A	1.46E-07	-2.26
REST	chr10	97151826-97152026	50.9	0	PDLIM1	286B	2.36E-03	-1.62
REST	chr10	105038678-105038878	68.91	0	INA	286B	0	-2.37
ZNF286A	chr10	105038696-105038896	10.35	0	INA	286B	0	-2.37
REST	chr11	65121709-65121909	19.23	0.08	TIGD3	286B	0.02	-1.97
ZNF286A	chr11	65121734-65121934	9.54	0	TIGD3	286B	0.02	-1.97
ZNF286A	chr11	109295713-109295913	7.82	0.13	ZC3H12C	286A	7.83E-05	1.66
ZNF286A	chr11	109295713-109295913	7.82	0.13	ZC3H12C	286B	0	1.71
REST	chr11	109295768-109295968	33.7	0	ZC3H12C	286A	7.83E-05	1.66
REST	chr11	109295768-109295968	33.7	0	ZC3H12C	286B	0	1.71
ZNF286A	chr11	109512334-109512534	6.35	1.39	ZC3H12C	286A	7.83E-05	1.66
ZNF286A	chr11	109512334-109512534	6.35	1.39	ZC3H12C	286B	0	1.71
REST	chr11	109512378-109512578	26.53	0	ZC3H12C	286A	7.83E-05	1.66
REST	chr11	109512378-109512578	26.53	0	ZC3H12C	286B	0	1.71
REST	chr11	116001521-116001721	31.93	0.13	CADM1	286B	5.07E-06	-2.05
ZNF286A	chr11	116250182-116250382	14.49	0	CADM1	286B	5.07E-06	-2.05
REST	chr11	116250231-116250431	67.97	0	CADM1	286B	5.07E-06	-2.05
ZNF286A	chr11	121425301-121425501	5.65	0.06	SORL1	286A	0	1.89
ZNF286A	chr11	121425301-121425501	5.65	0.06	SORL1	286B	0	1.89
REST	chr11	121526807-121527007	16.62	1.46	SORL1	286A	0	1.89
REST	chr11	121526807-121527007	16.62	1.46	SORL1	286B	0	1.89
REST	chr11	121636321-121636521	66.67	0	SORL1	286A	0	1.89
REST	chr11	121636321-121636521	66.67	0	SORL1	286B	0	1.89
ZNF286A	chr12	53472349-53472549	5.23	2.27	IGFBP6	286A	3.73E-05	-2.24
ZNF286A	chr12	53472349-53472549	5.23	2.27	IGFBP6	286B	4.53E-05	-2.23
ZNF286A	chr12	108775942-108776142	10.25	0	FICD	286B	0	1.93
REST	chr12	108776014-108776214	71.68	0	FICD	286B	0	1.93
ZNF286A	chr12	108797804-108798004	7.97	0.08	FICD	286B	0	1.93
REST	chr12	108797828-108798028	25.81	0	FICD	286B	0	1.93
ZNF286A	chr12	109284057-109284257	8.18	0.07	DAO	286B	0.02	-1.88
REST	chr12	109284072-109284272	18.25	0.66	DAO	286B	0.02	-1.88
ZNF286A	chr12	113375633-113375833	5.91	0.06	OAS3	286A	9.35E-05	1.77
ZNF286A	chr12	113375633-113375833	5.91	0.06	OAS3	286B	0	1.95
REST	chr12	119499103-119499303	42.07	0	HSPB8	286B	5.48E-03	2.32
REST	chr12	119499103-119499303	42.07	0	HSPB8	286A	4.63E-04	2.74
ZNF286A	chr12	119499119-119499319	8.33	0	HSPB8	286B	5.48E-03	2.32
ZNF286A	chr12	119499119-119499319	8.33	0	HSPB8	286A	4.63E-04	2.74
ZNF286A	chr13	110959830-110960030	5.15	0.1	COL4A2	286B	0	1.66
ZNF286A	chr13	110959830-110960030	5.15	0.1	COL4A2	286A	1.08E-05	1.72
ZNF286A	chr14	51027144-51027344	9.09	0	ATL1	286B	0	2.54
ZNF286A	chr15	48623686-48623886	5.39	0.13	DUT	286B	6.00E-03	1.85
REST	chr15	48623810-48624010	20.6	1.28	DUT	286B	6.00E-03	1.85
ZNF286A	chr15	51973516-51973716	7.41	0.07	SCG3	286B	0.04	-1.57
REST	chr15	51973623-51973823	41.61	0	SCG3	286B	0.04	-1.57
REST	chr15	70265567-70265767	37.83	0	TLE3	286B	0	-3.55
REST	chr15	70265567-70265767	37.83	0	TLE3	286A	0	-3.02
REST	chr15	70268812-70269012	48.08	0	TLE3	286B	0	-3.55
REST	chr15	70268812-70269012	48.08	0	TLE3	286A	0	-3.02
ZNF286A	chr15	70268858-70269058	14.69	0	TLE3	286B	0	-3.55
ZNF286A	chr15	70268858-70269058	14.69	0	TLE3	286A	0	-3.02

Supplemental Table S2.3. (Continued)

ChIPExpt	chr	SummitStart-End	FoldEnr.	FDR	siRNADEG	siExpt	BH Adj p-values	FC
REST	chr15	70389137-70389337	15.26	0.27	TLE3	286B	0	-3.55
REST	chr15	70389137-70389337	15.26	0.27	TLE3	286A	0	-3.02
ZNF286A	chr15	70766921-70767121	5.13	0.06	TLE3	286B	0	-3.55
ZNF286A	chr15	70766921-70767121	5.13	0.06	TLE3	286A	0	-3.02
ZNF286A	chr15	70798758-70798958	5.44	0.32	TLE3	286B	0	-3.55
ZNF286A	chr15	70798758-70798958	5.44	0.32	TLE3	286A	0	-3.02
REST	chr15	74557444-74557644	60.31	0	CCDC33	286A	3.61E-11	-2.48
REST	chr15	74557444-74557644	60.31	0	CCDC33	286B	0	-2.39
ZNF286A	chr15	74557525-74557725	12.5	0	CCDC33	286A	3.61E-11	-2.48
ZNF286A	chr15	74557525-74557725	12.5	0	CCDC33	286B	0	-2.39
ZNF286A	chr15	74583888-74584088	27.5	0	CCDC33	286A	3.61E-11	-2.48
ZNF286A	chr15	74583888-74584088	27.5	0	CCDC33	286B	0	-2.39
REST	chr15	74583902-74584102	121.79	0	CCDC33	286A	3.61E-11	-2.48
REST	chr15	74583902-74584102	121.79	0	CCDC33	286B	0	-2.39
ZNF286A	chr16	20034617-20034817	9.77	0	GPRC5B	286B	1.30E-06	-1.58
ZNF286A	chr16	20034617-20034817	9.77	0	GPRC5B	286A	0.01	-1.56
REST	chr16	20034630-20034830	51.44	0	GPRC5B	286B	1.30E-06	-1.58
REST	chr16	20034630-20034830	51.44	0	GPRC5B	286A	0.01	-1.56
ZNF286A	chr16	71928703-71928903	5.45	0.87	DHODH	286B	0.03	1.73
ZNF286A	chr17	8123957-8124157	6.08	0	AURKB	286B	3.64E-07	1.73
ZNF286A	chr17	8123957-8124157	6.08	0	AURKB	286A	4.84E-04	1.75
ZNF286A	chr17	19411366-19411566	6.25	1.22	SLC47A1	286B	8.67E-03	-1.62
ZNF286A	chr17	48715689-48715889	5.29	0	ABCC3	286B	1.21E-03	1.55
ZNF286A	chr17	79894652-79894852	7.3	1.56	MAFG	286B	0	1.51
ZNF286A	chr17	79894652-79894852	7.3	1.56	MAFG	286A	2.43E-06	1.53
ZNF286A	chr17	79894652-79894852	7.3	1.56	MAFG	286B	0.01	1.80
ZNF286A	chr17	79894652-79894852	7.3	1.56	MAFG	286A	2.66E-03	1.98
ZNF286A	chr19	1383561-1383761	8.18	0.07	NDUFS7	286B	3.06E-09	-1.55
ZNF286A	chr19	55728401-55728601	27.95	0	PTPRH	286A	0	-1.53
REST	chr19	55728412-55728612	48.58	0	PTPRH	286A	0	-1.53
REST	chr19	55731960-55732160	24.04	0.25	PTPRH	286A	0	-1.53
ZNF286A	chr19	55731997-55732197	5.23	1.39	PTPRH	286A	0	-1.53
ZNF286A	chr2	12316659-12316859	5.03	0.07	TRIB2	286A	2.65E-04	1.75
ZNF286A	chr2	12316659-12316859	5.03	0.07	TRIB2	286B	1.07E-03	1.85
REST	chr2	12858538-12858738	18.63	0.11	TRIB2	286A	2.65E-04	1.75
REST	chr2	12858538-12858738	18.63	0.11	TRIB2	286B	1.07E-03	1.85
ZNF286A	chr2	12858554-12858754	5.19	0	TRIB2	286A	2.65E-04	1.75
ZNF286A	chr2	12858554-12858754	5.19	0	TRIB2	286B	1.07E-03	1.85
ZNF286A	chr2	27061411-27061611	6.86	0.06	DPYSL5	286B	0.01	-2.10
ZNF286A	chr2	27061411-27061611	6.86	0.06	DPYSL5	286A	0.01	-1.72
REST	chr2	27061477-27061677	22.51	0	DPYSL5	286B	0.01	-2.10
REST	chr2	27061477-27061677	22.51	0	DPYSL5	286A	0.01	-1.72
REST	chr2	54683492-54683692	35.26	0	EML6	286A	1.75E-05	1.50
REST	chr2	54683492-54683692	35.26	0	EML6	286B	0.01	1.81
ZNF286A	chr2	54951549-54951749	5.89	1.42	EML6	286A	1.75E-05	1.50
ZNF286A	chr2	54951549-54951749	5.89	1.42	EML6	286B	0.01	1.81
REST	chr2	186108234-186108434	46.88	0	FSIP2	286A	0	-1.50
ZNF286A	chr2	186406312-186406512	5.91	0	FSIP2	286A	0	-1.50
ZNF286A	chr2	186554009-186554209	5.86	0.23	FSIP2	286A	0	-1.50
ZNF286A	chr2	212501543-212501743	5.23	0.06	ERBB4	286B	0.01	-2.42
REST	chr2	212867829-212868029	36.25	0	ERBB4	286B	0.01	-2.42
ZNF286A	chr2	213824019-213824219	5.17	0.9	ERBB4	286B	0.01	-2.42
REST	chr2	234821540-234821740	28.37	0	HJURP	286B	0.05	3.07
ZNF286A	chr2	234821558-234821758	8.47	0	HJURP	286B	0.05	3.07
ZNF286A	chr20	656021-656221	7.5	0.21	SRXN1	286A	0.03	2.48
ZNF286A	chr20	656021-656221	7.5	0.21	SRXN1	286B	2.58E-03	2.48
REST	chr20	656239-656439	74.5	0	SRXN1	286A	0.03	2.48
REST	chr20	656239-656439	74.5	0	SRXN1	286B	2.58E-03	2.48
ZNF286A	chr20	3767725-3767925	5.13	0.6	CDC25B	286A	1.86E-09	1.76
ZNF286A	chr20	3767725-3767925	5.13	0.6	CDC25B	286B	0	1.97
ZNF286A	chr22	20068669-20068869	5.33	0.2	DGCR8	286A	1.11E-06	1.98
ZNF286A	chr22	20068669-20068869	5.33	0.2	DGCR8	286B	4.37E-08	2.17
REST	chr3	48699190-48699390	49.92	0	SLC26A6	286B	5.75E-12	2.12
ZNF286A	chr3	48699210-48699410	19.54	0	SLC26A6	286B	5.75E-12	2.12
REST	chr3	126877750-126877950	24.83	0	PLXNA1	286B	0	1.70
REST	chr3	126877750-126877950	24.83	0	PLXNA1	286A	4.48E-06	1.81
REST	chr3	126881756-126881956	31.93	0	PLXNA1	286B	0	1.70
REST	chr3	126881756-126881956	31.93	0	PLXNA1	286A	4.48E-06	1.81
ZNF286A	chr3	126881857-126882057	8.64	0	PLXNA1	286B	0	1.70
ZNF286A	chr3	126881857-126882057	8.64	0	PLXNA1	286A	4.48E-06	1.81
REST	chr3	126952226-126952426	17.74	0.36	PLXNA1	286B	0	1.70
REST	chr3	126952226-126952426	17.74	0.36	PLXNA1	286A	4.48E-06	1.81
ZNF286A	chr3	126952250-126952450	6.59	0.36	PLXNA1	286B	0	1.70
ZNF286A	chr3	126952250-126952450	6.59	0.36	PLXNA1	286A	4.48E-06	1.81
ZNF286A	chr3	137916043-137916243	6.69	0.21	DBR1	286B	1.14E-03	2.02
REST	chr3	137916061-137916261	13.96	0	DBR1	286B	1.14E-03	2.02
ZNF286A	chr4	47211615-47211815	5.23	0	GABRB1	286B	0	4.22

Supplemental Table S2.3. (Continued)

ChIPExpt	chr	SummitStart-End	FoldEnr.	FDR	siRNADEG	siExpt	BH Adj p-values	FC
ZNF286A	chr4	47211615-47211815	5.23	0	GABRB1	286A	0	4.23
ZNF286A	chr4	87514616-87514816	5.08	0.22	MAPK10	286A	2.05E-14	-1.73
ZNF286A	chr4	184424759-184424959	5.23	0.57	ING2	286A	4.75E-11	1.70
ZNF286A	chr4	184424759-184424959	5.23	0.57	ING2	286B	3.63E-05	2.03
REST	chr5	14628569-14628769	22.93	0	FAM105B	286A	0.01	1.58
ZNF286A	chr5	14628570-14628770	17.64	0	FAM105B	286A	0.01	1.58
ZNF286A	chr5	35128791-35128991	5.41	0.07	AGXT2	286B	0.01	-1.66
REST	chr5	56670474-56670674	54.99	0	ACTBL2	286A	5.13E-03	-3.13
REST	chr5	56670474-56670674	54.99	0	ACTBL2	286B	0.01	-3.04
ZNF286A	chr5	56679497-56679697	5.23	0.42	ACTBL2	286A	5.13E-03	-3.13
ZNF286A	chr5	56679497-56679697	5.23	0.42	ACTBL2	286B	0.01	-3.04
REST	chr5	56682816-56683016	15.96	1.97	ACTBL2	286A	5.13E-03	-3.13
REST	chr5	56682816-56683016	15.96	1.97	ACTBL2	286B	0.01	-3.04
ZNF286A	chr5	56682892-56683092	6.06	0.2	ACTBL2	286A	5.13E-03	-3.13
ZNF286A	chr5	56682892-56683092	6.06	0.2	ACTBL2	286B	0.01	-3.04
ZNF286A	chr5	132901754-132901954	13.87	0	FSTL4	286A	0.03	-2.27
REST	chr5	132901771-132901971	68.17	0	FSTL4	286A	0.03	-2.27
REST	chr5	132987069-132987269	33.7	0.1	FSTL4	286A	0.03	-2.27
REST	chr6	19771064-19771264	44.35	0	ID4	286B	0.04	-2.01
ZNF286A	chr6	19803714-19803914	5.23	0.06	ID4	286B	0.04	-2.01
ZNF286A	chr6	19988879-19989079	8.12	0	ID4	286B	0.04	-2.01
REST	chr6	19988898-19989098	14.9	0.36	ID4	286B	0.04	-2.01
ZNF286A	chr6	28625879-28626079	5.24	0.13	SCAND3	286B	0	-1.63
ZNF286A	chr6	28806321-28806521	5.23	0	SCAND3	286B	0	-1.63
REST	chr6	33291220-33291420	40.25	0	KIFC1	286A	4.39E-10	3.03
REST	chr6	33291220-33291420	40.25	0	KIFC1	286B	1.41E-14	3.08
REST	chr6	33291220-33291420	40.25	0	KIFC1	286A	6.48E-07	3.27
REST	chr6	33291220-33291420	40.25	0	KIFC1	286B	1.80E-08	3.43
ZNF286A	chr6	33291238-33291438	8.31	0.07	KIFC1	286A	4.39E-10	3.03
ZNF286A	chr6	33291238-33291438	8.31	0.07	KIFC1	286B	1.41E-14	3.08
ZNF286A	chr6	33291238-33291438	8.31	0.07	KIFC1	286A	6.48E-07	3.27
ZNF286A	chr6	33291238-33291438	8.31	0.07	KIFC1	286B	1.80E-08	3.43
ZNF286A	chr6	33559035-33559235	5.23	1.24	ITPR3	286B	9.16E-11	1.51
ZNF286A	chr6	36647765-36647965	5.68	0.14	RAB44	286A	8.93E-07	-1.57
REST	chr6	36702029-36702229	21.29	0.08	RAB44	286A	8.93E-07	-1.57
ZNF286A	chr6	130683384-130683584	8.67	0	SAMD3	286A	1.25E-12	2.30
ZNF286A	chr6	130683384-130683584	8.67	0	SAMD3	286B	0	2.68
REST	chr6	130683424-130683624	50.05	0	SAMD3	286A	1.25E-12	2.30
REST	chr6	130683424-130683624	50.05	0	SAMD3	286B	0	2.68
ZNF286A	chr6	132270310-132270510	5.45	1.04	CTGF	286B	0	1.51
REST	chr6	132718680-132718880	38.46	0	CTGF	286B	0	1.51
REST	chr7	75831098-75831298	186.78	0	HSPB1	286B	0.03	1.57
ZNF286A	chr7	75831191-75831391	10.86	0	HSPB1	286B	0.03	1.57
REST	chr7	75854789-75854989	49.04	0	HSPB1	286B	0.03	1.57
ZNF286A	chr7	75854895-75855095	9.85	0	HSPB1	286B	0.03	1.57
ZNF286A	chr7	75910433-75910633	5.91	0.61	HSPB1	286B	0.03	1.57
ZNF286A	chr7	150949333-150949533	5.68	0.13	NUB1	286B	2.95E-09	-1.87
REST	chr8	26435706-26435906	20.98	3.68	DPYSL2	286A	0.02	4.18
REST	chr8	26435706-26435906	20.98	3.68	DPYSL2	286B	5.85E-03	4.43
ZNF286A	chr8	26435754-26435954	5.45	0	DPYSL2	286A	0.02	4.18
ZNF286A	chr8	26435754-26435954	5.45	0	DPYSL2	286B	5.85E-03	4.43
ZNF286A	chr8	31652922-31653122	18.35	0	NRG1	286A	5.58E-04	2.01
ZNF286A	chr8	31652922-31653122	18.35	0	NRG1	286B	7.42E-05	2.42
REST	chr8	31652925-31653125	44.96	0	NRG1	286A	5.58E-04	2.01
REST	chr8	31652925-31653125	44.96	0	NRG1	286B	7.42E-05	2.42
ZNF286A	chr8	32161512-32161712	6.36	0	NRG1	286A	5.58E-04	2.01
ZNF286A	chr8	32161512-32161712	6.36	0	NRG1	286B	7.42E-05	2.42
ZNF286A	chr8	32487166-32487366	5.21	1.24	NRG1	286A	5.58E-04	2.01
ZNF286A	chr8	32487166-32487366	5.21	1.24	NRG1	286B	7.42E-05	2.42
ZNF286A	chr8	80941232-80941432	5.23	1.83	HEY1	286B	0.04	-1.99
ZNF286A	chr9	2018039-2018239	5.23	2.48	SMARCA2	286A	9.79E-06	1.52
ZNF286A	chr9	2018039-2018239	5.23	2.48	SMARCA2	286B	1.90E-03	1.72
REST	chr9	17744869-17745069	27.47	0	ADAMTSL1	286B	6.66E-05	-2.21
REST	chr9	17744869-17745069	27.47	0	ADAMTSL1	286A	2.32E-04	-1.88
ZNF286A	chr9	17744892-17745092	7.43	0.14	ADAMTSL1	286B	6.66E-05	-2.21
ZNF286A	chr9	17744892-17745092	7.43	0.14	ADAMTSL1	286A	2.32E-04	-1.88
ZNF286A	chr9	91800991-91801191	9.14	0	CKS2	286B	1.64E-06	1.77
REST	chr9	91801009-91801209	35.37	0	CKS2	286B	1.64E-06	1.77
ZNF286A	chr9	123676102-123676302	7.66	0	PHF19	286A	2.88E-08	2.40
ZNF286A	chr9	123676102-123676302	7.66	0	PHF19	286B	1.15E-08	2.48
REST	chr9	123676175-123676375	25.49	0	PHF19	286A	2.88E-08	2.40
REST	chr9	123676175-123676375	25.49	0	PHF19	286B	1.15E-08	2.48
ZNF286A	chr9	126872665-126872865	49.63	0	NEK6	286A	0.03	1.69

