Formation of Chimeric Genes by Copy Number Variation as a Mutational Mechanism in Schizophrenia

Caitlin Fields Rippey

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2013

Reading Committee:

Mary-Claire King, Chair

Jon McClellan

John Neumaier

Program Authorized to Offer Degree:

Genome Sciences

University of Washington

Abstract

Formation of Chimeric Genes by Copy Number Variation as a Mutational Mechanism in Schizophrenia

Caitlin Fields Rippey

Chair of the Supervisory Committee:
Professor Mary-Claire King
Departments of Medicine (Medical Genetics) and Genome Sciences

Chimeric genes are caused by structural genomic rearrangements that fuse together portions of two different genes to create a novel gene. Chimeras may differ from their parent genes in localization, regulation, or function. We screened 122 individuals with schizophrenia and 120 controls for germline rearrangements anywhere in the genome leading to chimeras. Three cases and zero controls harbored such events: fusions of *MATK* to *ZFR2*, of *DNAJA2* to *NETO2*, and of *MAP3K3* to *DDX42*. Each fusion produces a stable protein when exogenously expressed in cultured cells. Temporal expression data indicate that the parent genes of all three chimeras are expressed in the brain during development. We detected the chimeric transcripts of DNAJA2-NETO2 and MAP3K3-DDX42 in patient lymphoblasts; parent genes of the MATK-ZFR2 chimera are expressed only in the brain. Formation of chimeras involved loss of critical domains of parent genes. Subcellular localizations of DNAJA2-NETO2 and MAP3K3-DDX42 are dramatically altered compared to their parent genes. The MATK-ZFR2 chimera includes a

novel, frame-shifting splice variant of the previously uncharacterized *ZFR2* gene. In contrast with the nuclear localization of full-length ZFR2, frameshifted ZFR2 localizes preferentially to dendritic branch sites, and its chimera is predicted to be highly overexpressed during development. Germline chimeric mutations in schizophrenia provide a new model for functional interpretation of structural variation in human illness, and implicate new genes and pathways in schizophrenia pathogenesis.

Table of Contents

List of Figures

Page

List of Tables

Page

Acknowledgments

Thank you to Mary-Claire and Jack for being wonderful mentors, scientifically, clinically, and personally. Thanks also to my committee, John Neumaier, Arno Motulsky, and John Stamatoyannopoulos for excellent suggestions and support. The King lab has been a great place to spend six years, with particular props to Cailyn, Tom, Sarah, Wendy, Alex, Susan, Molly, Suleyman, and Ming.

Thanks to all the patients and families who participated in this study; I hope it is a small step towards better diagnosis and treatment in the future.

My family has provided constant love, support, and child care, and I'm infinitely grateful: thanks Mom, Dad, Josh, Piers, Sayer, mi co-madre Patti, Sandy and Kathy.

The homes I've lived in and the housemates in them have been so important in keeping me centered and happy. Thanks in particular to the folks at 1812, Alex, Mark and Liz, you beautiful creatures.

And especially, to Jay and Phineas. You've made it all possible, and you've made it all part of a rich and meaningful life. This, and all the rest, is for you.

Chapter 1: Introduction

1.1 Schizophrenia Clinical Background

Schizophrenia is a complex and debilitating psychiatric illness affecting ~1% of the population worldwide. It is defined by three principal types of symptoms: positive (e.g. hallucinations and delusions), negative (e.g. lack of affect, social withdrawal) and cognitive (e.g. disturbances of memory and attention). The positive symptoms are often the most conspicuous, but functional outcome is more closely linked to negative [1; 2] and cognitive [3; 4] symptoms. Currently available antipsychotic medications can be effective against positive symptoms but provide little relief from negative and cognitive symptoms [5]. The pharmacological properties of antipsychotic medications are highly variable, but have in common an effect on dopaminergic transmission, with variation in binding affinity for individual dopamine receptor subtypes as well as variation in off-target binding accounting for differences in efficacy and side effect profiles. Discovery of novel drug targets and medications better able to manage the entire symptom spectrum is a major goal of schizophrenia research.

Although schizophrenia generally manifests in adolescence or early adulthood, multiple converging lines of evidence suggest that schizophrenia is a neurodevelopmental disorder, with earliest impacts on brain function arising during gestation [6]. Both genetic predisposition and environmental exposure, particularly during the perinatal period, have long been known to contribute to schizophrenia risk, but the genes and pathways involved have been elusive until recently.

1.2 Schizophrenia genetics

Family and twin studies of schizophrenia have long supported a strong genetic component. The lifetime risk of schizophrenia is 5 to 20 times higher in first-degree relatives of affected probands compared to the general population, and the proband-wise concordance is 40-65% for monozygotic twins versus 0-28% for dizygotic twins is 0-28% [7; 8]. Notably, even the relatively high concordance between monozygotic twins does not approach 100%, suggesting that environmental and/or epigenetic factors are likely to play a critical role in determining the penetrance of specific mutations.

The earliest evidence for specific genetic causes of schizophrenia emerged from cytogenetic studies. Recurrent microdeletions at 22q11.2 lead to variable cardiac, palate, and facial structural abnormalities, known also as DiGeorge syndrome or velocardiofacial syndrome. The long-standing observation that approximately 25% of deletion carriers also exhibit symptoms of psychosis established 22q11.2 as one of the earliest putative schizophrenia loci [9; 10]. A balanced translocation t(1;11)(q42.1;q14.3) in a large Scottish pedigree disrupts the gene *DISC1* on chromosome 1 as well as two non-coding RNAs on chromosome 11 [11; 12]. Of 29 adult family members who carry the translocation, 18 have a major mental illness, including 7 with schizophrenia. Of 37 family members screened who do not carry the translocation, none have a major mental illness.

