DEEP VERTEBRATE ROOTS FOR MAMMALIAN KRAB ZINC-FINGER TRANSCRIPTION FACTORS

BY

LI-HSIN CHANG

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, 2017

Urbana, Illinois

Doctoral Committee:

Associate Professor Craig A. Mizzen, Chair
Professor Lisa J. Stubbs, Director of Research
Associate Professor Stephanie S. Ceman
Associate Professor Alison M. Bell

ABSTRACT

KRAB-associated C2H2 zinc-finger (KRAB-ZNF) proteins are the products of a rapidly evolving gene family that traces back to early tetrapods, but which has expanded dramatically to generate an unprecedented level of species-specific diversity. While most attention has been focused on the more recently evolved primate KRAB-ZNF genes, the vertebrate roots of the KRAB-ZNF families have remained mysterious. We recently mined ZNF loci from seven sequenced genomes (opossum, chicken, zebra finch, lizard, frog, mouse, and human genome) and found hundreds of KRAB-ZNF proteins in every species we examined, but only three human genes were found with clear orthologs in non-mammalian vertebrates. These three genes, *ZNF777, ZNF282*, and *ZNF783*, are members of an ancient familial cluster and encode proteins with similar domain structures. These three genes, members of an ancient familial cluster, encode a noncanonical KRAB domain that is similar to an ancient domain which is prevalent in non-mammalian species. In contrast to the mammalian KRAB, which is thought to function as a potent repressor, this ancient domain serves as a transcriptional activator. Our evolutionary analysis confirmed the ancient provenance of this activating KRAB and revealed the independent expansion of KRAB-ZNFs in every vertebrate lineage. This finding led us to ask the question: what are the functions of these ancient family members and why, of such a large and diverse family group, were these three genes conserved so fastidiously over hundreds of millions of years?

In chapter 2, I report the regulatory function of ZNF777, combining chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-seq) with siRNA knockdown experiments to determine genome-wide binding sites, a distinct binding motif, and predicted targets for the protein in human BeWo choriocarcinoma cells. Genes neighboring ZNF777 binding sites can be either up- or down- regulated, suggesting a complex regulatory role. Our studies revealed that some of this complexity is due to the generation of HUB-containing and HUB-minus isoforms, which are predicted to have different regulatory activities. Based on these experiments, we hypothesize that ZNF777 regulates pathways best known for their roles in neurogenesis and axon pathfinding, but also recently shown to play critical roles in placental development.

Since *ZNF777* is also expressed in embryonic brain, we sought to further investigate the functional role of this ancient gene in neuron development. In chapter 3, I show that mouse *Zfp777* is expressed in neuronal stem cells (NSC) cultured from early mouse embryos, with a pattern that changes over the course of neuron differentiation *in vitro*. Using the NSC platform, I characterized the binding landscape of Zfp777 in undifferentiated NSC. To circumvent the roadblock posed by the lack of a ChIP-grade antibody for the mouse protein, I exploited the CRISPR-Cas9 technique to tag the endogenous Zfp777 protein with FLAG epitopes. Our results revealed a novel Zfp777 binding motif that bears significant similarity to a motif predicted in *in vitro* studies, and found that Zfp777 binds to promoters of genes encoding transcription factors, Wnt and TGF-beta pathways components, and proteins related to neuron development and axon guidance. Since these same functions were also found to be regulated by ZNF777 in BeWo cells, these results suggested that the mouse and human Zfp777 and ZNF777 proteins regulating similar genes and pathways, most classically associated with axon guidance, in diverse tissues.

ACKNOWLEDGEMENTS

This thesis wouldn't have been possible without my advisor, Dr. Lisa Stubbs' big heart. Thank you, Lisa, for accepting me into your lab, and for making the most important turning point so far in my journey of science. I appreciate greatly your guidance, patience, and all the supports along the years. I want to thank my committee Dr. Craig Mizzen, Dr. Stephanie Ceman, and Dr. Alison Bell, for your precious feedback on my project. My labmates Dr. Chase Bolt, Dr. Younguk-Calvin Sun, Dr. Derek Caetano-Anolles, Dr. Annie Weisner, Soumya Negi, Chris Seward, Huimin Zhang, Joseph Troy, Chih-Ying Chen, Dr. Xiaochen Lu, Dr. Michael Saul, and Bob Chen, for having countless informative and inspiring discussions and a lovely lab environment.

I want to thank all my dear friends I have met in Champaign: Sahand Hariri, Yu-Jen Hsu, Lana Šteković, Pei-Ci Wu, Meng-Jung Lee, Hsin-Yi Lin, Chih-Ting Kuo, Yu-Chieh Ho, Chantelle Hougland, Serhiy Potishuk, Robin Berthier, Judy Chiu, Christen Mercier, Shad Sharma, Louisa Xue, Chieh-Chun Chen, Jui-Ting Huang, and Yu-Ying Lee, for your moral support and warmest friendship. I am truly lucky to have you all awesomely multi-talented people in my life. My dearest friends Ming-Hsiang Lee, I-Jen Wang, Kuan-Yin Liu, Chinglin Tang, Kate Yang, Yvonne Yu, Yichen Kuo, and I-Yin Chen, for your constant support and friendship over more than a decade.

I would've never made it to this point without the unconditional love from my parents, Wei-Hua Chang and Man-Yi Chu. Thank you. I love you always.

TABLE OF CONTENTS

CHPATER 1: INTRODUCTION

Introduction of C2H2 Zinc Finger Transcription Factors

In eukaryotic cells, the transcriptional control is an extremely complex process involving a great number of transcription factors (TFs) and cofactors that regulate the assembly of transcription-initiation complexes and the rate at which transcription is initiated. There are a variety of enzymes which modify the chromatin structure via changes in histone modification, DNA methylation, and nucleosome positioning. The presence of specific DNA-binding domains (DBDs) which encode a sequence-specific DNA-binding module is an essential feature in the functioning of TFs, and TFs are often classified by the type of DNA-binding domain they contain. Other parts of the protein can contribute to and influence the intrinsic DNA-binding activity, including sequences that flank the DBD and that mediate dimerization. It is estimated that TFs constitute between 0.5 and 8% of the gene content of eukaryotic genomes, with both the absolute number and proportion of TFs in a genome roughly scaling with the complexity of the organism (Levine & Tjian 2003). Most eukaryotic TFs tend to recognize short, degenerate DNA sequence motifs, in contrast to the larger motifs preferred by prokaryotic TFs (Wunderlich & Mirny 2009). Cooperation among TFs, rather than highly-specific sequence preferences, is believed to be a pervasive feature of eukaryotic transcriptional regulation (Arnosti & Kulkarni 2005).

The distinguishing feature of TFs, relative to other transcriptional regulatory proteins, is that they interact with DNA in a sequence-specific manner (Karin 1990; Latchman 1997). In the vast majority of well-studied cases, these interactions are mediated by DNA binding domains (DBDs) (Luscombe et al. 2000), and TF families are typically defined on the basis of sequence similarity of their DBDs. One of the most abundant DBD in eukaryotic TFs is the zinc finger (ZNF). Different classes of zinc finger domains have been identified and characterized according to the nature and spacing of their zinc-chelating residues (Mackay & Crossley 1998). The canonical C2H2 ZNF motif, comprises 28 to 30 amino acid residues and its structure is stabilized by a zinc ion coordinated by four highly conserved residues, two cysteines and two histidines (Krishna et al. 2003). The stably folded structure consists of one alpha helix and two to three beta strands. The alpha helix mediates DNA binding through non-covalent interactions

between three of its amino acid residues and three adjacent bases within the DNA major groove (Wuttke et al. 1997). This zinc-dependent structure is required for the interaction between the finger motif and nucleic acids; in the absence of zinc, or if elements of conserved C2H2 structure are abolished through mutations, zinc fingers lose their ability to fold properly and to bind DNA (Pavletich & Pabo 1991; Pavletich & Pabo 1993; Brayer & Segal 2008).

The C2H2 zinc finger motif, first identified in studies of the *Xenopus* TF TIFIIA (Klug et al. 1986) is by far the most common protein domain in metazoan TFs. Most versions of this motif correspond to a subtype called the "*Krüppel* –type" named for the *Drosophila Krüppel* protein, a developmentally active TF that has the canonical C2H2 zinc-binding structure. C2H2 zinc finger, also called *Krüppel* –type (KZNF) proteins, contain from 3 to more than 30 zinc-finger motifs, which are arranged in tandem within the protein. The tandem arrangement of KZNF motifs permits the adjacent fingers to interact, to modulate each other's DNA binding, and to stabilize DNA binding of the protein at specific sites (Laity et al. 2001). In addition to the paired cysteine and histidine residues, KZNF motifs contain a highly conserved "spacer" within fingers, or H/C link sequence, a seven amino acid segment with the consensus sequence TGEKP(Y/F). Variations in the amino acid sequence of the finger domains and spacing, as well as in zinc finger number and higher-order structure, may increase the ability of these proteins to bind multiple different ligands such as RNA, DNA-RNA hybrids and even proteins, thus highlighting the structural and functional versatility of this protein family (Vissing et al. 1995; Tommerup & Vissing 1995).

Evolution and Structure of KRAB-ZNF Proteins

While zinc-fingers define binding site specificity and stability for KZNF proteins, most TFs of this type also require one of more "effector" domains to translate site-specific DNA binding into gene regulatory activities impacting neighboring genes. These include the BTB/POZ domain, the SCAN domain, and the KRAB domain (*Krüppel*-associated box) (Bellefroid et al. 1993; Collins et al. 2001). The KRAB-ZNF gene family represents a more recent evolutionary product and its expansion in the genome of tetrapod vertebrates could indicate the acquisition of new functions to sustain differentiation and

speciation. A comparative analysis of mammalian genomes revealed the existence of a large and highly conserved number of genes that originated through repeated cycles of duplications from a single ancestral gene. After their duplications, these new genes diversified their coding regions to produce novel proteins with new biological functions (Shannon et al. 2003; Emerson & Thomas 2009).

These genes have been found clustered at particular sites on chromosomes suggesting the existence of a common repertoire of regulatory sequences and a coordinated mechanism of their gene expression (Nowick et al. 2010; Huntley 2006). In the mammalian genome, the gene families encoding olfactory receptors, alpha-globins, KYR proteins and KRAB-ZNF are the most representative within this class of clustered genes (Mombaerts 1999; Uhrberg 2005). Interestingly, only the genes encoding KRAB-ZNFs are differentially expressed in various tissues during differentiation and development, indicating that these gens have functions unique to mammalian evolution and molecular processes that establish the phenotypic differences between vertebrates and other species (Vissing et al. 1995; Bellefroid et al. 1993; Lorenz et al. 2010). Therefore, expression of KRAB-ZNF genes is independent of their genome localization, as well as of nearby paralogs generated through gene duplications within the same gene cluster. These paralogs, as new members of the KRAB-ZNF family, show different expression patterns and novel non-redundant functions (Urrutia 2003).

While most vertebrate transcription factor families are largely conserved, the C2H2 zinc finger (ZNF) family stands out as a significant exception. Novel gene types have arisen to encode proteins in which DNA-binding ZNF motifs are tethered to different types of chromatin-interacting effector domains (Pearson et al. 2008). Some of the gene types have been expanded by duplication and diverged independently to yield many lineage-specific TF genes. For example, the evolutionary history of the KRAB-associated C2H2 zinc finger (KRAB-ZNF) family is distinct from that of other transcription factor types, involving an unprecedented level of species-specific diversity as a result of segmental duplication over the course of evolutionary history (Stubbs et al. 2011). Available data indicate that the process of generating new KRAB-ZNF genes is ongoing; for example, analysis of the human genome revealed more than 20 new genes generated

within the past 35-40 million years (My) (Nowick et al. 2010; Nowick et al. 2011), and at least 136 of the 394 identified human genes are primate-specific (Huntley 2006).

The larger class of ZNF genes primarily encode proteins that function as transcription factors, and typically contain an array of two or more tandemly arranged C2H2 zinc finger motifs. DNA binding of these zinc fingers is affected by specific interaction between four amino acids within a ZNF motif; each finger can bind three adjacent nucleotides at target sites with amino acids in positions -1, 2, 3, and 6 in the alpha-helical region (Pavletich & Pabo 1993; Kim & Berg 1996). We refer to these four amino acids as a protein's "fingerprint." The ZNF array winds around the DNA target site within the major groove on the DNA helix, such that the DNA-contacting amino acids in each finger interact directly with adjacent sets of target-site nucleotides. The human genome encodes hundreds of KRAB-ZNF genes, encoding proteins in which arrays of tandem ZNF motifs are tethered to an N-terminal effector domain called the *Krüppel*-associated box or KRAB (Bellefroid et al. 1993; Consiantinou-Deltas et al. 1992). The canonical mammalian KRAB A domain interacts with a universal cofactor, KAP1, which recruits histone deacetylase and methylation complexes to the ZNF-binding sites, and KRAB-ZNF proteins are thus thought to act as potent transcriptional repressors (Vissing et al. 1995; Margolin et al. 1994; Pengue et al. 1994; Witzgall et al. 1994).

The Deeply Conserved Gene Family Representing the Root of Mammalian ZNF Genes

While most attention has been focused on more recently evolved, primate-specific KRAB-ZNF genes, the origins and deeper vertebrate roots of the KRAB-ZNF family has remained mysterious. The evolutionary dynamic of this family severely complicates the identification of ZNF gene homologs, including those that remain functionally conserved. To alleviate that problem, we searched for vertebrate "DNA binding orthologs" by mining ZNF gene models from seven sequenced genomes (opossum, chicken, zebra finch, lizard, frog, mouse, and human genome) (Liu et al. 2014). From these models, we extracted and aligned the patterns of DNA-binding amino acids, or "fingerprints", to look for related patterns across species. Although this study identified all genes in which multiple, tandem ZNF motifs were encoded, the most interesting results were revealed in

analysis of the KRAB-ZNF genes. Surprisingly, of the nearly 400 human KRAB-ZNF genes, only three genes were found to recognize proteins with similar fingerprints in both mammalian and in non-mammalian vertebrates. These three genes, *ZNF777, ZNF282*, and *ZNF783*, are members of an ancient familial cluster, and are likely to represent the founders of the mammalian ZNF family. These ancient genes are unusual in mammals in that, like the single KRAB domain present in the original protein of this type, PRDM9, they encode a noncanonical KRAB domain that cannot bind to KAP1 and may function as a transcriptional activator (Okumura et al. 1997; Conroy et al. 2002). Evolutionary analysis confirmed the ancient provenance of this activating KRAB and revealed the independent expansion of KRAB-ZNFs in every vertebrate linage. In non-mammalian vertebrates, most KRAB-ZNF genes contain an activating KRAB domain, and the KAP1-binding version of this domain appears to have been selected for dominance particularly in mammalian lineages (Liu et al. 2014).

Since their first appearance in amniote species, *ZNF777, ZNF282* and *ZNF783* have given rise to new duplicates; *ZNF398, ZNF212* and *ZNF746* are present in marsupials and eutherians, while other duplicates are found only in eutherian species (Liu et al. 2014). These duplications occurred in tandem so that these closely related genes are located in a single cluster in mammalian genomes. A limited amount of data exists regarding the functions of these conserved KRAB-ZNF family members. In particular, ZNF282 has been shown to bind U5RE (U5 repressive element) on the long terminal repeat (LTR) of human T-cell leukemia virus type I (HTLV-I) and represses the HTLV-I LTR-mediated expression (Okumura et al. 1997). In the same report, the KRAB domain of ZNF282 was shown to function as a transcriptional activating domain unlike the canonical KRAB domain. The N-terminal exon encodes a domain which is specific to this conserved gene family, named “HUB” (HTLV-I U5RE binding) repressive domain, was demonstrated to repress transcription, and the amino acids 1-75 region of ZNF282 is indispensable for the repressive activity. In two more recent studies, ZNF282 was identified to interact with estrogen receptor α (ERα) and cooperate synergistically with CoCoA (Coiled-coil co-activator) to function as an ERα co-activator in breast cancer cells (Conroy et al. 2002). Also, ZNF282 is SUMOylated and the SUMOylation positively regulates the co-activator activity of ZNF282 by increasing the binding affinity

to ERα and CoCoA (Yu et al. 2012). The same group later demonstrated that ZNF282 functions as a coactivator for one of the key cell cycle-regulating transcription factors, E2F1, and is required for E2F1-mediated gene expression in Esophageal squamous cell carcinoma (ESCC) cells, which links ZNF282 to cell cycle control mechanisms (Yeo et al. 2014).

Another family member, ZNF398, has also been shown to interact with ERα. ZNF398 has two different isoforms that are generated by alternative splicing: the 71 kDa full-length isoform (p71), and a 52 kDa isoform lacking the HUB domain (p52). The p52 isoform interacts strongly with ERα in the presence of 17 β-estradiol, whereas the p71 isoform has a HUB domain that inhibits the interaction with ERα (Conroy et al. 2002). Both isoforms can activate transcription through the ZNF398 binding element; however, in the presence of ERα, transactivation by the p52 isoform is specifically repressed. Overexpression of the p52 isoform was able to abrogate activation by p71 isoform. Therefore, the regulation of transcription mediated by ZNF398, and possibly other family members, can be controlled by the relative level of expression of distinct isoforms (Conroy et al. 2002). A third family member, ZNF746 (PARIS), has been shown to accumulate in models of *parkin* inactivation and in human PD (Parkinson's disease) brain, and the levels of ZNF746 is regulated by *parkin* via the ubiquitin proteasome system (Shin et al. 2011). ZNF746 represses the expression of the transcriptional coactivator, PGC-1α (peroxisome proliferator-activated receptor gamma (PPARγ) coactivator-1α) and the PGC-1α target gene, NRF-1 by binding to insulin response sequences in the PGC-1α promoter (Shin et al. 2011).

Together these data suggest some common features and potential functions for members of this conserved family. First, this ancient, clustered KRAB-ZNF subfamily share a noncanonical KRAB domain which does not act as a repressor, in addition to a novel N-terminal HUB domain which may have repressive activity depending on possible cooperation with other proteins. Second, these members may have different isoforms including or excluding exons encoding HUB and KRAB domains, influencing protein-protein interactions and regulatory functions. Third, these related proteins commonly interact with nuclear receptors but may also interact with other TFs with strong influence on the expression of target genes. The fact that ZNF282 binds to and

silences retroviral LTRs has particularly interesting relevance to the human KRAB-ZNF family, since their dynamic evolution has been linked to an "arms race" to silence retroviral invasions (Emerson & Thomas 2009; Thomas & Schneider 2011; Wolf & Goff 2009; Jacobs et al. 2014) This connection raises questions regarding the possible interaction with other subfamily members and retroviral LTRs.

Aims of This Study

The focus of this study is to investigate the function of the most conserved founding members of the KRAB-ZNF transcription factor family, *ZNF777.* This gene stands out among this dynamic family for their deep conservation; the structure of the ZNF DNA binding domains suggests that the regulatory activities have been strictly maintained for hundreds of millions of years. These observations suggest that ZNF777 has adopted essential functions that are shared across amniotes. The central purpose of this study is to illuminate those regulatory functions and to understand the biological roles that have been adopted by mammalian *ZNF777*.

Chapter 2 is focused on addressing the regulatory function of ZNF777 in human placenta-derived choriocarcinoma cells. ZNF777 has been shown to be highly expressed in adult immune and reproductive tissues especially the placenta (Liu et al. 2014), indicating that it has been enlisted to regulate evolutionary divergent biological traits. In a recent study, Yuki et al. has shown that ZNF777 is involved in regulating cell cycle progression, as overexpression of ZNF777 inhibits proliferation at low cell density through down-regulation of FAM129A, and the induction of p21 activity (Yuki et al. 2015). This study provides the first examination of the global binding sites for ZNF777, focused on BeWo cells which are an established model of human placental trophoblasts. It also reveals the functions of genes affected by *ZNF777* depletion, in the form of siRNA knockdown, in BeWo cells. By correlating these two datasets, I was able to assess the effects of ZNF777 protein binding with direct and downstream transcriptomic outcomes. In the process, I have identified a clear and distinct binding motif for ZNF777 protein for the first time.

Chapter 3 describes the development and application of tools for studying the functions of Zfp777 and Zfp282 in the context of developing neurons, where our previous study

also showed (Liu et al., and this thesis) the genes and proteins are also highly expressed. Using a version of the CRISPR-Cas9 system to introduce C-terminal epitope tags ("CETCh-seq", Savic et al. 2015), we tagged endogenous Zfp282 and Zfp777 with FLAG sequences, and performed ChIP-seq to uncover the binding sites of the Zfp777 in mouse neural stem cells. A strong binding motif of Zfp777 was identified, with most of the binding sites at the promoter regions of protein-coding genes, therefore, possible functions of Zfp777 in mouse neural stem cells can be implicated by examining the genes whose promoters were bound by Zfp777.

Putting these pieces of information together allows a detailed model of the functions of ZNF777 to be developed, elucidating the genome-wide regulatory functions of this proteins, extant mammalian representatives of a large and ancient TF class, and the founders of the largest TF family in mammalian genomes.

References

Arnosti, D.N. & Kulkarni, M.M., 2005. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *Journal of Cellular Biochemistry*, 94(5), pp.890–898.

Bellefroid, E.J. et al., 1993. Clustered organization of homologous KRAB zinc-finger genes with enhanced expression in human T lymphoid cells. *The EMBO journal*, 12(4), pp.1363–1374.

Brayer, K.J. & Segal, D.J., 2008. Keep Your Fingers Off My DNA: Protein–Protein Interactions Mediated by C2H2 Zinc Finger Domains. *Cell Biochemistry and Biophysics*, 50(3), pp.111–131.

Collins, T., Stone, J.R. & Williams, A.J., 2001. All in the family: the BTB/POZ, KRAB, and SCAN domains. *Molecular and Cellular Biology*, 21(11), pp.3609–3615.

Conroy, A.T. et al., 2002. A Novel Zinc Finger Transcription Factor with Two Isoforms That Are Differentially Repressed by Estrogen Receptor. *Journal of Biological Chemistry*, 277(11), pp.9326–9334.

Consiantinou-Deltas, C.D. et al., 1992. The identification and characterization of KRAB-domain-containing zinc finger proteins. *Genomics*, 12(3), pp.581–589.

Emerson, R.O. & Thomas, J.H., 2009. Adaptive Evolution in Zinc Finger Transcription Factors S. Myers, ed. *PLoS Genetics*, 5(1), pp.e1000325–12.

Huntley, S., 2006. A comprehensive catalog of human KRAB-associated zinc finger genes: Insights into the evolutionary history of a large family of transcriptional repressors. *Genome Research*, 16(5), pp.669–677.

Jacobs, F.M.J. et al., 2014. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature*, 516, pp.242–245.

Karin, M., 1990. Too many transcription factors: positive and negative interactions. *The New biologist*, 2(2), pp.126–131.

Kim, C.A. & Berg, J.M., 1996. A 2.2 A resolution crystal structure of a designed zinc finger protein bound to DNA. *Nature structural biology*, 3(11), pp.940–945.

KLUG, A., MILLER, J. & McLACHLAN, A.D., 1986. Repetitive Zn 2+-binding domains in the protein transcription factor IIIA from Xenopusoocytes. *Biochemical Society Transactions*, 14(2), pp.221.2–221.

Krishna, S.S., Majumdar, I. & Grishin, N.V., 2003. Structural classification of zinc fingers: survey and summary. *Nucleic Acids Research*, 31(2), pp.532–550.

Laity, J.H., Lee, B.M. & Wright, P.E., 2001. Zinc finger proteins: new insights into structural and functional diversity. *Current opinion in structural biology*, 11(1), pp.39–46.