Note: This supplementary table only includes an excerpt of 228 of 7,093 entries of the originating data file. The full supplementary file is available as part of the submitted manuscript.

Supplemental Table S2.4.

IMAGEJ Raw Densitometry Data

Cell Line	ZNF286A	ZNF286B	REST	LaminB1
SH-SY5Y RA-0	166.385	456.092	805.092	9739.548
SH-SY5Y RA-3	752.113	650.456	594.506	11573.912
SH-SY5Y RA-5	1641.184	802.698	526.92	10719.376
SH-SY5Y RA-8	829.82	1043.527	83.728	7687.891

IMAGEJ Corrected Values Based On LaminB1

Cell Line	ZNF286A	ZNF286B	REST	LaminB1
SH-SY5Y RA-0	0.017083442	0.046828867	0.082662152	1
SH-SY5Y RA-3	0.064983473	0.056200185	0.051366038	1
SH-SY5Y RA-5	0.153104434	0.074882904	0.049155846	1
SH-SY5Y RA-8	0.107938575	0.135736446	0.010890893	1

Fold Change Of Retinoic Acid Day 0 vs Day 8

	ZNF286A	ZNF286B	REST	LaminB1
Fold Change:	6.432748538	2.893617021	0.13253012	

Supplemental Table S2.5. **(A)** After observing certain genes become up-regulated after knocking down ZNF286A using siRNA, and using ChIP-Seq to find that those genes lay next to potential ZNF286A binding sites, we designed primers specific to the genes and validated their enrichment using QPCR in the ChIP input gDNA, IgG control DNA, and supernatant DNA (as a secondary control). Each of the "Input DNA", "IgG", and "Supernatant" values are compared to the ZNF286A-WGA (Whole Gene Amplification) material using $2^{-\Delta CT}$ to determine relative gene expression. The "siRNA KD" values are the expected values from our *ZNF286A* knockdown microarray. **(B)** After siRNA knockdown of *ZNF286A* and *ZNF286B*, microarrays were done on the resulting cells to identify genes that were either up-regulated (implying that *ZNF286A* or *B* was (in)directly involved in *repressing* the gene) or down-regulated (implying that *ZNF286A* or *B* was (in)directly involved in *activating* the gene). ChIP-Seq then identified areas where *ZNF286A/B* had binding sites, and identified the genes on the following list.

A. Supplemental Table S2.5A. ZNF286A ChIP-seq peak fold enrichment ($2^{-\Delta Ct}$) validation

			MACS Fold Enr.		QPCR Fold Enr.	
Peak Coordinates	Nearest Genes	BP to Peak	MACS	FDR(%)	Input	IgG
chr11:72452433-72452633	PDE2A,ARAP1	-67036; 10915	3.86	0.96	2.416	3.974
chr7:54878443-54878643	EGFR,SEC61G	-208251;-50876	4.38	0	2.97	1.566
chr11:105543915-105544115	GRIA4,MSANTD4	63215; 348939	15.45	0	31.75	18.12
chr6:146349511-146349711	RAB23,GRM1	-515244; 829	13.63	0	40.369	21.542
chr5:78808148-78808348	HOMER1,JMY	1792; 276236	6.59	0	6.359	2.966
chr9:133060996-133061196	HMCN2,NCS1	-224852;236239	15.44	0	27.098	19.474
chr14:78986289-78986489	NRXN3	116296	6.02	0.2	12.445	9.196
chr2:163697224-163697424	KCNH7,FIGN	-2196; 895193	22.72	0	46.941	21.108
chr15:89905022-89905222	POLG,MIR9-3	-27044; 6744	10.68	0	12.005	4.561
chr8:31652922-31653122	NRG1,WRN	-726043;788368	18.35	0	2.318	1.429
chr5:95819672-95819872	CAST,PCSK1	-178169;-50788	12.78	0	6.882	4.226
chr8:50824149-50824349	SNTG1	16	17.2	0	26.004	4.421
chr15:70268858-70269058	RPLP1,TLE3	121887; 523835	14.69	0	15.898	12.178

B. Supplemental Table S2.5B. Regulated genes after knockdown, near ChIP binding sites

Gene	ΔFold siSc	ΔFold si0	FDR	Gene	ΔFold siSc	ΔFold si0	FDR
Up-regulated after ZNF286A KD				*Up-regulated after ZNF286B KD*			
NCS1	2.19	10.21	0	NCS1	2.16	9.26	0
GABRB1	4.23	3.43	0	GABRB1	4.22	4.35	0
NRG1	2.01	1.44	0	NRG1	2.42	1.61	0
SMARCA2	1.52	2.7	0	SMARCA2	1.72	3.51	0
EGFR	2.7	3.46	0.09	EGFR	2.95	4.48	0.09
HOMER1	1.14	3.04	0				
Down-regulated after ZNF286A KD				*Down-regulated after ZNF286B KD*			
ARAP1	-1.27	-1.04	0.75	NRXN3	-1.61	-1.84	0.16
NRXN3	-1.78	-1.79	0.16	NRXN3	-1.61	-1.84	0
NRXN3	-1.78	-1.79	0	GRM1	-1.1	-1.38	0
				GRIA4	-1.4	-2.07	0

CHAPTER 3

ZNF286 AND REST ALTERNATIVE SPLICING ISOFORMS IN HUMANS AND MICE

Introduction

As demonstrated in the previous chapters, the *ZNF286A* transcription factor codes for a *Krüppel*-type zinc finger transcription factor (TF) with a ten-finger long DNA-binding domain, tethered to a KRAB domain. We showed that, despite possessing a non-canonical KRAB, *ZNF286A* is still capable of interacting with the KRAB-interacting cofactor KAP-1. We also described how *ZNF286A* had experienced a human-specific duplication event, producing the *ZNF286B* gene that codes for a transcription factor that lacks *ZNF286A*'s KRAB-domain due to a pseudogene insertion event, but regulates the same genes, and in the same fashion as the originating TF. Indeed, our evidence suggests that despite their structural differences, both *ZNF286A* and *ZNF286B* serve primarily to repress genes involved in cell division and proliferation in neuroblastoma cells. In human tissues, *ZNF286A* and *ZNF286B* are both expressed in similar tissues, but at different levels – *ZNF286A* is most highly expressed in adult brain, while *ZNF286B* is most highly expressed in fetal brain – suggesting that both TFs serve different regulatory functions based on their independent regulation and differential expression.

One of the remaining issues that may bear on our interpretation of our experimental results is that *ZNF286A* also transcribes an alternative transcript encoding a protein that is similar to ZNF286B in that it lacks a KRAB domain; this transcript uses an alternative splice acceptor site in the spacer-finger exon that is 59 base pairs upstream of the site used in the KRAB-containing transcript (***Figure 3.1***). We refer to these "KRAB(+)" and "KRAB(-)" isoforms as *ZNF286A* (K+) and *ZNF286A* (K-), respectively in the following discussion. Subsequent sequencing of amplified PCR products revealed that the alternative (K-) transcript includes an intronic sequence that shifts the finger region out of frame if the upstream starting methionine is used. An internal methionine codon, located just inside the spacer-finger exon is thus presumably used as the start of translation for this protein.

As a truncated version of the full transcript, ZNF286A (K-) is theoretically capable of all the same DNA-binding functions as its full-length KRAB(+) isoform due to its possession of all the same DNA-binding zinc fingers. Our original hypothesis was that, therefore, this isoform would possess a distinct function, able to bind to the same

genomic binding sites but without the ability to repress due to the lack of KRAB(-) mediated silencing function. The functions of this (K-) isoform have not been studied, but these types of isoforms are commonly generated by ZNF transcription factors and have been generally presumed to encode "dominant negative" competitors that modulate the full-length protein's function (Hsu et al., 1992; Bellefroid et al., 1997).

The structure of *ZNF286A* is highly conserved in all marsupials and placental mammals, and in mice this unique gene is called *Zfp286*. The conservation between human and mouse genes also applies to splicing, meaning that the mouse *Zfp286* gene also gives rise to (K+) and *Zfp286* (K-) transcripts. This conservation aligns with previous studies that have shown that conserved human/mice alternative transcripts are enriched in regulatory genes and are frequently expressed in brain (Yeo et al., 2005).

For these reasons we hypothesized that *ZNF286A* (K-) and *ZNF286B*, both lacking KRAB domains, may act as competitors to the human ZNF286A (K+), and fine-tune the regulation of target genes by varying concentrations of each protein in different tissues at different times. This competition would be comparable to the interactions of *Zfp286* (K+) and Zfp286 (K-) in mice, with the caveat that mice do not have a duplicate of *Zfp286* like humans do. However, our studies with *ZNF286A* and *ZNF286B* siRNA knockdown (***Chapter 2***) throw this assumption into question since both human genes appear to function similarly in cells where they are co-expressed, most likely as repressors of the same genes.

Nevertheless, the fact that both humans and mice express both (K+) and (K-) isoforms still supports the idea that these isoforms may have distinct genetic functions. In alignment with this argument, ZNF286A and ZNF286B proteins may also function in subtly different ways. As our previous studies determined that *ZNF286A* and *ZNF286B* are expressed in human tissues at different levels, the human-specific duplicate may have evolved to fit a functional niche that is not occupied by the full-functioning ancestral gene. However, it is possible that a similar niche may in fact have been occupied by the ZNF286A KRAB(-) isoform in non-human species. Alternative splicing acts as a means of creating protein diversity (Keren et al., 2010; Nilsen & Graveley, 2010), so this suggests that it is possible for the function of KRAB-less ZNF286A has become part of the "official" non-alternative proteome in the form of the ZNF286B genetic duplicate.

In previous studies (***Chapter 2***) we showed that ZNF286A is enriched at binding sites of the RE-1 silencing transcription factor (REST) in human neuroblastoma cells and both ZNF286A and ZNF286B proteins can bind to the RE-1 motif *in vitro*. Therefore we hypothesize that ZNF286A (K+), (K-), and ZNF286B all play competitive and complementary roles to REST in transcriptional regulation. However these functions, while similar and perhaps cooperative and complementary, may nonetheless be distinct.

To provide further clues to isoform and paralog function, we determined the expression patterns of both *ZNF286A* (K+) and (K-) isoforms in humans and mice, together with *ZNG286B* in human tissues. *REST* is also known to generate functionally distinct splicing isoforms in neuronal cells (Hamelink et al., 2004; Raj et al., 2011) and since it is clear from our previous studies (***Chapter 2***) that ZNF286 and REST act in a coordinated fashion, we also sought to correlate the expression of these isoforms with ZNF286 genes and isoforms in human and mouse tissues. Finally, we developed a series of expression constructs and transfected cells that should be crucial to unraveling the coordinated REST and ZNF286 protein functions in future studies.

Materials and Methods

Cell line culture and neuroblast differentiation

SH-EP cells acquired from Dr. Martin Reick (Reick et al., 2001) were cultured and maintained in DMEM containing 10% fetal bovine serum (FBS). SH-SY5Y cells (*ATCC*® CCL-131) were grown in 1:1 DMEM/F12 media containing 10% FBS, and were induced to differentiate at 50-60% confluence by replacing growth media with DMEM media containing 15% FBS and 10 μM retinoic acid. Media was replaced every 2-3 days. On the fifth day of RA-treatment, cells were washed three times with DMEM, and 50 ng/mL BDNF was added to the media (without serum) to maintain division. SH-SY5Y cells are fully differentiated after 6-10 days (Encinas et al., 2002). Other cell lines cultured in this study were grown and maintained following the supplier's recommended protocols: JEG-3 (*ATCC*® HTB-36), HEK-293 (*ATCC*® CRL-1573), NCCIT (*ATCC*® CRL-2073), GM-Hu-248 (*Coriell Institute*), Ntera-2 (*ATCC*® CRL-1973), PANC-1

(*ATCC*® CRL-1469), HeLa (*ATCC*® CCL-2), LNCap (*ATCC*® CRL-1740), and MDA-MB-231 (*ATCC*® HTB-26).

QPCR of tissue or cell line RNA

RNA was collected by homogenizing tissue samples or cell pellets in TRIzol Reagent (*Invitrogen* Cat. 15596). Samples were purified and DNase-treated using the standard Qiagen protocol with RNeasy columns (*Qiagen* Cat. 74104), from which cDNA was synthesized using using M-MuLV Reverse Transcriptase (*New England Biolabs* Cat. M0253S). QPCR was performed on each sample using sequence-specific primers (***Table 3.1***) in Power SYBR Green PCR Master Mix (*ThermoFisher* Cat. 4367659). PCR reactions proceeded using the *Applied Biosystems* 7900HT Fast Real-Time PCR System.

Semi-QPCR of REST/REST4

Semi-QPCR was used to visualize the amplification of *REST* and *REST4* isoforms in differentiating SH-SY5Y. Unlike QPCR, which aims to fully amplify a product over the course of ~45 cycles, semi-QPCR seeks to minimize the number of cycles until the product is just barely able to visualized on an agarose gel. When a primer-set detects two products, then it is possible to determine which bands are brighter, and therefore have more starting RNA transcripts. The amount of starting template and number of PCR cycles is optimized according to the primers used. Primers used to test for *REST/REST4* were “h-p2-s” and “h-p8-as” as previously published and shown to detect both full-length *REST* and alternatively spliced *REST4* isoforms (Palm et al., 1999). Samples were run for 28 cycles at 55°C, and produced both *REST* and *REST4* products in SH-SY5Y, as predicted.