With the emergence of technologies to screen genome-wide for smaller and smaller variants, the evidence that rare mutations of all sizes are important in schizophrenia has continued to grow. Case-control studies of copy number variants (CNVs) in schizophrenia have firmly established that rare, gene-disrupting CNVs, particularly large CNVs, are collectively

more common in schizophrenia [13-15]. Our group reported a 3-fold enrichment of rare, large (>100 kb) genic CNVs in cases as compared to controls, with an even higher rate among individuals with juvenile onset schizophrenia [13]. Subsequent studies have confirmed and extended this finding. There is strong evidence that recurrent deletions at chr1q21.1, chr3q29, chr15q13.3, and chr2p16.3 (NRXN1) and duplications at chr16p11.2 and chr7q36.3 (VIPR2) contribute to schizophrenia (for review [16]). These recurrent events arise in regions whose genomic architecture predisposes them to rearrangement, for example by non-allelic homologous recombination of flanking segmental duplications. Family-based studies have further revealed a role for *de novo* copy number and single nucleotide variants [17-21].

The spectrum of schizophrenia mutations summarized here highlights two noteworthy features of psychiatric genetics. First, as mentioned above, many schizophrenia mutations exhibit incomplete penetrance, with known damaging mutations occasionally present in apparently healthy individuals. Second is the highly variable expressivity of individual mutations. Nearly every known schizophrenia CNV exhibits phenotypic variability among carriers, with significant genetic overlap between schizophrenia, autism, bipolar disorder, seizure disorders, intellectual disability, and other neuropsychiatric conditions (for review [16]). Evidence for a two-hit model of developmental delay, in which one rare CNV modulates expressivity of another, suggests that genetic epistasis may be one source of this variability [22]. Identifying genetic, epigenetic, and environmental modulators of penetrance and expressivity, both rare and common, may prove clinically relevant [23; 24].

This combination of features contributes to the incredible complexity of psychiatric genetics. The early 20th century psychopathologist Karl Jaspers, writing before the structure of DNA was known, predicted that “a psychic phenomenon is never a characteristic belonging to a

gene in the same sense as a physical phenomenon may be" [25]. In the current era of whole genomes and high throughput sequencing, Malhotra and Sebat conclude that "genes do not code for behavior" [16]. While devastating mutations clearly contribute to disordered mental processes, there is no one-to-one relationship between a given mutation and a given behavioral phenotype. Rather, a particular mutation more likely influences a basic cellular process or pathway, which when modulated by genetic background, stochastic variability, environmental exposures and social environment, emerges as a complex combination of ordered and disordered behaviors that can vary dramatically from individual to individual. Because neuroplasticity and adaptability are critical features of the central nervous system, it stands to reason that behavioral traits are exquisitely sensitive to the manifold influences on the brain. A major challenge now facing psychiatrists and geneticists alike is the integration of clinical, genetic and neurobiological data to establish diagnostic categories that are not defined only by behavioral symptom clusters, but more closely reflect the underlying neurobiological alterations. This may allow improved targeting of psychosocial and pharmacological treatments.

1.3 Neurodevelopmental pathways emerging from genetic studies

Functional analysis of emerging schizophrenia genes is beginning to reveal critical features and pathways, which include neurogenesis, glutamate signaling, and synaptic connectivity, as described below.

Proliferation of neuronal precursor cells in the ventricular zone of the developing human neocortex gives rise to post-mitotic neurons that then migrate to their final position in the brain. A small population of these cells persists in the adult subventricular zone and dentate gyrus,

contributing to ongoing adult neurogenesis[26]. Multiple schizophrenia genes have been shown to regulate these two critical developmental pathways, proliferation and migration. For example, knockdown of DISC1 in mice impairs precursor proliferation in adult and developing brain [27], and accelerates neuronal integration, resulting in mispositioning and aberrant morphological development of new neurons in the dentate gyrus [28]. Neuregulin-1 and its receptor ERBB4 interact to enhance neuronal precursor proliferation and regulate subsequent organization of cells in the subventricular zone [29; 30]. Glutamate is the primary excitatory neurotransmitter in the brain, and has been hypothesized based on clinical and neurobiological evidence to play a role in schizophrenia, particularly through its interactions with N-Methyl D-Aspartate receptors (NMDAR) (for recent review, see [31]). DISC1, Neuregulin-1, ERBB4 and VIPR2 have all been shown to modulate glutamate signaling [32-35]. Beyond glutamate, schizophrenia genes also impact overall synaptic connectivity, including neurite outgrowth and dendritic spine morphology [36-38].

Functional studies continue to refine the spectrum of experimental phenotypes resulting from mutation of schizophrenia genes, and provide a framework within which causality of novel candidate mutations can be evaluated.

1.4 Chimerism as a unique mutational mechanism

I have discussed the literature supporting a role for rare structural variants in schizophrenia and related neurodevelopmental disorders. The mechanisms by which these variants lead to disease are as heterogeneous as the mutations themselves, and likely include gene dosage effects - both overexpression [39] and haploinsufficiency [40; 41], unmasking of recessive alleles by a deletion on one chromosome [24], and epistatic interactions between multiple events [23;

[42]. In addition, some rearrangements have been shown to result in the formation of novel chimeric genes. Gene fusions are uniquely suited to play crucial roles in both evolution and disease because the novel genes that result can differ from parent genes in localization, regulation, and/or function [43].