Latchman, D.S., 1997. Transcription factors: an overview. *The international journal of biochemistry & cell biology*, 29(12), pp.1305–1312.

Levine, M. & Tjian, R., 2003. Transcription regulation and animal diversity. *Nature*, 424(6945), pp.147–151.

Liu, H. et al., 2014. Deep Vertebrate Roots for Mammalian Zinc Finger Transcription Factor Subfamilies. *Genome Biology and Evolution*, 6(3), pp.510–525.

Lorenz, P. et al., 2010. The ancient mammalian KRAB zinc finger gene cluster on human chromosome 8q24.3 illustrates principles of C2H2 zinc finger evolution associated with unique expression profiles in human tissues. *BMC genomics*, 11(1), p.206.

Luscombe, N.M. et al., 2000. An overview of the structures of protein-DNA complexes. *Genome biology*, 1(1), p.REVIEWS001.

Mackay, J.P. & Crossley, M., 1998. Zinc fingers are sticking together. *Trends in Biochemical Sciences*, 23(1), pp.1–4.

Margolin, J.F. et al., 1994. Krüppel-associated boxes are potent transcriptional repression domains. *Proceedings of the National Academy of Sciences*, 91(10), pp.4509–4513.

Mombaerts, P., 1999. Odorant receptor genes in humans. *Current Opinion in Genetics & Development*, 9(3), pp.315–320.

Nowick, K. et al., 2011. Gain, Loss and Divergence in Primate Zinc-Finger Genes: A Rich Resource for Evolution of Gene Regulatory Differences between Species M. A. Batzer, ed. *PLoS ONE*, 6(6), pp.e21553–11.

Nowick, K. et al., 2010. Rapid Sequence and Expression Divergence Suggest Selection for Novel Function in Primate-Specific KRAB-ZNF Genes. *Molecular Biology and Evolution*, 27(11), pp.2606–2617.

Okumura, K. et al., 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 Research*, 25(24), pp.5025–5032.

Pavletich, N.P. & Pabo, C.O., 1993. Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science*, 261(5129), pp.1701–1707.

Pavletich, N.P. & Pabo, C.O., 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science*, 252(5007), pp.809–817.

Pearson, R. et al., 2008. Krüppel-like transcription factors: a functional family. *The international journal of biochemistry & cell biology*, 40(10), pp.1996–2001.

Pengue, G. et al., 1994. Repression of transcriptional activity at a distance by the evolutionarily conserved KRAB domain present in a subfamily of zinc finger proteins. *Nucleic Acids Research*, 22(15), pp.2908–2914.

Savic, D. et al., 2015. CETCh-seq: CRISPR epitope tagging ChIP-seq of DNA-binding proteins. *Genome Research*, 25(10), pp.1581–1589.

Shannon, M. et al., 2003. Differential expansion of zinc-finger transcription factor loci in homologous human and mouse gene clusters. *Genome Research*, 13(6A), pp.1097–1110.

Shin, J.-H. et al., 2011. PARIS (ZNF746) Repression of PGC-1α Contributes to Neurodegeneration in Parkinson's Disease. *Cell*, 144(5), pp.689–702.

Stubbs, L., Sun, Y. & Caetano-Anolles, D., 2011. Function and Evolution of C2H2 Zinc Finger Arrays. *Sub-cellular biochemistry*, 52, pp.75–94.

Thomas, J.H. & Schneider, S., 2011. Coevolution of retroelements and tandem zinc finger genes. *Genome Research*, 21(11), pp.1800–1812.

Tommerup, N. & Vissing, H., 1995. Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs identify putative candidate genes for developmental and malignant disorders. *Genomics*, 27(2), pp.259–264.

Uhrberg, M., 2005. The KIR gene family: life in the fast lane of evolution. *European Journal of Immunology*, 35(1), pp.10–15.

Urrutia, R., 2003. KRAB-containing zinc-finger repressor proteins. *Genome biology*, 4(10), p.231.

Vissing, H. et al., 1995. Repression of transcriptional activity by heterologous KRAB domains present in zinc finger proteins. *FEBS Letters*, 369(2-3), pp.153–157.

Witzgall, R. et al., 1994. Genomic structure and chromosomal location of the rat gene encoding the zinc finger transcription factor Kid-1. *Genomics*, 20(2), pp.203–209.

Wolf, D. & Goff, S.P., 2009. Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature*, 458(7242), pp.1201–1204.

Wunderlich, Z. & Mirny, L.A., 2009. Different gene regulation strategies revealed by analysis of binding motifs. *Trends in genetics : TIG*, 25(10), pp.434–440.

Wuttke, D.S. et al., 1997. Solution structure of the first three zinc fingers of TFIIIA bound to the cognate DNA sequence: determinants of affinity and sequence specificity. *Journal of Molecular Biology*, 273(1), pp.183–206.

Yeo, S.-Y. et al., 2014. ZNF282 (Zinc finger protein 282), a novel E2F1 co-activator, promotes esophageal squamous cell carcinoma. *Oncotarget*, 5(23), pp.12260–12272.

Yu, E.J. et al., 2012. SUMOylation of ZFP282 potentiates its positive effect on estrogen signaling in breast tumorigenesis. *Oncogene*, 32(35), pp.4160–4168.

Yuki, R. et al., 2015. Overexpression of Zinc-Finger Protein 777 (ZNF777) Inhibits Proliferation at Low Cell Density Through Down-Regulation of FAM129A. *Journal of Cellular Biochemistry*, 116(6), pp.954–968.

CHAPTER 2: FUNCTIONS OF ZNF777, A GENE REPRESENTING THE ROOT OF THE MAMMALIAN KRAB ZINC FINGER FAMILY

Li-Hsin Chang[1,2], Joseph M. Troy[2,3], Huimin Zhang[1,2], Bob Chen[1,2], Xiaochen Lu[1,2], and Lisa Stubbs[1,2,3,4]

[1] Department of Cell and Developmental Biology,

[2] Carl R. Woese Institute for Genomic Biology,

[3] Illinois Informatics Institute,

University of Illinois at Urbana-Champaign, Urbana IL 61801

[4] Corresponding author

Running Title: Functional analysis of ZNF777

Abstract

The evolutionary history of the KRAB-associated C2H2 zinc-finger (KRAB-ZNF) family is distinct from that of other transcription factor (TF) types, involving an unprecedented level of species-specific diversity. We recently showed that most land vertebrates carry hundreds of KRAB-ZNF genes; however, of the 394 human KRAB-ZNF genes only three have been conserved throughout amniote history. These three genes, members of an ancient familial cluster, encode a noncanonical KRAB domain that is similar to an ancient domain which is prevalent in non-mammalian species. In contrast to the mammalian KRAB, which is thought to function as a potent repressor, this ancient domain serves as a transcriptional activator. Here we report the regulatory functions of the most deeply conserved member in this family, *ZNF777,* using chromatin immunoprecipitation (ChIP-seq) and siRNA knockdown experiments. We used human choriocarcinoma cells for these experiments to model functions in placental trophoblasts, where *ZNF777* is most highly expressed. Of the genes flanking *ZNF777* binding regions, many were down-regulated after *ZNF777* depletion consistent with a transcriptional activator role. However, a significant number of bound genes were oppositely regulated, suggesting a more complex relationship. Investigating further, we show that this discrepancy is likely linked to the fact that *ZNF777* encodes both full-length (HUB-KRAB-ZNF) and ZNF-only isoforms, which can be predicted to display different regulatory activities. Together the data suggest roles in regulation of genes such as semaphorins, ephrins and related proteins with known roles in placenta angiogenesis and in the embryonic brain, where *ZNF777* is also highly expressed.

Introduction

Although most vertebrate transcription factor families are relatively conserved, the C2H2 zinc finger (ZNF) family stands out as a significant exception. In particular, the KRAB-associated C2H2 zinc finger (KRAB-ZNF) subfamily displays an unprecedented level of evolutionary diversity, driven by repeated series of gene duplications accompanied by gene loss (Huntley et al. 2006; Nowick et al. 2010). For example, of the 394 KRAB-ZNF genes in the human genome, fewer than 100 genes are conserved as 1:1 orthologs in mouse and at least 136 are found only in primate genomes.

The KRAB-ZNF gene family encodes proteins with two primary structural domains: a C-terminal DNA binding domain (DBD) composed of a tandem array of zinc fingers, and one or more copies of an effector domain, called the *Krüppel*-associated box (KRAB). DNA binding is mediated by specific interaction between four amino acids within each ZNF motif (amino acids in positions -1, 2, 3, and 6 relative to the alpha helix) and three adjacent nucleotides at the DNA target sites (Pavletich & Pabo 1991; Pavletich & Pabo 1993; Kim & Berg 1996; Wolfe et al. 2000). This pattern of four DNA-binding amino acids in each ZNF unit thus defines a protein's DNA binding capabilities. As we have in previous reports (Liu et al. 2014), we will refer to this pattern as a protein's "fingerprint" in the following discussion. After the ZNF motifs select the target DNA site based on fingerprint specificity, the canonical mammalian KRAB domain, called KRAB A, interacts with a universal cofactor, KAP1, to recruit histone deacetylase and methylation complexes to the ZNF-binding sites. For this reason, KRAB-ZNF proteins are thus typically thought to act as potent transcriptional repressors (Margolin et al. 1994; Pengue et al. 1994; Witzgall et al. 1994; Vissing et al. 1995).

While most attention has been focused on the more recently evolved primate KRAB-ZNF genes (Nowick et al. 2011; Lupo et al. 2013), the vertebrate roots of the KRAB-ZNF families has remained mysterious. To address questions regarding the pre-mammalian history of the KRAB-ZNF family, we recently mined ZNF loci from seven sequenced genomes (opossum, chicken, zebra finch, lizard, frog, mouse, and human genome) and compared DBD sequence and fingerprints looking for predicted "DNA binding orthologs" across species (Liu et al., 2014). Interestingly, we found hundreds of

KRAB-ZNF proteins in every species we examined, but only three human genes were found with clear orthologs in non-mammalian vertebrates. These three genes, *ZNF777, ZNF282*, and *ZNF783*, are members of an ancient familial cluster and encode proteins with similar domain structures. Our evolutionary analysis confirmed the ancient provenance of this activating KRAB and revealed the independent expansion of KRAB-ZNFs in every vertebrate lineage. This finding led us to ask the question: what are the functions of these ancient family members and why, of such a large and diverse family group, were these three genes conserved so fastidiously over hundreds of millions of years?

The existing literature offers a few functional clues. For example, *ZNF282* has been shown to bind U5RE (U5 repressive element) on the LTR of human T-cell leukemia virus type I (HTLV-I) and to repress HTLV-I LTR-mediated expression (Okumura et al. 1997). This same report offered the first evidence that the KRAB domain of ZNF282 functions as an activator and does not bind KAP1. The repressive function of ZNF282 is derived instead from an N-terminal domain specific to this conserved gene cluster, named "HUB" (HTLV-I U5RE binding). In two more recent studies, ZNF282 was identified to interact with estrogen receptor α (ERα) (Yu et al. 2012), and E2F1, linking ZNF282 to cell cycle control (Yeo et al. 2014). With a pointer to some common functions, a recent study also implicated ZNF777 as a cell cycle regulator (Yuki et al. 2015). We demonstrated high levels of human *ZNF777* expression in placenta and mouse *Zfp777* in embryonic brain, suggesting that the protein has adopted lineage-specific functions in mammals (Liu et al. 2014). However, regulatory functions of these ancient proteins have not been further explored.

Here we report the regulatory function of ZNF777, combining chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-seq) with siRNA knockdown experiments to determine genome-wide binding sites, a distinct binding motif, and predicted targets for the protein in human BeWo choriocarcinoma cells. Genes neighboring ZNF777 binding sites can be either up- or down-regulated, suggesting a complex regulatory role. Our studies revealed that some of this complexity is due to the generation of HUB-containing and HUB-minus isoforms, which are predicted to have different regulatory activities. Based on these experiments, we hypothesize that ZNF777

regulates pathways best known for their roles in neurogenesis and axon pathfinding, but also recently shown to play critical roles in placental development.

Results

ZNF777 and the members of a deeply conserved family cluster on human chromosome 7

The genes representing the deepest vertebrate roots of the mammalian KRAB-ZNF family, *ZNF282, ZNF777*, and *ZNF783*, cluster together in mammalian species including the distal end of chromosome 7q36.1 in the human genome (**Figure 2.1A**). The proteins encoded by genes in this region each possess distinct ZNF DNA binding regions, suggesting that they bind different DNA sequences; on the other hand, the homologs for a particular gene in different species possess tightly conserved DNA binding domains (Liu et al., 2014).

In each zinc finger region, four amino acids, at positions -1, 2, 3, and 6 relative to the alpha-helix, bind specifically to cognate DNA sequences; this pattern of amino acids thus defines a ZNF protein's DNA binding preferences uniquely, and is generally conserved throughout evolution. We have referred to the amino acid sequences in these DNA binding positions as "fingerprints" in a previous study (Liu et al., 2014) and will use that abbreviation in this study. The fingerprints of human, mouse, platypus, opossum, bird, and lizard ZNF777 proteins share strikingly similarity, as illustrated by the alignment of the *ZNF777* orthologs (**Figure 2.1B**). Given the fact that so few KRAB-ZNF proteins are conserved in this respect, this very high level of conservation is especially remarkable. The data indicate a high level of selection for the DNA-binding specificities that are represented in these deeply conserved, ancestral genes. Among the members in this family, only ZNF777 was found to have conserved fingerprint in mammalian, avian, and reptilian genomes, indicating that ZNF777 is the most conserved member in this clustered group.

Comparison of the HUB domains of ZNF777 and ZNF282 suggests distinct functions

The predicted ZNF777 protein is comprised of a N-terminal domain (the HUB domain) from amino acids 1-282, a KRAB A-like domain from amino acids 283-324, a "tether" region, and nine zinc fingers at the C-terminus (**Figure 2.2A, top**). ZNF282 has been shown to act in transcriptional repression, with two domains within amino acids 1-75 and amino acids 96-184 of the protein, both required for repression (Okumura et al. 1997). As mentioned above, we have already shown that ZNF777, ZNF282 and ZNF783 have distinct fingerprint profiles (Liu et al. 2014). To ask whether the HUB domains of the clustered family members were similar enough in structure to share common function, we aligned the HUB domain protein sequences of ZNF777 and ZNF282 (**Figure 2.2B**), and other members within this subfamily (**Supplemental Figure 2.1**). At 282 amino acids in length, the HUB domain of ZNF777 is almost twice the length of that in ZNF282 (195 amino acids); other family members have even shorter HUB domains (108-140 amino acids).

One of the repressive domains identified in the ZNF282 HUB domain (amino acids 96 -184) shares high sequence similarity with the HUB domain of ZNF777 and all other members of this subfamily. However, ZNF777 lacks homology to the second region shown to be required for full repressive activity in ZNF282, spanning amino acids 1-73 (Okumura et al. 1997). Instead, amino acids 1-177 of the ZNF777 HUB domain are novel and not shared by ZNF282 or other cluster neighbors (**Supplemental Figure 2.1**). Although the mechanism of ZNF282 repression has not been clearly defined, these data suggest that ZNF777 and ZNF282 could have different functions, perhaps through recruitment of different binding partners. The status of ZNF777 as an activating or repressive TF is therefore not clear.

KRAB-ZNF genes frequently give rise to alternative splicing isoforms with various combinations of ZNF and effector domains (Huntley 2006). Several family members within the ZNF777 cluster are also known to be alternatively spliced, giving rise to HUB-containing (HUB+) and HUB-less (HUB-) isoforms. These alternative protein isoforms are of special interest, since they are likely to have distinct regulatory functions.

To investigate whether *ZNF777* also produces alternative isoforms, we used primers flanking the exons encoding HUB, KRAB A, and ZNF domains in reverse transcript PCR (qRT-PCR) experiments (**Figure 2.2A, bottom left**). In addition to the full-length *ZNF777* transcript, we also detected a PCR band of the length expected of a HUB minus and KRAB A minus isoform (ZNF-only). Concordant with these results, we also detected a protein isoform with size corresponding to ZNF-only isoform with a ZNF777 antibody in BeWo cell nuclear protein extracts (**Figure 2.2A, bottom right**).

ZNF777 is expressed in human placenta and other tissues

Analysis of publicly available RNA-seq data revealed high levels of expression of *ZNF777* and cluster relatives in human placenta (Liu et al. 2014). To map the expression of *ZNF777* more extensively, we employed quantitative real-time reverse transcript PCR (qRT-PCR) to measure the expression of *ZNF777* in human tissues and cell lines. These experiments confirmed that *ZNF777* is expressed placenta, in addition to a variety of human tissues, including lung, thymus, brain, pancreas, uterus, and fetal brain (**Figure 2.3A**). We also measured expression of *ZNF777* with immunohistochemistry (IHC) in a human tissue array (**Figure 2.3B**). The ZNF777 protein is expressed widely in a pattern that is consistent with the qRT-PCR results. In those tissues, the protein was identified in both nuclear and cytoplasmic compartments, depending on the cell type.

ZNF777 localization was further investigated by Immunocytochemistry (ICC) in cultured BeWo cells, a cell line derived from human choriocarcinoma that is used to model placental trophoblast functions (**Figure 2.4E**). The ZNF777 antibody (labeled in red) detected protein in the nucleus as well as in the perinuclear region in BeWo cells. These data suggest either that the protein has functions outside the nucleus, or that it may be mobilized to the nucleus under certain conditions, perhaps due to protein modifications, as is true for many TFs (Ziegler & Ghosh 2005). Expression in human cell lines was also measured in Western blots, confirming that both long (approximately 85 kDa) and short (~55 kDa) ZNF777 isoforms are expressed in human cell lines such as BeWo (human placenta choriocarcinoma), HEK293 (human embryonic kidney), JEG-3 (human placenta choriocarcinoma), SHEP (human neuroblastoma), human trophoblast stem cells (hTSC), U2OS (human osteosarcoma). In contrast, the shorter ZNF-only

isoform was the only protein detected in HUVEC (human umbilical vein/vascular endothelium) (**Figure 2.4A**).

We chose BeWo cells for further experiments to model activity in placenta. In addition to the short and long isoforms described above, we detected a faint protein band of larger size, possibly corresponding to a modified form of the protein in BeWo cells (**Figure 2.4C**). These data are in agreement with the study of Yuki and colleagues (Yuki et al. 2015), who tested the expression of ZNF777 protein in HCT116 cells. We tested the specificity of the antibody by siRNA knock down followed by a western blot with protein from the siRNA-treated cells; we detected a decrease of the ZNF777 protein when *ZNF777* transcript expression was depleted by treating BeWo cells with two different siRNA molecules, which we will call Si1 and Si4 (**Figure 2.4C**). Interestingly, the two different siRNAs, which target distinct *ZNF777* exons (**Figure. 2.4C**) had different effects on the protein profile. Specifically, Si1, which binds ZNF777 mRNA at the HUB domain, reduced the quantity of the full-length isoform only (**Figure. 2.4D**). On the other hand, Si4 binds ZNF777 mRNA at the spacer exon shared by full-length and ZNF-only isoforms, and it knocked down both short and long protein isoforms similarly (**Figure. 2.4D**).

The binding motif of ZNF777 identified by ChIP-Seq has sequence similarity with GRHL1 binding motif

To identify genomic binding sites, we performed Chromatin immunoprecipitation (ChIP) followed by Illumina sequencing (ChIP-seq) in chromatin from BeWo cells. After alignment of ChIP-enriched fragments we used MACS software to identify 1979 peaks. Of these, 709 peaks were detected at a minimal false discovery rate (fdr ~ 0). We found that ZNF777 binds to its own family members, including ZNF398, ZNF212, and ZNF282 (**Figure 2.5A**), suggesting regulation by ZNF777 of the expression of these members. We used the summits of 118 peaks with the highest level of ChIP enrichment to search for a potential ZNF777 binding motif. The predicted motif (**Figure 2.5B**) was identified in 113 out of the 118 peaks and with an unusually strong enrichment (p= 7.9e-103); some other less enriched motifs were found but mostly not centrally located and more degenerated, which suggested that ZNF777 might interact with different binding partners on different

binding sites depending on the contexts. The most enriched predicted motif has significantly similarity to that identified for another conserved TF, GRHL1 (grainyhead like 1). Interestingly, ZNF777 was found to bind to the promoter region of GRHL1 gene (**Figure 2.5A**), suggesting the regulation of the expression of GRHL1 by ZNF777.

Previous studies have shown that ZNF282 can bind to the long terminal repeat (LTR) or human T cell leukemia virus type I (HTLV-I) and represses its LTR-mediated expression (Okumura et al. 1997). To ask whether ZNF777 also interacts with LTRs, we examined the overlapping of the peaks and repeat elements in the human genome. We intersected the peaks and repeat elements and performed a randomization test to filter out the peaks that intersect with repeat elements by chance. Although the majority of ZNF777 peaks are unique, we found certain subfamilies of repeat elements were at ZNF777 peaks. These include MER31A, MER31B, MER39B, MER9B, MER65C, all of which belong to ERV1 family (**Supplemental Table 2.1**). These data suggest that like ZNF282, ZNF777 may also originally have evolved to bind endogenous retrovirus LTRs and may play a role in regulating ERV expression, in particular those specific human MER subfamilies. These are older, established ERV element and ZNF777 motif may have been carried by ancient elements, some of which are too degenerate to be detected as retroviral elements in modern genomes. We hypothesized that these ancient elements might have been coopted in mammalian genomes as regulatory elements for nearby genes. To examine this possibility, we looked at gene expression after depleting BeWo cells for ZNF777 protein, as described in the following section.

Gene Expression after ZNF777 knockdown reveals a role in extracellular matrix interactions and axon pathfinding during differentiation

To elucidate the biological functions of ZNF777, we performed siRNA knockdown by transfecting BeWo cells with ZNF777 siRNA_Si1, siRNA_Si4, or negative control siRNA (Si-Neg), and compared gene expression in the treated cells by RNA-sequencing (RNA-seq). We focused in 915 differentially expressed genes (DEG) that were identified as similarly up- or down-regulated in Si1 and Si4 treated cells (by at least 1.5-fold change), including 566 up-regulated and 349 downregulated genes (**Figure 2.6C**). These DEGs are expected to include both direct ZNF777 regulatory targets as well as

downstream genes. This overlapping gene set should enrich for genes affected by knockdown of the full-length protein since Si1 is specific to that isoform.

To identify potential direct targets of full-length ZNF777, we intersected the consistently up- or down-regulated DEGs with ChIP peaks and found 54 DEGs either containing or flanking the 709 fdr=0 ZNF777 peaks. These putative peak-associated DEGs also showed a mixed pattern of up- or down-regulation (34 compared to 20 genes, respectively) after siRNA treatment, reminiscent of the pattern of total DEGs. To validate the RNA-seq results, we tested 20 DEGs with QPCR in repeat knockdown experiments, including some genes with overlapping patterns and some with opposite patterns of expression after Si1 or Si4 treatment, and we saw the same patterns of differential expression in both QPCR duplicates (**Table 2.1**). We found most of the peak-associated DEGs from the two knock-downs share similar trends, in which the DEGs are either both up-regulated or down-regulated in both Si1 and Si4, but showed more differential expression in one Si as opposed to the other **(Figure 2.6A**).