Inducible tagged Zfp286 vector

The *pTRE-tight* vector (Clontech Cat. 631059) is designed with a tetracycline response element (TRE) promoter that is highly expressed in the presence of doxycycline, and a SV40 poly-A signal fragment to signal translation. This plasmid was edited to incorporate an two strep-tags and three hemagglutinin (HA) tags for easier detection of the activated protein, as well as a novel 89bp multiple cloning site that adds

40 additional restriction sites to the plasmid, making it easier to modify after construction (***Table 3.2***). Plasmids were purified from E. coli using a Midiprep Plasmid Purification system (Qiagen Cat. 12145). A *Zfp286* sequence was isolated from an IMAGE clone (*Mammalian Gene Collection* IMAGE ID 5365187) and modified with two double-stranded linker oligonucleotides on either side (***Table 3.5***), allowing for the insert to be ligated into the modified *pTRE-tight* vector.

Inducible small hairpin RNA vector

Small-interfering RNA (siRNA) oligos were designed (***Table 3.3***) and tested for knockdown efficiency in SH-EP cells. SH-EP cells were grown under normal growth conditions, and transfection proceeded using 5 nM concentrations of siRNA in HiPerFect Transfection Reagent (*Qiagen* Cat. 301704), following the standard protocol. Cells were incubated under normal growth conditions with Opti-MEM I Reduced Serum Medium (*Gibco* Cat. 31985) for 32 hours, then immediately collected for RNA and protein collection. Once siRNA were validated, the same designs were used to create shRNA-hairpins with BglII and HindIII recognition sites (***Table 3.4***) for ligation into the *pSUPERIOR.puro* vector (*OligoEngine* Cat. VEC-IND-0006) using the manufacturer's protocol and restriction enzymes from *New England Biolabs*. Samples were sequenced and returned with 100% identity (***Table 3.5***). Next, the plasmids were transfected into the Neuro2A cell line using Lipofectamine 2000 Transfection Reagent (*ThermoFisher* Cat. 11668027) with the manufacturer's protocol.

Results

Pattern of ZNF286A isoform expression compared to ZNF286B and REST

The previous study into *ZNF286A* and *ZNF286B* expression in tissues (***Chapter 2***) generated a pattern of expression using human tissue RNA. However, the *ZNF286A* primer sequences used to do those studies amplified both KRAB(+) and KRAB(-) isoforms, offering little insight into *ZNF286A*'s more nuanced activity. To remedy this, primers were designed to target regions specific to each isoform – the 5' UTR in the KRAB(+) transcript, and the 59 bp unique alternative splice region that was identified in the KRAB(-) transcript (***Figure 3.1***).

Surprisingly, *ZNF286A* (K+) and *ZNF286A* (K-) display remarkably similar patterns of expression in human tissues (***Figure 3.2***). The highest expression for both isoforms was detected in adult brain with slightly less expression in fetal brain. This contrasts with *ZNF286B*, which is overwhelmingly expressed in fetal brain, with significantly lower expression in adults. Consistent with previous studies (Jones and Meech, 1999), *REST* is expressed much more highly in non-neuronal tissues (***Figure 3.2)***.

In contrast, these genes and isoforms are differently expressed in established human cell lines (***Figure 3.3***). For example, *ZNF286A* (K+) and *REST* are both significantly expressed in SH-EP neuroblastoma and expressed at lower levels in JEG-3 placental cells, while this pattern was reversed for *ZNF286A* (K-) and *ZNF286B*. It is thus clear that while *ZNF286* isoforms may be co-expressed, in tissues, in individual cells – at least in immortalized cell lines – they are expressed in a more independent manner.

Zfp286 and REST expression during mouse development

Human tissue expression data suggested that *ZNF286A* and *ZNF286B* are expressed at high levels in fetal tissues (*Fig 3.2).* We were therefore interested to examine isoform expression over the course of development. As this task is complicated in human tissues, where both *ZNF286A* and *ZNF286B* can be monitored, measuring gene expression in mouse tissues throughout embryonic development is relatively straightforward. Therefore, *Zfp286* (K+), *Zfp286* (K-), and *Rest* transcript levels were

examined in mouse embryos from E13.5 to E18.5, in RNA from whole dissected head and body (***Figure 3.4***). Expression for all three genes was detected as highest in the head at embryonic day 14.5 (E14.5) and E15.5, a time during which a massive level of neurogenesis, and the first waves of neuronal differentiation, are occurring in the brain. During this period, the hippocampus reaches its peak of neurogenesis, and the hippocampal commissure appears (Ashwell et al., 1996; Finlay and Darlington, 1995). The cortex, overall, also growing and differentiating rapidly during this same period. The very high levels of *Zfp286* expression at this timepoints is thus consistent with a role in neural differentiation, something that has already been confirmed with *Rest* (Ernsberger, 2012; Rockowitz and Zheng, 2015). Our data indicate that in mouse, the *Zfp286* (K-) isoform is dominantly expressed throughout mouse development, correlating well with the expression of *Rest*; similarly, although the (K-) isoform of *Zfp286*, was expressed at very low levels throughout mouse development, this transcript is also expressed most highly during the E14.5-E15.5 interval. The data also suggested a similar temporal pattern of expression for *Zfp286* isoforms and *Rest* in these tissue samples.

ZNF286 and REST expression during human neuronal differentiation in vitro

These data suggest a dominant role for the *Zfp286* (K+) isoform during mouse brain development, but also showed that the (K-) isoform is present in brain at the same developmental timepoints. However, since the embryonic samples include cell types other than neurons, questions regarding the specific roles of the *Zfp286* isoforms during neuronal differentiation could not be addressed. Furthermore, since *ZNF286B* does not exist in mice, the similarities and differences between the *ZNF286A* (K-) isoform and the naturally (K-) human duplicate cannot be examined in mice.

Fortunately, neuronal differentiation can be modeled in the human SH-SY5Y neuroblastoma cell line. SH-SY-5Y cells represent an undifferentiated neuronal progenitor state, and these cells can be induced to differentiate into mature neurons through retinoic acid treatment (Sidell et al., 1983; Encinas et al., 2000). We therefore tested expression of *ZNF286A, ZNF286B,* and *REST* in SH-SY-5Y samples over the course of neuronal differentiation (***Figure 3.5A***).

We observed that *REST* expression increases over the course of SH-SY5Y differentiation (***Figure 3.5B***). At first glance, this appears to contradict the observed decrease in REST protein levels by Western blots over the course of SY-SY5Y differentiation (***Chapter 2***). However, this result is actually expected, as it has been shown that *REST* transcription levels increase after neuroblast differentiation but are degraded before translation, leading to decreasing protein levels over the course of differentiatiation than what would be predicted based solely on the RNA expression (Singh et al., 2011). Meanwhile, expression of *ZNF286A* (K+) increases over differentiation (***Figure 3.5B***) consistent with our previous western blots (***Chapter 2***), but *ZNF286A* (K-) was detected at highest levels in undifferentiated SH-SY5Y cells and dropped in abundance over differentiation. The (K-) isoform is thus more similar to *REST* in that regard. In contrast, *ZNF286B* increased throughout differentiation, very much like the *(K+)* isoform of the *ZNF286A* gene.

These results showed that – in this cell line, at least – the KRAB(+) and KRAB(-) isoforms of *ZNF286A* are not co-expressed throughout differentiation, despite the fact their co-expression during mouse development (***Figure 3.4***) and human fetal brain (***Figure 3.2)***. These data must be interpreted with some caution since gene expression in the SH-SY-5Y cells may not necessarily be identical to isolated neurons differentiating *in vivo*. However, the contrast may be explained by the fact that the tissue samples include multiple cell types including neurons in different states of differentiation, whereas the neurons differentiated *in vitro* represent a single cell type differentiating in a coordinated time course and state.

These data suggest that, although *ZNF286A* (K-) and *ZNF286B* share very similar sequences and domain structures, their functional role during neuron differentiation is likely to be distinct. In this model, the (K-) version of the ZNF286A protein is co-expressed with REST, while the K+ isoform and *ZNF286B* are up-regulated as neuron differentiation proceeds. The data also suggest a functional cooperation between *ZNF286A* (K+) and *ZNF286B,* a hypothesis that is also supported by siRNA knockdown and Western blot data (***Chapter 2***), which also suggest similar, and perhaps cooperative roles for the two human proteins.

Alternative splicing of REST over the course of neuronal differnentiation in vitro

One further part of this overall puzzle relates to the alternative splicing of *REST* throughout neuron differentiation. The *REST4* truncated isoform retains the first five zinc fingers and nuclear localization signal of full-length REST (Palm et al., 1999; Coulson et al., 2000), and has unique expression patterns from REST in different cell types (Spencer et al., 2006; Uchida et al, 2010). While REST4 does not actually repress transcription, or does so very weakly (Magin et al., 2002; Lee et al., 2000), it functions as a full-length REST inhibitor by competing for NRSE binding sites (Shimojo et al., 1999; Tabuchi et al., 2002).

We therefore tested the expression of full-length *REST* and *REST4* isoforms throughout neuron differentiation at the same stages in which *ZNF286* gene and isoform expression was tested (***Figure 3.5B***) using previously published primers demonstrated to amplify both isoforms of REST (Palm et al., 1999). We also included RNA from the SH-EP cell line, in which functional assays for *ZNF286A* and *ZNF286B* were conducted (***Chapter 2***). These results confirm that *REST* alternative splicing is differentially regulated over neuron differentiation, with both full-length *REST* and truncated *REST4* isoforms being expressed at early stages, and *REST4* being expressed exclusively when SH-SY-5Y cells are fully differentiated (at day 10, ***Figure 3.5C***), consistent with published reports (Palm et al., 1999). Importantly, the data show that in the SH-EP cells we tested functionally, only the full-length *REST* isoform is detectably expressed, which is also consistent with published reports (Raj et al., 2011; Hamelink et al., 2004). This information may be critical in terms of understanding the relationships between *REST, ZNF286A* isoforms, and *ZNF286B*.

Tools for further study

Inducible over-expression Zfp286 (K-) construct

In order to better test the cellular functions of Zfp286 as they relate to cellular function and neuronal differentiation, we began construction of a Tet-inducible plasmid **(*Table 3.2*)** capable of over-expressing Zfp286 in vitro or (ultimately) wild type mice transfected with the vector. This plasmid utilizes a truncated *Zfp286* insert sequence derived from the Mammalian Gene Collection that lacks the KRAB domain – unfortunately no IMAGE clones currently exist that can provide *Zfp286* sequences with the KRAB domain included, and so restriction sites were added to the *pTRE-tight* vector to allow artificially synthesized KRAB oligos to be inserted in frame with the *Zfp286* plasmid sequence, effectively recreating the naturally occurring gene. The *pTRE-tight* vector was further modified to include two strep-tags and three HA-tags, allowing for the Zfp286 proteins induced from this plasmid to be detected using antibodies – as of now, there are no reliable Zfp286 antibodies that are commercially available, and this construct allows us to experiment on our proteins without possessing such an antibody.

With this plasmid, we will be able to conduct future experiments allowing us to over-express Zfp286 (K-) in a variety of cell lines (ie. Neuro2A before and after differentiation with retinoic acid) and thereby evaluate the functions of the Zfp286 (K-) isoform.

Inducible small hairpin RNA (shRNA)

In order to better understand the functions of *ZNF286A*, *ZNF286B*, and *REST*, the expression of each of these genes was knocked down using specific siRNAs (***Table 3.3***). The results of these knockdowns in *ZNF286A* and *ZNF286B* were discussed in length in Chapter 2 of this thesis. However, this form of non-stable interference can lead to data that is difficult or impossible to modulate. However, inducible shRNA vectors provide a much more convenient method by which to regulate protein expression in stable mammalian cell lines (Szulc et al., 2006; Laurenti et al., 2010).

Using the same siRNA designs that were previously validated, we designed shRNA hairpin constructs, which would be ligated as oligonucleotides into the *pSUPERIOR.puro* vector at the location of a HindIII restriction site (***Table 3.4***). *pSUPERIOR.puro* allows for the inducible expression of genes using the doses of tetracycline, meaning that the activation of the shRNA hairpins could be carefully regulated in order to knock down the expression of *ZNF286A, ZNF286B*, and *REST* in cell lines that have been stably transfected with the vector.

All eight shRNA were correctly ligated into *pSUPERIOR.puro* and confirmed by sequencing (***Table 3.5***). One of these vectors, *pSUPERIOR.puro_sh-REST-1*, was stably transfected into the Neuro2A cell line, which develops neuron-like properties when differentiated with retinoic acid (Mao et al., 2000). The same vector was also transfected successfully into the human U-2 OS (ATCC® HTB-96) osteosarcoma.

Discussion

This analysis of KRAB(+) and KRAB(-) isoforms of human *ZNF286A* and mouse *Zfp286* has revealed several important findings. First, while the two isoforms that they are similarly expressed in human and mouse whole-tissue RNA, the human isoforms, at least are differentially expressed in isolated cell types, including during neuronal differentiation *in vitro.* In contrast, *ZNF286B,* which is structurally very similar to *ZNF286A* (K-), is closely co-expressed with *ZNF286A* (K+) and appears to function very similarly to, and perhaps cooperatively with, the ancestral full-length protein. While understanding the functions of these proteins definitively will require further study, our data suggest that the (K+), (K-) and human-specific ZNF286B protein serve distinct functional roles.

The constructs I have generated to tag individual isoforms and to knockdown the individual human proteins stably in human and mouse cells provide new tools to address these hypotheses definitively. For example, future experiments can utilize the inducible shRNA vectors that have been constructed as a part of this project to test the roles of each protein at specific steps in neuronal differentiation. Inducible knockdowns of *REST* can now be carried out in a human non-neuronal (U-2 OS) and a mouse neuronal (Neuro2A)

cells lines, and these constructs, in addition to the ZNF286/Zfp286 vectors can now be introduced into SH-SY5Y or SK-N-SH neuroblasts for temporal control of gene knockdown during differentiation.

Data from Chapter 2 of this thesis suggested that the inclusion or omission of the KRAB domains in ZNF286A and ZNF286B proteins does not necessarily determine distinct functions. Indeed, ZNF286B appears to function as a repressor, even without the KRAB domain, at least in SH-EP cells. This function may indeed be context-dependent, as has been demonstrated for REST (Perera et al., 2015; Spencer et al., 2006). How the KRAB-less duplicate of the human protein carries out its repressive function is presently not understood. It is possible that both ZNF286A and B proteins include repressive domains aside from the KRAB, for example, within their shared "tether" regions. In this case, one would expect the (K-) versions of ZNF286A to also possess intrinsic repressor function. A second possibility is that the additional protein segment co-opted by ZNF286B from its *FOX03B* insertion has conferred a novel repressive role. However, it is also possible that the ZNF286B protein (and by analogy, possibly the protein encoded by the *ZNF286A* (K-) isoform) interacts with ZNF286A (K+) protein as a dimer, with the repressive function of K- proteins dependent on the KRAB-containing ancestral partner. The very similar functions detected after siRNA knockdown of the two human genes is consistent with such a cooperative function.

Whatever the specific mode of interaction, together our data do strongly suggest that, in contrast to serving as a "dominant negative" version of the (K+) full-length ZNF286A protein with opposite interactions to REST, ZNF286B and the (K-) isoform of the ancestral proteins may serve more complex functions than previously imagined. Since the isoforms of ZNF286A are not distinguishable on the protein level and cannot be targeted specifically, the functions of these individual proteins may be best addressed through tagged protein expression constructs, such as the ZNF286A (K-) isoform expression construct described in this study.

Further studies, for example, with tagged constructs, should focus on identification of binding partners for the (K+) and (K-) versions of ZNF286A or its unique mouse counterpart, to better understand the regulatory functions and possible differential interactions with REST during neuron development. The data and tools

described here will streamline this analysis permitting a new understanding of ZNF286 gene function and the novel human functions that arose through creation of *ZNF286B*.

References Cited

Ashwell KW, Waite PM, Marotte L. **1996**. Ontogeny of the projection tracts and commissural fibres in the forebrain of the tammar wallaby (Macropus eugenii): timing in comparison with other mammals. *Brain Behav Evol* 47:8–22.

Bellefroid E, Bourguignon C, Bouwmeester T, Rausch O, Blumberg B, Pieler T. **1997**. Transcription regulation and alternative splicing of an early zygotic gene encoding two structurally distinct zinc finger proteins in Xenopus laevis. *Mech Dev* 63:99-108.

Cheunga YT, Laua WK, Yua MS, Laia CS, Yeunga SC, Soa KF, Chang RC. **2009**. Effects of all-trans-retinoic acid on human SH-SY5Y neuroblastoma as in vitro model in neurotoxicity research. *NeuroToxicology* 30(1):127–135.

Coulson JM, Edgson JL, Woll PJ, Quinn JP. **2000**. A splice variant of the neuron-restrictive silencer factor repressor is expressed in small cell lung cancer: a potential role in derepression of neuroendocrine genes and a useful clinical marker. *Cancer Res* 60(7):1840-1844.

Encinas M, Iglesias M, Liu Y, Wang H, Muhaisen A, Ceña V, Gallego C, Comella JX. **2000**. Sequential Treatment of SH-SY5Y Cells with Retinoic Acid and Brain-Derived Neurotrophic Factor Gives Rise to Fully Differentiated, Neurotrophic Factor-Dependent, Human Neuron-Like Cells. *Journal of Neurochemistry* 75: 991–1003.

Ernsberger U. **2012**. Regulation of gene expression during early neuronal differentiation: evidence for patterns conserved across neuron populations and vertebrate classes. *Cell Tissue Res* 348(1):1-27.