Adaptive chimeric genes have been incorporated into genomes across many phyla, including a handful identified in the hominid lineage. For example, *PMCHL1* and *PMCHL2* are primate-specific chimeric genes expressed in human testis and developing brain[44]. Both are derived from the melanin-concentrating hormone (*MCH*) gene through a complex process involving exon shuffling, retrotransposition, and emergence of novel splice sites, and have undergone selection in the primate lineage. Both genes are transcribed into non-coding RNAs whose functions are not yet known[45]. The human polyubiquitination coeffector gene *UEV1* is fused to the gene *Kua* by transcriptional read through of adjacent genes. While UEV1 is a nuclear protein, Kua as well as Kua-UEV1 fusion proteins are located in the cytoplasm; this change in localization may alter UEV1's ubiquitination targets. A final example is the evolutionarily recent insertion of a retrotransposable element into intron 5 of the chromatin remodeling gene ERCC6, which codes for the protein CSB [46]. This insertion leads to translation of an in-frame CSB-transposase fusion protein that is as abundant as CSB in many human cell lines, and is evolutionarily conserved down to marmoset. The fusion protein has some functions in common with its full length counterpart, CSB. For example, both CSB and the fusion protein promote DNA repair after UV damage, although presumably by different mechanisms as the fusion protein lacks critical domains[47]. Conversely, after oxidative damage, only the full length protein promotes DNA repair, while the fusion protein is inhibitory[47]. These fusion events illustrate the unique properties of chimeric genes in the emergence of novel protein functions.

Because chimeric genes can combine functional features of their parent genes, they are able to enhance phenotypic diversity while preserving domains necessary for protein-protein interactions. Whereas other mutations proceed step by step through functional space, the dramatic nature of chimeric mutations effectively allows them to leap through functional space [43]. A recent study compared the phenotypic impacts of different classes of engineered mutations in the eleven genes of the yeast mating pathway[48]. The authors found that recombination of functional domains into chimeric genes led to a striking diversification in mating response while whole gene duplications, single domain duplications, and co-expression of distinct functional domains did not. A survey of 14 conserved chimeric genes in Drosophila melanogaster revealed that these genes are frequently at the center of selective sweeps [49]. Parent genes and chimeras had striking regulatory differences, at the level both of subcellular localization and tissue-specific expression. The authors also note that many apparently functional chimeras result from mid-domain breaks, and suggest that functional units of peptide modularity may be smaller than generally thought.

Despite their potential for novel and beneficial functions, the majority of newly arising chimeras are likely detrimental. Analysis of the formation and preservation of chimeric genes in the genome of *Drosophila melanogaster* revealed that over evolutionary time, chimeric genes are formed at a lower rate than simple duplications[50]. By modeling the evolutionary dynamics of chimera preservation versus decay, the authors conclude that the vast majority of newly emerging chimeras are eliminated by natural selection, with only1.4% of chimeras preserved. By comparison, they estimate that simple duplications are preserved at a higher rate of 4.1% , although this difference was not statistically significant Somatic formation of oncogenic fusion genes provides another example of their deleterious potential. The best characterized of these in

humans is the *Bcr-Abl* fusion in chronic myelogenous leukemia, which acts as a constitutively active form of the *Abl* kinase [51]. Subsequent studies have revealed that many cancers may in fact result from structural rearrangements leading to diverse fusion genes (for summary see [52], [53]).

Less is known about the role in disease of chimeras that arise in the germline. Our group previously reported a rare deletion in a patient with juvenile onset schizophrenia resulting in expression of a chimera of *SKP2- SLC1A3* predicted to alter the expression and function of SLC1A3, a glial glutamate transporter that regulates neurotransmitter concentration at excitatory synapses [13]. The well-documented t(1;11) translocation described previously disrupts the gene *DISC1* on chromosome 1 and the non-coding RNA *Boymaw* on chromosome 11 [12]. This leads to a decrease in *DISC1* transcript level, and also produces multiple chimeric transcripts fusing *DISC1* and *Boymaw* [54]. Recent evidence suggests that the resulting fusion proteins have altered localization compared with *DISC1* and lead to severe mitochondrial dysfunction when exogenously expressed [55]. These mutations provide anecdotal evidence that germline chimerism may play a role in schizophrenia pathogenesis. A systematic survey of copy number variation in persons with autism revealed potential candidate chimeric events but no difference in frequency between cases and controls [56].

1.5 Outline of Dissertation

Genetic studies of schizophrenia have firmly established that rare CNVs, collectively, are implicated in schizophrenia pathogenesis. Some CNVs have been shown to result in formation of chimeric genes. Chimeras are uniquely prone to producing dramatic functional changes which are often deleterious. These observations led me to ask a series of questions. To what extent do

rare CNVs result in formation of chimeric genes? Are CNVs in DNA of individuals with schizophrenia more likely to result in chimeras than CNVs in controls? And finally, what are the functional effects of chimeric genes arising in individuals with schizophrenia? My doctoral research aimed to address these questions. To do so, I focus on newly arisen chimeric genes in patients with schizophrenia and in controls. I hypothesize that brain-expressed chimeric genes in these subjects contribute to schizophrenia by acting as damaging mutations, and that we can gain insight into critical neuronal pathways by studying the genes involved.

In the first part of this study, I detected and characterized at the genomic and transcript level chimeric genes resulting from CNVs in 124 individuals with schizophrenia and 122 matched controls. Focusing on those chimeras most likely to have a detrimental impact on the brain, I asked whether there are more such events in cases than in controls. For all candidate events I tested whether chimeric RNA was present in lymphoblasts of carriers. I also compared spatial and temporal expression profiles of parent genes using publicly available transcriptome data from brain, both in brain subregions and across developmental time points. Because of the significant genetic overlap between neurodevelopmental illnesses, I then searched for these same candidate chimeras in individuals with autism spectrum disorder. Finally, I genetically characterized additional chimeras detected in our case-control series that are not likely to be involved in schizophrenia but may contribute to other phenotypes. These are potential targets of ongoing natural selection in the human lineage.

In the second part of this study, I functionally characterized fusion proteins resulting from three candidate schizophrenia chimeras. By exogenously expressing fusions and parent genes in mammalian cells, I evaluated protein stability and subcellular localization for all events. When possible, I tested fusion protein function in targeted assays of known parent protein functions.