To uncover the pathways regulated by ZNF777, we submitted the total DEGs from Si1 and Si4 knock-downs separately and the overlapping DEGs to DAVID (**Supplemental Table 2.2**). We found several significantly enriched GO categories, mostly derived from the up-regulated DEGs. Despite the differences in Si1 and Si4, the DEGs associated with both knockdowns were enriched in the same functional categories, including extracellular matrix organization, heparin binding, cell adhesion molecules, synapse, and axon guidance (**Supplemental Table 2.2**). Looking only at the 915 genes affected similarly by Si1 and Si4, we found striking enrichment in categories related to differentiation and neuron development, including semaphorin activity and synapse assembly for down-regulated genes; up-regulated genes, by contrast, were highly enriched in a variety of categories including virus receptors and immune function and nuclear hormone receptor pathways (**Table 2.2, Table 2.3**). Interestingly, DEGs flanking ZNF777 binding sites were particularly highly enriched in the related semaphorin pathways and PI3K-Akt3 signaling (**Table 2.4**). These pathways are central to both neurological and placental development (Jongbloets & Pasterkamp 2014; Liao et al. 2010; Dun & Parkinson 2017; Andermatt et al. 2014; Stoeckli 2017).

Discussion

Is ZNF777 and activator, repressor, or both?

The data presented here identify two transcript isoforms of *ZNF777* encoding proteins with different domain structures, and suggest that the may play distinct roles in the regulation of target genes. Another member of the same conserved, clustered gene family, *ZNF398* (Conroy et al. 2002) was similarly shown to express two isoforms: one called p52, which does not contain a HUB domain, and another called p71, which gives rise to the full-length HUB+KRAB+ZNF protein. In this published study, p52 was shown to interact with estrogen receptor α (ERα) via its zinc finger region; in the presence of estradiol, ERα binds to p52 and inhibits its gene activating role. In contrast, the HUB domain present in p71 inhibits ERα interaction, and the full-length protein can activate transcription without estrogen receptor interference. Therefore, the HUB domain does not interact with the interacting partner *per se*, but instead interfere with this interaction.

Here, we demonstrate for the first time that ZNF777 also gives rise to two isoforms, including a full-length protein and short isoform lacking the HUB and KRAB domains. siRNA knockdowns followed by RNA-seq data suggested possibly opposing regulatory activities were asserted by the full-length and ZNF-only isoforms on similar sets of genes. This inference will require further analysis but would be an intriguing result. It is well-known that antagonists of transcriptional activators can be useful in certain developmental situations, for example, to silence inappropriate gene expression in defined spatial or temporal domains, to down-regulate gene expression induced by transient stimuli, and to fine-tune transcriptional responses to complex developmental cues (Mitchell & Tjian 1989). There are many documented cases of activator or repressor isoforms produced from the same genes by alternative splicing (Foulkes & Sassone-Corsi 1992), and if this situation holds for ZNF777 it would not be an unusual one. If developmentally controlled, the production of such isoforms could permit finely tuned quantitative regulation of a defined set of genes or even opposite regulatory outcomes in response to different combinations of developmental cues. Further investigation of the detailed mechanism and interplay between the two isoforms of ZNF777 can be addressed by finding possible different binding partners for the HUB+ and HUB- isoforms; by

analogy with ZNF398 for example, the possible interaction with ERα or other nuclear receptors could be a fruitful avenue of future study.

The role of ZNF777 in placenta

Both DEGs more generally, and those located closest to ZNF777 binding sites in particular were enriched for genes that are best known for their roles in neuron differentiation; semaphorin genes, PI3k/AKT and related signaling pathways were especially highlighted in functional enrichments. We therefore hypothesize that ZNF777, the vertebrate root for the KRAB-ZNF family, is involved in the regulation of genes related to axon pathfinding, which is mostly well studied in neurogenesis. This pathway is an ancient one, active in brain development across the evolutionary spectrum and could be served by a TF stringently conserved as ZNF777 is known to be.

However, our experiments were completed not in neurons, but in a choriocarcinoma cell line that serves as a model for placental trophoblasts, a much more recently evolved cell type with a mammalian-specific function. Intriguingly, the semaphorin-plexin signaling pathway is also extensively involved in the development of a variety of tissues including cardiac and bone (Jongbloets & Pasterkamp 2014), and in mammals it has been coopted to serve as an important role in placental angiogenesis (Liao et al. 2010). The finding that ZNF777 is involved in regulation of this process is intriguing, and suggests that the expression of this transcription factor in placenta may have played a role in coopting the pathway for a mammalian-specific purpose.

The interaction between ZNF777 and repeat elements

There is growing evidence to suggest that the mammalian KRAB-ZNF evolved in an "arms race" to silence endogenous retroviruses (ERVs) (Jacobs et al. 2014). ERV insertions can create insertional mutations, the insertion of strong LTR promoters can also give rise to disease-causing mutations by inappropriately activating nearby genes (Jern & Coffin 2008). One way to battle these potentially harmful effects would be through the selection of TFs that could bind to viral LTR sequences and 'shut down' those potent enhancers where their expression would be harmful (Friedli & Trono 2015; Cordaux & Batzer 2009). However, like all infective agents, ERVs can evolve in

response to new mechanisms of suppression, through selection of mutations within the TFs' LTR binding sites. To keep up with the rapidly evolving ERVs, the TFs responsible for the task of silencing them might also be expected to evolve quickly, particularly in regions of the proteins that bind directly to the LTR DNA. The *Krüppel*-type zinc-finger (KZNF) TF family displays just this pattern of remarkable sequence divergence (Emerson & Thomas 2009). The creation of new gene copies and sequence divergence in this family both track LTR evolution remarkably well. Most TFs are deeply conserved, encoding proteins with highly similar structures and regulatory functions in a diverse array of species. However, specific classes of KZNF genes stand out from the rest, displaying a rapid pace of sequence, expression, and copy-number divergence. KZNF genes that encode proteins with multiple, tandemly arrayed ZNF motifs are particularly prone to this type of rapid divergence, reflecting unique properties of the genes. A co-evolution between KRAB-ZNFs and ERVs has been shown by the evidence that the number and age of newly emerged KRAB-ZNF genes and ERVs share striking correlation (Thomas & Schneider 2011).

In support of this idea, several KRAB-ZNFs have been demonstrated to bind to and regulate ERV LTR sequences. For example, rodent-specific Zfp809 has been shown to silence the moloney murine leukemia virus (MMLV) expression in mouse ES cells (Wolf & Goff 2009). Although *Zfp809* does not exist in humans, the mouse KRAB-ZNF protein can also bind to LTR regions of human T-cell lymphotrophic virus (HTLV-1), which shares the binding site found in MMLV. Presumably, silencing of HTLV-1 is accomplished through a different set of TFs in human cells. Recent studies also have shown that KRAB-ZNF genes ZNF91/93 interact with SVA/L1 retrotransposons and repress the expression of the two distinct retrotransposon families shortly after they began to spread in our ancestral genome (Jacobs et al. 2014).

Most relevant to this study, the human ZNF282 protein also binds to HTLV-1 LTR sequences and silences viral gene expression (Okumura et al. 1997). Since the *ZNF282* and *ZNF777* are close relatives, and both have been implicated as "roots" of the mammalian KRAB-ZNF family (Lui et al., 2014), and since ERVs play a critical role in placental development (Chuong et al. 2013), a relationship between ZNF777 and ERV sequences in the human genome presented an intriguing possibility.

Indeed, our data suggest that ZNF777 may also interact with certain ERV families in placental chromatin. Specifically, we found an enrichment for ERV1 and ERVL element sequences among the ZNF777 binding peaks. Most ZNF777 binding peaks are unique, as expected from the deep conservation of this protein since the original elements that may have driven its early divergence are certainly inactive in mammalian genomes. Therefore, the fact that we found the binding peaks to be enriched in recognizable, mammalian ERV sequences is especially interesting. The data suggest a more modern role in managing ERV-driven gene expression, likely one beyond the simple "arms race" silencing function that has been suggested for recent human and mouse gene duplicates. Given the flexible structure of ZNF777, which may generate silencing, activating, or other types of regulatory functions depending on alternative splicing, we speculated that this protein and its closest relatives may have had a more nuanced relationship to bioactive transposable elements like ERVs over the course of evolution. As these elements age, we hypothesize that they have left behind binding sites for ZNF777 and other TF proteins that are retained to regulate cellular genes. Further testing of this hypothesis will require additional experimentation, including tagging of endogenous proteins in species for which antibodies are not available; these kinds of approaches, made possible by the development of CRISPR technology (Ran et al. 2013; Savic et al. 2015) will open new doors to discovery of ZNF evolution and the functions of these deeply conserved TFs in the very near future.

Materials and Methods

Ethics Statement

This investigation has been conducted in accordance with the ethical standards and according to the Declaration of Helsinki and according to national and international guidelines.

RNA preparation and quantitative RT-PCR

Total RNA was isolated from cell lines and tissues using TRIzol (Invitrogen) followed by 30 minutes of RNase-free DNaseI treatment (NEB) at 37^{o}C and RNA Clean &

Concentrator™-5 (Zymo Research). 2 μg of total RNA was used to generate cDNA using Superscript III Reverse Transcriptase (Invitrogen) with random hexamers (Invitrogen) according to manufacturer's instructions.

Resulting cDNAs were analyzed of transcript-specific expression through quantitative reverse-transcript PCR (qRT-PCR) using Power SYBR Green PCR master mix (Applied Biosystems) with custom-designed primer sets purchased from Integrated DNA Technology. Relative expression was determined by normalizing the expression of all genes of interest to either human or mouse Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide (*YWHAZ*) expression (ΔCt) as described (Eisenberg and Levanon, 2003).

Cell culture and transfections

BeWo (ATCC, CCL-98) cell line and other cell lines were obtained from the American Type Culture Collection. BeWo cells in DMEM/F12K containing 2 mM L-glutamine, 10% FBS, 1X NEAA, 1X Pen Strep, incubated at 37 °C in 5% CO2.

For siRNA knockdown, approximately $4.5x10^5$ BeWo cells were seeded to 6-well plates 24 hours before transfection. Cells were treated with 10 nM of siRNA specific to ZNF777 (si1: SI04152729, si4: SI00458024, Qiagen) with a scrambled negative control (Alexa-siRNA, Ambion) for 48 hours using Lipofectamine RNAiMAX transfection reagent (Invitrogen) according to manufacturer's instructions.

RNA-Seq and computational analysis

48 hours after siRNA treatment, total RNA was prepared and tested for quality using an Agilent BioAnalyzer and Illumina libraries generated using the KAPA Stranded mRNA-Seq kit with mRNA Capture Beads (Kapa Biosystems, KK8420). Sequencing was performed on an Illumina Hi-Seq 2000 instrument at the University of Illinois Roy J. Carver Biotechnology sequencing facility, to yield 60-65 million reads per sample. The data have been submitted to the Gene Expression Omnibus database (accession numbers, in progress).

RNA-seq data were analyzed using the Tophat-Cufflinks Suite of tools (Trapnell et al. 2012). For *ZSCAN5A* knockdown, expression results from si4 and si5 were analyzed as a

group in comparison with the scrambled control. Genes identified as differentially expressed with $p \leq 0.05$ (after Benjamini-Hochberg correction for multiple testing) compared to the negative control-treated samples were considered for further analysis. For *ZSCAN5B* knockdown, which was effective only for a single siRNA design, we considered all genes with expression levels of at least 1 FPKM in at least one sample and considered genes with $\geq$ 1.5 X fold change relative to scrambled control as DEGs. siRNA up-regulated and down-regulated genes were analyzed for function separately using the DAVID (Huang et al. 2009)functional clustering algorithm with default settings.

Protein preparation, Western blots, and antibodies

Nuclear Extracts were prepared with NucBuster™ Protein Extraction Kit (Novagen) and measured by Bradford-based assay (BioRad). The extracts were stored at -80°C and thawed on ice with the addition of protease inhibitor Cocktail (Roche) directly before use. 15 µg of nuclear extracts were run on 10% acrylamide gels and transferred to hydrophobic polyvinylidene difluoride (PVDF) membrane (GE-Amersham, 0.45 µm) using BioRad Semi-dry system, then visualized by exposure to MyECL Imager (Thermo Scientific).

ZNF777 rabbit polyclonal antibody (ARP32659) was purchased from Aviva Systems Biology. The antibody is generated from an epitope on exon 5. (Epitope: LPQHLQSLGQLSGRYEASMYQTPLPGEMSPEGEESPPPLQLGNPAVKRLA).

Chromatin immunoprecipitation

Chromatin immunoprecipitation was carried out as essentially as described (Kim et al., 2003) with modifications for ChIP-seq. Chromatin was prepared from BeWo cell lines. About 1.0 x 10^6 Cells were fixed in PBS with 1% formaldehyde for 10 minutes. Fixing reaction was stopped with addition of Glycine to 0.125M. Fixed cells were washed 3x with PBS+Protease inhibitor cocktail (PIC, Roche) to remove formaldehyde. Washed cells were lysed to nuclei with lysis solution – 50 mM Tris-HCl (pH 8.0), 2 mM EDTA, 0.1% v/v NP-40, 10% v/v glycerol, and PIC – for 30 minutes on ice. Cell debris was washed away with PBS with PIC. Nuclei were pelleted and flash-frozen on dry ice.

Cross-linked chromatin was prepared and sonicated using Bioruptor UCD-200 in ice water bath to generate DNA fragments 200-300 bp in size. Twenty micrograms of each antibody preparation, or 20 μg IgG for mock pulldown controls, were incubated with chromatin prepared from nuclei of approximately 5 million cells.

DNA was released and quantitated using Qubit 2.0 (Life Technologies) with dsDNA HS Assay kit (Life Technologies, Q32854), and 15 ng of DNA was used to generate libraries for Illumina sequencing. ChIP-seq libraries were generated using KAPA LTP Library Preparation Kits (Kapa Biosystems, KK8232) to yield two independent ChIP replicates for each antibody. We also generated libraries from sonicated genomic input DNA from the same chromatin preparations as controls. Libraries were bar-coded with Bioo Scientific index adapters and sequenced to generate 15-23 million reads per duplicate sample using the Illumina Hi-Seq 2000 instrument at the University of Illinois W.M. Keck Center for Comparative and Functional Genomics according to manufacturer's instructions. Separate ChIP preparations were generated for qRT-PCR validation experiments; in this case, released DNA was amplified by GenomePlex® Complete Whole Genome Amplification (WGA) Kit (Sigma, WGA2).

ChIP-Seq data analysis

Human ZNF777 ChIP-enriched sequences as well as reads from the input genomic DNA were mapped to the HG19 human genome build using Bowtie 2 software (Langmead et al. 2009) allowing 1 mismatch per read but otherwise using default settings. Bowtie files were used to identify peaks in human ChIP samples using MACS software (version 14.2) (Zhang et al. 2008), with default settings. After comparison of the individual files, sequence reads from the two separate ChIP libraries were pooled and a final peak set determined in comparison to genomic-input controls. Peaks were mapped relative to nearest transcription start sites using the GREAT program (McLean et al. 2010).

Repetitive elements overlap analysis

To identify enrichment or under-representation of repetitive element types or families in the ChIP-peak datasets, we used a method modified from that described by Cuddapah

and colleagues (Cuddapah et al. 2009). Human repeat data were retrieved on the RepeatMasker Table (www.repeatmasker.org) from USCS's table browser (genome.ucsc.edu) (Karolchik et al. 2004) with the following parameters: assembly= 'Feb. 2009 (GRCh37/hg19)', group= 'Variation and Repeats', track= 'RepeatMasker', table= 'rmsk', region= 'genome', output format= 'BED – Browser extensible data'. The human chromosome sizes required for the analysis were retrieved from the hg19.chromInfo table of the UCSC public database (Kuhn et al. 2013). We examined overlap between genome coordinates of repeat element features and 100 bp intervals surrounding the summits of peaks determined by MACS software from ZNF777 ChIP experiments using the BEDTools intersect function (Quinlan & Hall 2010).

For each peak set 500 random sets of the same number and peak size were generated by the BEDTools random function, and overlaps between these random peak sets and repeats were counted for each of the 500 random sets. For each repeat element and family the average overlap count of the random sets and the standard deviation was determined. Then for each repeat element and family a Z-score was calculated using the overlap count of the peak set, and the average overlap count and standard deviation of the random sets.

If the overlap count of the peak set was less than or equal to the average of the random sets z was calculated as: $z = \frac{(overlap\ count\ of\ peak\ set) - (average\ overlap\ count\ of\ random\ sets)}{standard\ deviation\ of\ the\ overlap\ count\ of\ random\ sets}$. If the overlap count of the peak set was greater: $z = \frac{(average\ overlap\ count\ of\ random\ sets) - (overlap\ count\ of\ peak\ set)}{standard\ deviation\ of\ the\ overlap\ count\ of\ random\ sets}$. The R function pnorm(z) was used to calculate a p-value to indicate if the overlap count was significantly under-represented or enriched in a ChIP-peak set when compared to the overlap counts of the random sets. Repeat families or specific elements that were significantly enriched in at least one of the ChIP peak sets, along with p-values determined for enrichments or under-representation of that family or element type in each peak set.

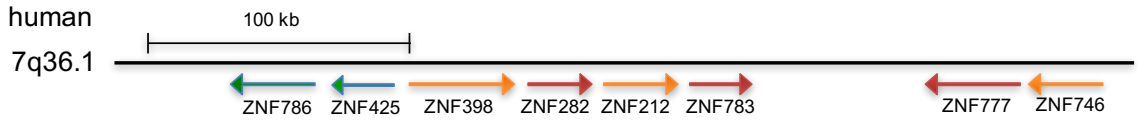
Motif analysis

To identify enriched motifs, we used sequence from a 200 bp region surrounding the predicted summits of selected peaks for analysis with MEME-ChIP with default

parameters (Machanick & Bailey 2011). Motifs displayed in Fig. 2.5 were identified from peaks with the following cutoffs: MACS ef $\geq$ 20, fdr=0 peaks from ZNF777 ChIP in BeWo chromatin; the identified motif occurred in 113 out of total 118 peaks submitted peaks, with p value = 7.9e-103.

Figures and Tables

A



B

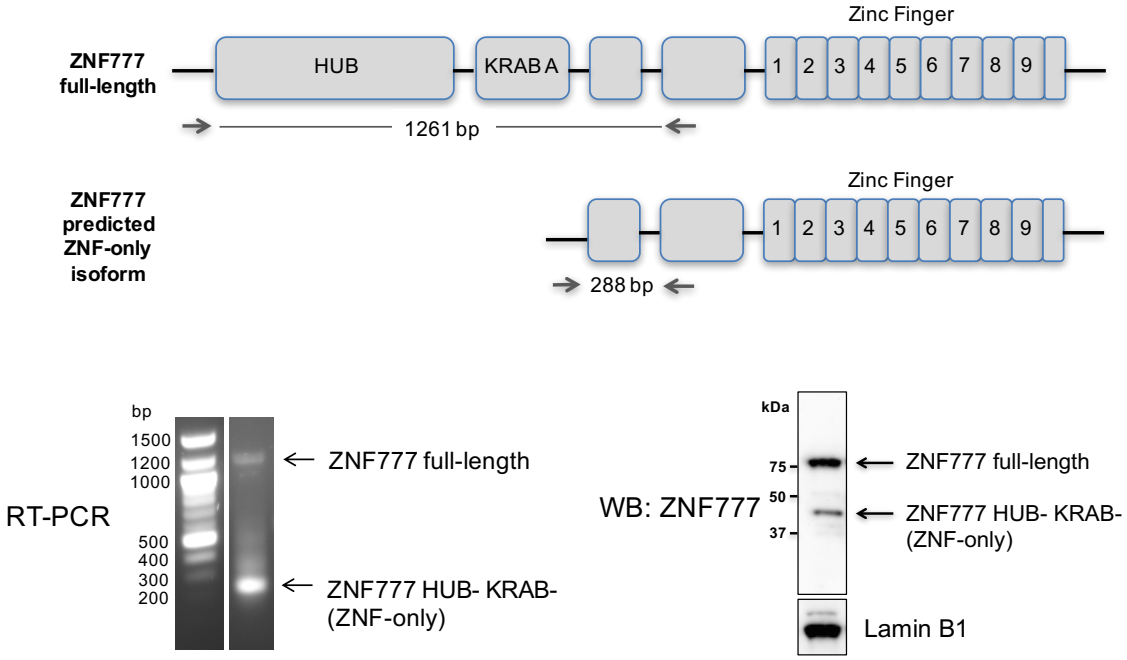
	1	2	3	4	5	6	7	8	9
Zebrafinch	LNI	NQL	HSK	LSM	RHR	EKN	RHE	QHE	YSY
Lizard	LNI	IQL	HSK	LSM	RHR	EKN	RHE	QHE	YSY
Platypus	-	-	HSK	LSI	RHR	EKN	RHE	QHE	YSY
Opossum	LNI	HQL	HSK	LSI	RHR	EKN	RHE	QHE	YSY
Human	LNI	HQL	HSK	LSI	RHR	EKN	RHE	QHE	YSY
Mouse	-	HQL	HSK	LSI	RHR	EKN	RHE	QHE	YSY

Figure 2.1 ZNF777 and neighboring genes are members of a deeply conserved gene cluster.

(A) The ZNF777 family members locate in a cluster on human chromosome 7. (hg19 sequence build, 7q36.1) ZNF398, ZNF212, ZNF783, and ZNF746 are HUB- and KRAB-containing ZNF genes that are closely related to ZNF282 and ZNF777. ZNF786 and ZNF425 are the most recent members lacking the HUB domain. (B) Fingerprint alignment of ZNF777 orthologs in mammalian and non-mammalian vertebrate species shows the conservation of DNA-binding amino acids and the ZNFs of this gene. Each column contains the DNA-binding amino acids (shown here the positions -1, 3, and 6 relative to the α helix in each finger) and rows correspond to the sequence in each species. ZNF are numbered at top in N- to C- terminal orientation in the protein. The platypus sequence in this region is incomplete, allowing on a partial protein sequence to be deduced.

Figure 2.2

A



B

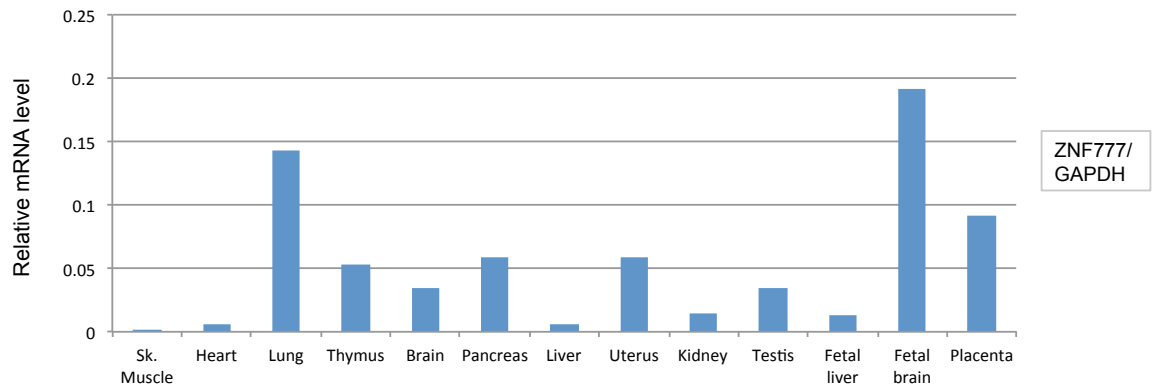
```
ZNF777      MENQRSSPLSFPSVPQEETLRQAPAGLPRETLFQSRILPPKEIPSLSPTIPRQASLPQTS    60
ZNF282      ----------------------------------------MQFVSTRPQPQQLGIQG    17
                                                     .:  * *:  .*   .

ZNF777      SAPKQETSGWMPHVLQKGPSLLCSAASEQETPLQGPLASQEGTQYPPPAAAEQEISLLSH    120
ZNF282      LGLDSGSWSWAQA---LPPEEVCH----QEPALRGEMA----EGMPPMQAQEWDMD--AR    64
             . .. : .*        *. :*     **  *:* :*       **  * * ::.  ::

ZNF777      SPHHQEAPVHSPETPEKDPLTLSPTVPETDGDPLLQSPVSQKDTPFQISSAVQKEQPLPT    180
ZNF282      ----RPMPFQFP-----------------------PFPDR--APVFPDRMMREPQLPT    93
                :  *.: *                        *. ::     : . : :*  ***

ZNF777      AEITRLAVWAAVQAVERKLEAQAMRLLTLEGRTGTNEKKIADCEKTAVEFANHLESKWVV    240
ZNF282      AEISLWTVVAAIQAVERKVDAQASQLLNLEGRTGTAEKKLADCEKTAVEFGNHMESKWAV    153
            ***:  :* **:******::*** :**.******* ***:**********.**:****.*

ZNF777      LGTLLQEYGLLQRRLENMENLLKNRNFWILRLPPGSNGEVPK    282
ZNF282      LGTLLQEYGLLQRRLENLENLLRNRNFWVLRLPPGSKGEAPK    195
            ******************:****:*****:*******:**.**
```

Figure 2.2 (cont.) The protein structure of ZNF777 and the HUB domain alignment of ZNF777 and ZNF282.