Finlay BL, Darlington RB. **1995**. Linked regularities in the development and evolution of mammalian brains. *Science* 268:1578–1584.

Hamelink C, Hahm SH, Huang H, Eiden LE. **2004**. A restrictive element 1 (RE-1) in the VIP gene modulates transcription in neuronal and non-neuronal cells in collaboration with an upstream tissue specifier element. *J Neurochem* 88(5):1091-101.

Hsu T, Gogos JA, Kirsh SA, Kafatos FC. **1992**. Multiple zinc finger forms resulting from developmentally regulated alternative splicing of a transcription factor gene. *Science* 257(5078):1946-1950.

Jones FS, Meech R. **1999**. Knockout of REST/NRSF shows that the protein is a potent repressor of neuronally expressed genes in non-neural tissues. *Bioessays* 21(5):372-6.

Keren H, Lev-Maor G, Ast G. **2010**. Alternative splicing and evolution: diversification, exon definition and function. *Nat Rev Genet* 11:345-355.

Laurenti E, Barde I, Verp S, Offner S, Wilson A, Quenneville S, Wiznerowicz M, MacDonald HR, Trono D, Trumpp A. **2010**. Inducible Gene and shRNA Expression in Resident Hematopoietic Stem Cells In Vivo. *Stem Cells* 28(8):1390–1398.

Lee JH, Chai YG, Hersh LB. **2000**. Expression patterns of mouse repressor element-1 silencing transcription factor 4 (REST4) and its possible function in neuroblastoma. *J Mol Neurosci* 15(3):205-14.

Livak KJ, Schmittgen TD. **2001**. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods* 25:402–408.

Magin A, Lietz M, Cibelli G, Thiel G. **2002**. RE-1 silencing transcription factor-4 (REST4) is neither a transcriptional repressor nor a de-repressor. *Neurochem Int* 40(3): 195–202.

Mao AJ, Bechberger J, Lidington D, Galipeau J, Laird DW, Naus CCG. **2000**. Neuronal differentiation and growth control of Neuro-2a cells after retroviral gene delivery of Connexin43. *J Biol Chem* 275:34407-34414.

Nilsen TW, Graveley BR. **2010**. Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463:457-463.

Palm K, Metsis M, Timmusk T. **1999**. Neuron-specific splicing of zinc finger transcription factor RESTrNRSFrXBR is frequent in neuroblastomas and conserved in human, mouse and rat. *Mol Brain Res* 72:30-39.

Perera A, Eisen D, Wagner M, Laube SK, Künzel AF, Koch S, Steinbacher J, Schulze E, Splith V, Mittermeier N, Müller M, Biel M, Carell T, Michalakis S. **2015**. TET3 is recruited by REST for context-specific hydroxymethylation and induction of gene expression. *Cell Rep* 11(2):283-94.

Raj B, O'Hanlon D, Vessey JP, Pan Q, Ray D, Buckley NJ, Miller FD, Blencowe BJ. **2011**. Cross-regulation between an alternative splicing activator and a transcription repressor controls neurogenesis. *Mol Cell* 43(5):843-50.

Reick M, Garcia J, Dudley C, McKnight SL. **2001**. NPAS2: An Analog of Clock Operative in the Mammalian Forebrain. *Science* 293(5529):506-509.

Rockowitz S, Zheng D. **2015**. Significant expansion of the REST/NRSF cistrome in human versus mouse embryonic stem cells: potential implications for neural development. *Nucleic Acids Res* 43(12):5730-43.

Shimojo M, Paquette AJ, Anderson DJ, Hersh LB. **1999**. Protein Kinase A Regulates Cholinergic Gene Expression in PC12 Cells: REST4 Silences the Silencing Activity of Neuron-Restrictive Silencer Factor/REST. *Mol Cell Bio* 19(10):6788–6795.

Sidell N, Altman A, Haussler MR, Seeger RC. **1983.** Effects of retinoic acid (RA) on the growth and phenotypic expression of several human neuroblastoma cell lines. *Exp Cell Res* 148(1): 21-30.

Singh A, Rokes C, Gireud M, Fletcher S, Baumgartner J, Fuller G, Stewart J, Zage P, Gopalakrishnan V. **2011**. Retinoic acid induces REST degradation and neuronal differentiation by modulating the expression of SCF(β-TRCP) in neuroblastoma cells. *Cancer* 117(22):5189-202.

Spencer EM, Chandler KE, Haddley K, Howard MR, Hughes D, Belyaev ND, Coulson JM, Stewart JP, Buckley NJ, Kipar A, Walker MC, Quinn JP. **2006**. Regulation and role of REST and REST4 variants in modulation of gene expression in in vivo and in vitro in epilepsy models. *Neurobiol Dis* 24(1):41-52.

Szulc J, Wiznerowicz M, Sauvain MO, Trono D, Aebischer P. **2006**. A versatile tool for conditional gene expression and knockdown. *Nat Methods* 3(2):109-16.

Tabuchi A, Yamada T, Sasagawa S, Naruse Y, Mori N, Tsuda M. **2002**. REST4-Mediated Modulation of REST/NRSF-Silencing Function during BDNF Gene Promoter Activation. *Biochem and Biophys Res Commun* 290:415–420.

Uchida S, Hara K, Kobayashi A, Funato H, Hobara T, Otsuki K, Yamagata H, McEwen BS, Watanabe Y. **2010**. Early Life Stress Enhances Behavioral Vulnerability to Stress through the Activation of REST4-Mediated Gene Transcription in the Medial Prefrontal Cortex of Rodents. *J Neurosci* 30(45):15007-15018.

Yeo GW, Van Nostrand E, Holste D, Poggio T, Burge CB. **2005**. Identification and analysis of alternative splicing events conserved in human and mouse. *Proc Natl Acad Sci USA* 102:2850–2855.

FIGURES

Figure 3.1. Structure of human ZNF286 genes. **(A)** In mice, *Zfp286A* codes for a *Krüppel*-type zinc finger transcription factor with a ten-finger long DNA-binding domain tethered to its KRAB domain. **(B)** This structure is conserved in mammals and marsupials, along with a "KRAB-minus" alternatively-spliced isoform that lacks the KRAB domain. **(C)** *ZNF286A* is the human paralog of *Zfp286*, including a KRAB-A and KRAB-B domain. **(D)** Like the mouse paralog, *ZNF286A* also codes for a KRAB-minus isoform. **(E)** *ZNF286B* similarly lacks the KRAB-coding domain that is present in *ZNF286A*, incorporating a novel sequence in its place. The locations of the 5' UTR ("5"), pre-KRAB region ("P"), KRAB-A domiain ("A"), the tether region ("T"), zinc finger domain ("F"), and 3' UTR ("3") regions are represented with letters above each exon. In addition, the locations of primers used in this study are outlined in red, with an asterisk used to indicate the location of primer sequences that are isoform-specific.

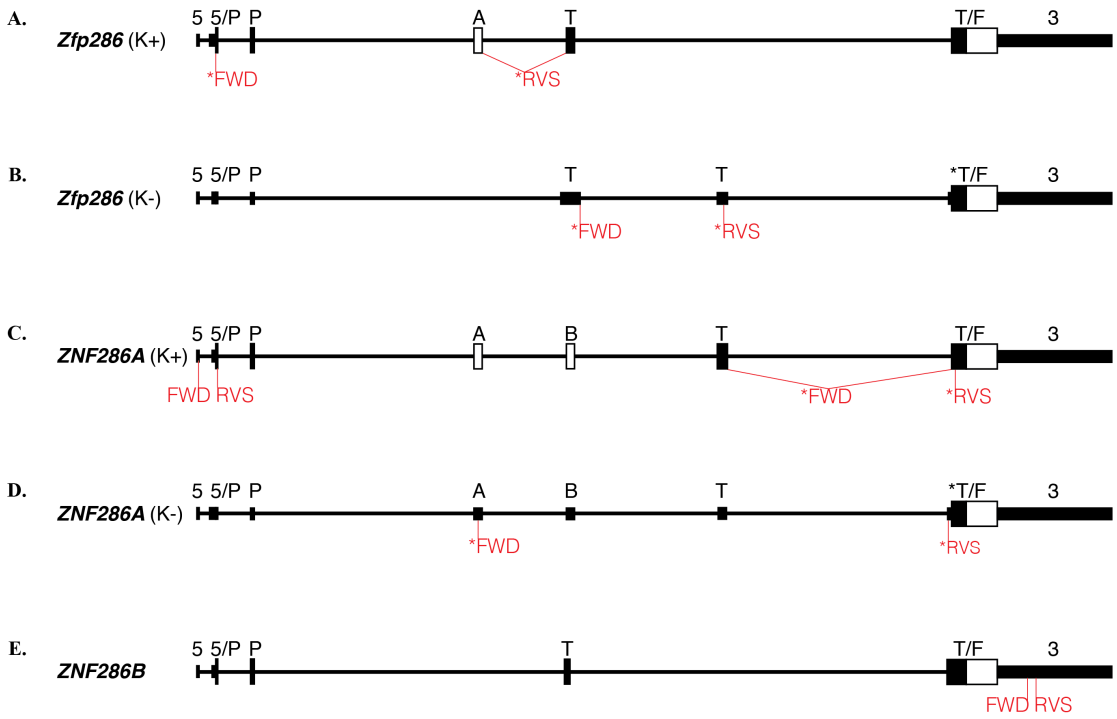


Figure 3.2. Relative expression of ZNF286 transcripts in adult and fetal human tissues. The expression of both **(A)** *ZNF286A*(K+) and **(B)** *ZNF286A* (K-) transcripts in various human tissues was determined by performing QPCR on each sample using sequence-specific primers for each gene. The $2^{-\Delta Ct}$ values for each sample were calculated in relation to *GAPDH* positive controls (Livak and Schmittgen, 2001) and then normalized against the highest-expressed tissue for each primer set, in order to minimize primer-specific effects when comparing both genes. The pattern for both isoforms is mostly comparable across tissues. **(C)** This is in contrast to the human duplicate *ZNF286B*, which is expressed much higher in fetal tissues than in adult tissues. **(D)** REST expression is higher in non-neuronal tissues.

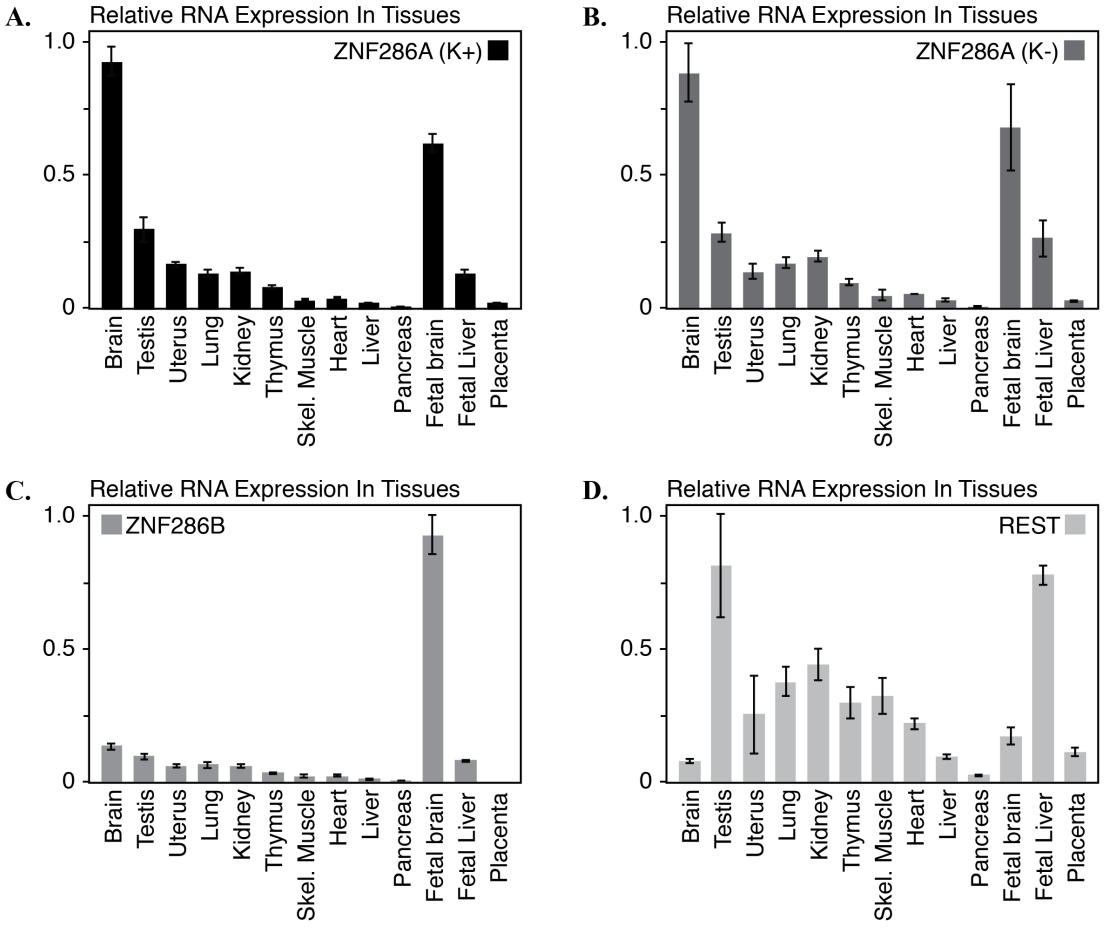


Figure 3.3. Relative expression of ZNF286 and REST transcripts in human cell lines. The relative expression of select gene transcripts in various human cell lines, as calculated through QPCR and subsequent analysis using *GAPDH* as a positive control. Relative gene expression was calculated for **(A)** KRAB-plus *ZNF286A* isoform, **(B)** KRAB-minus *ZNF286A* isoform, **(C)** *ZNF286B*, and **(D)** *REST* transcripts. Each data set has been normalized against the cell line with the highest-expression in that set. For each gene, expression is highest in either SHEP or JEG-3 cell lines.

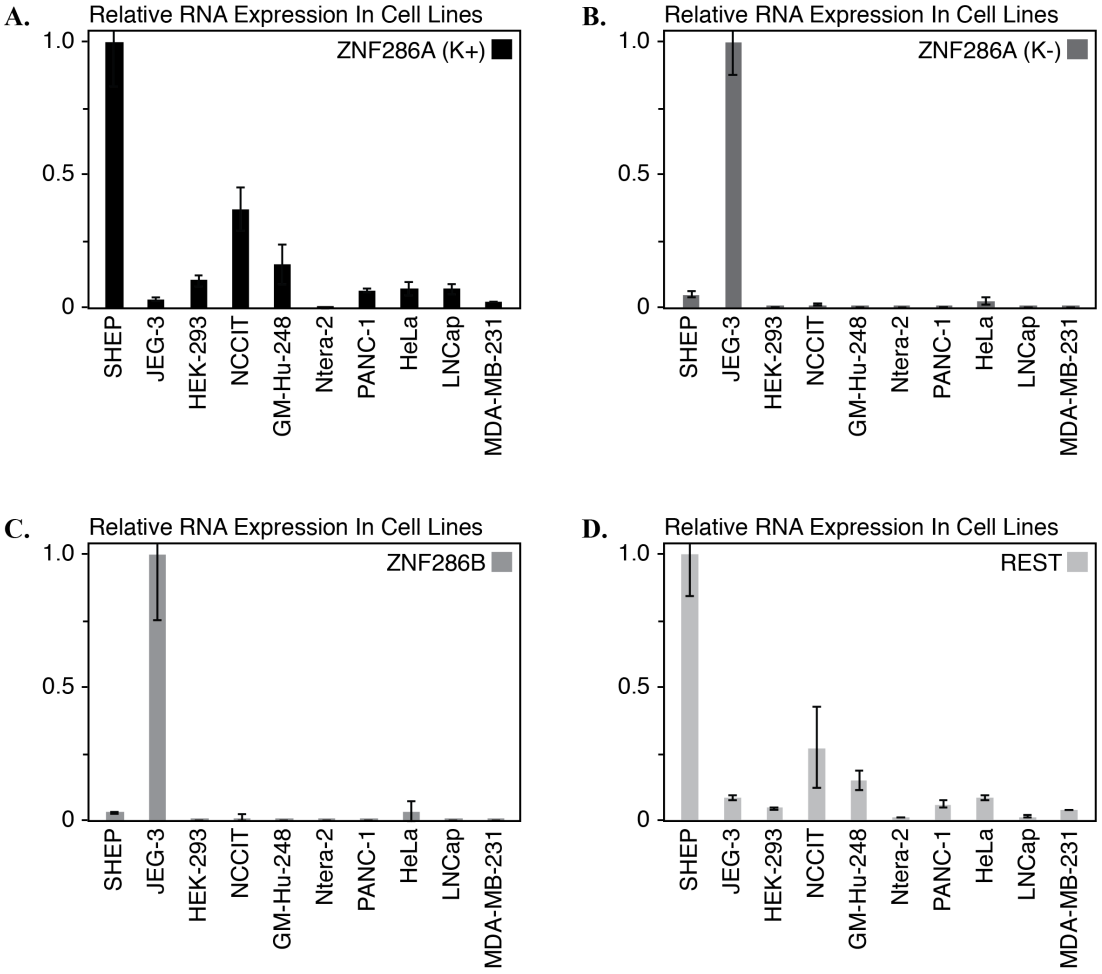


Figure 3.4. Relative expression of Zfp286 and Rest transcripts in embryonic mouse tissues. The relative expression of **(A)** *Zfp286* (K+), **(B)** *Zfp286* (K-), and **(C)** *Rest* was calculated using whole tissue RNA collected from whole mouse heads and bodies collected at various stages of embryonic development from E13.5 to E18.5. The patterns of expression between each transcript very closely coincide.