When chimeras involved poorly characterized genes, I used bioinformatics tools including protein-protein interactions, developmental co-expression, and sequence homology to predict likely gene functions, and resulting impact of fusion events.

To my knowledge, this is the most comprehensive study to date of germline chimerism and its potential role in disease.

Chapter 2: Discovery and genetic characterization of chimeric genes in individuals with schizophrenia

2.1 Introduction

Structural rearrangements arise continuously in the human genome, and while many are benign, a subset contributes to disease, particularly neurodevelopmental illnesses such as schizophrenia [16; 57]. Functional analysis of strong candidate genes impacted by recurrent and/or *de* novo rearrangements continues to reveal new cellular functions and pathways in schizophrenia pathogenesis. Still, only a fraction of likely causal mutations have so far been discovered and characterized; efforts to demonstrate causality of individual rare events are of critical importance to expanding our understanding of this devastating illness.

As outlined in the introduction, a subset of genomic rearrangements results in the formation of novel chimeric genes. A wealth of evidence has demonstrated that gene fusions are uniquely suited to play crucial roles both in evolution and, when they arise somatically, in disease. Anecdotal reports suggest that germline chimeras may play a role in schizophrenia, but a comprehensive survey has yet to be carried out.

We hypothesize that brain-expressed chimeric genes contribute to schizophrenia by acting as damaging mutations, and that we can gain insight into critical neuronal pathways by studying the genes involved. In the first part of this project, we test this hypothesis by asking whether cases have a higher burden of rare chimeras with potential to result in expression of functional fusion proteins in the brain. We then go on to define genetic attributes of these chimeric events that may contribute to emergent functions.

2.2 Results

2.2.1 Discovery of Chimeric Genes

We screened DNA from 124 individuals with schizophrenia and 122 matched controls using array comparative genomic hybridization (aCGH) then scanned genome-wide for copy number variants (CNVs) longer than 30kb using a sliding window method described previously [58]. We eliminated events present in the Database of Genomic Variants [59] or in 122 additional unaffected controls run on the same platform. We detected 158 rare, gene-disrupting CNVs in cases and 137 in controls. From these events, I selected those predicted to delete or duplicate both the 5' end of one gene and the 3' end of another, as these CNVs should produce chimeras. I also visually inspected the aCGH data for all rare CNVs predicted to disrupt only one gene in order to find chimeric genes that were overlooked due to a mis-called breakpoint. To detect chimeras formed by inverted tandem duplications I designed breakpoint primers to test for an inverted orientation when a duplication disrupted two genes on opposite strands. Neither of these latter two approaches yielded additional putative chimeras. Breakpoints of all candidate chimeric events were obtained by PCR with diagnostic primers followed by Sanger sequencing.

Several filters were applied to each validated event: (1) breakpoints occurred in introns rather than mid-exon so that transcript structure could be predicted, (2) the predicted protein product included >5% of the protein coding sequence of at least one gene, (3) the 5' parent gene was expressed in brain, as determined by RT-PCR or scientific literature, and (4) the event was absent in 2,972 NHLBI control individuals (ref?).

Three candidate chimeric events passed through all the filters, all of which were in cases.

These are fusions of:

- Megakaryocyte-associated tyrosine kinase (*MATK)* to Zinc-finger RNA-binding protein 2 (*ZFR2*) (MATK-ZFR2) (Fig. 1A)
- DNAJ homolog subfamily A member 2 (*DNAJA2*) to Neuropilin and tolloid-like 2 (*NETO2*) (DNAJA2-NETO2) (Fig. 1B)
- Mitogen-activated protein kinase kinase kinase 3 (*MAP3K3*) to DEAD box protein 42 (*DDX42*) (MAP3K3-DDX42). (Fig. 1C)

Our previous study of 150 cases and 268 controls screened with a lower resolution aCGH platform revealed chimeras meeting passing these filters in three other cases (including the event discussed in the introduction) and none in controls [13]. Overall, matching on evaluation platform, six chimeric events meeting our criteria were detected in 172 cases versus zero in 390 controls (p=0.00023 by two-tailed Fisher's exact test). Characteristics of the candidate chimeric genes are indicated in Table 1.

Coordinates (hg19)	Gain/ Loss	Size (kb)	5'GENE (size in a.a.)	3'GENE (size in a.a.)	same frame	Length of predicted fusion proteins	Brain Expression		Race	Age at onset
							5' gene	3' gene		
chr19:3800538-3847280	Gain	46.7	MATK (508)	ZFR2 (939)	Y	0 (*5'-UTR)* + 839	Y	Y	White	20
chr16:47004235-47124878	Gain	121	DNAJA2 (412)	NETO2 (525)	N	120 + 11 novel	Y	Y	White	26
chr17:61737993-61887860	Loss	150	MAP3K3 (657)	DDX42 (938)	Y	128 + 554	Y	Y	White	16
chr5:36154948-36657630	Loss	503	SKP2 (424)	SLC1A3 (542)	Y	93 + 436	Y	Y	White	13
chr7:100,460,221-116,179,210	Gain	15668	CAV1 (178)	SLC12A9 (914)	N	65 + 23 novel	Y	Y	Black	18
chr11:33,304,314-33,583,526	Gain	279	C11orf41 (1855)	HIPK3 (1215)	Y	1068 + 1215	Y	Y	White	22

Table 1. Chimeric genes in persons with schizophrenia

No chimeras satisfying criteria laid out in text were found in 390 controls. Events in italics are from [13]. MAP3K3-DDX42 breakpoints sequenced to nearby *Alu*. "Same frame" indicates whether the second gene is predicted to be in the same frame as the first one when fused. Length of fusion proteins is written as amino acids contributed by 5' gene + amino acids contributed by 3' gene. Brain expression determined by RT-PCR of human brain cDNA or scientific literature.