(A) Human ZNF777 contains 9 zinc fingers, a N-terminal HUB domain, a KRAB A-like box, and a tether region between the KRAB A-like box and the zinc fingers. The HUB domain and the KRAB A-like box are encoded on different exons and could be spliced separately into different isoforms. Lower left panel: a HUB minus, KRAB A minus isoform (ZNF-only) (288 bp) was detected by RT-PCR in addition to full length (1261 bp) ZNF777 transcript by the primers (indicated by grey arrows) flanking the two domains. Lower right panel: both ZNF777 full-length and ZNF-only proteins were detected by Western Blot using ZNF777 antibody (AVIVA, ARP32569). (B) HUB domain alignment of ZNF777 and ZNF282. The sequences from amino acid 1 at N-terminus to the end of the exon encoding the HUB domain (exon 1 for ZNF777, exon 2 for ZNF282) were aligned using Clustal Omega.

Figure 2.3

A


Relative mRNA level
0.25
0.2
0.15
0.1
0.05
0
Sk. Muscle
Heart
Lung
Thymus
Brain
Pancreas
Liver
Uterus
Kidney
Testis
Fetal liver
Fetal brain
Placenta
ZNF777/ GAPDH

B

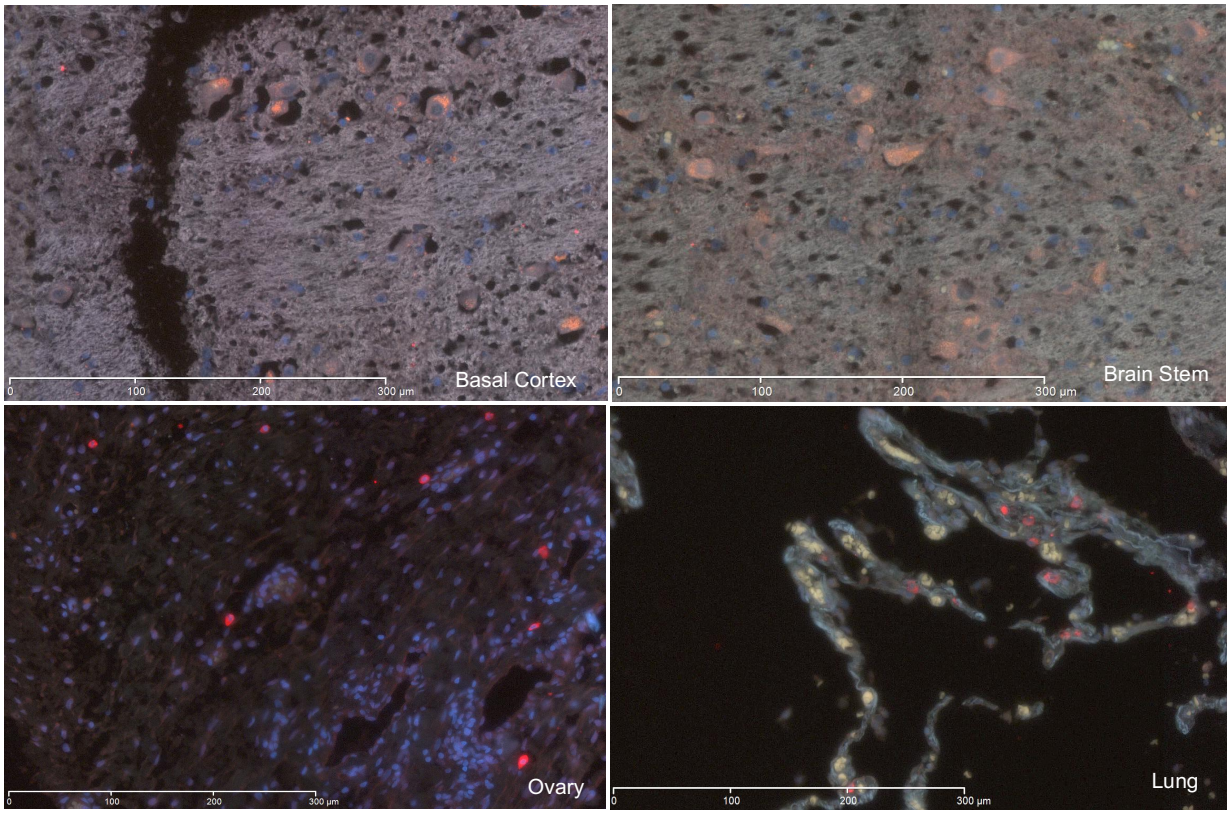
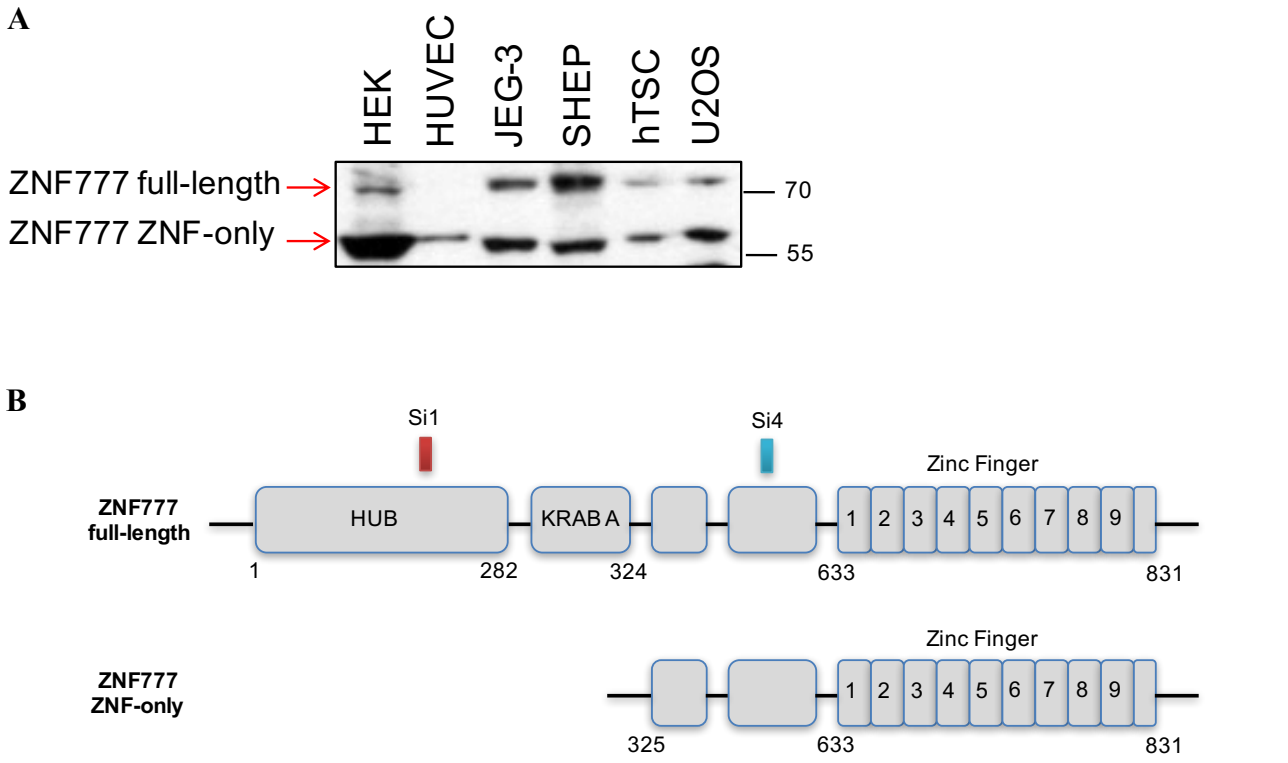

Basal Cortex
Brain Stem
Ovary
Lung
0
100
200
300 µm

Figure 2.3 (cont.) Extensive expression of ZNF777 in human tissues.

(A) Real-time PCR (QPCR) was performed to detect the expression of ZNF777 at mRNA level in different human tissues. ZNF777 is extensively expressed in many different human tissues, with relatively higher expression in lung, thymus, brain, pancreas, uterus, fetal brain and placenta. (B) The protein level of ZNF777 (red) was detected by immunohistochemistry (IHC) using ZNF777 antibody (AVIVA, ARP32569). The protein was shown to be expressed in many different tissues, including placenta, brain, lung, and ovary.

Figure 2.4


A
HEK
HUVEC
JEG-3
SHEP
hTSC
U2OS
ZNF777 full-length
ZNF777 ZNF-only
70
55
B
Si1
Si4
Zinc Finger
ZNF777
full-length
HUB
KRAB A
1
2
3
4
5
6
7
8
9
1
282
324
633
831
Zinc Finger
ZNF777
ZNF-only
325
633
831
C
ZNF777 siRNA
Si-Alexa
si1
si4
kDa
130
100
70
55
40
ZNF777 full-length
ZNF777 ZNF-only
Lamin B1
70
D
Relative protein level
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Full-length
ZNF-only
Alexa
si1
si4

E

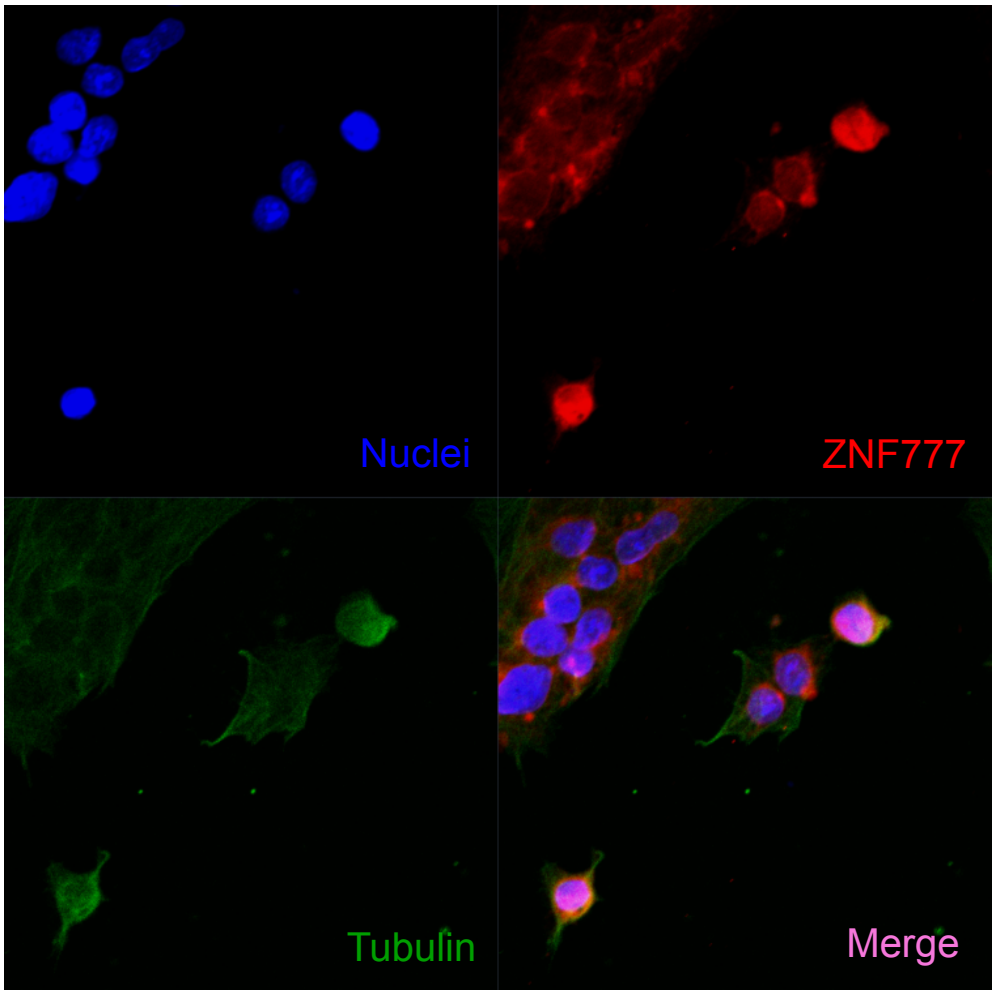
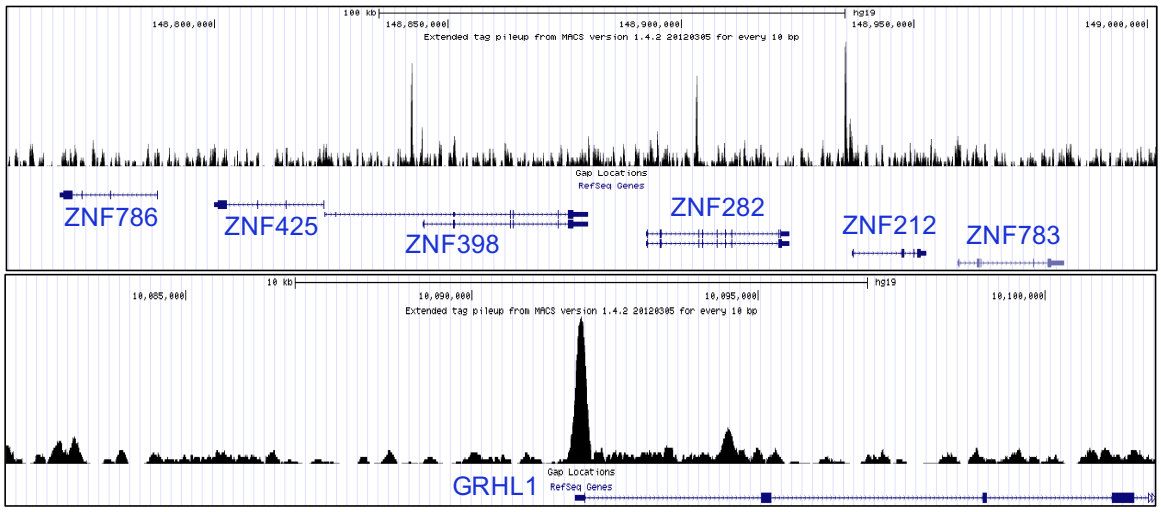


Figure 2.4 (cont.) Expression of ZNF777 in human and mouse cells.
(A) Western Blot was performed on human and mouse cell lines by ZNF777 antibody against human ZNF777. The full-length and ZNF-only proteins were found to be expressed in HEK293 (human embryonic kidney), HUVEC (human umbilical vein/vascular endothelium), JEG-3 (human placenta choriocarcinoma), SHEP (human neuroblastoma), human trophoblast stem cells (hTSC), U2OS (human osteosarcoma). (B) Locations of the binding of two ZNF777 siRNAs on the ZNF777 mRNA isoforms. Si1 binds ZNF777 mRNA at the HUB domain, therefore it can only knock down the full-length isoform of ZNF777. Si4 binds ZNF777 mRNA at the spacer exon existing in both

Figure 2.4 (cont.)

full-length and ZNF-only isoforms, thus it knocks down both isoforms. (C) BeWo cells were transfected with different siRNA against ZNF777 (siRNA_Si1, and siRNA_Si4) or a negative control siRNA-Alexa (SiAlexa) for 72 hours and then nuclear extracts were collected for Western Blot analysis. (D) Quantification of the Western Blot. Full-length ZNF777 were shown to be reduced in cells treated with both Si1 and Si4 while ZNF-only isoform was reduced only in Si4 treated cells. (E) Immunocytochemistry (ICC) was performed on BeWo cells using the ZNF777 antibody and anti-tubulin antibody. Secondary antibody used were Alexa594 anti rabbit for ZNF777 primary antibody recognition, and Alexa486 anti mouse for tubulin primary antibody recognition. Nuclei were counterstained by Hoeschst. ZNF777-red, Tubulin-green, Nuclei-blue, n=3.

A



B

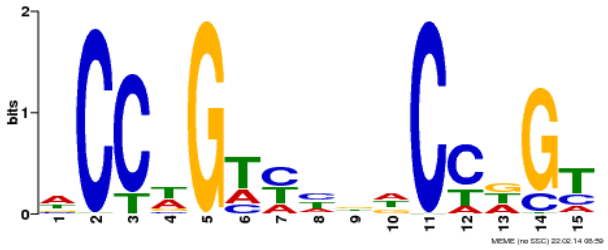
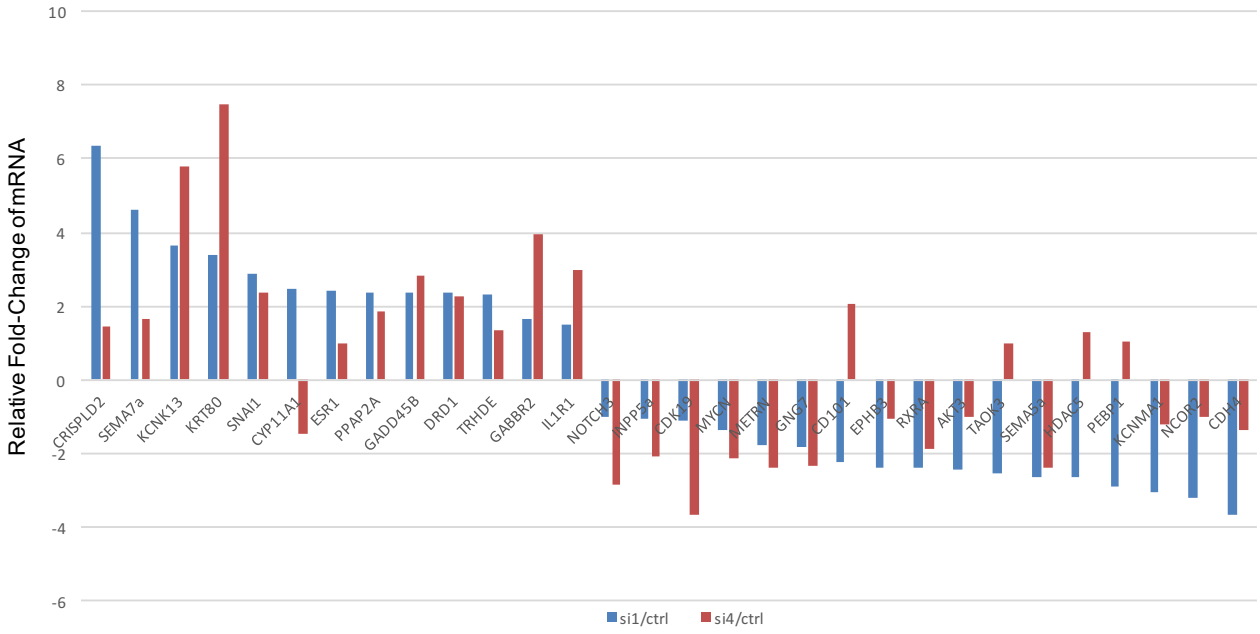


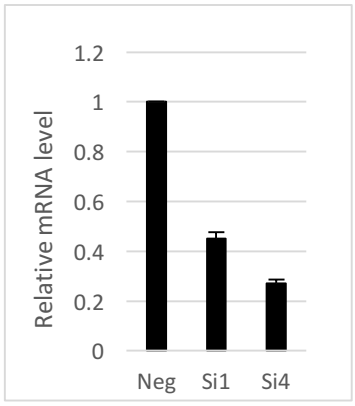
Figure 2.5 ChIP-seq of ZNF777. We performed Chromatin immunoprecipitation (ChIP)-seq in BeWo cells. We sequenced the ChIP libraries prepared from BeWo cells, the input DNA from the same cell preparations, and the DNA isolated after a mock pull-down (Immunoglobin G, IgG) experiments using an Illumina GA-II Sequencer (University of Illinois Keck Biotechnology Center). After alignment of pull-down fragments to genome by Bowtie program, and comparisons to genomic-input controls by MACS software, 1979 peaks were identified. (A) Examples of ChIP peaks identified by MACS. The peaks located near the promoters of some nearby genes were shown. The nearby genes are the family members ZNF212, ZNF282, ZNF298 and GRHL1. (B) The binding motif of ZNF777 was predicted by MEME. Motif can be found centrally located in 113 of the 118 fdr=0 peaks submitted.

Figure 2.6

A



B



C

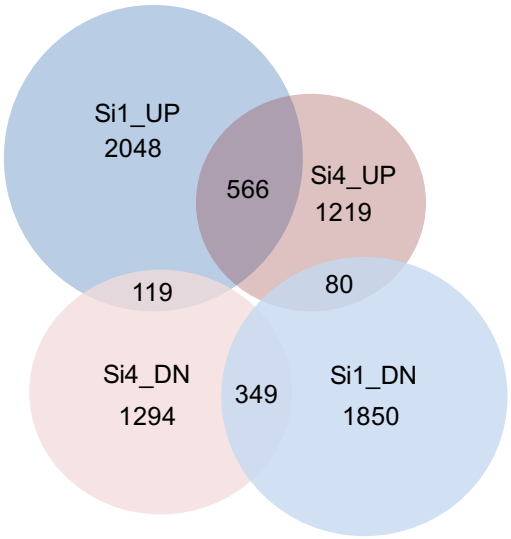


Figure 2.6 (cont.) Differentially expressed genes (DEGs) identified by RNA-seq in BeWo cells treated with ZNF777 siRNA_Si1 or siRNA_Si4.

BeWo cells were harvested 72 hours after transfection.

(A) Relative fold-change of 30 DEGs from RNA-Seq. Blue: Si1/control; red: Si4/control. Control: GAPDH. (B) siRNA knock down efficiency mediated by ZNF777-Si1 and ZNF777-Si4, each knocked down the mRNA level of ZNF777 by 55% and 73% (n = 3). (C) Distribution of numbers of DEGs. Si1_UP: DEGs with greater than 1.5-fold change (up-regulated) in Si1 treated cells. Si1_DN: DEGs with greater than negative 1.5-fold change (down-regulated) in Si1 treated cells. Same with Si4_UP and Si4_DN.