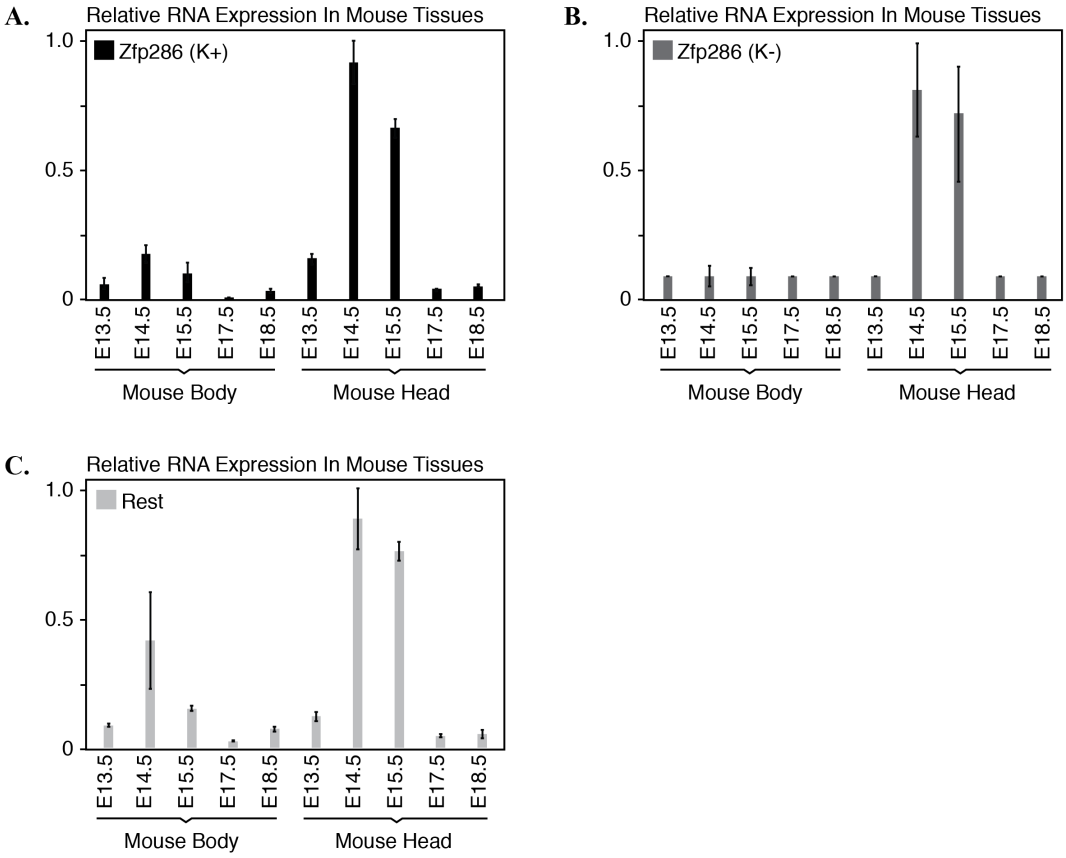
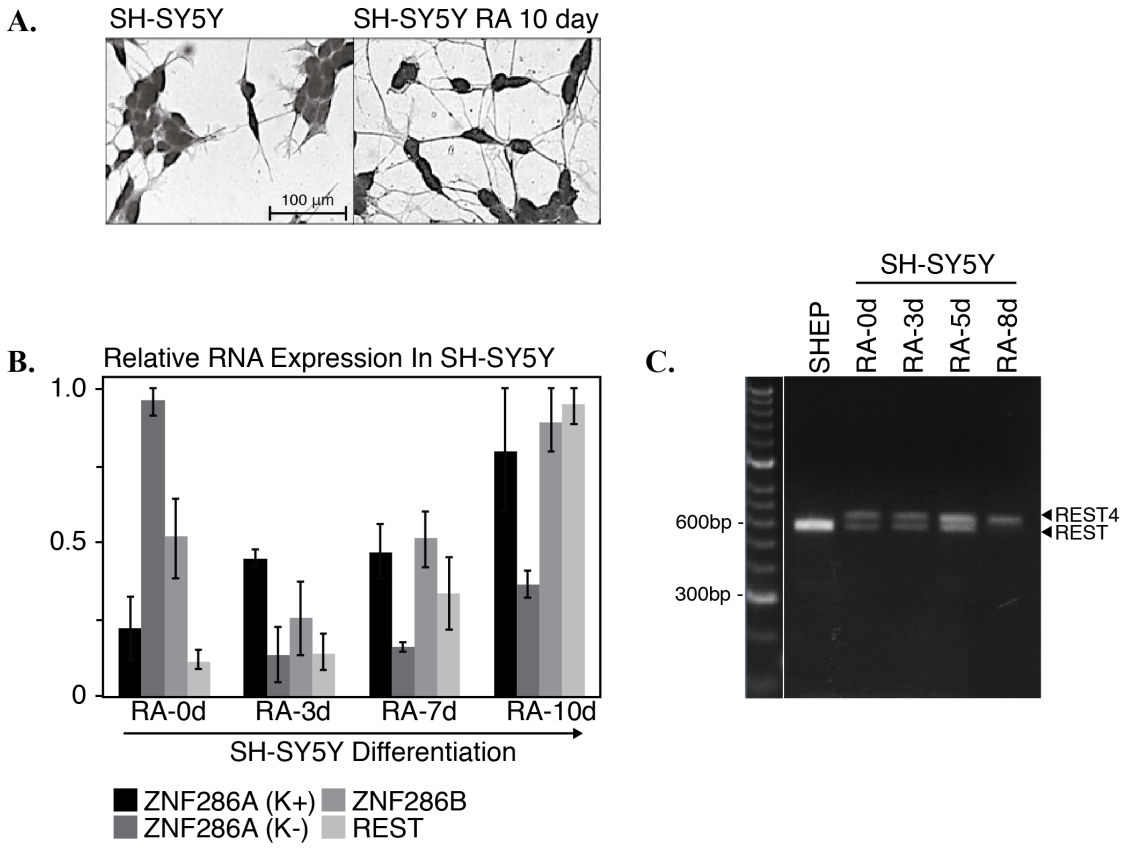


Figure 3.5. Relative expression of ZNF286 transcripts during neural differentiation. **(A)** SH-SY5Y neuroblastoma are induced to differentiate after the addition of retinoic acid (RA). Cells are considered fully differentiated after treatment for 7-10 days. **(B)** The relative expression of *ZNF286A* (K+), *ZNF286A* (K-), *ZNF286B*, and *REST* transcripts was calculated in the differentiating SH-SY5Y cell line over the course of differentiation using retinoic acid. **(C)** Semi-QPCR show the expression of *REST* and *REST4* transcripts in differentiating SH-SY5Y, using primers from Palm et al. (1999). As REST expression goes down after differentiation, REST4 remains. Note that SHEP cells only contain the full-length REST.



TABLES

Table 3.1. Primer sequences for human (Hu) and mouse (Ms).

Gene Target	FWD Primer (5′-to-3′)	RVS Primer (5′-to-3′)
Ms-Gapdh	Qiagen Cat. QT01658692	
Ms-Rest	TGGCCACCCAGGTGATGGGG	TGAGCTGAGGGGCTGCCAGT
Ms-Zfp286 (K+)	ACGGGTTTGGCCGAAGTGCC	TGGGAAGCCAGAGCGAGACC
Ms-Zfp286 (K-)	GCAGTGACTCAGCTATTCGTCT	CTGACGCTCTTACTGGATGGT
Hu-GAPDH	Qiagen Cat. QT01192646	
Hu-REST	TGGGGCCTGCTCCCACAGAG	GCAGGCCGTATCTGGGCACC
Hu-ZNF286A	TGTGAGGCCCGGGATGGGAG	TGCCTCAGTCCTACGGCGCT
Hu-ZNF286A (K+)	GGAAGCTGGATCCTGCACAA	TGTGCTCAACAAAAAGGCAGT
Hu-ZNF286A (K-)	GGAAGCTGGATCCTGCACAA	TGTGCTCAACAAAAAGGCAGT
Hu-ZNF286B	GCCATTCAGTGCATATTCAACACCAG	TACACTCATAGGGTTTCTCTCCAGTGTGA
Hu-REST/REST4	Palm et al. (1999) h-p2-s GTGACCGCTGCGGCTACAATACTAA	Palm et al. (1999) h-p8-as GGACAAGTAGGATGCTTAGATTTGA

Table 3.2. pTTS plasmid outline.

pTTS-Zfp286(KRAB-); 3897bp

```
CTCGAGTTTACTCCCTATCAGTGATAGAGAACGTATGTCGAGTTTACTCCCTATCAGTGATAGAGAACGATGTCGAGTTTACTCCCTATCAGTGATAGAGAA
CGTATGTCGAGTTTACTCCCTATCAGTGATAGAGAACGTATGTCGAGTTTACTCCCTATCAGTGATAGAGAACGTATGTCGAGTTTATCCCTATCAGTGATA
GAGAACGTATGTCGAGTTTACTCCCTATCAGTGATAGAGAACGTATGTCGAGGTAGGCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAAC
CGTCAGATCGCCTGGAGAATTCGAGCTCGGTACCCGGGGATCCTCTAGTCAGCTGACGCGTGCTAGCGCGGCCGCATCGAT^

      ^AGCTggttgctttattgctgcgATCTCCTAACTCCACAAAGGCATGCTTCCATCTTGATTCATCTTCACCTGTTATAGGACTGTGTTACTGTT
      GAGCACAGAAAACAAAGAACATGTGCATTTGCAATGTCTTTTCAGACATGGAAACCAGACCTGAGAGCAAGGAGTCAACTTCAGTGCTAGACTT
      TTCCAAAGAAGAGTCAAACAGACTGACAGAAAATACTGCTTCTGACTCCGACATTGAAACTGCTCTTGAATGTGAGCGTTGGCTAGAACATCGT
      CAAAGAAATCGGGAGAGGTACCCAAGACAAATGTCTGCTCACAGGAACTCCCTCCCTGCGCAAGAAGATGGTGAACGTGATGTATATTGGAAAA
      GCTTCAGTCAGAAACCTCTCCTTACAACTGATGACAGAGACCCCAAAGGATCCTGGGACGTTCACACGCTTGGGAAAAGACTGAAGCAGAAATC
      AAAGTTAATGACAAAGCAGAGACCTTATAAAGAGAAAAAACCTCATAAGTGTGATGAGTGTGGCGAACTCTTCACTTACCGTTCGGTACTTATC
      CGACACCAGAGAGTTCACACTGGAGAGAAACCTTATACCTGCAGTGAGTGCGGGAAGTCCTTCAGCCACAGAGCCAATCTGACTAAGCACCAGA
      GAACCCACACTAGGATTTCTCTTCGAGTGTGGCGAGTGCAAGAAAACCTTCATAGAAAGTTCATCCCTTGAGGTACATCAGAGAATCCACATCGG
      GGAGAGACCTTACGAATGCGCCGAGTGTGGGAAAGGCTTCAACCGAAGTACACACCTTGCCCAGCATCAGCTGATTCATACTGGGGTGAAGCCT
      TATGAATGTAACGAATGTGACAAAGCTTTCATTCACTCATCAGCTCTCATTAAGCATCAAAGAACTCATACTGGAGAGAAGCCCTATAAATGTC
      AAGACTGTGGGAAAGCCTTCAGCCACTGCTCATCTCTTACGAAGCACCAGAGGGTTCACACGGGGGAAAGGCCATATGAGTGCAGTGAGTGTGG
      GAAGACTTTCAGTCAGAGCACGCACCTTGTTCAGCACCAGAGAATCCACACTGGAGAAAAGCCCTACGAATGTAACGAATGTGGGAAAACATTC
      AGCCGGgaggtgcaagtgagtgttGAT^

^ATCTCTAGATCTAGACCATATGGGTGCATGGAGCCACCCGCAGTTCGAAAAGTGGAGCCACCCGCAGTTCGAAAAGGGTGCATACCCATACGATGTTCCAG
ATTACGCTTACCCATACGATGTTCCAGATTACGCTTACCCATACGATGTTCCAGATTACGCTTAATCTAGACAGCCATACCACATTTGTAGAGGTTTTACTT
GCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAA
AGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCCTCGAGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCG
GTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGT
AAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTA
TAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGC
GTGGCGCTTTCTCAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGC
TGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTAT
GTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGA
AAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAA
GAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAG
ATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCA
GCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATA
CCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATC
CAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCG
TCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCG
ATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTG
ACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGA
ACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCC
AACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGT
TGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAA
ATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCC
TTTCGTCTTCA
```

In order from start to end, the following regions are outlined in the above sequence:

TRE	Tet-response Element
MCS	Multiple Cloning Site with EcoRV (GAT^ATC) site
XbaI site	T^CTAGA (insert removal marker)
NdeI site	CA^TATG (cloning selection marker)
Gly-Ala spacer	GGTGCA
Strep-tag x2	TGGAGCCACCCGCAGTTCGAAAAG
Gly-Ala spacer	GGTGCA
HA-tag x3	TACCCATACGATGTTCCAGATTACGCT
Stop codon	TAA
XbaI site	T^CTAGA (insert removal marker)

The indented region represents the MGC Image Clone insert amplified out of mouse cDNA. The following "linker" sequences were added to each side of the insert to allow it to ligate to the vector:

```
IMAGE Clone FWD Linker:   5'-pTTS-   (HindIII)-AGCTTgattgctttattgctgA            -Zfp286-3'
                          3'-pTTS-                Actaacgaaataacgac TCTAG-(BglII)-Zfp286-5'

IMAGE Clone RVS Linker:   5'-Zfp286-(HpaII)-CGGgaggtgcaagtgagtgttGAT             -pTTS-3'
                          3'-Zfp286-          Cctccacgttcactcacaa CTA-(EcoRV)     -pTTS-5'
```

In addition, the following primers were used for sequencing and ligation confirmation:

pTTS_seq_FWD	5'-AGGCGTATCACGAGGCCCTTTCGT-3'
pTTS_seq_RVS	5'-TCAGCTCACTCAAAGGCGGTAATA-3'
pQE-For	5'-CCCGAAAAGTGCCACCTG-3' (Sequencing Primer)

Table 3.3. shRNA oligos for ligation into pSUPERIOR plasmid.

siRNA oligo name	Company, product ID	siRNA sequence (5'-to-3')
SI-286A-1	Qiagen Cat. SI04135600	CACCTACCATTCAGTGCTTAT
SI-286A-5	Qiagen Cat. SI03227448	TAGCAGTACGAACATTGTGAA
SI-286B-1	Qiagen Cat. SI04868115	CAGGCCCTTATAAGGGCACTA
SI-286B-2	Qiagen Cat. SI05457697	TACCATCAATTAAGTTTCATT
SI-286B-3	Qiagen Cat. SI05457704	TAACTCTTACCTATTACAGAA
SI-NRSF-1-1597	Proligo, 1300782-H	GTGCATACAGGAAGCAATTCAGA
SI-NRSF-2-2612	Proligo, 1300784-H	TAGTGCCTGTTAAAGATAGCTGG
SI-NRSF-3-2902	Proligo, 1300786-H	CAGACTGACAGTATAGTTTGTGA
SI-Zfp286-1	Not used in this study.	AACCATCCAGTAAGAGCGTCAGA
SI-Zfp286-2	Not used in this study.	CTCCTAACTCCACAAAGGCATGC
SI-Zfp286-3	Not used in this study.	TCCACAAAGGCATGCTTCCATCT
SI-Zfp286-4	Not used in this study.	ATCTTCACCTGTTATAGGACTGT

Table 3.4. shRNA oligos for ligation into pSUPERIOR plasmid.