2.2.2 Evaluation of transcripts of parent and chimeric gene mRNA

I next obtained lymphoblast RNA from each individual and designed primers to detect the predicted chimeric mRNA. For *DNAJA2-NETO2* and *MAP3K3-DDX42*, the full predicted chimeric transcript and additional isoforms resulting from alternate splicing of one or both parent genes were amplified and sequenced (Fig. 2A and B). *MAP3K3-DDX42* had two stable transcripts. *MAP3K3-DDX42* transcript 2 is spliced at novel sites, with the donor site in the middle of *MAP3K3* exon 3 and the acceptor site in *DDX42* exon 17. Both *MAP3K3-DDX42* fusions are in frame. *DNAJA2-NETO2* also had two stable transcripts, both out-of-frame, adding 11 (transcript 1) or 8 (transcript 2) novel frameshifted codons before a premature stop.

The chimeric *MATK-ZFR2* transcript could not be amplified from patient lymphoblasts, consistent with the expression profiles of the two parent genes, both of which were expressed only in adult and fetal brain (Fig. 2C and D). Evaluation of the parent gene *ZFR2* revealed a previously undocumented transcript, characterized by deletion of *ZFR2* exon 15 (95 base pairs), and therefore an altered reading frame. Both *ZFR2* transcripts were consistently detected in all brain regions (Fig. 2E) and may therefore be included in brain-expressed *MATK-ZFR2* chimeras.

2.2.3 Spatial and temporal brain expression patterns of parent genes

One potential pathogenic mechanism for chimeric genes is alteration of the temporal and spatial expression patterns of a parent gene as the result of fusion to new regulatory sequences. We compared temporal expression patterns of parent genes forming each chimera using RNAseq

data from human prefrontal cortex from the BrainSpan Atlas of Developing Human Brain [60] and microarray data from mouse brain [61; 62], both at multiple developmental time points (Fig. 3 and Fig. 4). Parent genes *MAP3K3* and *DDX42* both had relatively constant and even expression levels throughout development. *DNJA2* had a similar pattern, while its fusion partner *NETO2* was substantially higher in the prenatal period than in the postnatal period in humans. Interestingly, developmental expression of NETO2 was reversed in mice and appeared to rise over the course of development. A particularly striking finding was the upregulation of *MATK* in human brain between the late fetal period and early infancy while *ZFR2* expression remains at a relatively constant low level (Fig. 3C). We observed a similar pattern in developing mouse brain (Fig. 4C). Fusion of *MATK* and its upstream regulatory elements to most of the coding sequence of *ZFR2* may lead to aberrant overexpression of ZFR2 isoforms 1 and 2 at critical developmental times.

We also compared expression patterns of parent genes across brain sub-regions and in whole adult and fetal brain using publicly available data from a meta-analysis of more than 1000 published microarray experiments [63] (Fig. 5). The relative patterns were different for each parent pair. Both *DDX42* and *MAP3K3* were expressed fairly evenly across all brain regions, with expression of *DDX42* consistently higher than that of *MAP3K3* (Fig. 5A). *DNAJA2* was expressed at the same or higher levels than its 3' partner *NETO2* in all adult brain regions, but this pattern was sharply reversed in fetal brain (Fig. 5B). This result is consistent with the overall pattern we observed in human RNAseq data for these two genes (Fig. 3B), and differs from that seen in mouse brain (Fig. 4B). And finally, the observed expression for *MATK* and *ZFR2* in whole adult and whole fetal brain mirrors that observed above for these two genes (Fig. 5C). Data from sub-regions indicates that the striking difference in expression between the two may

be driven primarily by high *MATK* expression in just a few areas, notably the cingulate cortex, pons, caudate nucleus, and globus pallidus.

Temporal and spatial expression of parent genes in the brain may partially determine functional consequences of these fusion events.

2.2.4 Putative schizophrenia chimeras in children with autism spectrum disorder

Because of the significant genetic overlap between schizophrenia and other neurodevelopmental illnesses, I sought to determine if these or similar events were also present in exomes from 411 families identified via a proband with simplex autism spectrum disorder (ASD) (Niklas Krumm, personal communication). A duplication of most of *ZFR2* was detected in an individual with ASD, inherited from an apparently healthy parent and also transmitted to a healthy sibling (Fig. 8A). I note however that the *MATK* exon duplicated in the chimeric event in our series is not targeted by exome capture and as such the chimera would not be detected in the autism family. Furthermore, the breakpoints of the duplication in our series occur at 43-base pair regions of microhomology, and as a result this region may be moderately prone to recurrent rearrangements. Unfortunately additional DNA was not available from the child with autism to determine if the event is actually a *MATK-ZFR2* chimera. Additionally, an intragenic *ZFR2* deletion was present in one autism proband, also inherited from a healthy parent and transmitted to a healthy sibling (Fig. 8B). Neither of the other two putative schizophrenia-associated events (*MAP3K3-DDX42* or *DNAJA2-NETO2*) was found in probands with ASD.

2.2.5 Additional chimeric genes in our study

Events that did not pass through our filters are unlikely to be schizophrenia risk genes. Some may nonetheless form chimeric genes with novel functions, and are potentially interesting from the perspective of human diversity and ongoing evolution. Characteristics of validated rare chimeras eliminated by filters described previously are enumerated in Table 2.