	Si1/Ctrl			Si4/Ctrl		
	RNA-seq	QPCR1	QPCR2	RNA-seq	QPCR1	QPCR2
CRISPLD2	6.36	5.97	7.01	1.45	1.01	0.62
SEMA7a	4.6	6.05	5.81	1.64	1.88	1.21
KCNK13	3.67	3.02	3.98	5.79	2.74	3.96
SNAI1	2.87	2.65	2.44	2.37	2.21	1.1
PPAP2A	2.39	3.53	3.52	1.84	1.83	1.97
GADD45B	2.37	4.07	1.52	2.82	3.41	1.96
DRD1	2.36	2.87	2.21	2.29	1.93	1.51
TRHDE	2.32	2.27	3.28	1.36	1.31	1.21
GABBR2	1.66	2.77	2.32	3.95	4.75	3.78
IL1R1	1.5	2.51	4.58	3.01	2.4	2.65
NOTCH3	0.99	1.15	2.35	0.35	0.36	0.73
INPP5a	0.94	1.27	0.94	0.49	0.51	0.56
CDK19	0.9	1.35	1.28	0.27	0.38	0.45
AKT3	0.41	0.61	0.49	0.54	0.49	0.58
PEBP1	0.34	0.48	0.66	1.04	1.51	0.98
NCOR2	0.31	0.37	0.63	0.99	0.91	0.89
HDAC5	0.38	0.61	0.76	1.3	1.41	1.05
CD101	0.45	0.49	0.76	2.07	1.82	1.96
CYP11A1	2.47	2.75	2.51	0.69	0.75	0.41

Table 2.1 The relative mRNA level of differentially expressed genes (DEGs) by RNA-seq and QPCR. The QPCR analyses replicated the trends of the fold-changes for DEGs identified by RNA-seq. Ctrl: control, GAPDH.

	ZNF777 Si1 and Si4 concordant[1] DAVID enrichment factor[2]
DAVID Functional Cluster	**Up-regulated DEGs[3]**
Glycoprotein	8.7
Membrane	5.6
Signal peptide	5.1
Synapse/cell junction	3.7
Virus receptor activity	3.1
Antigen binding	2.9
Fibronectin	2.8
Nuclear receptor	1.8
Lysosome	1.7
Growth factor	1.5
Ion transport	1.4

Table 2.2 Gene Ontology (GO) clusters identified as significantly enriched in gene sets up- regulated in both ZNF777 Si1 and Si4 siRNA-knockdowns
[1] Clusters enriched in both Si1 and Si4 knock-downs
[2] David enrichment scores are calculated as the geometric mean of –log transformed P-values of GO terms within a cluster based on content of similar genes
[3] Clusters associated with up-regulated differentially expressed genes (DEGs)

	ZNF777 Si1 and Si4 concordant[1] DAVID enrichment factor[2]
DAVID Functional Cluster	**Down-regulated DEGs[3]**
Neurogenesis	2.6
Differentiation	2.6
Semaphorin-plexin signaling pathway	2.3
Metalloprotease	2.2
Cell junction	1.4
Cadherin	1.4
Glycolysis	1.3

Table 2.3 Gene Ontology (GO) clusters identified as significantly enriched in gene sets down- regulated in both ZNF777 Si1 and Si4 siRNA-knockdowns
[1] Clusters enriched in both Si1 and Si4 knock-downs
[2] David enrichment scores are calculated as the geometric mean of –log transformed P-values of GO terms within a cluster based on content of similar genes
[3] Clusters associated with down-regulated differentially expressed genes (DEGs)

	ZNF777 Si1 and Si4 concordant[1] ChIP peak-associated DEGs DAVID enrichment factor[2]
DAVID Functional Cluster	**Up or Down[3]**
Axon guidance	1.7
Semaphorin-plexin signaling pathway	1.7
Focal adhesion	1.6
PI3K-Akt signaling pathway	1.4
Cell adhesion	1.4
Wnt signaling pathway	1.3

Table 2.4 Gene Ontology (GO) clusters identified as significantly enriched in gene sets flanking or within ZNF777 binding sites and up- or down-regulated after ZNF777 siRNA-knockdown

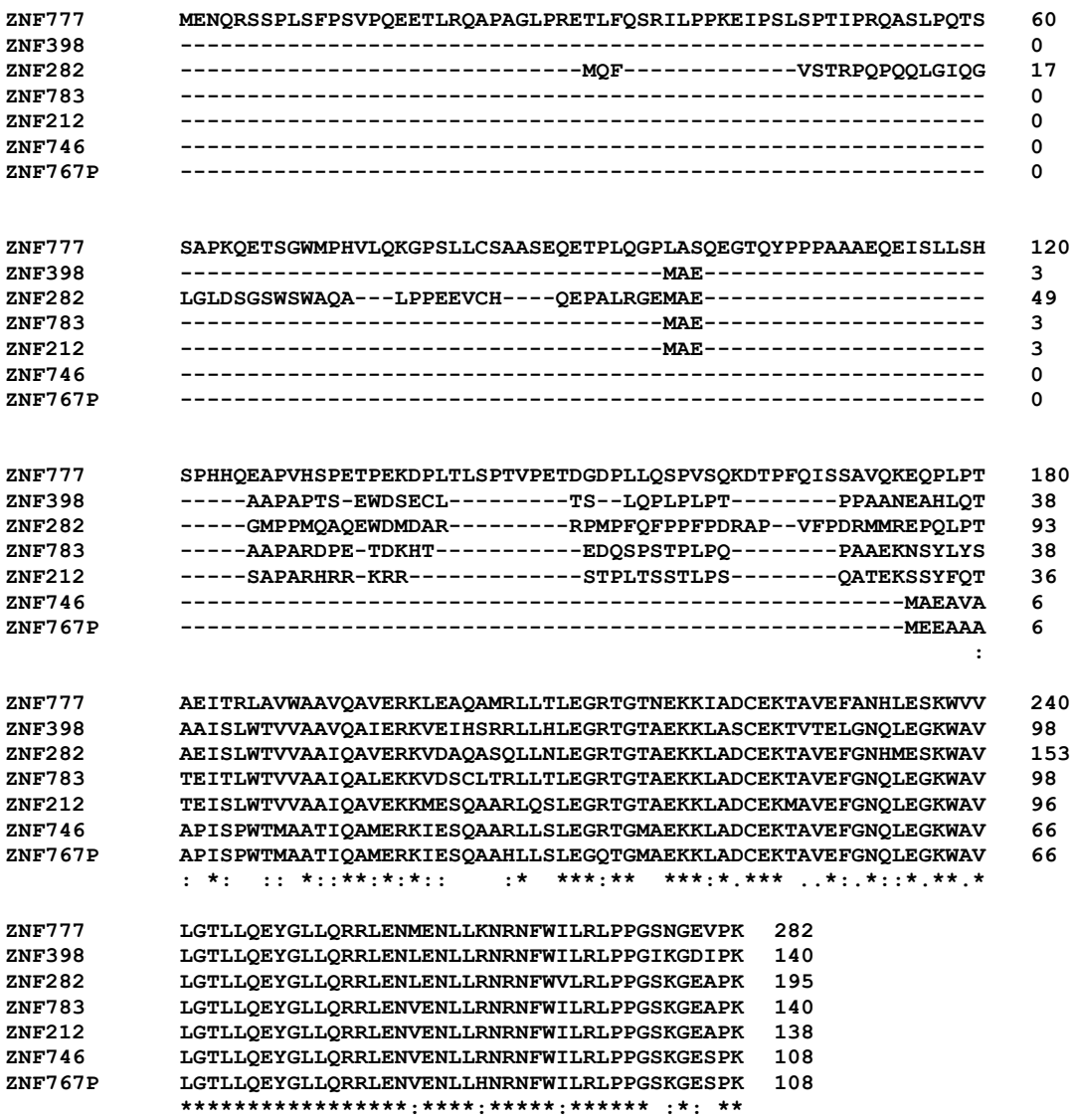
[1] Clusters enriched in both Si1 and Si4 knock-downs

[2] David enrichment scores are calculated as the geometric mean of –log transformed P-values of GO terms within a cluster based on content of similar genes

[3] Clusters associated with up- or down-regulated genes, combined

Supplemental Figure 2.1

A

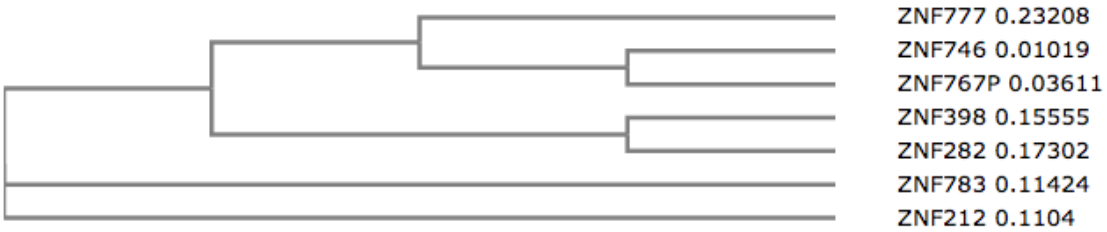


B

Percent Identity Matrix - created by Clustal2.1

1: ZNF777	100.00	58.57	51.79	64.29	63.04	71.30	69.44
2: ZNF398	58.57	100.00	67.14	70.07	68.89	75.00	71.30
3: ZNF282	51.79	67.14	100.00	67.86	68.12	76.85	75.00
4: ZNF783	64.29	70.07	67.86	100.00	77.54	78.70	75.93
5: ZNF212	63.04	68.89	68.12	77.54	100.00	82.41	79.63
6: ZNF746	71.30	75.00	76.85	78.70	82.41	100.00	95.37
7: ZNF767P	69.44	71.30	75.00	75.93	79.63	95.37	100.00

C



Supplemental Figure 2.1 (cont.) The HUB domain alignment of members in ZNF777 subfamily.

(A) The sequences from amino acid 1 at N-terminus to the end of the exon encoding the HUB domain (exon 1 for ZNF777, exon 2 for other members) were aligned using Clustal Omega. (B) Percentage Identity Matrix of the HUB domain sequences of the members in ZNF777 subfamily created by Clustal 2.1. (C) Phylogenetic Tree based on sequences of HUB domains from the members in ZNF777 subfamily.

```
CONSENSUS    F DV   F  EEW  L   Q  LY  VMLENY  L

ZNF10     MVTFKDVFVDFTREEWKLLDTAQQIVYRNVMLENYKNLVSL

ZNF777    PVTFDDVAVHFSEQEWGNLSEWQKELYKNVMRGNYESLVSM
ZNF282    PVTFVDIAVYFSEDEWKNLDEWQKELYNNLVKENYKTLMSL
ZNF783    PVTFDDVAVYFSELEWGKLEDWQKELYKHVMRGNYETLVSL
ZNF398    PVAFDDVSIYFSTPEWEKLEEWQKELYKNIMKGNYESLISM
ZNF212    SRSLENDGVCFTEQEWENLEDWQKELYRNVMESNYETLVSL
ZNF746    PVTFDDVAVYFSEQEWGKLEDWQKELYKHVMRGNYETLVSL
ZNF425    TVTFDDVALYFSEQEWEILEKWQKQMYKQEMKTNYETLDSL
ZNF786    PLTFEDVAIYFSEQEWQDLEAWQKELYKHVMRSNYETLVSL
```

Supplemental Figure 2.2 KRAB A box alignment of the members in ZNF777 subfamily.

The ZNF10 is an example of zinc finger protein with canonical consensus sequence of KRAB A box. The DV at positions 6,7 and MLE at positions 36-38 in human consensus have been shown to be essential for KAP1 binding. The KRAB A-like box of all the members lack the LE consensus amino acids, suggesting the KRAB A-like box of this family may not interact with KAP1.

Over-represented elements

Repeat Element	P Value
MER31B (ERV1)	5.64E-93
MER31A (ERV1)	8.42E-60
MER9B (ERVK)	2.75E-56
MER39B (ERV1)	3.20E-18
MER65C (ERV1)	1.67E-16
LTR45 (ERV1)	2.36E-15
HERVL32-int (ERVL)	2.36E-15
MER39 (ERV1)	3.25E-09
MER70-int (ERVL)	2.50E-07
LTR81 (Gypsy)	2.50E-07
Charlie17a (hAT-Charlie)	2.79E-07
MER61E (ERV1)	4.94E-07
MLT1J-int (ERVL-MaLR)	1.93E-06
Charlie14a (hAT-Charlie)	2.69E-06
MER74A (ERVL)	1.26E-05
Kanga1b (TcMar-Tc2)	1.45E-05
MLT1M (ERVL-MaLR)	1.74E-05
LTR62 (ERVL)	2.84E-05
LTR84a (ERVL)	3.83E-05
L5 (RTE)	3.84E-05
MamGypLTR1d (Gypsy)	3.84E-05
MER21A (ERVL)	5.22E-05
Charlie13a (hAT-Charlie)	6.60E-05

Under-represented elements

Repeat Element	P Value
L1MEc (L1)	0.340660893
FLAM_C (Alu)	0.302680256
L1MC4 (L1)	0.278736929
MER5B (hAT-Charlie)	0.278338464
MLT1B (ERVL-MaLR)	0.2548366
L1ME3A (L1)	0.236542705
L2a (L2)	0.215780855
L1MC1 (L1)	0.200023409
MLT1D (ERVL-MaLR)	0.192961197
L3 (CR1)	0.171284683
AluJo (Alu)	0.142569658
AluSg (Alu)	0.109512995
AluSc (Alu)	0.077898077
MIR (MIR)	0.071162415
AluSq2 (Alu)	0.051313966
L1ME1 (L1)	0.034178961
AluSp (Alu)	0.030438076
L1M5 (L1)	0.012366321
AluSx1 (Alu)	0.00845582
AluJr (Alu)	0.007746328
AluJb (Alu)	0.007586817
AluSz (Alu)	0.002133738
AluSx (Alu)	0.000157561

Supplemental Table 2.1 Repeat elements intersected with ZNF777 ChIP-seq peaks that are over-representative or under-representative. 709 ChIP-seq peaks were analyzed by a randomization test using hypergeometric distribution. Random peaks were created to intersect with the repeats and compared with the repeats intersected with ChIP-seq peaks for the enrichment of intersected ChIP-seq peaks that are specific to certain repeat elements.

	ZNF777 Si1[1] (full-length only) DAVID enrichment factor[3]		ZNF777 Si4[2] (both isoforms) DAVID enrichment factor	
DAVID Functional Cluster	**Up[4]**	**Down[5]**	**Up**	**Down**
Glycoprotein	11.6		5.8	
Signal peptide	9.1		7.4	
Cell membrane	5.6	1.3	5.8	1.1
Extracellular matrix organization	5.6		3.1	
Integrin binding			3.1	
Heparin binding	3.3		2.2	
Antigen binding	3.1		2.4	
Cell adhesion molecules	2.7	1.3	3.1	
Growth factor activity	2.3		3.1	
Positive regulation of cell division	2.3		3.1	
Cell junction	2.1		1.7	
Synapse	2.1	1.3	1.1	
Fibronectin	2.0		1.2	
Axon guidance	1.6			
Protein kinase		2.2		
Bi-cellular tight junction			1.7	
EGF-like domain		1.2		
Signaling pathway			2.2	

Supplemental Table 2.2: Gene Ontology (GO) clusters identified as significantly enriched in gene sets up- or down-regulated after ZNF777 gene siRNA knockdown
[1] Clusters enriched in Si1 knock-down. Si1 knocks down full-length of ZNF777 only.
[2] Clusters enriched in Si4 knock-down. Si4 knocks down both full-length and ZNF-only isoforms
[3] David enrichment scores are calculated as the geometric mean of –log transformed P-values of GO terms within a cluster based on content of similar genes
[4, 5] Clusters associated with Up- or Down-regulated genes, respectively

References

Andermatt, I. et al., 2014. Semaphorin 6B acts as a receptor in post-crossing commissural axon guidance. *Development*, 141(19), pp.3709–3720.

Chuong, E.B. et al., 2013. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nature Genetics*, 45(3), pp.325–329.

Conroy, A.T. et al., 2002. A Novel Zinc Finger Transcription Factor with Two Isoforms That Are Differentially Repressed by Estrogen Receptor. *Journal of Biological Chemistry*, 277(11), pp.9326–9334.

Cordaux, R. & Batzer, M.A., 2009. The impact of retrotransposons on human genome evolution. *Nature Publishing Group*, 10(10), pp.691–703.

Cuddapah, S. et al., 2009. Native chromatin preparation and Illumina/Solexa library construction. *Cold Spring Harbor protocols*, 2009(6), pp.pdb.prot5237–pdb.prot5237.

Dun, X.-P. & Parkinson, D., 2017. Role of Netrin-1 Signaling in Nerve Regeneration. *International Journal of Molecular Sciences*, 18(3), pp.491–22.

Emerson, R.O. & Thomas, J.H., 2009. Adaptive Evolution in Zinc Finger Transcription Factors S. Myers, ed. *PLoS Genetics*, 5(1), pp.e1000325–12.

Foulkes, N.S. & Sassone-Corsi, P., 1992. More is better: activators and repressors from the same gene. *Cell*, 68(3), pp.411–414.

Friedli, M. & Trono, D., 2015. The developmental control of transposable elements and the evolution of higher species. *Annual review of cell and developmental biology*, 31(1), pp.429–451.

Huang, D.W., Sherman, B.T. & Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1), pp.44–57.

Huntley, S., 2006. A comprehensive catalog of human KRAB-associated zinc finger genes: Insights into the evolutionary history of a large family of transcriptional repressors. *Genome Research*, 16(5), pp.669–677.

Jacobs, F.M.J., Greenberg, D., Nguyen, N., Haeussler, M., Ewing, A.D., Katzman, S., Paten, B., Salama, S.R. & Haussler, D., 2014b. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature*, 516, pp.242–245.

Jern, P. & Coffin, J.M., 2008. Effects of Retroviruses on Host Genome Function. *Annual Review of Genetics*, 42(1), pp.709–732.

Jongbloets, B.C. & Pasterkamp, R.J., 2014. Semaphorin signalling during development. *Development*, 141(17), pp.3292–3297.

Karolchik, D. et al., 2004. The UCSC Table Browser data retrieval tool. *Nucleic Acids Research*, 32(Database issue), pp.D493–6.

Kim, C.A. & Berg, J.M., 1996. A 2.2 A resolution crystal structure of a designed zinc finger protein bound to DNA. *Nature structural biology*, 3(11), pp.940–945.

Langmead, B. et al., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, 10(3), p.R25.

Liao, W.-X. et al., 2010. Perspectives of SLIT/ROBO signaling in placental angiogenesis. *Histology and Histopathology*, 25, pp.1181–1190.

Liu, H. et al., 2014. Deep Vertebrate Roots for Mammalian Zinc Finger Transcription Factor Subfamilies. *Genome Biology and Evolution*, 6(3), pp.510–525.

Lupo, A. et al., 2013. KRAB-Zinc Finger Proteins: A Repressor Family Displaying Multiple Biological Functions. *Current genomics*, 14(4), pp.268–278.

Machanick, P. & Bailey, T.L., 2011. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics (Oxford, England)*, 27(12), pp.1696–1697.

Margolin, J.F. et al., 1994. Krüppel-associated boxes are potent transcriptional repression domains. *Proceedings of the National Academy of Sciences*, 91(10), pp.4509–4513.

McLean, C.Y. et al., 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology*, 28(5), pp.495–501.

Mitchell, P.J. & Tjian, R., 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science*, 245(4916), pp.371–378.

Nowick, K. et al., 2011. Gain, Loss and Divergence in Primate Zinc-Finger Genes: A Rich Resource for Evolution of Gene Regulatory Differences between Species M. A. Batzer, ed. *PLoS ONE*, 6(6), pp.e21553–11.

Nowick, K. et al., 2010. Rapid Sequence and Expression Divergence Suggest Selection for Novel Function in Primate-Specific KRAB-ZNF Genes. *Molecular Biology and Evolution*, 27(11), pp.2606–2617.

Okumura, K. et al., 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 Research*, 25(24), pp.5025–5032.

Pavletich, N.P. & Pabo, C.O., 1993. Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science*, 261(5129), pp.1701–1707.

Pavletich, N.P. & Pabo, C.O., 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science*, 252(5007), pp.809–817.

Pengue, G. et al., 1994. Repression of transcriptional activity at a distance by the evolutionarily conserved KRAB domain present in a subfamily of zinc finger proteins. *Nucleic Acids Research*, 22(15), pp.2908–2914.

Quinlan, A.R. & Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, 26(6), pp.841–842.

Ran, F.A. et al., 2013. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, 8(11), pp.2281–2308.

Savic, D. et al., 2015. CETCh-seq: CRISPR epitope tagging ChIP-seq of DNA-binding proteins. *Genome Research*, 25(10), pp.1581–1589.

Stoeckli, E., 2017. Where does axon guidance lead us? *F1000Research*, 6, pp.78–8.

Thomas, J.H. & Schneider, S., 2011. Coevolution of retroelements and tandem zinc finger genes. *Genome Research*, 21(11), pp.1800–1812.

Top, 2014. ZNF282 (Zinc nger protein 282), a novel E2F1 co-activator, promotes esophageal squamous cell carcinoma. pp.1–13.

Vissing, H. et al., 1995. Repression of transcriptional activity by heterologous KRAB domains present in zinc finger proteins. *FEBS Letters*, 369(2-3), pp.153–157.

Witzgall, R. et al., 1994. The Krüppel-associated box-A (KRAB-A) domain of zinc finger proteins mediates transcriptional repression. *Proceedings of the National Academy of Sciences*, 91(10), pp.4514–4518.

Wolf, D. & Goff, S.P., 2009. Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature*, 458(7242), pp.1201–1204.

Wolfe, S.A., Nekludova, L. & Pabo, C.O., 2000. DNA recognition by Cys2His2 zinc finger proteins. *Annual review of biophysics and biomolecular structure*, 29(1), pp.183–212.

Yeo, S.-Y. et al., 2014. ZNF282 (Zinc finger protein 282), a novel E2F1 co-activator, promotes esophageal squamous cell carcinoma. *Oncotarget*, 5(23), pp.12260–12272.

Yu, E.J. et al., 2012. SUMOylation of ZFP282 potentiates its positive effect on estrogen signaling in breast tumorigenesis. *Oncogene*, 32(35), pp.4160–4168.

Yuki, R. et al., 2015. Overexpression of Zinc-Finger Protein 777 (ZNF777) Inhibits Proliferation at Low Cell Density Through Down-Regulation of FAM129A. *Journal of Cellular Biochemistry*, 116(6), pp.954–968.

Zhang, Y. et al., 2008. Model-based analysis of ChIP-Seq (MACS). *Genome biology*, 9(9), p.R137.

Ziegler, E.C. & Ghosh, S., 2005. Regulating inducible transcription through controlled localization. *Science's STKE : signal transduction knowledge environment*, 2005(284), pp.re6–re6.