shRNA oligo name	5'- BglII-SenseSequence-Hairpin-AntisenseSequence-HindIII -3'
sh-ZNF286A-1-S	GATCCCCACCTACCATTCAGTGCTTATTCAAGAGATAAGCACTGAATGGTAGGTTTTTTA
sh-ZNF286A-1-AS	AGCTTAAAAAACCTACCATTCAGTGCTTATCTCTTGAATAAGCACTGAATGGTAGGTGGG
sh-ZNF286A-5-S	GATCCCCAGCAGTACGAACATTGTGATTCAAGAGATCACAATGTTCGTACTGCTTTTTTA
sh-ZNF286A-5-AS	AGCTTAAAAAAGCAGTACGAACATTGTGATCTCTTGAATCACAATGTTCGTACTGCTGGG
sh-ZNF286B-1-S	GATCCCCAGGCCCTTATAAGGGCACTTTCAAGAGAAGTGCCCTTATAAGGGCCTTTTTTA
sh-ZNF286B-1-AS	AGCTTAAAAAAGGCCCTTATAAGGGCACTTCTCTTGAAAGTGCCCTTATAAGGGCCTGGG
sh-ZNF286B-2-S	GATCCCCACCATCAATTAAGTTTCATTTCAAGAGAATGAAACTTAATTGATGGTTTTTTA
sh-ZNF286B-2-AS	AGCTTAAAAAACCATCAATTAAGTTTCATTCTCTTGAAATGAAACTTAATTGATGGTGGG
sh-ZNF286B-3-S	GATCCCCAACTCTTACCTATTACAGATTCAAGAGATCTGTAATAGGTAAGAGTTTTTTTA
sh-ZNF286B-3-AS	AGCTTAAAAAAACTCTTACCTATTACAGATCTCTTGAATCTGTAATAGGTAAGAGTTGGG
sh-NRSF-1-S	GATCCCCGCATACAGGAAGCAATTCATTCAAGAGATGAATTGCTTCCTGTATGCTTTTTA
sh-NRSF-1-AS	AGCTTAAAAAGCATACAGGAAGCAATTCATCTCTTGAATGAATTGCTTCCTGTATGCGGG
sh-NRSF-2-S	GATCCCCGTGCCTGTTAAAGATAGCTTTCAAGAGAAGCTATCTTTAACAGGCACTTTTTA
sh-NRSF-2-AS	AGCTTAAAAAGTGCCTGTTAAAGATAGCTTCTCTTGAAAGCTATCTTTAACAGGCACGGG
sh-NRSF-3-S	GATCCCCGACTGACAGTATAGTTTGTTTCAAGAGAACAAACTATACTGTCAGTCTTTTTA
sh-NRSF-3-AS	AGCTTAAAAAGACTGACAGTATAGTTTGTTCTCTTGAAACAAACTATACTGTCAGTCGGG
sh-Zfp286-1-S	GATCCCCCCATCCAGTAAGAGCGTCATTCAAGAGATGACGCTCTTACTGGATGGTTTTTA
sh-Zfp286-1-AS	AGCTTAAAAACCATCCAGTAAGAGCGTCATCTCTTGAATGACGCTCTTACTGGATGGGGG
sh-Zfp286-2-S	GATCCCCCCTAACTCCACAAAGGCATTTCAAGAGAATGCCTTTGTGGAGTTAGGTTTTTA
sh-Zfp286-2-AS	AGCTTAAAAACCTAACTCCACAAAGGCATTCTCTTGAAATGCCTTTGTGGAGTTAGGGGG
sh-Zfp286-3-S	GATCCCCCACAAAGGCATGCTTCCATTTCAAGAGAATGGAAGCATGCCTTTGTGTTTTTA
sh-Zfp286-3-AS	AGCTTAAAAACACAAAGGCATGCTTCCATTCTCTTGAAATGGAAGCATGCCTTTGTGGGG
sh-Zfp286-4-S	GATCCCCCTTCACCTGTTATAGGACTTTCAAGAGAAGTCCTATAACAGGTGAAGTTTTTA
sh-Zfp286-4-AS	AGCTTAAAAACTTCACCTGTTATAGGACTTCTCTTGAAAGTCCTATAACAGGTGAAGGGG

Table 3.5. shRNA construct sequences confirmed by sequencing.

pSUPERIOR.puro (with no shRNA inserts)

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCA
AAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGAC
TCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTG
CCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGA
AGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCG
CTACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGC
GAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCGCG
CGTAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCATGGGGTC
GTGCGCTCCTTTCGGTCGGGCGCTGCGGGTCGTGGGGCGGGCGTCAGGCACCGGGCTTGCGGGTCATGCACCAGGTGCGCGGTCCTTCG
GGCACCTCGACGTCGGCGGTGACGGTGAAGCCGAGCCGCTCGTAGAAGGGGAGGTTGCGGGGCGCGGAGGTCTCCAGGAAGGCGGGCAC
CCCGGCGCGCTCGGCCGCCTCCACTCCGGGGAGCACGACGGCGCTGCCCAGACCCTTGCCCTGGTGGTCGGGCGAGACGCCGACGGTGG
CCAGGAACCACGCGGGCTCCTTGGGCCGGTGCGGCGCCAGGAGGCCTTCCATCTGTTGCTGCGCGGCCAGCCGGGAACCGCTCAACTCG
GCCATGCGCGGGCCGATCTCGGCGAACACCGCCCCCGCTTCGACGCTCTCCGGCGTGGTCCAGACCGCCACCGCGGCGCCGTCGTCCGC
GACCCACACCTTGCCGATGTCGAGCCCGACGCGCGTGAGGAAGAGTTCTTGCAGCTCGGTGACCCGCTCGATGTGGCGGTCCGGATCGA
CGGTGTGGCGCGTGGCGGGGTAGTCGGCGAACGCGGCGGCGAGGGTGCGTACGGCCCTGGGGACGTCGTCGCGGGTGGCGAGGCGCACC
GTGGGCTTGTACTCGGTCATGGTAAGCTAGCTTGGGCTGCAGGTCGAAAGGCCCGGAGATGAGGAAGAGGAGAACAGCGCGGCAGACGT
GCGCTTTTGAAGCGTGCAGAATGCCGGGCCTCCGGAGGACCTTCGGGCGCCCGCCCCGCCCCTGAGCCCGCCCCTGAGCCCGCCCCCGG
ACCCACCCCTTCCCAGCCTCTGAGCCCAGAAAGCGAAGGAGCAAAGCTGCTATTGGCCGCTGCCCCAAAGGCCTACCCGCTTCCATTGC
TCAGCGGTGCTGTCCATCTGCACGAGACTAGTGAGACGTGCTACTTCCATTTGTCACGTCCTGCACGACGCGAGCTGCGGGGCGGGGGG
GAACTTCCTGACTAGGGGAGGAGTAGAAGGTGGCGCGAAGGGGCCACCAAAGAACGGAGCCGGTTGGCGCCTACCGGTGGATGTGGAAT
GTGTGCGAGGCCAGAGGCCACTTGTGTAGCGCCAAGTGCCCAGCGGGGCTGCTAAAGCGCATGCTCCAGACTGCCTTGGGAAAAGCGCC
TCCCCTACCCGGTAGAATTCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTGTCACTAGGCGGGAACACCCAGCG
CGCGTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGGAGTGGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCA
CCATAAACGTGAAATGTCTTTGGATTTGGGAATCTTATAAGTTCCCTATCAGTGATAGAGATCT**A^AGCTT**ATCGATACCGTCGACCTC
GAGGGGGGGCCCGGTACCCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGT
GAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTC
ACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAG
AGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTC
ACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAAC
CGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCG
AAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGAT
ACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCC
AAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACA
CGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGC
CTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGA
TCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCC
TTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCA
CCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATC
AGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGG
CTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAA
GGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCA
GTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTC
CCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGT
TGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGT
GAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACA
TAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGA
TGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCC
GCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTG
TCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC

The digested *HindIII* site (**A^AGCTT**) is shown in bold and highlighted. It is into this restriction site that the shRNA constructs are ligated.

Table 3.5. (Continued)

pSUPERIOR.puro_sh-286A-1

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCA
AAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGAC
TCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTG
CCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGA
AGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCG
CTACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGC
GAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCGCG
CGTAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCATGGGGTC
GTGCGCTCCTTTCGGTCGGGCGCTGCGGGTCGTGGGGCGGGCGTCAGGCACCGGGCTTGCGGGTCATGCACCAGGTGCGCGGTCCTTCG
GGCACCTCGACGTCGGCGGTGACGGTGAAGCCGAGCCGCTCGTAGAAGGGGAGGTTGCGGGGCGCGGAGGTCTCCAGGAAGGCGGGCAC
CCCGGCGCGCTCGGCCGCCTCCACTCCGGGGAGCACGACGGCGCTGCCCAGACCCTTGCCCTGGTGGTCGGGCGAGACGCCGACGGTGG
CCAGGAACCACGCGGGCTCCTTGGGCCGGTGCGGCGCCAGGAGGCCTTCCATCTGTTGCTGCGCGGCCAGCCGGGAACCGCTCAACTCG
GCCATGCGCGGGCCGATCTCGGCGAACACCGCCCCCGCTTCGACGCTCTCCGGCGTGGTCCAGACCGCCACCGCGGCGCCGTCGTCCGC
GACCCACACCTTGCCGATGTCGAGCCCGACGCGCGTGAGGAAGAGTTCTTGCAGCTCGGTGACCCGCTCGATGTGGCGGTCCGGATCGA
CGGTGTGGCGCGTGGCGGGGTAGTCGGCGAACGCGGCGGCGAGGGTGCGTACGGCCCTGGGGACGTCGTCGCGGGTGGCGAGGCGCACC
GTGGGCTTGTACTCGGTCATGGTAAGCTAGCTTGGGCTGCAGGTCGAAAGGCCCGGAGATGAGGAAGAGGAGAACAGCGCGGCAGACGT
GCGCTTTTGAAGCGTGCAGAATGCCGGGCCTCCGGAGGACCTTCGGGCGCCCGCCCCGCCCCTGAGCCCGCCCCTGAGCCCGCCCCCGG
ACCCACCCCTTCCCAGCCTCTGAGCCCAGAAAGCGAAGGAGCAAAGCTGCTATTGGCCGCTGCCCCAAAGGCCTACCCGCTTCCATTGC
TCAGCGGTGCTGTCCATCTGCACGAGACTAGTGAGACGTGCTACTTCCATTTGTCACGTCCTGCACGACGCGAGCTGCGGGGCGGGGGG
GAACTTCCTGACTAGGGGAGGAGTAGAAGGTGGCGCGAAGGGGCCACCAAAGAACGGAGCCGGTTGGCGCCTACCGGTGGATGTGGAAT
GTGTGCGAGGCCAGAGGCCACTTGTGTAGCGCCAAGTGCCCAGCGGGGCTGCTAAAGCGCATGCTCCAGACTGCCTTGGGAAAAGCGCC
TCCCCTACCCGGTAGAATTCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTGTCACTAGGCGGGAACACCCAGCG
CGCGTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGGAGTGGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCA
CCATAAACGTGAAATGTCTTTGGATTTGGGAATCTTATAAGTTCCCTATCAGTGATAG**A^**GATCCCCACCTACCATTCAGTGCTTATTC
AAGAGATAAGCACTGAATGGTAGGTTTTTTA**^AGCTT**ATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAGCTTTTGTTCCCTT
TAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACAT
ACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC
AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCC
TCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATC
AGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATA
GGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTT
CCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGC
GCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGC
CCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAAC
AGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGG
TATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTT
TTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGG
AACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAA
ATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTC
GTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCG
CGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATC
CGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTA
CAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATG
TTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGC
ACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTA
TGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAA
CGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGC
ATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTT
GAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAG
AAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC

The digested *HindIII* site (**A^AGCTT**) is shown in bold, with the inserted shRNA construct located in between them. The sequence of interest is highlighted in yellow, and was sequenced to confirm proper insertion of the shRNA construct.

Table 3.5. (Continued)

pSUPERIOR.puro_sh-286A-5

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCA
AAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGAC
TCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTG
CCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGA
AGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCG
CTACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGC
GAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCGCG
CGTAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCATGGGGTC
GTGCGCTCCTTTCGGTCGGGCGCTGCGGGTCGTGGGGCGGGCGTCAGGCACCGGGCTTGCGGGTCATGCACCAGGTGCGCGGTCCTTCG
GGCACCTCGACGTCGGCGGTGACGGTGAAGCCGAGCCGCTCGTAGAAGGGGAGGTTGCGGGGCGCGGAGGTCTCCAGGAAGGCGGGCAC
CCCGGCGCGCTCGGCCGCCTCCACTCCGGGGAGCACGACGGCGCTGCCCAGACCCTTGCCCTGGTGGTCGGGCGAGACGCCGACGGTGG
CCAGGAACCACGCGGGCTCCTTGGGCCGGTGCGGCGCCAGGAGGCCTTCCATCTGTTGCTGCGCGGCCAGCCGGGAACCGCTCAACTCG
GCCATGCGCGGGCCGATCTCGGCGAACACCGCCCCCGCTTCGACGCTCTCCGGCGTGGTCCAGACCGCCACCGCGGCGCCGTCGTCCGC
GACCCACACCTTGCCGATGTCGAGCCCGACGCGCGTGAGGAAGAGTTCTTGCAGCTCGGTGACCCGCTCGATGTGGCGGTCCGGATCGA
CGGTGTGGCGCGTGGCGGGGTAGTCGGCGAACGCGGCGGCGAGGGTGCGTACGGCCCTGGGGACGTCGTCGCGGGTGGCGAGGCGCACC
GTGGGCTTGTACTCGGTCATGGTAAGCTAGCTTGGGCTGCAGGTCGAAAGGCCCGGAGATGAGGAAGAGGAGAACAGCGCGGCAGACGT
GCGCTTTTGAAGCGTGCAGAATGCCGGGCCTCCGGAGGACCTTCGGGCGCCCGCCCCGCCCCTGAGCCCGCCCCTGAGCCCGCCCCCGG
ACCCACCCCTTCCCAGCCTCTGAGCCCAGAAAGCGAAGGAGCAAAGCTGCTATTGGCCGCTGCCCCAAAGGCCTACCCGCTTCCATTGC
TCAGCGGTGCTGTCCATCTGCACGAGACTAGTGAGACGTGCTACTTCCATTTGTCACGTCCTGCACGACGCGAGCTGCGGGGCGGGGGG
GAACTTCCTGACTAGGGGAGGAGTAGAAGGTGGCGCGAAGGGGCCACCAAAGAACGGAGCCGGTTGGCGCCTACCGGTGGATGTGGAAT
GTGTGCGAGGCCAGAGGCCACTTGTGTAGCGCCAAGTGCCCAGCGGGGCTGCTAAAGCGCATGCTCCAGACTGCCTTGGGAAAAGCGCC
TCCCCTACCCGGTAGAATTCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTGTCACTAGGCGGGAACACCCAGCG
CGCGTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGGAGTGGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCA
CCATAAACGTGAAATGTCTTTGGATTTGGGAATCTTATAAGTTCCCTATCAGTGATAG**A^**GATCCCCAGCAGTACGAACATTGTGATTC
AAGAGATCACAATGTTCGTACTGCTTTTTTA**^AGCTT**ATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAGCTTTTGTTCCCTT
TAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACAT
ACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC
AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCC
TCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATC
AGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATA
GGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTT
CCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGC
GCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGC
CCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAAC
AGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGG
TATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTT
TTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGG
AACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAA
ATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTC
GTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCG
CGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATC
CGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTA
CAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATG
TTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGC
ACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTA
TGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAA
CGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGC
ATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTT
GAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAG
AAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC

The digested *HindIII* site (**A^AGCTT**) is shown in bold, with the inserted shRNA construct located in between them. The sequence of interest is highlighted in yellow, and was sequenced to confirm proper insertion of the shRNA construct.

Table 3.5. (Continued)

pSUPERIOR.puro_sh-286B-1

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCA
AAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGAC
TCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTG
CCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGA
AGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCG
CTACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGC
GAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCGCG
CGTAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCATGGGGTC
GTGCGCTCCTTTCGGTCGGGCGCTGCGGGTCGTGGGGCGGGCGTCAGGCACCGGGCTTGCGGGTCATGCACCAGGTGCGCGGTCCTTCG
GGCACCTCGACGTCGGCGGTGACGGTGAAGCCGAGCCGCTCGTAGAAGGGGAGGTTGCGGGGCGCGGAGGTCTCCAGGAAGGCGGGCAC
CCCGGCGCGCTCGGCCGCCTCCACTCCGGGGAGCACGACGGCGCTGCCCAGACCCTTGCCCTGGTGGTCGGGCGAGACGCCGACGGTGG
CCAGGAACCACGCGGGCTCCTTGGGCCGGTGCGGCGCCAGGAGGCCTTCCATCTGTTGCTGCGCGGCCAGCCGGGAACCGCTCAACTCG
GCCATGCGCGGGCCGATCTCGGCGAACACCGCCCCCGCTTCGACGCTCTCCGGCGTGGTCCAGACCGCCACCGCGGCGCCGTCGTCCGC
GACCCACACCTTGCCGATGTCGAGCCCGACGCGCGTGAGGAAGAGTTCTTGCAGCTCGGTGACCCGCTCGATGTGGCGGTCCGGATCGA
CGGTGTGGCGCGTGGCGGGGTAGTCGGCGAACGCGGCGGCGAGGGTGCGTACGGCCCTGGGGACGTCGTCGCGGGTGGCGAGGCGCACC
GTGGGCTTGTACTCGGTCATGGTAAGCTAGCTTGGGCTGCAGGTCGAAAGGCCCGGAGATGAGGAAGAGGAGAACAGCGCGGCAGACGT
GCGCTTTTGAAGCGTGCAGAATGCCGGGCCTCCGGAGGACCTTCGGGCGCCCGCCCCGCCCCTGAGCCCGCCCCTGAGCCCGCCCCCGG
ACCCACCCCTTCCCAGCCTCTGAGCCCAGAAAGCGAAGGAGCAAAGCTGCTATTGGCCGCTGCCCCAAAGGCCTACCCGCTTCCATTGC
TCAGCGGTGCTGTCCATCTGCACGAGACTAGTGAGACGTGCTACTTCCATTTGTCACGTCCTGCACGACGCGAGCTGCGGGGCGGGGGG
GAACTTCCTGACTAGGGGAGGAGTAGAAGGTGGCGCGAAGGGGCCACCAAAGAACGGAGCCGGTTGGCGCCTACCGGTGGATGTGGAAT
GTGTGCGAGGCCAGAGGCCACTTGTGTAGCGCCAAGTGCCCAGCGGGGCTGCTAAAGCGCATGCTCCAGACTGCCTTGGGAAAAGCGCC
TCCCCTACCCGGTAGAATTCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTGTCACTAGGCGGGAACACCCAGCG
CGCGTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGGAGTGGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCA
CCATAAACGTGAAATGTCTTTGGATTTGGGAATCTTATAAGTTCCCTATCAGTGATAG**A^**GATCCCCAGGCCCTTATAAGGGCACTTTC
AAGAGAAGTGCCCTTATAAGGGCCTTTTTTA**^AGCTT**ATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAGCTTTTGTTCCCTT
TAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACAT
ACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC
AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCC
TCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATC
AGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATA
GGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTT
CCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGC
GCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGC
CCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAAC
AGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGG
TATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTT
TTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGG
AACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAA
ATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTC
GTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCG
CGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATC
CGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTA
CAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATG
TTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGC
ACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTA
TGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAA
CGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGC
ATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTT
GAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAG
AAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC

The digested *HindIII* site (**A^AGCTT**) is shown in bold, with the inserted shRNA construct located in between them. The sequence of interest is highlighted in yellow, and was sequenced to confirm proper insertion of the shRNA construct.