Coordinates (hg19)	Gain/ Loss	Size (kb)	5'GENE (size in a.a.)	3'GENE (size in a.a.)	same frame	Length of predicted fusion proteins	Expression in brain		Frequency in controls	Race	Age at onset
							5' gene	3' gene			
Cases											
chr14:69,921,835-69,979,453	Gain	57.6	PLEKHD1 (506)	SLC39A9 (307)	Y	186 + 121	Y	Y	3/2972	White	16
Controls											
chr9:33,140,885-33,262,114	Gain	121	B4GALT1 (398)	BAG1 (345)	Y	139 + 153	Y	Y	4/2972	Black	NA
chr17:9,981,338-10,416,002	Gain	140	GAS7 (476)	MYH1 (1939)	NA	lacks splice acceptor site	Y	N	0/2972	White	NA
chr5:59,639,346-59,986,710	Loss	347	DEPDC1B (529)	PDE4B (644)	N	15 + 5 novel	N	Y	0/2972	White	NA
chr3:101,005,281-101,108,239	Gain	103	IMPG2 (1241)	SENP7 (1050)	N	167 + 26 novel	N	Y	0/2972	White	NA

Table 2. Additional chimeras in cases and controls
"Same frame" indicates whether the second gene is predicted to be in the same frame as the first one when fused. Length of fusion proteins is written as amino acids contributed by 5' gene + amino acids contributed by 3' gene. Brain expression determined by RT-PCR of human brain cDNA (PLEKHD1, SLC39A9, B4GALT1, BAG1, GAS7) or scientific literature (remaining genes). Frequency in controls is based on exome data from 2972 NHLBI controls.

Of these, we believe *GAS7-MYH1* and *DEPDC1B-PDE4B* are unlikely to be functional. Breakpoints of *GAS7-MYH1* fall in intron 1 of *GAS7*, which is more than 100kb in length, and in the middle of an exon of *MYH1*. As a result the predicted chimeric pre-mRNA lacks a splice acceptor site, and is unlikely to be spliced appropriately. Indeed, we were unable to detect either parent gene or the chimeric transcript in lymphoblasts. *DEPDC1B-PDE4B* is an out of frame fusion that would include only 15 of the 529 amino acids of *DEPDC1B*, and none of *PDE4B*.

Even if the resulting mRNA escaped nonsense-mediated decay (NMD), the resulting protein lacks substantial coding sequence and is unlikely to be functional. *IMPG2-SENP7* is also out of frame, but includes >10% of the coding sequence of *IMPG2* and may therefore be functional, if not targeted for NMD. *IMPG2* expression is limited to the retina [64] and fusion proteins are unlikely to be expressed in other tissues.

The two remaining chimeras are more likely to be functional. The in-frame *PLEKHD1-SLC39A9* fusion (Fig. 9A) was present in one case and in 3/2972 controls. It was also observed in the exome of one ASD proband, inherited from a healthy parent and also transmitted to a healthy sibling. We detected two chimeric transcripts in lymphoblasts of the individual carrying this event in our series, one encoding each alternate 3'-UTR of *SLC39A9* (Fig. 9B). Furthermore, when exogenously expressed in Human Embryonic Kidney 293 (HEK293) cells, both transcripts were translated into stable proteins (Fig 9C and D). PLEKHD1 and both fusion proteins appeared as appropriately sized bands by western blotting. SLC39A9 is predicted to have a molecular weight of 32.4 kD, but runs close to the 15 kD marker. (Fig. 9D) suggesting that the protein product may be cleaved. In other replicates of this experiment (not shown) no band was detectable in lysates from cells transfected with *SLC39A9*. *SLC39A9* codes for the zinc transporter ZIP9, which was recently shown to enhance AKT and ERK phosphorylation in immune cells by regulating intracellular zinc concentration [65]. *PLEKHD1* is an uncharacterized gene that has an N-terminal pleckstrin homology (PH) domain. PH domains exist in a wide range of proteins (including, coincidentally, AKT) and are generally involved in intracellular signaling by binding to phosphatidylinositol lipids and recruitment of target proteins to the cell membrane [66]. Both genes are expressed in the brain as well as in many other tissues, and the fusion proteins likely are as well. The fusions contain both the PH domain of PLEKHD1 and the Zinc

transporter domain of ZIP9. Although this fusion gene is unlikely to be involved in schizophrenia, it may still have a newly evolved function in the brain or in other tissues.

The in-frame B4GALT1-BAG1 fusion (fig. 10A) was present in one control in our series and in 4/2972 additional controls. One control also carried the reciprocal deletion. As with the *PLEKHD1-SLC39A9* chimera, the duplication was observed in one ASD proband and their healthy parent and sibling. Chimeric transcripts were present in lymphoblasts of the control in our series. *B4GALT1* codes for a metabolic enzyme with two main isoforms. One localizes to the Golgi apparatus and catalyzes lactose biosynthesis in the mammary gland [67; 68]. The other localizes to the cell surface and is involved in a variety of cell-cell and cell-matrix interactions, particularly during development. Recessive mutations in *B4GALT1* cause congenital disorder of glycosylation type 2D (CDG2D). Disorders of glycosylation caused by mutations in other genes include profound psychomotor and mental retardation, but individuals with *B4GALT1* glycosylation disorders generally have no neurological symptoms, consistent with low B4GALT1 expression in brain [69; 70]. The 3' gene of this chimera, BAG1, It has many alternately translated isoforms, including a short isoform that includes the same protein domains that are included in the fusion gene. It interacts with the oncogene BCL2 and enhances its anti-apoptotic effect and as such is found to be overexpressed and otherwise dysregulated in many cancers [71-73]. In the brain, it regulates neuronal differentiation [74] and promotes axonal outgrowth [75]. Multiple studies have now demonstrated that overexpression of BAG1 in the brain is neuroprotective in diverse models of brain injury and disease [76-78]. Because this fusion results from a duplication, it likely results in overexpression of BAG1 protein domains, particularly in tissues where B4GALT1 is expressed (Fig. 10B). This may include low-level overexpression in the brain, in

which case it may be neuroprotective. In other tissues it may modulate cancer risk or cancer progression. Its presence at low frequency in the population makes it an interesting target for further study.