CHAPTER 3: BINDING LANDSCAPE AND FUNCTION OF ZFP777 IN MOUSE NEURAL STEM CELLS

Li-Hsin Chang[1,2], Christopher Seward[1,2], Huimin Zhang[1,2], and Lisa Stubbs[1,2,3]

[1] Department of Cell and Developmental Biology,

[2] Carl R. Woese Institute for Genomic Biology,

University of Illinois at Urbana-Champaign, Urbana IL 61801

[3] Corresponding author

Running Title: Genomic binding analysis of Zfp777

Abstract

KRAB-associated C2H2 zinc-finger (KRAB-ZNF) proteins are the products of a rapidly evolving gene family that traces back to early tetrapods, but which has expanded dramatically to generate an unprecedented level of species-specific diversity. Most land vertebrates carry hundreds of KRAB-ZNF genes, but remarkably, only three of the 394 human KRAB-ZNF genes have been conserved throughout amniote history. These three genes, *ZNF777*, *ZNF282*, and *ZNF783*, are members of an ancient cluster, and encode proteins with a noncanonical KRAB domain and a unique HUB domain of unknown function. We recently reported the functions of ZNF777 in human choriocarcinoma cells (BeWo) to model placental trophoblasts, where *ZNF777* and relatives are highly expressed (Chang et al. 2017). That these ancient genes should be expressed so highly in trophoblasts, which are found only in mammals, posed an interesting puzzle. However, we showed that ZNF777 regulates semaphorins and related genes, with known roles in placenta angiogenesis and in embryonic brain, where *ZNF777* is also highly expressed. Here, we describe the investigation of this neuronal function in mouse neural stem cells (NSCs). We tagged endogenous *Zfp777* with FLAG epitopes using the CRISPR-Cas9 system, and performed chromatin immunoprecipitation (ChIP-seq) in mouse NSC chromatin. Zfp777 binds near promoters of genes involved in transcriptional regulation, Wnt and TGF-beta signaling pathways, neuron development and axon guidance - functions also regulated by ZNF777 in BeWo cells. The results indicate that Zfp777 and ZNF777 regulate similar pathways in diverse cell types, and suggest that a conserved role in neuron development was coopted for novel ZNF777 functions in a mammalian-specific tissue.

Introduction

C2H2 zinc finger (ZNF) proteins represent the largest single class of eukaryotic transcription factors, and while many vertebrate transcription factor families are conserved, the C2H2 zinc finger family stands out as a particularly significant exception. Over the course of evolution, distinct ZNF families have emerged independently in different lineages, through exon shuffling events that bring DNA sequences encoding zinc-finger arrays together with different types of protein-interaction or chromatin-modifying "effector" domains (Collins et al. 2001; Stubbs et al. 2011). In mammalian lineages particularly, one major ZNF subfamily has diverged very rapidly and dramatically: the KRAB-ZNF family, which comprises over 400 genes in human and over 500 genes in mouse (Consiantinou-Deltas et al. 1992; Bellefroid et al. 1993; Huntley 2006; Liu et al. 2014). By their sheer numbers, this single subfamily of ZNF proteins dominates the mammalian transcription-factor landscape, comprising up to one-fourth of all predicted TF genes (Vaquerizas et al. 2009). Most intriguingly, although all mammals have roughly equal numbers of KRAB-ZNF genes, the number of 1:1 orthologous pairs is remarkably small (Huntley 2006). About one-third of human KRAB-ZNF genes are primate specific, and about 30 human KRAB-ZNF genes have arisen through segmental duplication since the divergence of old world monkeys, creating novel transcriptional regulators that exist only in higher primates (Nowick et al. 2010).

The KRAB-ZNF gene subfamily encodes proteins with two primary structural domains: one or more copies of an effector domain, called the *Krüppel*-associated box (KRAB), and a C-terminal DNA binding domain (DBD) composed of a tandem array of zinc fingers. DNA binding is mediated by specific interaction between four amino acids -1, 2, 3, and 6 relative to the alpha helix within each ZNF (which we call the "fingerprints"), with three adjacent nucleotides at the DNA target sites (Pavletich & Pabo 1991; Pavletich & Pabo 1993; Kim & Berg 1996; Wolfe et al. 2000). The canonical mammalian KRAB domain, called KRAB A, has been shown in specific cases to interact with a universal cofactor, KAP1, to recruit histone deacetylase and methylation complexes to the ZNF-binding sites. For this reason, KRAB-ZNF proteins are thus typically thought to act as potent transcriptional repressors (Margolin et al. 1994; Pengue

et al. 1994; Witzgall et al. 1994; Tommerup & Vissing 1995). Since KAP1 is involved in silencing both exogenous retroviruses and endogenous retroelements (EREs) (Rowe et al. 2010; Rowe & Trono 2011) and since the evolutionary pattern of KRAB-ZNFs has been linked to that of retroviral invasions (Jacobs et al. 2014; Thomas & Schneider 2011), it has been proposed that KRAB-ZNF diversity stems from an "arms race" to silence EREs.

In a previous study, we mined a number of existing amniote genomes to identify the vertebrate roots of the mammalian KRAB-ZNF family (Liu et al. 2014). We found hundreds of KRAB-ZNF proteins in every species, but only three human genes with clear orthologs in non-mammalian vertebrates. These three genes, *ZNF777, ZNF282*, and *ZNF783*, are members of an ancient familial cluster and encode proteins with similar domain structures. This finding led us to several questions, including: what are the functions of these ancient family members and why, of such a large and diverse family group, were these three genes conserved so fastidiously over hundreds of millions of years?

Some intriguing properties have been observed for this cluster of genes, which we will refer to here as the "ZNF777 subfamily". All members encode a non-canonical KRAB domain that does not interact with KAP1, but rather, has activating activity (Okumura et al. 1997). Another unique characteristics of this family is the presence of a 5' exon encoding a novel domain, the HUB domain (Okumura et al. 1997). At least one cluster member, *ZNF398,* gives rise to isoforms with or without the HUB domain, which vary significantly in terms of their regulatory activity (Conroy et al. 2002). We also recently found that ZNF777 gives rise to HUB-plus and HUB-minus isoforms in choriocarcinoma cells (Chang et al. 2017). We uncovered the binding landscape of ZNF777 in human BeWo choriocarcinoma cells using chromatin immunoprecipitation followed by deep sequencing (ChIP-seq), revealing a role in regulation of genes involved in axon guidance, which have been coopted in placenta to regulate angiogenesis (Liao et al. 2010).

Since *ZNF777* is also expressed in embryonic brain (Liu et al. 2014), we sought to further investigate the functional role of this ancient gene in neuron development. Here we show that mouse *Zfp777* is expressed in neuronal stem cells (NSC) cultured from early mouse embryos, with a pattern that changes over the course of neuron differentiation *in vitro.* Using the NSC platform, we characterized the binding landscape

of Zfp777 in undifferentiated NSC. To circumvent the roadblock posed by the lack of a ChIP-ready antibody for the mouse protein, we exploited the CRISPR-Cas9 technique (Ran et al. 2013; Savic et al. 2015) to tag the endogenous Zfp777 protein with FLAG epitopes. Because we are interested in comparing the two proteins, we also tagged mouse Zfp282 using the same procedure. Our results revealed a novel Zfp777 binding motif that bears significant similarity to a motif predicted in *in vitro* studies (Isakova et al. 2017), and found that Zfp777 binds to promoters of genes encoding transcription factors, Wnt and TGF-beta pathways components, and proteins related to neuron development and axon guidance. Since these same functions were also found to be regulated by ZNF777 in BeWo cells (Chang et al., 2017), these results suggested that the mouse and human Zfp777 and ZNF777 proteins regulating similar genes and pathways, most classically associated with axon guidance, in diverse tissues. Recent studies have suggested that ZNF777 and ZNF282 interact closely at many promoters (Imbeault et al. 2017). The mouse NSC cell lines in which Zfp282 protein has also been successfully tagged provide an important resource for us to investigate this role in the future.

Results

Zfp282 and Zfp777 are expressed in mouse neural stem cells as they differentiate in culture

Both *Zpf777* and *Zfp282* have been shown to be expressed at high levels in embryonic brain (Chang et al. 2017; see Chapter 2). We thus wished to determine whether both genes would also be expressed in cultured primary mouse neuroblasts, or neural stem cells (NSCs). We tested the expression of Zfp777 and Zfp282 in mouse NSCs and cell types resulting from their differentiation *in vitro* by real-time reverse transcript PCR (qRT-PCR). We generated RNA samples from undifferentiated NSC and neurons, oligodendrocytes, and astrocytes differentiated from NSCs for this purpose, and used primers developed previously for qRT-PCR (Liu et al. 2014) (**Figure 3.1**). The results showed that both *Zfp777* and *Zfp282* are expressed in the undifferentiated NSCs, and continue to be expressed in each the differentiated cell type. Expression slightly decreased after 2 days of differentiation into neurons compared to the expression of NSCs, and continued decreasing during later stages of differentiation. The same types of expression pattern were also observed in differentiating astrocytes and oligodendrocytes. These data suggest that both *Zfp777* and *Zfp282* are active in each of the three derivatives of NSC, although expression is highest in the undifferentiated cells. We therefore used NSCs for the following experiments.

Engineering FLAG-tagged *Zfp282 and Zfp777* genes in mouse neural stem cells

To circumvent the obstacles caused by the limitations of the availability of suitable ChIP-grade antibodies, we exploited the powerful CRISPR-Cas9 genome editing tools to "knock-in" epitope sequences in frame to the 3' ends of endogenous *Zfp777* and *Zfp282* genes. We inserted three tandem FLAG tag sequences into each locus using the CETCh-seq approach (Savic et al. 2015). This method introduces the FLAG tag sequences, followed by a self-cleavage 2A peptide sequence (P2A) and neomycin resistance (NeoR) gene, which allows cells in which successful editing has taken place be selected (**Figure 3.2A**). Briefly, in successfully edited cells the neomycin resistance gene is cotranscribed with the FLAG tagged transcription factor, and separate tagged TF and neomycin

resistance proteins are generated through amino-acid peptide cleavage and ribosomal skipping at the P2A sequence.

After three weeks of G418 selection, we obtained a mixed pool of NSCs carrying tagged TFs and untagged TFs, with cells carrying tagged TFs significantly enriched in this population. PCR-based genotyping (**Figure 3.2B**) with primers flanking the regions outside of FLAG-P2A-NeoR showed an almost equal intensity of bands corresponding to the tagged genes (3069 bp for *Zfp282* and 2804 bp for *Zfp777*) versus the untagged genes (2063 bp for *Zfp282* and 1801 bp for *Zfp777*) in cell populations selected with 50 μg/ml of G418 (see **Supplemental Table 3.1** for sgRNA and primer sequences; **Supplemental Table 3.2** for sequences of HOM1 and HOM2); this result indicated efficient selection since most cells will carry the FLAG insertion in heterozygous form (Savic et al. 2015). We thereafter used NSC cell pools selected under 50 μg/ml of G418 for following experiments.

The PCR products were sent for Sanger-sequencing and successful knock-in of the three tagged cell pools (Zfp282-B3-2, Zfp777-C2-1, and Zfp777-C2-2) was confirmed. We tested Zfp282-FLAG and Zfp777-FLAG protein expression in the NSCs by western blot analysis, and found both the endogenously expressed tagged proteins could be detected by the FLAG antibody (**Figure 3.2C**). Interestingly, the Zfp282-FLAG protein showed a doublet pattern, which may indicate post-transcriptional modification as has been previously described for ZNF282 (Yu et al. 2012). The tagged Zfp777 protein was detected around 130 kDa, which is larger than the predicted size (~ 85 kDa), also suggesting the protein is possibly modified in NSCs. These results were consistent with a previous study, which showed that human ZNF777 migrates as a ~ 130 kDa band in western blots prepared from HCT116 cell protein extracts (Yuki et al. 2015). We also detected a protein of similar size in human BeWo cells, together with a shorter protein band that we confirmed as a HUB-minus, KRAB-minus isoform of the protein (Chang et al. 2017). However importantly, we did not observe any evidence of the shorter isoform either NSC cell pools expressing FLAG-tagged Zfp777; these data, along with western blot data from other human cell types (Chang et al. 2017, Chapter 2) suggest that the balance of isoforms arising from this protein may be dependent on cellular context.

Zfp777 binding landscape in mouse neural stem cells

To investigate the binding landscape of Zfp777 protein under endogenous conditions, we performed Chromatin immunoprecipitation (ChIP) followed by Illumina sequencing (ChIP-seq) in chromatin from tagged NSC pools using a well-tested FLAG antibody. After sequencing and alignment of ChIP-enriched fragments, we used HOMER software to identify 1245 peaks with peak scores of 5 or higher. Among these peaks, 685 peaks (55%) were located within 5 kb from a transcription start site (TSS) (648 peaks are within 2.5 kb). These results suggested Zfp777 is mostly involved in the regulation of the transcription initiation of the protein-coding genes through the interaction at the promoter (265 peaks) or the 5'UTR (218 peaks) regions (**Figure 3.3A)**.

ZNF282 has been reported to bind to the long-terminal repeat (LTR) regions of human retroviruses (Okumura et al. 1997), and KRAB zinc-finger proteins have been implicated to have evolved in an "arms race" with endogenous LTRs and other types of bioactive transposable elements (Jacobs et al. 2014). Furthermore, we found an enrichment for human ERV1 endogenous retroviral elements in the binding sites for ZNF777 in human BeWo choriocarcinoma cells (Chang et al. 2017; Chapter 2). For that reason, we examined the potential overlap between Zfp777 NSC binding sites and repetitive elements. Only 10 (< 1%) of the peaks identified in these experiments overlapped with endogenous retroviruses (ERVs), 18 with LINEs (11 intergenic; 7 intragenic) and 12 with SINEs (4 intergenic; 8 intragenic). These results suggested that regulation of repeats is not a major function for Zfp777, at least in NSC.

We used the summits of the most highly enriched Zfp777 peaks (at least 10-fold enrichment compared to input control; see Methods) and identified a highly enriched (p= 7.7e-23) centrally located motif (**Figure 3.3B**). This motif shares very high resemblance to a motif predicted for ZNF777 by the SMiLE-seq technique (Selective microfluidics-based ligand enrichment followed by sequencing) (Isakova et al. 2017). The concordance between this *in vitro* predicted and our *in vivo* validated motif provides strong evidence for this motif as the *bona fide* binding site for Zfp777. Interestingly, the motif we identified is identical to the predicted motif in a central core sequence (consensus CCGTGG) but differs from the SMILE-seq prediction in the less well-defined flanking sequences. This difference could reflect a difference in binding in the *in vivo* and *in vitro*

contexts, or could possibly vary depending on cellular context. The human ZNF777 motif we identified (Chang et al. 2017; Chapter 2) has fewer well-defined nucleotide positions but nevertheless aligns well with the spacing of Zfp777 and SMiLE motifs.

Among the 685 peaks located within 5 kb from annotated TSS, 17 of Zfp777 peaks were identified at a promoter for a ncRNA, and one peak was found on the promoter region of an annotated pseudogene; the remaining 667 peaks lie near the TSS of protein-coding genes. We found that the promoter of the *Zfp777* gene is bound by its own protein product with very high enrichment (**Figure 3.3C**), suggesting autoregulation. We also found several other interesting examples in this Zfp777-FLAG ChIP-seq dataset including many genes also regulated by ZNF777 in human BeWo cells. For example, a robust peak was identified at the promoter of the *Grhl1* (Grainy head-like 1) gene (**Figure 3.3C**); ZNF777 also binds to the GRHL1 promoter in human cells (Chang et al. 2017). Another peak was found at the promoter of *Sema6a*, which encodes a ligand involved in axon guidance; ZNF777 binds to human paralogs *SEMA5A* and *SEMA7A* in BeWo cells.

To gain a more global view of Zfp777 function, we analyzed all genes with promoter-associated Zfp777 peaks for functional analysis using the functional clustering option in the DAVID suite (Huang et al. 2009). The most enriched functional cluster identified was transcription regulation, with 117 of the peaks located adjacent to transcription factor genes (**Table 3.1**). Interestingly, genes within the Wnt and TGF-beta signaling pathways were also very highly enriched as were genes related to neuron development and differentiation, cell junctions, and synapses (**Table 3.1**). These same functions were also highly enriched in our ZNF777 study, suggesting that although in different species and tissues, with not exactly identical binding motifs, Zfp777 and ZNF777 regulate similar pathways, in mouse neural stem cells and human BeWo cell lines respectively.

Discussion

Here we define, for the first time, the functions of the ancient Zfp777 KRAB-ZNF transcription factor protein in neuronal cells. As we described previously (Liu et al., 2014; Chang et al. 2017; Chapter 2), ZNF777 and its closest relatives are very highly expressed in placenta tissues, but also were found to be active in embryonic brain. Similarly, mouse Zfp777 is expressed at high levels in embryonic brain, around the time when the brain is growing rapidly in mammals. As we demonstrate here, *Zfp777* and its close paralog, *Zfp282,* are both also expressed in cultured NSC, providing an excellent platform for functional exploration.

Because antibodies for the mouse proteins were not available, we used a recently developed system based on CRISPR-Cas9, called CETCh-seq (Savic et al. 2015) to engineer epitope tags into the endogenous *Zfp777* and *Zfp282* loci to enable investigation of the chromatin binding landscape for each protein in NSC. In this paper, we report results for chromatin analysis of Zfp777 using these CRISPR-engineered cells.

Zfp777 binding are enriched in CpG islands at promoters and 5'UTRs of coding genes, (**Supplemental Table 3.3**), suggesting a correlation with transcription initiation (Deaton & Bird 2011). This result is also in agreement with the recent findings of Imbeault and colleagues (Imbeault et al. 2017) who showed that ZNF777, ZNF282 and another family members ZNF398 bind in close proximity at many shared promoters. The Zfp777 binding peaks are enriched in a sequence motif with very high similarity to one predicted for ZNF777 in *in vitro* assays (Isakova et al. 2017), confirming the accuracy of the ChIP results.

In our recent study, we identified a binding motif of ZNF777 with less "information content" – that is, fewer distinct, and more degenerated nucleotide positions (Chang et al. 2017). Though low in information content, this motif was detected in virtually all the high-scoring human peaks and was thus identified with very high probability. Interestingly, the most distinct nucleotide positions in the human ZNF777 binding motif align well with the mouse Zfp777 motif we identified here with the same nucleotides positioned and the same spacing (**Figure 3.3B**). The human motif did not include six more weakly defined nucleotides at the 5' end of the mouse and the predicted motifs. Furthermore, we found very few overlaps in binding sites in comparisons of the human

BeWo and mouse NSC ChIP experiments. This difference could have several explanations. However, we conjecture that the disparity between the binding properties of human and mouse proteins is due to both species- and tissue-specific factors, similar to factors that impact binding of mammalian TFs more generally, as discussed in sections below.

An increasing number of studies have compared experimentally determined TF-DNA interactions between species (Kunarso et al. 2010; Mikkelsen et al. 2010; Schmidt et al. 2012; Cotney et al. 2013). In particular, Odom and colleagues compared binding profiles for several TFs across species, and concluded that tissue-specific transcriptional regulation has diverged significantly between human and mouse (Odom et al. 2007). They carried out chromatin immunoprecipitation followed by array hybridization (ChIP-chip) using specially designed proximal promoter microarrays with a set of liver-specific TFs (FOXA2, HNF1A, HNF4A, HNF6) whose function, amino acid sequence, and targeted binding motif are highly conserved throughout mammals. Their result confirmed that despite high levels of conservation in DNA-binding domains, the homologous TFs differed in both their global binding locations and their potentially targeted genes. This study was limited to the proximal promoters around 4,000 transcription start sites in human and mouse, due to the technical limitations imposed on the experiments by the extant microarray densities. However, in a follow-up study, this same group performed ChIP-seq for two TFs (CEBPA and HNF4A) in five vertebrates, unambiguously revealing tens of thousands of binding events that are unique to each evolutionary lineage. Although these TF displayed similar DNA binding preferences in terms of motifs, most binding events were species-specific, and aligned binding events present in all five species were rare (Shmidt et al. 2010). It is also well known that TFs binding is also highly tissue-specific (Badis et al. 2009; Blow et al. 2010; Jolma et al. 2013; Neph et al. 2012; Pique-Regi et al. 2011). More specifically, C2H2 zinc-finger protein binding sites were found to be enriched in many cell type-specific DNaseI hypersensitive regions, suggesting a role in regulation of cell type-specific transcriptional programs (Najafabadi et al. 2015). It is therefore perhaps not surprising to find binding differences between human choriocarcinoma cells and mouse NSC.

Finally, and perhaps most importantly, we note that the *ZNF777* gene gives rise to two clear isoforms in human BeWo cells, corresponding to HUB-plus and HUB-minus proteins, respectively (Chang et al. 2017; Chapter 2). For ZNF777 family member *ZNF398*, which produces similar isoforms, inclusion of the HUB domain prevents interaction of the protein's zinc finger domain with interacting partner, ER-α, and significantly alters ZNF398's binding sites and regulatory activities (Conroy et al. 2002). Similarly, we speculate that the HUB-minus version of ZNF777 may possess different, and possibly much more promiscuous, binding properties than does the full-length protein.

Since BeWo cells include both protein isoforms and the antibody we used for ChIP detects both protein versions equally well, we hypothesize that the less distinct motif we detected for human ZNF777 may represent a "composite" motif that reflects binding sites of both proteins. In mouse NSC, on the other hand, only the full-length, HUB+ protein isoform was detected; this protein would be similar to that used for *in vitro* protein predictions (Isakova et al. 2017), with more nucleotide positions well defined in the predicted motif.

Despite these differences, it is striking to find that genes related to neuron development, axon guidance, and synapses were enriched near binding site for both ZNF777 and Zfp777 in these vastly different types of cells. Interestingly, many of the genes are best known for their functions in neurogenesis, but have also shown to play a non-neuronal role in development of the placenta (Liao et al. 2010; Jongbloets & Pasterkamp 2014). Genes within the semaphorin pathways, which were particularly enriched in both the human ZNF777 and the mouse Zfp777 results, present one of the most striking examples. We speculate that this regulatory activity is representative of the "root" function of this ancient protein, and that this activity has been captured in mammals for a novel biological role in placental trophoblasts.

In closing, we should again note that recent studies have suggested interacting roles for ZNF777 and ZNF282, including co-binding at many promoters (Imbeault et al. 2017). The two proteins are expressed in very similar patterns in both humans and mice, and could well interact *in vivo.* Such interaction would be interesting and might relate to the deep conservation of these unique KRAB-ZNF "root" proteins. The NSC cells we have

engineered to expression a tagged Zfp282 protein should be ideally suited to test this interaction and the notion of a coupled functional role.

Materials and Methods

Ethics Statement

This investigation has been conducted in accordance with the ethical standards and according to the Declaration of Helsinki and according to national and international guidelines.

Cell culture

Mouse neural stem cells (NSC) were obtained from E14.5 mouse embryos. NSCs were cultured by NeuroCult™ Proliferation Kit (STEMCELL ™) incubated at 37 °C in 5% CO2. The cell culture plates were coated with Poly-D-Lysine for 2 hours at room temperature and Laminin solution for 2 hours at 37 °C before plating the cells.

Generation of FLAG tagged Zfp282 and Zfp777 genes in mouse NSC

The CRISPR procedures used in our lab follows the protocol published by Ran et al. (Ran et al. 2013) with some modifications. We use http://crispr.mit.edu/ website created by Zhang lab for the design of the small guide RNAs (sgRNAs). We test at least 2-3 sgRNAs with the highest scores for each site to be modified. The plasmid pSpCas9-2A-GFP (PX458; Addgene plasmid ID: 48138) encoding Cas9, a sgRNA cloning site under U6 promoter, and a GFP gene is used to express our sgRNAs. Cells are transfected with Neon™ transfection system for all our CRISPR plasmids and the pFETCh template plasmid encoding HOM1 and HOM2 according to the manufacturer's instructions. The transfection conditions of Neon™ system were optimized on different cell lines. Neural stem cells were transfected with 1150 pulse voltage, 30 ms pulse width, and 2 pulses. At 24 h post transfection, transfection efficiency can be estimated from the fraction of fluorescent cells. We typically got 30-50% transfection efficiency with neuronal stem cells. At 48 h post transfections, cells were treated with 10, 20, or 50 μg/ ml G418 for selection. Cells were then incubated for ~3 weeks, resuspended in Accutase

(STEMCELL ™), 20% of the cells from one well can be harvested for genotyping and sequencing while the remaining cells were plated onto a new 6-well plate. The positive clones were further expended and harvested for downstream analysis, e.g. Western Blots and ChIP.