Table 3.5. (Continued)

pSUPERIOR.puro_sh-286B-2

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCA
AAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGAC
TCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTG
CCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGA
AGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCG
CTACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGC
GAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCGCG
CGTAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCATGGGGTC
GTGCGCTCCTTTCGGTCGGGCGCTGCGGGTCGTGGGGCGGGCGTCAGGCACCGGGCTTGCGGGTCATGCACCAGGTGCGCGGTCCTTCG
GGCACCTCGACGTCGGCGGTGACGGTGAAGCCGAGCCGCTCGTAGAAGGGGAGGTTGCGGGGCGCGGAGGTCTCCAGGAAGGCGGGCAC
CCCGGCGCGCTCGGCCGCCTCCACTCCGGGGAGCACGACGGCGCTGCCCAGACCCTTGCCCTGGTGGTCGGGCGAGACGCCGACGGTGG
CCAGGAACCACGCGGGCTCCTTGGGCCGGTGCGGCGCCAGGAGGCCTTCCATCTGTTGCTGCGCGGCCAGCCGGGAACCGCTCAACTCG
GCCATGCGCGGGCCGATCTCGGCGAACACCGCCCCCGCTTCGACGCTCTCCGGCGTGGTCCAGACCGCCACCGCGGCGCCGTCGTCCGC
GACCCACACCTTGCCGATGTCGAGCCCGACGCGCGTGAGGAAGAGTTCTTGCAGCTCGGTGACCCGCTCGATGTGGCGGTCCGGATCGA
CGGTGTGGCGCGTGGCGGGGTAGTCGGCGAACGCGGCGGCGAGGGTGCGTACGGCCCTGGGGACGTCGTCGCGGGTGGCGAGGCGCACC
GTGGGCTTGTACTCGGTCATGGTAAGCTAGCTTGGGCTGCAGGTCGAAAGGCCCGGAGATGAGGAAGAGGAGAACAGCGCGGCAGACGT
GCGCTTTTGAAGCGTGCAGAATGCCGGGCCTCCGGAGGACCTTCGGGCGCCCGCCCCGCCCCTGAGCCCGCCCCTGAGCCCGCCCCCGG
ACCCACCCCTTCCCAGCCTCTGAGCCCAGAAAGCGAAGGAGCAAAGCTGCTATTGGCCGCTGCCCCAAAGGCCTACCCGCTTCCATTGC
TCAGCGGTGCTGTCCATCTGCACGAGACTAGTGAGACGTGCTACTTCCATTTGTCACGTCCTGCACGACGCGAGCTGCGGGGCGGGGGG
GAACTTCCTGACTAGGGGAGGAGTAGAAGGTGGCGCGAAGGGGCCACCAAAGAACGGAGCCGGTTGGCGCCTACCGGTGGATGTGGAAT
GTGTGCGAGGCCAGAGGCCACTTGTGTAGCGCCAAGTGCCCAGCGGGGCTGCTAAAGCGCATGCTCCAGACTGCCTTGGGAAAAGCGCC
TCCCCTACCCGGTAGAATTCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTGTCACTAGGCGGGAACACCCAGCG
CGCGTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGGAGTGGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCA
CCATAAACGTGAAATGTCTTTGGATTTGGGAATCTTATAAGTTCCCTATCAGTGATAG**A^**GATCCCCACCATCAATTAAGTTTCATTTC
AAGAGAATGAAACTTAATTGATGGTTTTTTA**^AGCTT**ATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAGCTTTTGTTCCCTT
TAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACAT
ACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC
AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCC
TCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATC
AGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATA
GGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTT
CCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGC
GCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGC
CCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAAC
AGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGG
TATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTT
TTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGG
AACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAA
ATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTC
GTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCG
CGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATC
CGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTA
CAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATG
TTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGC
ACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTA
TGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAA
CGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGC
ATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTT
GAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAG
AAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC

The digested *HindIII* site (**A^AGCTT**) is shown in bold, with the inserted shRNA construct located in between them. The sequence of interest is highlighted in yellow, and was sequenced to confirm proper insertion of the shRNA construct.

Table 3.5. (Continued)

pSUPERIOR.puro_sh-286B-3

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCA
AAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGAC
TCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTG
CCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGA
AGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCG
CTACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGC
GAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCGCG
CGTAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCATGGGGTC
GTGCGCTCCTTTCGGTCGGGCGCTGCGGGTCGTGGGGCGGGCGTCAGGCACCGGGCTTGCGGGTCATGCACCAGGTGCGCGGTCCTTCG
GGCACCTCGACGTCGGCGGTGACGGTGAAGCCGAGCCGCTCGTAGAAGGGGAGGTTGCGGGGCGCGGAGGTCTCCAGGAAGGCGGGCAC
CCCGGCGCGCTCGGCCGCCTCCACTCCGGGGAGCACGACGGCGCTGCCCAGACCCTTGCCCTGGTGGTCGGGCGAGACGCCGACGGTGG
CCAGGAACCACGCGGGCTCCTTGGGCCGGTGCGGCGCCAGGAGGCCTTCCATCTGTTGCTGCGCGGCCAGCCGGGAACCGCTCAACTCG
GCCATGCGCGGGCCGATCTCGGCGAACACCGCCCCCGCTTCGACGCTCTCCGGCGTGGTCCAGACCGCCACCGCGGCGCCGTCGTCCGC
GACCCACACCTTGCCGATGTCGAGCCCGACGCGCGTGAGGAAGAGTTCTTGCAGCTCGGTGACCCGCTCGATGTGGCGGTCCGGATCGA
CGGTGTGGCGCGTGGCGGGGTAGTCGGCGAACGCGGCGGCGAGGGTGCGTACGGCCCTGGGGACGTCGTCGCGGGTGGCGAGGCGCACC
GTGGGCTTGTACTCGGTCATGGTAAGCTAGCTTGGGCTGCAGGTCGAAAGGCCCGGAGATGAGGAAGAGGAGAACAGCGCGGCAGACGT
GCGCTTTTGAAGCGTGCAGAATGCCGGGCCTCCGGAGGACCTTCGGGCGCCCGCCCCGCCCCTGAGCCCGCCCCTGAGCCCGCCCCCGG
ACCCACCCCTTCCCAGCCTCTGAGCCCAGAAAGCGAAGGAGCAAAGCTGCTATTGGCCGCTGCCCCAAAGGCCTACCCGCTTCCATTGC
TCAGCGGTGCTGTCCATCTGCACGAGACTAGTGAGACGTGCTACTTCCATTTGTCACGTCCTGCACGACGCGAGCTGCGGGGCGGGGGG
GAACTTCCTGACTAGGGGAGGAGTAGAAGGTGGCGCGAAGGGGCCACCAAAGAACGGAGCCGGTTGGCGCCTACCGGTGGATGTGGAAT
GTGTGCGAGGCCAGAGGCCACTTGTGTAGCGCCAAGTGCCCAGCGGGGCTGCTAAAGCGCATGCTCCAGACTGCCTTGGGAAAAGCGCC
TCCCCTACCCGGTAGAATTCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTGTCACTAGGCGGGAACACCCAGCG
CGCGTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGGAGTGGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCA
CCATAAACGTGAAATGTCTTTGGATTTGGGAATCTTATAAGTTCCCTATCAGTGATAG**A^**GATCCCCAACTCTTACCTATTACAGATTC
AAGAGATCTGTAATAGGTAAGAGTTTTTTTA**^AGCTT**ATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAGCTTTTGTTCCCTT
TAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACAT
ACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC
AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCC
TCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATC
AGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATA
GGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTT
CCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGC
GCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGC
CCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAAC
AGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGG
TATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTT
TTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGG
AACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAA
ATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTC
GTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCG
CGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATC
CGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTA
CAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATG
TTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGC
ACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTA
TGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAA
CGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGC
ATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTT
GAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAG
AAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC

The digested *HindIII* site (**A^AGCTT**) is shown in bold, with the inserted shRNA construct located in between them. The sequence of interest is highlighted in yellow, and was sequenced to confirm proper insertion of the shRNA construct.

Table 3.5. (Continued)

pSUPERIOR.puro_sh-REST-1

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCA
AAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGAC
TCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTG
CCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGA
AGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCG
CTACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGC
GAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCGCG
CGTAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCATGGGGTC
GTGCGCTCCTTTCGGTCGGGCGCTGCGGGTCGTGGGGCGGGCGTCAGGCACCGGGCTTGCGGGTCATGCACCAGGTGCGCGGTCCTTCG
GGCACCTCGACGTCGGCGGTGACGGTGAAGCCGAGCCGCTCGTAGAAGGGGAGGTTGCGGGGCGCGGAGGTCTCCAGGAAGGCGGGCAC
CCCGGCGCGCTCGGCCGCCTCCACTCCGGGGAGCACGACGGCGCTGCCCAGACCCTTGCCCTGGTGGTCGGGCGAGACGCCGACGGTGG
CCAGGAACCACGCGGGCTCCTTGGGCCGGTGCGGCGCCAGGAGGCCTTCCATCTGTTGCTGCGCGGCCAGCCGGGAACCGCTCAACTCG
GCCATGCGCGGGCCGATCTCGGCGAACACCGCCCCCGCTTCGACGCTCTCCGGCGTGGTCCAGACCGCCACCGCGGCGCCGTCGTCCGC
GACCCACACCTTGCCGATGTCGAGCCCGACGCGCGTGAGGAAGAGTTCTTGCAGCTCGGTGACCCGCTCGATGTGGCGGTCCGGATCGA
CGGTGTGGCGCGTGGCGGGGTAGTCGGCGAACGCGGCGGCGAGGGTGCGTACGGCCCTGGGGACGTCGTCGCGGGTGGCGAGGCGCACC
GTGGGCTTGTACTCGGTCATGGTAAGCTAGCTTGGGCTGCAGGTCGAAAGGCCCGGAGATGAGGAAGAGGAGAACAGCGCGGCAGACGT
GCGCTTTTGAAGCGTGCAGAATGCCGGGCCTCCGGAGGACCTTCGGGCGCCCGCCCCGCCCCTGAGCCCGCCCCTGAGCCCGCCCCCGG
ACCCACCCCTTCCCAGCCTCTGAGCCCAGAAAGCGAAGGAGCAAAGCTGCTATTGGCCGCTGCCCCAAAGGCCTACCCGCTTCCATTGC
TCAGCGGTGCTGTCCATCTGCACGAGACTAGTGAGACGTGCTACTTCCATTTGTCACGTCCTGCACGACGCGAGCTGCGGGGCGGGGGG
GAACTTCCTGACTAGGGGAGGAGTAGAAGGTGGCGCGAAGGGGCCACCAAAGAACGGAGCCGGTTGGCGCCTACCGGTGGATGTGGAAT
GTGTGCGAGGCCAGAGGCCACTTGTGTAGCGCCAAGTGCCCAGCGGGGCTGCTAAAGCGCATGCTCCAGACTGCCTTGGGAAAAGCGCC
TCCCCTACCCGGTAGAATTCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTGTCACTAGGCGGGAACACCCAGCG
CGCGTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGGAGTGGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCA
CCATAAACGTGAAATGTCTTTGGATTTGGGAATCTTATAAGTTCCCTATCAGTGATAG**A^**GATCCCCGCATACAGGAAGCAATTCATTC
AAGAGATGAATTGCTTCCTGTATGCTTTTTA**^AGCTT**ATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAGCTTTTGTTCCCTT
TAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACAT
ACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC
AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCC
TCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATC
AGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATA
GGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTT
CCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGC
GCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGC
CCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAAC
AGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGG
TATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTT
TTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGG
AACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAA
ATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTC
GTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCG
CGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATC
CGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTA
CAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATG
TTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGC
ACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTA
TGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAA
CGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGC
ATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTT
GAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAG
AAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC

The digested *HindIII* site (**A^AGCTT**) is shown in bold, with the inserted shRNA construct located in between them. The sequence of interest is highlighted in yellow, and was sequenced to confirm proper insertion of the shRNA construct.

Table 3.5. (Continued)

pSUPERIOR.puro_sh-REST-2

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCA
AAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGAC
TCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTG
CCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGA
AGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCG
CTACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGC
GAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCGCG
CGTAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCATGGGGTC
GTGCGCTCCTTTCGGTCGGGCGCTGCGGGTCGTGGGGCGGGCGTCAGGCACCGGGCTTGCGGGTCATGCACCAGGTGCGCGGTCCTTCG
GGCACCTCGACGTCGGCGGTGACGGTGAAGCCGAGCCGCTCGTAGAAGGGGAGGTTGCGGGGCGCGGAGGTCTCCAGGAAGGCGGGCAC
CCCGGCGCGCTCGGCCGCCTCCACTCCGGGGAGCACGACGGCGCTGCCCAGACCCTTGCCCTGGTGGTCGGGCGAGACGCCGACGGTGG
CCAGGAACCACGCGGGCTCCTTGGGCCGGTGCGGCGCCAGGAGGCCTTCCATCTGTTGCTGCGCGGCCAGCCGGGAACCGCTCAACTCG
GCCATGCGCGGGCCGATCTCGGCGAACACCGCCCCCGCTTCGACGCTCTCCGGCGTGGTCCAGACCGCCACCGCGGCGCCGTCGTCCGC
GACCCACACCTTGCCGATGTCGAGCCCGACGCGCGTGAGGAAGAGTTCTTGCAGCTCGGTGACCCGCTCGATGTGGCGGTCCGGATCGA
CGGTGTGGCGCGTGGCGGGGTAGTCGGCGAACGCGGCGGCGAGGGTGCGTACGGCCCTGGGGACGTCGTCGCGGGTGGCGAGGCGCACC
GTGGGCTTGTACTCGGTCATGGTAAGCTAGCTTGGGCTGCAGGTCGAAAGGCCCGGAGATGAGGAAGAGGAGAACAGCGCGGCAGACGT
GCGCTTTTGAAGCGTGCAGAATGCCGGGCCTCCGGAGGACCTTCGGGCGCCCGCCCCGCCCCTGAGCCCGCCCCTGAGCCCGCCCCCGG
ACCCACCCCTTCCCAGCCTCTGAGCCCAGAAAGCGAAGGAGCAAAGCTGCTATTGGCCGCTGCCCCAAAGGCCTACCCGCTTCCATTGC
TCAGCGGTGCTGTCCATCTGCACGAGACTAGTGAGACGTGCTACTTCCATTTGTCACGTCCTGCACGACGCGAGCTGCGGGGCGGGGGG
GAACTTCCTGACTAGGGGAGGAGTAGAAGGTGGCGCGAAGGGGCCACCAAAGAACGGAGCCGGTTGGCGCCTACCGGTGGATGTGGAAT
GTGTGCGAGGCCAGAGGCCACTTGTGTAGCGCCAAGTGCCCAGCGGGGCTGCTAAAGCGCATGCTCCAGACTGCCTTGGGAAAAGCGCC
TCCCCTACCCGGTAGAATTCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTGTCACTAGGCGGGAACACCCAGCG
CGCGTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGGAGTGGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCA
CCATAAACGTGAAATGTCTTTGGATTTGGGAATCTTATAAGTTCCCTATCAGTGATAG**A^**GATCCCCGTGCCTGTTAAAGATAGCTTTC
AAGAGAAGCTATCTTTAACAGGCACTTTTTA**^AGCTT**ATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAGCTTTTGTTCCCTT
TAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACAT
ACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC
AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCC
TCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATC
AGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATA
GGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTT
CCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGC
GCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGC
CCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAAC
AGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGG
TATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTT
TTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGG
AACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAA
ATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTC
GTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCG
CGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATC
CGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTA
CAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATG
TTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGC
ACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTA
TGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAA
CGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGC
ATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTT
GAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAG
AAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC

The digested *HindIII* site (**A^AGCTT**) is shown in bold, with the inserted shRNA construct located in between them. The sequence of interest is highlighted in yellow, and was sequenced to confirm proper insertion of the shRNA construct.