2.3 Discussion

Array CGH data from the present study revealed that a small subset of CNVs (~3% in this series) produces chimeric genes. Furthermore, combining these data with our previous study, I found an increased burden of highly rare, likely brain-expressed chimeric genes in cases, with 6 in 172 cases compared with 0 in 390 controls. This finding substantially expands on previous anecdotal evidence supporting a role for individual chimeric genes in schizophrenia. Because each of these events arises in only a single affected individual in our series, these findings do not demonstrate causality for any one chimera. They do suggest, however, that rare chimeric genes as a group contribute to schizophrenia. This finding is consistent with multiple lines of evidence suggesting that newly emerging chimeric genes are uniquely likely to be deleterious.

A previous study of similar events in individuals with autism did not show a similarly increased burden of chimeric events among cases [56]. This may reflect a fundamental difference between the two illnesses, or may result from differences in event calling and filtering. That study reported a higher average number of chimeric genes per individual, both cases and controls, suggesting a difference in rates of false positive or false negative calls between our event calling algorithms. The filters we applied to prioritize likely functional chimeras were not used in that study and may have further narrowed down candidate events; we note however that even without those filters we see more chimeric events in cases than in controls.

Of four case-specific chimeras tested (three from the present study) three were expressed in patient lymphoblasts; the fourth is expected to be brain-specific based on parent gene expression patterns. These events – particularly *MATK-ZFR2* and *DNAJA2-NETO2* - fuse genes with strikingly divergent temporal and/or spatial expression patterns. Regulatory dysregulation is

likely to be a critical component of functional impact of chimeric genes in the context of the whole organism.

I also identified two likely functional chimeras present at low frequency in controls. These are unlikely to play a role in schizophrenia, but may code for proteins with novel functions that are involved in other illnesses or in the emergence of phenotypic diversity.

2.4 Materials and Methods

Subjects

Our case series is 124 patients with schizophrenia or schizoaffective disorder whose diagnostic status was assessed by chart review and confirmed by diagnostic interview, using the Structured Clinical Interview for DSM-IV disorders (SCID) for adults, or the KID-SCID for youth. Subjects consented to a blood draw. DNA was extracted from whole blood for array CGH, and lymphoblastoid cell lines were also generated in our lab for each subject. These cell lines were used for RNA and protein analyses.244 unrelated, race-matched controls were drawn from 4000 individuals older than 30 years, without mental illness, from NIMH distribution 5 (http://zork.wustl.edu/nimh/home/index.html). These individuals have been screened against a wide range of psychiatric disorders, including schizophrenia, bipolar disorder, and major depression.

Demographic characteristics

By self-report, ethnicities of 124 cases were: 87 white (70.2%), 20 black (16.1%), 10 mixed-ancestry (8.1%), 4 unknown (3.2%), 1 Native American (0.8%), 1 Chinese (0.8%), and 1 Filipino (0.8%). 244 NIMH controls were: 196 white (80.3%), and 48 black (19.7%). Average age of onset of schizophrenia in our series was 19 years (standard deviation =5.0 years; range=9 - 35 years).

Array Comparative Genomic Hybridization

DNA from 124 individuals with schizophrenia and 244 controls was hybridized to Nimblegen HD2 (Roche Nimblegen, Madison, WI, USA) microarray slides against a well-characterized reference subject, SKN1. The HD2 array features 2.1 million probes spaced at an average of 1.2 kb across the genome. aCGH data was normalized and CNVs were called using a sliding-window algorithm with a 10-probe (~10 kb) minimum size threshold. Controls were split into two groups: half were used for the identification of normal variation and half were used to test for rare chimeric genes. Rare CNVs were defined as having <30% overlap with a CNV identified in our control sample or in the Database of Genomic Variants [59].

Exomes

To identify events in autism cases and in controls, we used exome sequencing data from 2,972 individuals from the NHLBI Exome Sequencing project (approximately 2/3 white, 1/3 black) and 411 families with autism spectrum disorder proband, mother, father, and unaffected sibling) from the Simons Simplex Collection. CNVs were derived using CoNIFER [79], and were screened for events involving either gene of each chimera.

Breakpoint validation

In order to determine genomic breakpoints, PCR primers were designed from approximately 1kb boundaries surrounding each CNV based on HD2 probe locations on the SignalMap track on the

Genome Browser. Long-range PCR amplification was performed on genomic DNA from the appropriate subject, using Takara LA Taq (TaKaRa Bio, Otsu, Shiga, Japan) as described in the manufacturers protocol. PCR products were electrophoresed on 0.8% agarose gels, extracted, and sequenced with BigDyeV3.1 (Life Technologies, Carlsbad CA) chemistry on an ABI 3130XL capillary instrument. Sequencing reactions were primed with breakpoint primers and additional sequencing primers as necessary. Experimentally derived sequences were aligned to the UCSC Genome Browser and exact genomic breakpoints determined. Primers sequences and PCR conditions are in included in Appendix A.

Transcript analysis

First strand cDNA synthesis of patient lymphoblast cell line RNA or of total RNA from different human tissues (Clontech Laboratories Inc, Mountain View CA) was performed with SuperScript III (Life Technologies, Carlsbad CA) primed with random hexamers and/or gene-specific primers. Chimeric transcripts or parent genes were amplified with PCR primers targeted to flanking exons. Products were purified after gel electrophoresis and Sanger sequenced.

Expression levels of chimeric genes in different human tissues and brain sub-regions

For analysis of parent gene expression across developmental time points in human brain, we used RNAseq data from the BrainSpan Atlas of Developing Human Brain (http://www.brainspan.org). Normalized gene expression levels for 26 different brain tissues in 10 different developmental periods were obtained from the BrainSpan RNAseq dataset v3 of

developing human brain. Tissue qualification, processing and dissection, and experimental and bioinformatics procedures are described in the technical white paper of the dataset (http://help.brain-map.org/display/devhumanbrain/Documentation). Normalized values of expression of parent genes were obtained from 149 frontal cortex samples at different developmental age. Developmental ages were grouped into 10 developmental periods: early fetal (8-12 weeks gestation), early-mid fetal (13-18 weeks gestation), late-mid fetal (19-24 weeks gestation), late fetal (25-38 weeks gestation), early infancy (birth-5 months), late infancy (6-18 months), early childhood (19 months-5 years), late childhood (6-11 years), adolescence (12-19 years), and adulthood (20-60+ years). Median values for each gene in each period were calculated and plotted.