RNA preparation and quantitative RT-PCR

Total RNA was isolated from cell lines and tissues using TRIzol (Invitrogen) followed by 30 minutes of RNase-free DNaseI treatment (NEB) at 37°C and RNA Clean & Concentrator™-5 (Zymo Research). 2 μg of total RNA was used to generate cDNA using Superscript III Reverse Transcriptase (Invitrogen) with random hexamers (Invitrogen) according to manufacturer's instructions.

Resulting cDNAs were analyzed of transcript-specific expression through quantitative reverse-transcript PCR (qRT-PCR) using Power SYBR Green PCR master mix (Applied Biosystems) with custom-designed primer sets purchased from Integrated DNA Technology. Relative expression was determined by normalizing the expression of all genes of interest to either mouse Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide (*YWHAZ*) expression (ΔCt) as described (Eisenberg and Levanon, 2003).

Protein preparation, Western blots, and antibodies

Nuclear Extracts were prepared with NucBuster™ Protein Extraction Kit (Novagen) and measured by Bradford-based assay (BioRad). The extracts were stored at -80°C and thawed on ice with the addition of protease inhibitor Cocktail (Roche) directly before use. 15 μg of nuclear extracts were run on 10% acrylamide gels and transferred to hydrophobic polyvinylidene difluoride (PVDF) membrane (GE-Amersham, 0.45 μm) using BioRad Semi-dry system, then visualized by exposure to MyECL Imager (Thermo Scientific). FLAG rabbit monoclonal antibody (F1804) was purchased from Sigma® .

Chromatin immunoprecipitation

Chromatin immunoprecipitation was carried out as essentially as described (Kim et al., 2003) with modifications for ChIP-seq. Chromatin was prepared from NSC cell lines.

About 1.0 x 10^6 Cells were fixed in PBS with 1% formaldehyde for 10 minutes. Fixing reaction was stopped with addition of Glycine to 0.125M. Fixed cells were washed 3x with PBS+Protease inhibitor cocktail (PIC, Roche) to remove formaldehyde. Washed cells were lysed to nuclei with lysis solution – 50 mM Tris-HCl (pH 8.0), 2 mM EDTA, 0.1% v/v NP-40, 10% v/v glycerol, and PIC – for 30 minutes on ice. Cell debris was washed away with PBS with PIC. Nuclei were pelleted and flash-frozen on dry ice.

Cross-linked chromatin was prepared and sonicated using Bioruptor UCD-200 in ice water bath to generate DNA fragments 200-300 bp in size. 15 micrograms of FLAG antibody preparations were incubated with chromatin prepared from nuclei of approximately 10 million cells.

DNA was released and quantitated using Qubit 2.0 (Life Technologies) with dsDNA HS Assay kit (Life Technologies, Q32854), and 15 ng of DNA was used to generate libraries for Illumina sequencing. ChIP-seq libraries were generated using KAPA LTP Library Preparation Kits (Kapa Biosystems, KK8232) to yield two independent ChIP replicates for each antibody. We also generated libraries from sonicated genomic input DNA from the same chromatin preparations as controls. Libraries were bar-coded with Bioo Scientific index adapters and sequenced to generate 15-23 million reads per duplicate sample using the Illumina Hi-Seq 2000 instrument at the University of Illinois W.M. Keck Center for Comparative and Functional Genomics according to manufacturer's instructions. Separate ChIP preparations were generated for qRT-PCR validation experiments; in this case, released DNA was amplified by GenomePlex® Complete Whole Genome Amplification (WGA) Kit (Sigma, WGA2).

ChIP-Seq data analysis

Human ZNF777 ChIP-enriched sequences as well as reads from the input genomic DNA were mapped to the HG19 human genome build using Bowtie 2 software (Langmead et al. 2009) allowing 1 mismatch per read but otherwise using default settings. Bowtie files were used to identify peaks in ChIP samples using the HOMER software package (http://homer.salk.edu/homer/ngs/index.html) using default conditions for the TF setting and false discovery rate cutoffs of 0.1. After comparison of the individual files, sequence reads from the two separate ChIP libraries were pooled and a final peak set determined in

comparison to genomic-input controls. Peaks were mapped relative to nearest transcription start sites using the GREAT program (Mclean et al. 2010).

Motif analysis

To identify enriched motifs, we used sequence from a 200 bp region surrounding the predicted summits of selected peaks for analysis with MEME-ChIP with default parameters (Machanick et al. 2011). Motifs displayed in Fig. 3.3B were identified from peaks with the following cutoffs: HOMER ef > 10, fdr=0 peaks from Zfp777 ChIP in NSC chromatin; the identified motif occurred in 84 out of total 178 peaks submitted peaks, with p value = p= 7.7e-23.

Figures and Tables

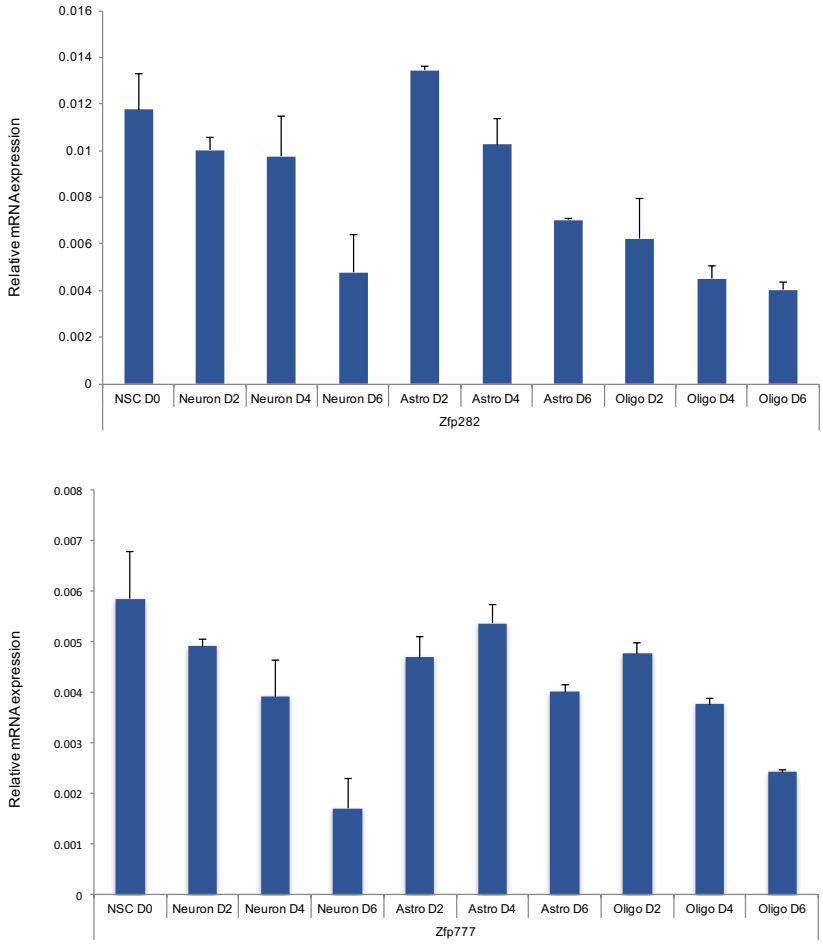
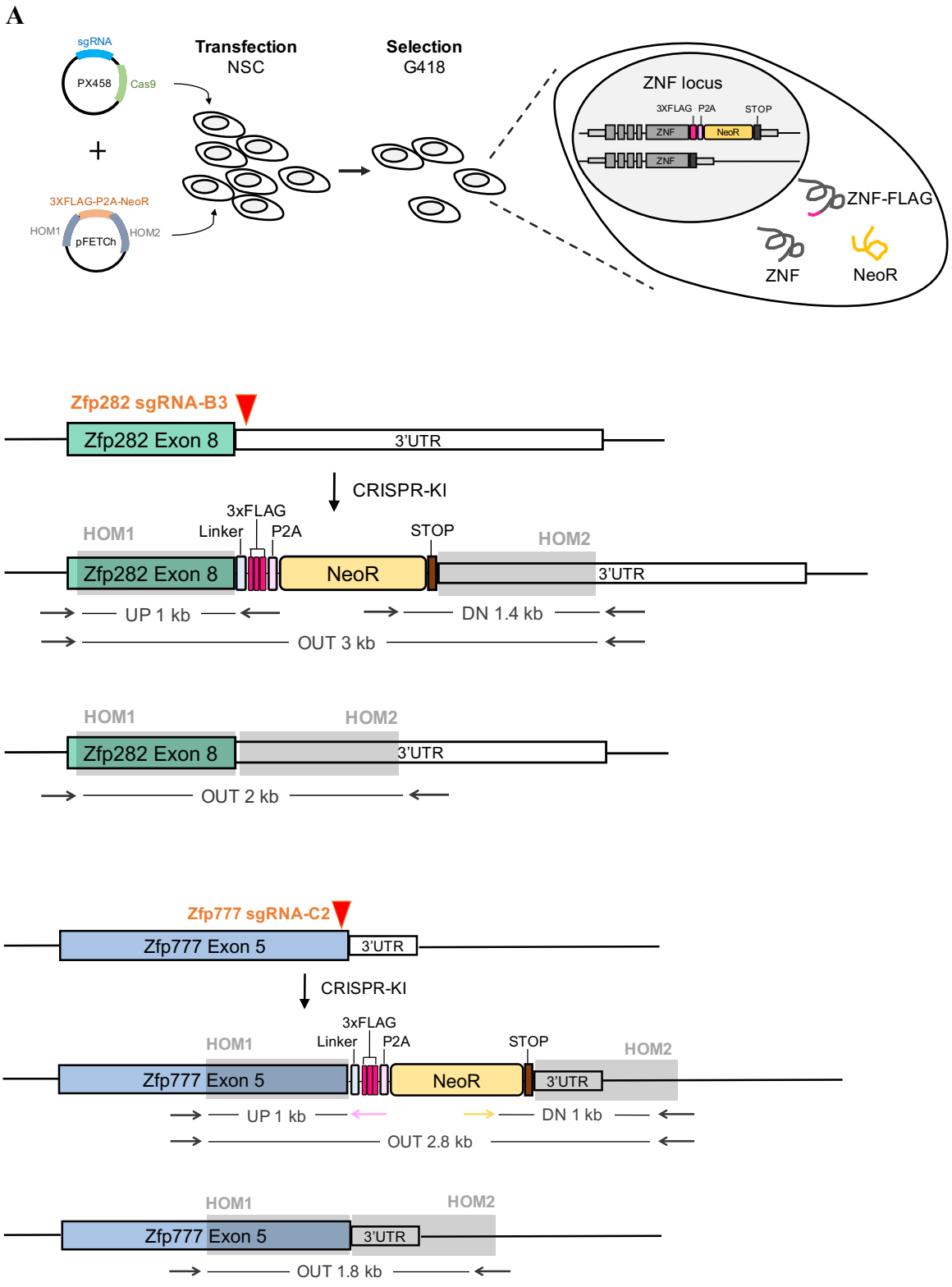
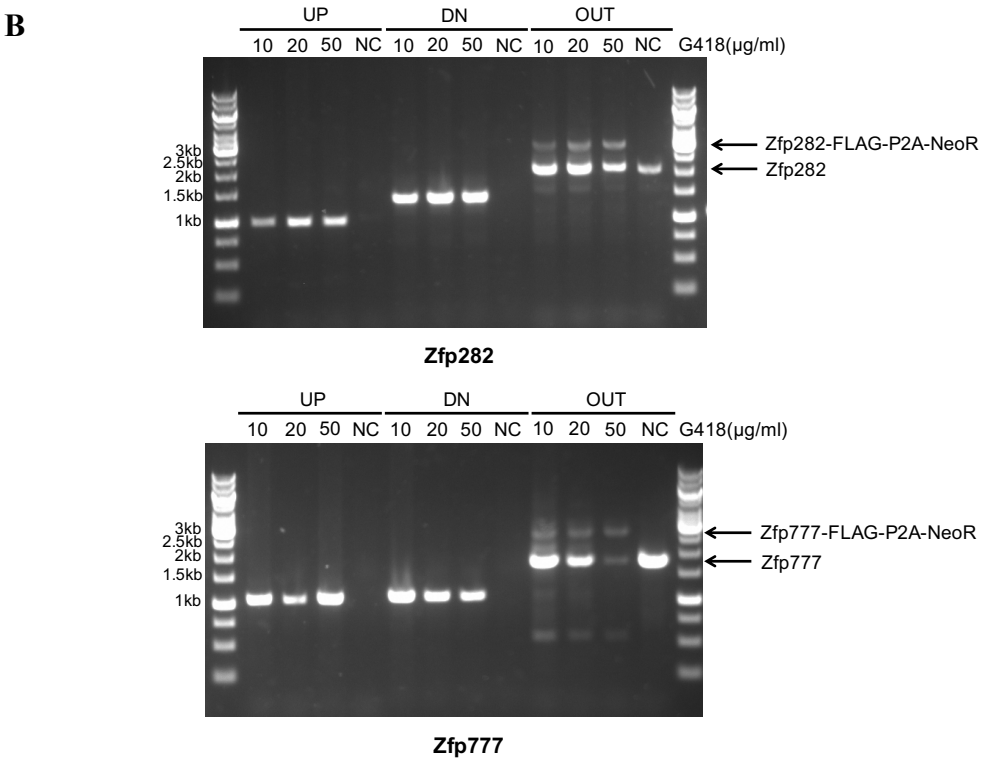


Figure 3.1 Expression levels of Zfp282 and Zfp777 in mouse neural stem cells (NSC), neurons, astrocytes, and oligodendrocytes by qRT-PCR.

The relative mRNA expression levels were detected in all cell types. Astro: astrocytes. Oligo: oligodendrocytes. D0: undifferentiated. D2, D4, D6: 2, 4, 6 days after differentiation.

Figure 3.2


A
sgRNA
PX458
Cas9
+
3XFLAG-P2A-NeoR
HOM1
pFETCh
HOM2
Transfection
NSC
Selection
G418
ZNF locus
3XFLAG
P2A
STOP
ZNF
NeoR
ZNF-FLAG
ZNF
NeoR
Zfp282 sgRNA-B3
Zfp282 Exon 8
3'UTR
CRISPR-KI
3xFLAG
Linker
P2A
HOM1
STOP
HOM2
NeoR
UP 1 kb
DN 1.4 kb
OUT 3 kb
OUT 2 kb
Zfp777 sgRNA-C2
Zfp777 Exon 5
UP 1 kb
DN 1 kb
OUT 2.8 kb
OUT 1.8 kb



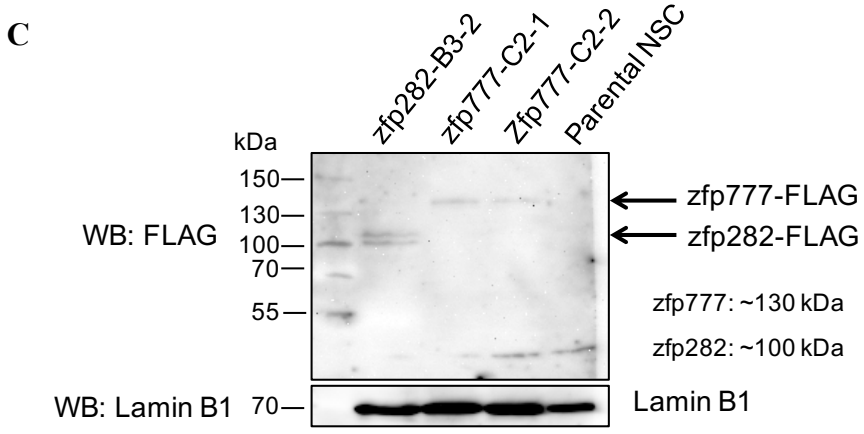


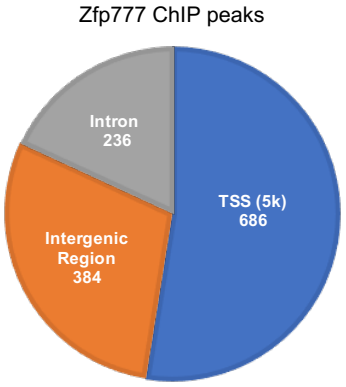
Figure 3.2 (cont.) Generation of endogenous FLAG-tagged Zfp282 and Zfp777 in mouse neural stem cells (NSCs) by CETCh-seq method.

(A) Schemes of FLAG-tagged Zfp282 and Zfp777. sgRNAs were designed to target near the stop codon of each gene, and cloned into the cloning site under the U6 promoter of PX458 plasmid which also encodes Cas9 protein. A template plasmid (pFETCh) was constructed to include two homology arms (HOM1 and HOM2) containing around 800

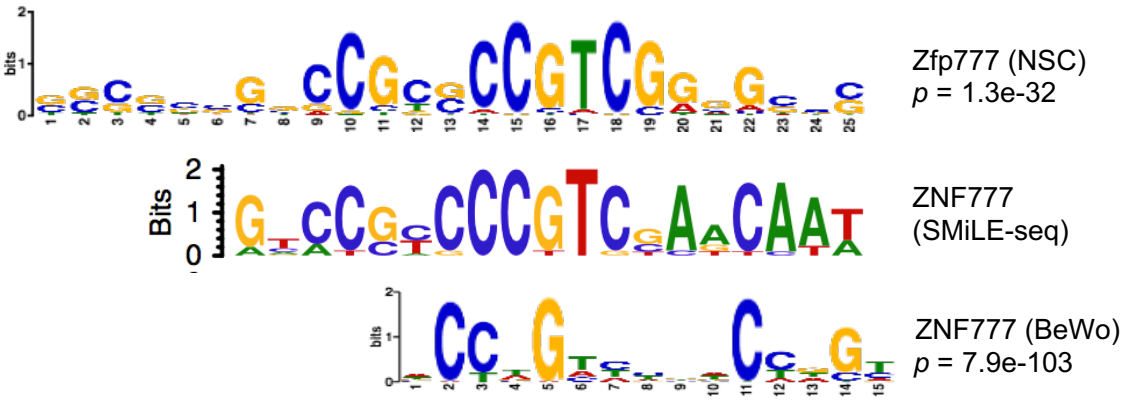
Figure 3.2 (cont.) bp of the genomic sequences of Zfp282 or Zfp777 immediate upstream of the stop codon (HOM1) and downstream of the stop codon (HOM2). HOM1 and HOM2 were cloned into the template plasmid to flank the GS linker, 3 FLAG tags, P2A sequence, and the neomycin resistant gene (NeoR). NSCs were co-transfected with PX458 containing sgRNA and the pFETCh template plasmid for 2 days, and selected by G418 for 3 weeks. After the homologous recombination in the NSCs, the original sequences near the stop codon were replaced by the sequences provided by the template plasmid, resulting in the knock-in of 3 FLAG sequences and the NeoR gene. The NSCs that were successfully knocked-in at the Zfp282 or Zfp777 genes can survive the selection of G418. After selection, the heterozygous modified cells would carry one allele of un-tagged Zfp282 or Zfp777, thus producing a smaller band (~1 Kb difference) compared to the band produced by the tagged allele in PCR validation. The primer locations for PCR validation of homologous recombination are depicted by the arrows. UP: primer pairs flanking the upstream (HOM1) region, which only produce PCR amplicon in tagged NSCs. DN: primer pairs flanking the downstream (HOM2) region. OUT: primer pairs flanking the entire region including both HOM1 and HOM2. These pairs can produce PCR products in both tagged and untagged cells, thus provide the information of the ratio of tagged/untagged gene in the cell pools. (B) PCR validation of homologous recombination. NSCs were selected under 10, 20, or 50 μg/ml of G418 for 3 weeks. UP and DN primer pairs confirmed the tagging of Zfp282 and Zfp777. OUT primer pairs provided information of the ratio of tagged/untagged Zfp282 and Zfp777 in the mixed cell pools by comparing the intensity of the upper band/lower band (3 kb/2 kb in Zfp282; 2.8kb/1.8 kb in Zfp777). The NSCs selected by 50 μg/ml G418 had more enriched cell pools with tagged genes. NC (negative control): parental NSCs without modification. (C) Western blot validation of FLAG tagged Zfp282 and Zfp777 in NSCs. Three NSC CRISPR-modified stable cell lines (Zfp282-B3-2: modified by Zfp282 sgRNA-B3, and Zfp777-C2-1, Zfp777-C2-2: modified by Zfp777 sgRNA-C2) were harvested, and the nuclear extracts were collected for western blot analysis. The Zfp777-FLAG and Zfp282-FLAG proteins were detected by FLAG monoclonal antibody at ~130 kDa and ~100 kDa, respectively.

Figure 3.3

A



B



C

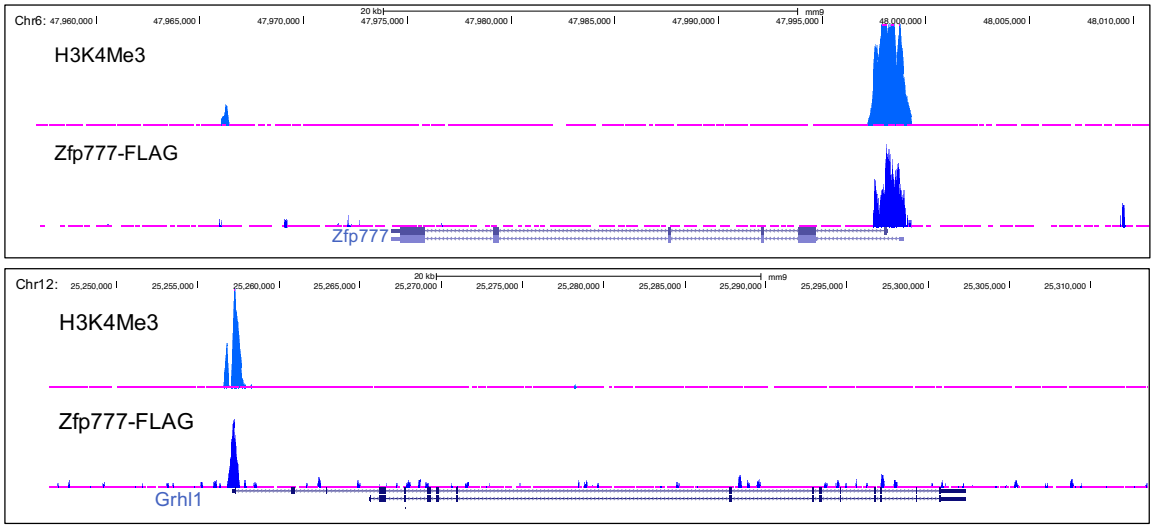


Figure 3.3 (cont.) Zfp777 binding landscape in mouse NSC.

(A) Distribution of Zfp777 peaks identified by ChIP-seq in mouse NSCs. 1245 peaks (with peak score higher than 5) were identified by HOMER software. 685 peaks (55.1%) were found located within 5 kb from a transcription start site (TSS), 384 peaks (30.8%) in intergenic regions, and 236 peaks (18.9%) in introns (5 kb away from TSS).

(B) Alignment of binding motifs of Zfp777 in mouse NSCs, ZNF777 predicted by SMiLE-seq, and ZNF777 in human BeWo cell lines. The consensus sequence of ZNF777 binding motif defined by SMiLE-seq is: G**CCGTCG**AACAT, with the core CCGTCG being found in the mouse Zfp777 binding motif identified in our FLAG-ChIP assay in mouse NSC. Both Zfp777 and ZNF777 motifs were identified by MEME software.