Table 3.5. (Continued)

pSUPERIOR.puro_sh-REST-3

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCA
AAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGAC
TCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTG
CCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGA
AGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCG
CTACAGGGCGCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGC
GAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAGCGCG
CGTAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCATGGGGTC
GTGCGCTCCTTTCGGTCGGGCGCTGCGGGTCGTGGGGCGGGCGTCAGGCACCGGGCTTGCGGGTCATGCACCAGGTGCGCGGTCCTTCG
GGCACCTCGACGTCGGCGGTGACGGTGAAGCCGAGCCGCTCGTAGAAGGGGAGGTTGCGGGGCGCGGAGGTCTCCAGGAAGGCGGGCAC
CCCGGCGCGCTCGGCCGCCTCCACTCCGGGGAGCACGACGGCGCTGCCCAGACCCTTGCCCTGGTGGTCGGGCGAGACGCCGACGGTGG
CCAGGAACCACGCGGGCTCCTTGGGCCGGTGCGGCGCCAGGAGGCCTTCCATCTGTTGCTGCGCGGCCAGCCGGGAACCGCTCAACTCG
GCCATGCGCGGGCCGATCTCGGCGAACACCGCCCCCGCTTCGACGCTCTCCGGCGTGGTCCAGACCGCCACCGCGGCGCCGTCGTCCGC
GACCCACACCTTGCCGATGTCGAGCCCGACGCGCGTGAGGAAGAGTTCTTGCAGCTCGGTGACCCGCTCGATGTGGCGGTCCGGATCGA
CGGTGTGGCGCGTGGCGGGGTAGTCGGCGAACGCGGCGGCGAGGGTGCGTACGGCCCTGGGGACGTCGTCGCGGGTGGCGAGGCGCACC
GTGGGCTTGTACTCGGTCATGGTAAGCTAGCTTGGGCTGCAGGTCGAAAGGCCCGGAGATGAGGAAGAGGAGAACAGCGCGGCAGACGT
GCGCTTTTGAAGCGTGCAGAATGCCGGGCCTCCGGAGGACCTTCGGGCGCCCGCCCCGCCCCTGAGCCCGCCCCTGAGCCCGCCCCCGG
ACCCACCCCTTCCCAGCCTCTGAGCCCAGAAAGCGAAGGAGCAAAGCTGCTATTGGCCGCTGCCCCAAAGGCCTACCCGCTTCCATTGC
TCAGCGGTGCTGTCCATCTGCACGAGACTAGTGAGACGTGCTACTTCCATTTGTCACGTCCTGCACGACGCGAGCTGCGGGGCGGGGGG
GAACTTCCTGACTAGGGGAGGAGTAGAAGGTGGCGCGAAGGGGCCACCAAAGAACGGAGCCGGTTGGCGCCTACCGGTGGATGTGGAAT
GTGTGCGAGGCCAGAGGCCACTTGTGTAGCGCCAAGTGCCCAGCGGGGCTGCTAAAGCGCATGCTCCAGACTGCCTTGGGAAAAGCGCC
TCCCCTACCCGGTAGAATTCGAACGCTGACGTCATCAACCCGCTCCAAGGAATCGCGGGCCCAGTGTCACTAGGCGGGAACACCCAGCG
CGCGTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGGGGAGTGGCGCCCTGCAATATTTGCATGTCGCTATGTGTTCTGGGAAATCA
CCATAAACGTGAAATGTCTTTGGATTTGGGAATCTTATAAGTTCCCTATCAGTGATAG**A^**GATCCCCGACTGACAGTATAGTTTGTTTC
AAGAGAACAAACTATACTGTCAGTCTTTTTA**^AGCTT**ATCGATACCGTCGACCTCGAGGGGGGGCCCGGTACCCAGCTTTTGTTCCCTT
TAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACAT
ACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC
AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCC
TCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATC
AGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATA
GGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTT
CCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGC
GCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGC
CCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAAC
AGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGG
TATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTT
TTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGG
AACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAA
ATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTC
GTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCG
CGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATC
CGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTA
CAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATG
TTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGC
ACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTA
TGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAA
CGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGC
ATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTT
GAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAG
AAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC

The digested *HindIII* site (**A^AGCTT**) is shown in bold, with the inserted shRNA construct located in between them. The sequence of interest is highlighted in yellow, and was sequenced to confirm proper insertion of the shRNA construct.

CHAPTER 4

CONCLUSIONS

In summary

ZNF286A is a unique member of the mammalian KRAB-ZNF gene family. Originally arising in metatherian lineages, it has remained conserved for hundreds of millions of years. We demonstrate here that sometime during very recent primate evolution, and most probably in the hominid lineage, *ZNF286A* was involved in a large duplication event that led to the creation of *ZNF286B*. Concomitantly or shortly after, a *FOXO3B* pseudogene inserted into the duplicate in place of a functional KRAB domain, creating the human-specific version of this gene with structural elements and regulatory sequences that distinguish it clearly from its metatherian ancestor.

Putative regulatory targets, defined as differentially expressed genes (DEGs) that are flanked by or contain ChIP-enriched binding sites, implicate ZNF286A as a regulator of cell cycle function and cell proliferation. It is well established that the cell cycle is carefully regulated during neurogenesis, and differences in cell cycle regulation can very likely explain many aspects of the evolution of the larger primate brain (Calegari et al., 2005). Although this hypothesis will require additional testing in the future, we speculate that ZNF286A and ZNF286B may play important roles in brain development, and that the generation of the ZNF286B gene duplicate had an impact on evolution of the human brain.

Explorations of RNA expression in cells and tissues confirm that *ZNF286A* and *ZNF286B* are highly expressed in the brain and differentiated neural cells, and yet their overall patterns of expression are not identical in all issues. This, as well as the structural differences between the two paralogs, leads us to believe that their functions are closely related, but distinct from one another. It is likely that, given their shared zinc finger domains, the two paralogs interact at the same DNA-binding sites. Whether they compete or co-operate in binding at those sites remains unanswered, and an important question for further study.

Close analysis of ZNF286 binding regions as discovered through ChIP-seq revealed a high degree of overlap with the known recognition motif of REST, as well as a secondary G-rich motif to which REST also binds in neuronal cells. All three proteins of interest – ZNF286A, ZNF286B, and REST – were shown to bind to these two motifs *in*

vitro, as evidenced through EMSA. We could not find evidence, however, that the ZNF286 paralogs interact directly with REST, and we surmise that it is much more likely that they compete for DNA-binding sites, or possibly, prefer different "half-sites" without direct protein:protein contact.

We predict that the ZNF286 proteins inhibit the proliferation and stimulation of neuronal differentiation – a function that is opposite to the known functions of REST in neuronal cells (Schoenherr & Anderson, 1995). We hypothesize, therefore, that the ZNF286 paralogs evolved as competitive modulators of REST activities in the brain, with the duplication and divergence of *ZNF286B* having allowed for separate regulation of this deeply conserved activity, that is, adding a novel layer of REST modulation in fetal and adult brain, as well as other human tissues.

To further complicate the interpretation of these functions, *ZNF286A* additionally codes for a KRAB(-) isoform that is conserved across species. We show that the expression of both KRAB(+) and KRAB(-) transcripts is mostly comparable across tissues, which consist of multiple cell types, but differs in individual cell lines, including neuroblastoma cell lines that can be induced towards neuronal differentiation. This suggests that, at least under certain conditions, *ZNF286A* (K+) and (K-) can serve different functions. Since we were unable to target the (K-) isoform specifically, our current study cannot definitively address this assumption. However, the tagged expression constructs that have been generated for the (K-) isoform may help solve this puzzle in the future.

As such, we have identified three distinct proteins that can be predicted to bind to the same binding sites as REST. One might expect that ZNF286 (K+), possessing what we have shown to be a functional KAP-1 interacting KRAB domain, would act as a repressor at these sites (Friedman et al., 1996). Indeed, our data suggest this is most likely true, at least for a majority of the "target" genes. However, it is most surprising that, after knockdown of ZNF286A by siRNA, we do find genes – including some closely neighboring ones – that appear to be up-regulated by the protein. This suggests that KRAB may not be a simple transcriptional repressor (Margolin et al., 1994; Friedman et al., 1996; Urratia, 2003; Sripathy et al., 2006), but might have more complex functions that have not yet been fully described. There are hundreds of KRAB-ZNF proteins, and

yet studies into KRAB domains have predominantly focused on KRAB activity using the KOX1 "consensus KRAB" as a model to represent all KRAB domains across the proteome (Moosman et al., 1996). Given what we know about ZNF286A and its KRAB domain, it is very likely that these other studies into KRAB are lacking in scope.

From our EMSA competition data, it appears that ZNF286B has a slight preference for the "left" REST half-site, and it is possible that ZNF286A and ZNF286B could both occupy the REST site. Such "cooperative binding" could elicit a stronger regulatory effect, and might explain the very similar results we obtained when we knocked down each gene specifically. Furthermore, complete blockage of the RE-1 site could more effectively block REST's ability to bind, and have a stronger effect on adjacent genes. Indeed, it is possible the ZNF286 proteins' regulatory function results primarily from the blockage of REST. However, both ZNF286 paralogs have functionally identical zinc finger domains, and it is difficult to explain how they might have distinct half-site preference. The hypothesis of cooperative binding thus remains to be tested and further explored.

As such, it is likely that the *FOXO3B* region that has inserted into *ZNF286B* is introducing some additional function into the human-specific transcription factor that would affect its binding affinity (either to DNA or to the ZNF286A protein) in some way. Fully understanding this would require more research into ZNF286B interactions, for example, through co-immunoprecipitation followed by protein mass spectroscopy or through other types of co-IP experiments. Co-IP like this would require the use of tagged proteins, as the only effective ZNF286A and ZNF286B antibodies that we have found were derived in rabbit, and co-IP requires antibodies from two different species. Already, the groundwork for tagged ZNF286 construct has been laid out by the work done in this thesis, as we now have a *pTRE-tight* inducible vector with a strep and HA-tags, along with a modified multiple cloning site to allow for easy insertion of a gene sequence. Further utilizing these vectors in human cell lines would open up a large number of possibilities for new experiments into this transcription factor family. Future research should provide insight into the full range of cofactors interacting with ZNF286A and B, as well as the different contexts in which ZNF286A and B functions; we have examined

functions in neuroblastoma, but we do not have a clear picture of whether these results can predict functions in different types of cells.

Additionally, humans possess two versions of *ZNF286* as we have seen, but other mammals like mice only possess one – *Zfp286*. By focusing on how ZNF286A and B function in human cells, and contrasting that with how Zfp286 functions by itself in mice, we should have a much better understanding of these systems than is presently possible. The inducible shRNA constructs that have been designed as a part of this thesis will be an indispensible tool for this type of study in the future.

Closing remarks

This thesis has attempted to elucidate the functions and regulatory activities of both ZNF286A isoforms, ZNF286B, and their potential interplay with REST. ZNF286 represents an interesting case study into a recent KRAB-ZNF segmental duplication, demonstrating how novel functions can arise over an evolutionarily short period of time (Lynch & Conery, 2000). There is still much that we do not know about these TFs, and future studies may help to expand the scope of our understanding. However, we now know that these TFs play a role in neural development.

Neural development is human development – our species has spent a lot of energy over the course of evolution in developing our brains in a manner that is unique to other species. The regulation of cell division in neural progenitor cells, like the ones examined in this thesis, has played crucial evolutionary roles in affecting brain size (Fish et al., 2008), with the number and complexity of neural networks being reflected by the outgrowth of cortical neurons (Hofman, 2014). Indeed, the mechanisms that govern neurite outgrown have been linked to neuronal plasticity, as regulated by REST (Lepagnol-Bestel et al., 2006) and possibly further regulated by the ZNF286 proteins.

This thesis represents a small and focused study into a very specialized set of genes, singled out from a genome of thousands. There are countless examples of segmental duplications and novel functions that have evolved in countless species. As such, it may be tempting to dismiss these genes as insignificant to the greater picture – here we have yet another study of miniscule mechanisms, which is certainly of miniscule

importance to the field as it is unlikely to cure any diseases or revolutionize the field. And yet, each new discovery is revolutionary in its own way. Each new gene that we investigate provides us with novel perspectives, and allows us to use the tools we have at our disposal to tackle problems in new and creative ways. Each answer that we articulate helps us to more coherently respond to questions that have already been answered, and more confidently respond to those we will ask in the future. While it is certainly true that the mechanisms studied in this work are miniscule, the wealth of understanding we gain in their pursuit is anything but.

References Cited

Aiello LC. **1997**. Brains and guts in human evolution: The Expensive Tissue Hypothesis. *Braz J Genet* 20(1):141-148.

Calegari F, Haubensak W, Haffner C, Huttner WB. **2005**. Selective Lengthening of the Cell Cycle in the Neurogenic Subpopulation of Neural Progenitor Cells during Mouse Brain Development. *J Neurosci* 25(28):6533-6538.

Fish JL, Dehay C, Kennedy H, Huttner WB. **2008**. Making bigger brains – the evolution of neural-progenitor-cell division. *J Cell Sci* 121(17):2783-93.

Friedman JR, Fredericks WJ, Jensen DE, Speicher DW, Huang XP, Neilson EG, Rauscher FJ 3rd. **1996**. KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes Dev* 10(16):2067-78.

Hofman MA. **2014**. Evolution of the human brain: when bigger is better. *Front Neuroanat* 8(15):1-12.

Lepagnol-Bestel AM, Maussion G, Ramoz N, Moalic JM, Gorwood P, Simonneau M.. **2007**. Nrsf silencing induces molecular and subcellular changes linked to neuronal plasticity. *Neuroreport* 18(5):441-6.

Lynch M, Conery JS. **2000**. The evolutionary fate and consequences of duplicate genes. *Science* 290(5494):1151-5.

Margolin JF, Friedman JR, Meyer WK, Vissing H, Thiesen HJ, Rauscher FJ 3rd. **1994**. Kruppel-associated boxes are potent transcriptional repression domains. *Proc Natl Acad Sci USA* 91:4509–4513.

Moosmann P, Georgiev O, Le Douarin B, Bourquin JP, Schaffner W. **1996**. Transcriptional repression by RING finger protein TIF1 beta that interacts with the KRAB repressor domain of KOX1. *Nucleic Acids Res* 24:4859-4867.

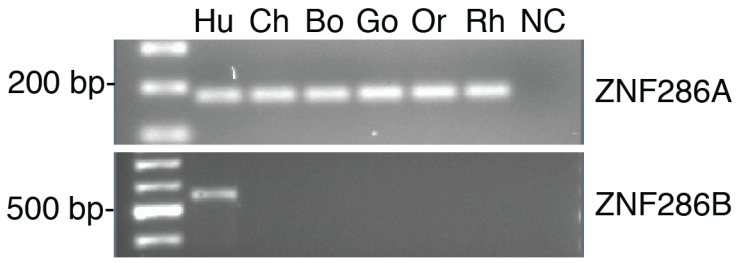
Urrutia R. **2003**. KRAB-containing zinc-finger repressor proteins. *Genome Biol* 4:231.

Schoenherr CJ, Paquette AJ, Anderson DJ. **1996**. Identification of potential target genes for the neuron-restrictive silencer factor *Proc Natl Acad Sci USA* 93(18):9881-9886.

Sripathy SP, Stevens J, Schultz DC. **2006**. The KAP1 corepressor functions to coordinate the assembly of de novo HP1-demarcated microenvironments of heterochromatin required for KRAB zinc finger protein-mediated transcriptional repression. *Mol Cell Biol* 26:8623–8638.

APPENDIX

Appendix Figure A.1. Confirming the human specificity of the ZNF286B gene. *ZNF286B*'s role as a human-specific duplicate of *ZNF286A* was first confirmed using PCR, with the forward primer targeting the first finger that distinguishes the two paralogs to amplify the *ZNF286B* gene sequences in genomic DNA from six primates: human (Hu), Chimpanzee (Ch), Bonobo (Bo), Gorilla (Go), Orangutan (Or), and Rhesus macaque (Rh). A size standard ladder and no-template negative control (NC) are also included. The *ZNF286B*-specific PCR product was generated only in human templates (lower panel). These same DNA preparations were tested with control PCR primer sets designed against several other genes, including *ZNF286A*, which is known to be present in all species (upper panel). The production of clear PCR products for this and other shared genes confirmed the quality of the non-human primate DNA. *ZNF286B* panel reproduced from Nowick et al.[1]



[1]Nowick K, Fields C, Gernat T, Caetano-Anollés D, Kholina N, Stubbs L. 2011. Gain, Loss and Divergence in Primate Zinc-Finger Genes: A Rich Resource for Evolution of Gene Regulatory Differences between Species. *PLOS One* 6(6):e21553.

Appendix Figure A.2. Full expanse of the ch17 duplicon, including genes. UCSC Genome Browser including the full expanse of the ch17 duplicon, in which *ZNF286A* and a number of surrounding genes duplicated. Here, the duplicated region around *ZNF286A* (top panel) can be seen, while the corresponding sequence around *ZNF286B* (bottom panel) can be seen.