(C) Examples of Zfp777 ChIP peaks. The Zfp777 peaks and H3K4Me3 peaks (identified in mouse frontal cortex tissues) were show, indicating promoter regions. Zfp777 was found to bind to its own promoter (top panel). Zfp777 binds to Grhl1 gene (bottom panel) at the promoter region.

DAVID Functional Cluster	Enrichment Score[1]
Transcription regulation	10.2
Zinc-finger	6.0
Wnt Signaling Pathway	4.1
Neuron development and differentiation	3.6
C2 calcium dependent membrane targeting	3.1
TGF-beta signaling pathway	3.0
Axon guidance	2.8
Chromatin organization	2.5
Helix-loop-helix domain	2.5
Limb development	2.1
C2H2-type	2.0
Retinoid X receptor	2.0
PAS domain	1.9
BTB/POZ domain	1.9
Cell-cell adhesion	1.8
Protein kinase C-like	1.7

Table 3.1: Gene Ontology (GO) clusters identified as significantly enriched in gene sets with Zfp777 bound within 5 kb of their transcription start sites (TSS).

[1] David enrichment scores are calculated as the geometric mean of –log transformed P-values of GO terms within a cluster based on content of similar genes

Supplemental Figure 3.1 Target sites of Zfp282-sgRNA-B3 and Zfp777-sgRNA-C2

The sgRNAs were designed using the http://crispr.mit.edu/ website created by Zhang lab. The sgRNAs consisting of a 20-nt guide sequence (orange box) were designed near the stop codon of Zfp282 and Zfp777 genes, directly upstream of a requisite 5'-NGG adjacent motif (PAM; red underline). The Cas9 nuclease is targeted to genomic DNA by the sgRNAs, mediates a double strand break ~3 bp upstream of the PAM, indicated by the red arrow heads.

Gene	Primer pair	Forward primer	Reverse primer	Product size (bp)
Zfp282	UP	GCGGTATCAGCGTGTCACTT	CAGCAGGCTGAAGTTAGTAGC	1010
	DN	GGCCGCTTTTCTGGATTCAT	GTCCATGTCCGTGAGCACAA	1411
	OUT	GCGGTATCAGCGTGTCACTT	GTCCATGTCCGTGAGCACAA	2063+3069
Zfp777	UP	GGTGAGAACCGTGGGAACTC	CAGCAGGCTGAAGTTAGTAGC	1062
	DN	GGCCGCTTTTCTGGATTCAT	CCACAGACCACACTAGAGGC	1094
	OUT	GGTGAGAACCGTGGGAACTC	CCACAGACCACACTAGAGGC	1801+2804

Gene	sgRNA	
Zfp777	C1	CGCATGCTCACTCGCCCGTG
	C2	**GCATGCTCACTCGCCCGTGT**
	C3	CGCCCGTGTGGGTCCGCAGG
Zfp282	B2	GCCCAACCCTAGTCTCTTTC
	B3	**GCACCAGAAAGAGACTAGGGT**

Supplemental Table 3.1 Primers and sgRNA sequences for Zfp282 and Zfp777 FLAG tagging mediated by homologous recombination of pFETCh template plasmid (CETCh-seq method). All sgRNAs were tested, the best KI efficiencies resulted from Zfp282 sgRNA-B3 and Zfp777 sgRNA- C2.

Supplemental Table 3.2

zfp777 HOM1 (5'HOM)

ACCAGCGTAACCACATCAAGGAGGGGCCCTACGAGTGTGCCGAGTGTGAGATCAGCTTC
CGCCACAAGCAGCAACTCACGCTGCACCAGCGCATCCACCGGGTACGCAGCGGCTATGC
CTCCCCTGAGCGCGGGTCAGCCTTCAATCCCAAGCACTCGCTCAAACCACGTCCCAAAT
CGCCCAGCTCAGGCAGTGGCGGCGGCCCCAAACCCTACAAATGCCCTGAGTGTGACAGC
AGCTTCAGCCACAAGTCAAGCTTGACCAAGCACCAGATCACACACACGGGTGAGCGGCC
CTACACGTGCCCAGAATGCAAGAAGAGCTTCCGCTTGCATATCAGTCTGGTGATCCACC
AGCGTGTGCATGCAGGCAAGCACGAAGTCTCCTTCATTTGCAGTCTGTGCGGCAAGAGT
TTCAGCCGCCCGTCGCATCTGCTGCGCCACCAGCGGACTCATACTGGTGAACGGCCTTT
TAAGTGCCCGGAGTGCGAGAAGAGCTTCAGTGAGAAATCTAAGCTCACCAACCACTGCC
GCGTGCACTCCCGCGAGCGGCCGCACGCCTGCCCTGAGTGCGGCAAGAGCTTCATCCGC
AAGCACCACTTGCTGGAACACCGGCGCATCCACACGGGTGAGCGGCCCTACCACTGCGC
CGAGTGTGGCAAGCGCTTCACGCAGAAGCACCACCTGCTGGAGCATCAGCGTGCGCACA
CAGGCGAGCGGCCATACCCCTGCACGCACTGCGCCAAGTGCTTCCGCTACAAACAGTCG
CTCAAGTACCACCTGCGGACCCACACGGGCGAG

Zfp777 HOM2 (3'HOM)

Gcatgcgccccgcccctgccccggccagatgtgcagccaggtgcaagggtctaaggccc
ctgggacaggtacctcggctgcccccagactgagctcagtgggcgggagcggggcgccc
caagcccttctgctgtgaaccctccttccctcccgtcccttcttccccaggacggggta
gtgagaccaggtcgcttcttgcctgcttccccagggccccaggggggagtgcttgggcc
tggggaacccccttcaggctgttaatttccttgacaataaaatggatgaaaacaatctgc
acgggggcagtgatttggctgccagccactcgcaggcgcgatgcagggccatttagtcg
gggatagaactttctaattacctttggatactgtggttctatttgataataatagagt
aatttttaaaagacgagtgtttcctgtttgctgttcttgtttggtttgtaaggggaggg
gctaaggtggcccaagggacatgtgccccagtattagctgacagacaaccaagttcctt
tctcaaactcattgtctgctggatgatggagaaataagatactgcttataaatttaaaa
ggagtgatgctgacaaacttaaaggagagaaatctggggaagggtaaaagcacctgcct
gccagccttcctgtgccctcttgtctacctgcggggtggtctctccgtaggtaatactg
tcgtcccagctgcccagagtgatcagggagaaagtgatggtcgggctgagaatggtctg
aaaagatggtcctatggcaaaagctgggggctt

Zfp282 HOM1 (5'HOM)

TAAAGAACCCACCCCCATCTTCTGCGCAGCCCCAAACCCAACCTCATCAGCAGAGCCTG
CCCGCCTTGGCTGTGCCGGAGAACCCTGGCGGACCCGGGAGCCGTAGCCTGCTGGAGGA
TGGCTTCCCTGCTCTTCCAGGCGAGCGCAGTACCGGAGGCGAGGCTCAGCCCACCGGAG
AAGGCAGTGCAGGCGGTGGCGGTGGTGGTGGCAGCGGCGGCGGCGGCGGCACTGGTGCG
GGTAGTGGCAATAGTACCGGTGCTGGTGCGGGCAGTGGCTGCGGTAGCTGCTGCCCAGG
CGGCCTGCGGCGGAGCCTCCTTGCTCACGGCGCGCGCAGCAAGCCCTACTCTTGCCTGG
AATGCGGCAAGACCTTCGGCGTGCGAAAGAGCCTCATCATTCATCACCGCAGCCACACC
AAGGAACGACCATACGAGTGCGCAGAGTGCGAGAAGAGCTTCAACTGCCACTCTGGCCT
CATCCGCCACCAGATGACGCACCGCGGTGAGCGGCCCTACAAATGCTCCGAGTGTGAGA
AGACCTACAGCCGCAAGGAGCACCTGCAGAACCACCAGCGGCTGCACACGGGCGAGCGG
CCCTTCCAGTGCGCGCTCTGCGGCAAGAGCTTCATCCGAAAGCAGAACCTGCTAAAGCA
CCAGCGGATCCACACGGGCGAGCGGCCCTACACATGTGGCGAGTGCGGCAAGAGCTTCC

GCTACAAGGAGTCACTCAAGGACCACCTGCGCGTGCACAATGGCCCGGGCCTGGGGGCC
CCGCGGCCACTCCAGGTGCCACCAGAAAGAGAC

Zfp282 HOM2 (3'HOM)
ggttgggctgggggcggggaggaatggactggagttggggcgggttagggttctcctg
ccccaccgttctcagcgcacctccccgccctctcctcacctcctgctggaaatcggcac
aggaattgcactccagacagggtattccaaggggtggacctgggtaccccagtactgtc
caactctagtggacagtccagctcatctcatagggtggacccagtggccagggaaggtc
ccagagggacagcaaggcagcaggaatcgttgggacacacctcggacacaccactggcc
actggggttacagattctgatcagggaaagcgaccagagagtctccaacccccttctgag
aaaaggaaatatgatccatcctgaaggtgaggagacatcctgaaaaggagagcaaatct
gcggtgtggaagctgagggaagcgctaagggtaacatcctcatgacaacactgcctcgc
gctctaatagcgctttatacttttttaaaaagtgttttctatccgttatctatttacac
ccttagcttatcccttcgagttaggtggggtagggttttcctgatgtggtaactgagga
gagtgagacacaggtgagatagttgtttagcaagaccacatgagaacagtgcggccaag
ccgcaccagggctccagcccagtgcagtgtccccaccgcacacactgcctacctctgcc
ggtctcagaccgatctcacctggcctttctggtctctctcctctcccacttccctccct
ggacccccccaaatcctctcagaagcaacagggg

Supplemental Table 3.2 (cont.) HOM sequences for Zfp282 and Zfp777 template plasmids (pFETCh) construction.

Each HOM arm is 800 bp in length. HOM1 contains the gnomic sequence immediately upstream of the stop codon of the target gene while HOM2 contains the sequence downstream of the stop codon.

P-value	Log P-value	Annotation	#features	Coverage(bp)	AvgFeatureSize	Overlap(#peaks)
1e-990	-2277.69	cpgIsland	16026	10496250	654	627
1.00E-235	-539.97	utr5	45216	4336491	95	317
1.00E-231	-530.49	Simple_repeat|Simple_repeat	1062306	64001831	60	565
1.00E-231	-530.49	Simple_repeat	1062306	64001831	60	565
1.00E-221	-508.05	TGGAn|Simple_repeat|Simple_repeat	3850	431127	111	125
1.00E-219	-502.93	promoters	34938	28145798	805	353
1.00E-198	-455.69	protein-coding	367832	61167268	166	444
1.00E-195	-447.82	exons	380955	65359233	171	450
1.00E-168	-384.66	TCCAn|Simple_repeat|Simple_repeat	3788	423808	111	104

Supplemental Table 3.3 Genome Ontology enrichment analysis of Zfp777 ChIP peaks

References

Badis, G. et al., 2009. Diversity and complexity in DNA recognition by transcription factors. *Science*, 324(5935), pp.1720–1723.

Bellefroid, E.J. et al., 1993. Clustered organization of homologous KRAB zinc-finger genes with enhanced expression in human T lymphoid cells. *The EMBO journal*, 12(4), pp.1363–1374.

Blow, M.J. et al., 2010. ChIP-Seq identification of weakly conserved heart enhancers. *Nature Genetics*, 42(9), pp.806–810.

Chang, L.-H. et al., 2017. Functions of ZNF777, a gene representing the root of the mammalian KRAB zinc finger family. *Submitted*.

Collins, T., Stone, J.R. & Williams, A.J., 2001. All in the family: the BTB/POZ, KRAB, and SCAN domains. *Molecular and Cellular Biology*, 21(11), pp.3609–3615.

Conroy, A.T. et al., 2002. A Novel Zinc Finger Transcription Factor with Two Isoforms That Are Differentially Repressed by Estrogen Receptor. *Journal of Biological Chemistry*, 277(11), pp.9326–9334.

Consiantinou-Deltas, C.D. et al., 1992. The identification and characterization of KRAB-domain-containing zinc finger proteins. *Genomics*, 12(3), pp.581–589.

Cotney, J. et al., 2013. The evolution of lineage-specific regulatory activities in the human embryonic limb. *Cell*, 154(1), pp.185–196.

Deaton, A.M. & Bird, A., 2011. CpG islands and the regulation of transcription. *Genes & Development*, 25(10), pp.1010–1022.

Huang, D.W., Sherman, B.T. & Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1), pp.44–57.

Huntley, S., 2006. A comprehensive catalog of human KRAB-associated zinc finger genes: Insights into the evolutionary history of a large family of transcriptional repressors. *Genome Research*, 16(5), pp.669–677.

Imbeault, M., Helleboid, P.-Y. & Trono, D., 2017. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature*, 543(7646), pp.550–554.

Isakova, A. et al., 2017. SMiLE-seq identifies binding motifs of single and dimeric transcription factors. *Nature Methods*, 14(3), pp.316–322.

Jacobs, F.M.J. et al., 2014. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature*, pp.1–18.

Jolma, A. et al., 2013. DNA-binding specificities of human transcription factors. *Cell*, 152(1-2), pp.327–339.

Jongbloets, B.C. & Pasterkamp, R.J., 2014. Semaphorin signalling during development. *Development*, 141(17), pp.3292–3297.

Kim, C.A. & Berg, J.M., 1996. A 2.2 A resolution crystal structure of a designed zinc finger protein bound to DNA. *Nature structural biology*, 3(11), pp.940–945.

Kunarso, G. et al., 2010. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature Genetics*, 42(7), pp.631–634.

Liao, W.-X. et al., 2010. Perspectives of SLIT/ROBO signaling in placental angiogenesis. *Histology and Histopathology*, 25, pp.1181–1190.

Liu, H. et al., 2014. Deep Vertebrate Roots for Mammalian Zinc Finger Transcription Factor Subfamilies. *Genome Biology and Evolution*, 6(3), pp.510–525.

Margolin, J.F. et al., 1994. Krüppel-associated boxes are potent transcriptional repression domains. *Proceedings of the National Academy of Sciences*, 91(10), pp.4509–4513.

Mikkelsen, T.S. et al., 2010. Comparative epigenomic analysis of murine and human adipogenesis. *Cell*, 143(1), pp.156–169.

Najafabadi, H.S. et al., 2015. C2H2 zinc finger proteins greatly expand the human regulatory lexicon. *Nature Biotechnology*, 33(5), pp.555–562.

Neph, S. et al., 2012. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature*, 489(7414), pp.83–90.

Nowick, K. et al., 2010. Rapid Sequence and Expression Divergence Suggest Selection for Novel Function in Primate-Specific KRAB-ZNF Genes. *Molecular Biology and Evolution*, 27(11), pp.2606–2617.

Odom, D.T. et al., 2007. Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nature Genetics*, 39(6), pp.730–732.

Okumura, K. et al., 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 Research*, 25(24), pp.5025–5032.

Pavletich, N.P. & Pabo, C.O., 1993. Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science*, 261(5129), pp.1701–1707.

Pavletich, N.P. & Pabo, C.O., 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science*, 252(5007), pp.809–817.

Pengue, G. et al., 1994. Repression of transcriptional activity at a distance by the evolutionarily conserved KRAB domain present in a subfamily of zinc finger proteins. *Nucleic Acids Research*, 22(15), pp.2908–2914.

Pique-Regi, R. et al., 2011. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Research*, 21(3), pp.447–455.

Ran, F.A. et al., 2013. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, 8(11), pp.2281–2308.

Rowe, H.M. & Trono, D., 2011. Dynamic control of endogenous retroviruses during development. *Virology*, 411(2), pp.273–287.

Rowe, H.M. et al., 2010. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature*, 463(7278), pp.237–240.

Savic, D. et al., 2015. CETCh-seq: CRISPR epitope tagging ChIP-seq of DNA-binding proteins. *Genome Research*, 25(10), pp.1581–1589.

Schmidt, D. et al., 2012. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell*, 148(1-2), pp.335–348.

Shmidt, D., Odom, D.T. & Wilson, M.D., 2010. Five-Vertebrate ChIP-seq Reveals the Evolutionary Dynamics of Transcription Factor Binding. *Science*, 328(5981), pp.1036–1040.

Stubbs, L., Sun, Y. & Caetano-Anolles, D., 2011. Function and Evolution of C2H2 Zinc Finger Arrays. *Sub-cellular biochemistry*, 52, pp.75–94.

Thomas, J.H. & Schneider, S., 2011. Coevolution of retroelements and tandem zinc finger genes. *Genome Research*, 21(11), pp.1800–1812.

Tommerup, N. & Vissing, H., 1995. Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs identify putative candidate genes for developmental and malignant disorders. *Genomics*, 27(2), pp.259–264.

Vaquerizas, J.M. et al., 2009. A census of human transcription factors: function, expression and evolution. *Nature Reviews Genetics*, 10(4), pp.252–263.

Witzgall, R. et al., 1994. Genomic structure and chromosomal location of the rat gene encoding the zinc finger transcription factor Kid-1. *Genomics*, 20(2), pp.203–209.

Wolfe, S.A., Ramm, E.I. & Pabo, C.O., 2000. Combining structure-based design with phage display to create new Cys(2)His(2) zinc finger dimers. *Structure (London, England : 1993)*, 8(7), pp.739–750.

Yu, E.J. et al., 2012. SUMOylation of ZFP282 potentiates its positive effect on estrogen signaling in breast tumorigenesis. *Oncogene*, 32(35), pp.4160–4168.

Yuki, R. et al., 2015. Overexpression of Zinc-Finger Protein 777 (ZNF777) Inhibits Proliferation at Low Cell Density Through Down-Regulation of FAM129A. *Journal of Cellular Biochemistry*, 116(6), pp.954–968.

CHAPTER 4: CONCLUSIONS

In this thesis, I characterized the binding landscapes and functions of ZNF777 and Zfp777, and vertebrate roots of mammalian KRAB zinc finger family. In chapter 2, I reported the binding sites of ZNF777 in human choriocarcinoma cells. Intersecting the binding sites and the differentially expressed genes identified by siRNA knockdowns followed by transcriptome analysis (RNA-seq), we revealed that ZNF777 is involved in regulating genes related to axon guidance, a mechanism well-known to be involved in neuronal development, but also recently shown to play critical roles in placental development (Liao et al. 2010; Jongbloets & Pasterkamp 2014). The finding that ZNF777 is involved in regulation of this process is intriguing, and suggests that the expression of this transcription factor in placenta may have played a role in coopting the pathway for a mammalian-specific purpose. Since *ZNF777* is also expressed in embryonic brain (Liu et al. 2014), we sought to further investigate the functional role of this ancient gene in neuron development. In chapter 3, I showed that mouse *Zfp777* is expressed in neuronal stem cells (NSC) cultured from early mouse embryos. Using the NSC platform, I characterized the binding landscape of Zfp777 in undifferentiated NSC. To circumvent the roadblock posed by the lack of a ChIP-grade antibody for the mouse protein, I exploited the CRISPR-Cas9 technique (Ran et al. 2013; Savic et al. 2015) to tag the endogenous Zfp777 protein with FLAG epitopes. Because we are interested in comparing the two proteins, Zfp282 was also tagged using the same procedure. The ChIP-seq results revealed a novel Zfp777 binding motif that bears significant similarity to a motif predicted in *in vitro* studies (Isakova et al. 2017), and found that Zfp777 binds to promoters of genes encoding transcription factors, Wnt and TGF-beta pathways components, and proteins related to neuron development and axon guidance. Since these same functions were also found to be regulated by ZNF777 in BeWo cells (Chang et al. 2017), these results suggested that the mouse and human Zfp777 and ZNF777 proteins regulating similar genes and pathways, most classically associated with axon guidance, in diverse tissues.

Our discoveries have led to several interesting questions. Recent studies have implicated interacting roles for ZNF777 and ZNF282, suggested by the observation that they bind at many promoters in very close proximity (Imbeault et al. 2017). The two proteins are expressed in very similar patterns in both humans and mice (**Figure 3.1**). These results led us to ask many questions, such as, do ZNF777 and ZNF282 interact with each other? Do they co-regulate similar pathways? What are their interacting partners in specific cell types? To address these issues, the mouse NSC cell lines in which Zfp777 and Zfp282 proteins were successfully tagged provide an important resource for the investigation. The obvious next step would be to uncover the binding landscape of Zfp282 in mouse NSC and compare that with Zfp777 binding sites. Co-immunoprecipitation can reveal if these two founding members from the same ancient subfamily interact with each other and what would be the interplay between the interaction and their regulation roles. Previous studies have reported ZNF282 and another family member, ZNF398, interact with estrogen receptor ERα (Yeo et al. 2014; Conroy et al. 2002), and the interaction altered the regulating activity of these two TF proteins. We are interested in knowing if ZNF777 also interacts with ERα or other possible binding partners. This can be addressed by an unbiased approach, using a recently developed protocol, RIME (Rapid immunoprecipitation mass spectrometry of endogenous proteins) (Mohammed et al. 2016), which is designed specifically for studying protein complexes bound to the chromatin. The FLAG antibody was tested in this method, thus our CRISPR engineered mouse NSC cell lines that express FLAG tagged Zfp777 and Zfp282 serve as a perfect platform for this analysis, and these experiments are currently in progress.

Furthermore, questions like: what is the relationship between retroviral sequences and ZNF777 and ZNF282? ZNF282 has been shown to regulate modern-day extant human viruses (Okumura et al. 1997); does it regulate human ERVs? Is there possibly a cytoplasmic virally- related role? Also, would deletion of Zfp777, Zfp282, or both affect neurogenesis in cultured NSC or *in vitro*? These are important issues to resolve in the future. With the FLAG tagged Zfp777 and Zfp282 NSC platform I developed, more physiologically-relevant characteristics of these vertebrate roots of mammalian KRAB-ZNF can be unraveled in the near future.

References

Chang, L.-H. et al., Functions of ZNF777, a gene representing the root of the mammalian KRAB zinc finger family. *Submitted.*

Conroy, A.T. et al., 2002. A Novel Zinc Finger Transcription Factor with Two Isoforms That Are Differentially Repressed by Estrogen Receptor. *Journal of Biological Chemistry*, 277(11), pp.9326–9334.

Imbeault, M., Helleboid, P.-Y. & Trono, D., 2017. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature*, 543(7646), pp.550–554.

Isakova, A. et al., 2017. SMiLE-seq identifies binding motifs of single and dimeric transcription factors. *Nature Methods*, 14(3), pp.316–322.

Jongbloets, B.C. & Pasterkamp, R.J., 2014. Semaphorin signalling during development. *Development*, 141(17), pp.3292–3297.

Liao, W.-X. et al., 2010. Perspectives of SLIT/ROBO signaling in placental angiogenesis. *Histology and Histopathology*, 25, pp.1181–1190.

Liu, H. et al., 2014. Deep Vertebrate Roots for Mammalian Zinc Finger Transcription Factor Subfamilies. *Genome Biology and Evolution*, 6(3), pp.510–525.

Mohammed, H. et al., 2016. Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. *Nature Protocols*, 11(2), pp.316–326.

Okumura, K. et al., 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 Research*, 25(24), pp.5025–5032.

Ran, F.A. et al., 2013. Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, 8(11), pp.2281–2308.

Savic, D. et al., 2015. CETCh-seq: CRISPR epitope tagging ChIP-seq of DNA-binding proteins. *Genome Research*, 25(10), pp.1581–1589.

Yeo, S.-Y. et al., 2014. ZNF282 (Zinc finger protein 282), a novel E2F1 co-activator, promotes esophageal squamous cell carcinoma. *Oncotarget*, 5(23), pp.12260–12272.