Analysis of developmental brain expression in mouse was based on two published microarray datasets using the Affymetrix Mouse Genome 430 2.0 Array. Dataset GSE8091 [62] was used for ages: E9.5 (6 biological replicates), E11.5 (4 biological replicates), and E13.5 (6 biological replicates). Dataset GSE16675 [61] was used for ages: E14.5, P1, P7, and P90 (9 biological replicates each from 3 mouse strains). The dataset also included liver samples which were not included in the analysis. GeneSpring 12.1 Biological Signficance module was used for normalization. A total of 52 CEL files were used as input. 45,101 probes were matched. Developmental groups were defined for each CEL file. 41,183/45,101 probes were filtered with 20th percentile cut-off. Medians of filtered, normalized probe values were used to plot expression figures.

For analysis of expression patterns in different human brain regions, we used a published dataset in which expression levels of genes in 128 human tissues were curated from 1,068 gene expression microarrays [63]. Mean expression levels of each parent gene in available brain regions

were extracted from the dataset and plotted.

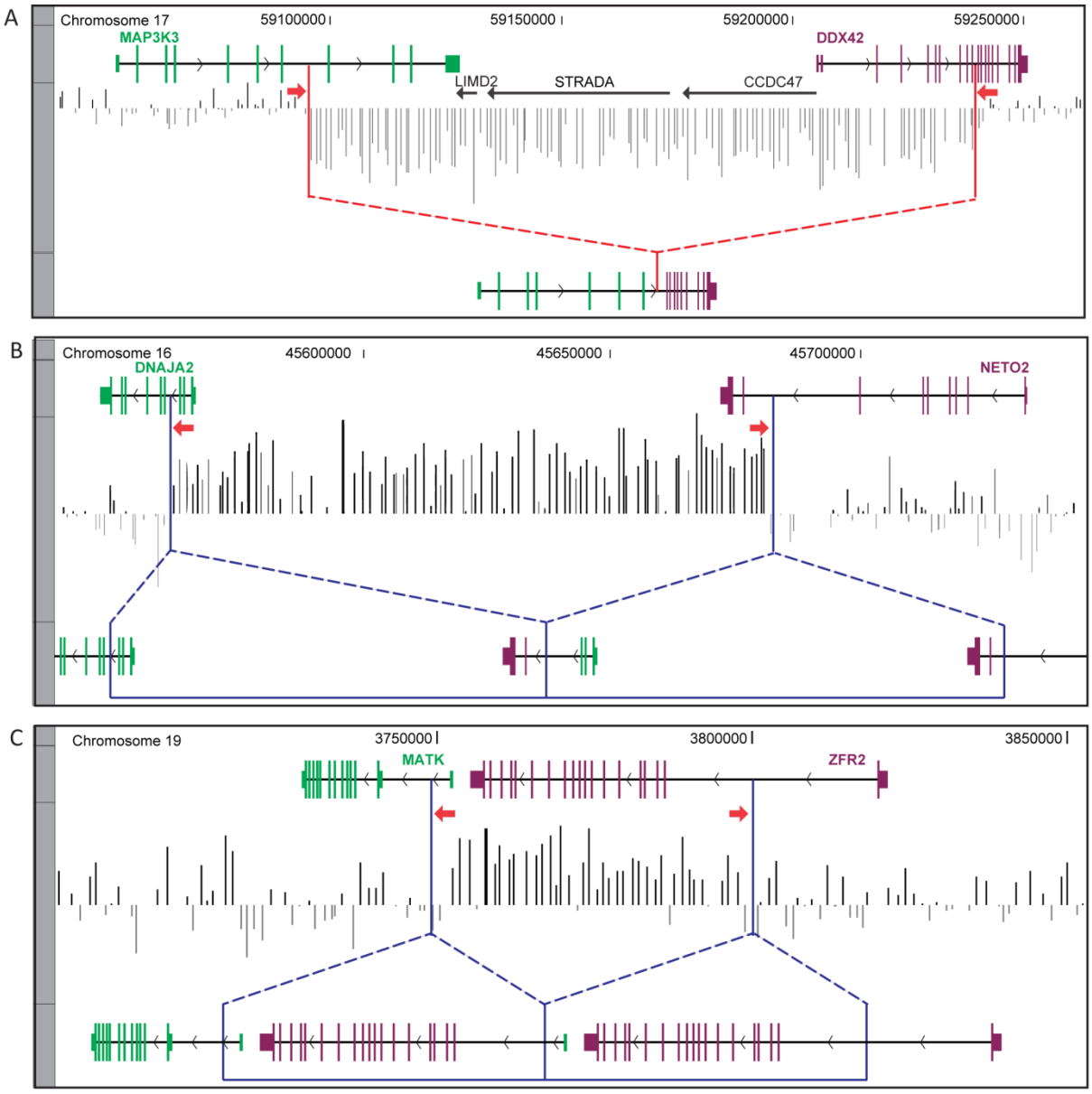

Figure 1. Array Comparative Genomic Hybridization data showing formation of chimeric genes in individuals with schizophrenia

Deletion (red dashed lines) and duplications (blue dashed lines) result in the formation of MAP3K3-DDX42 (A), DNAJA2-NETO2 (B), and MATK-ZFR2 (C) chimeric genes. Each box shows: genomic coordinates, impacted genes with direction of transcription, regions targeted by PCR primers for breakpoint validation (red arrows), histograms of aCGH hybridization Z-scores, and resulting chimeric gene. 5' parent genes are green, 3' parent genes are purple.

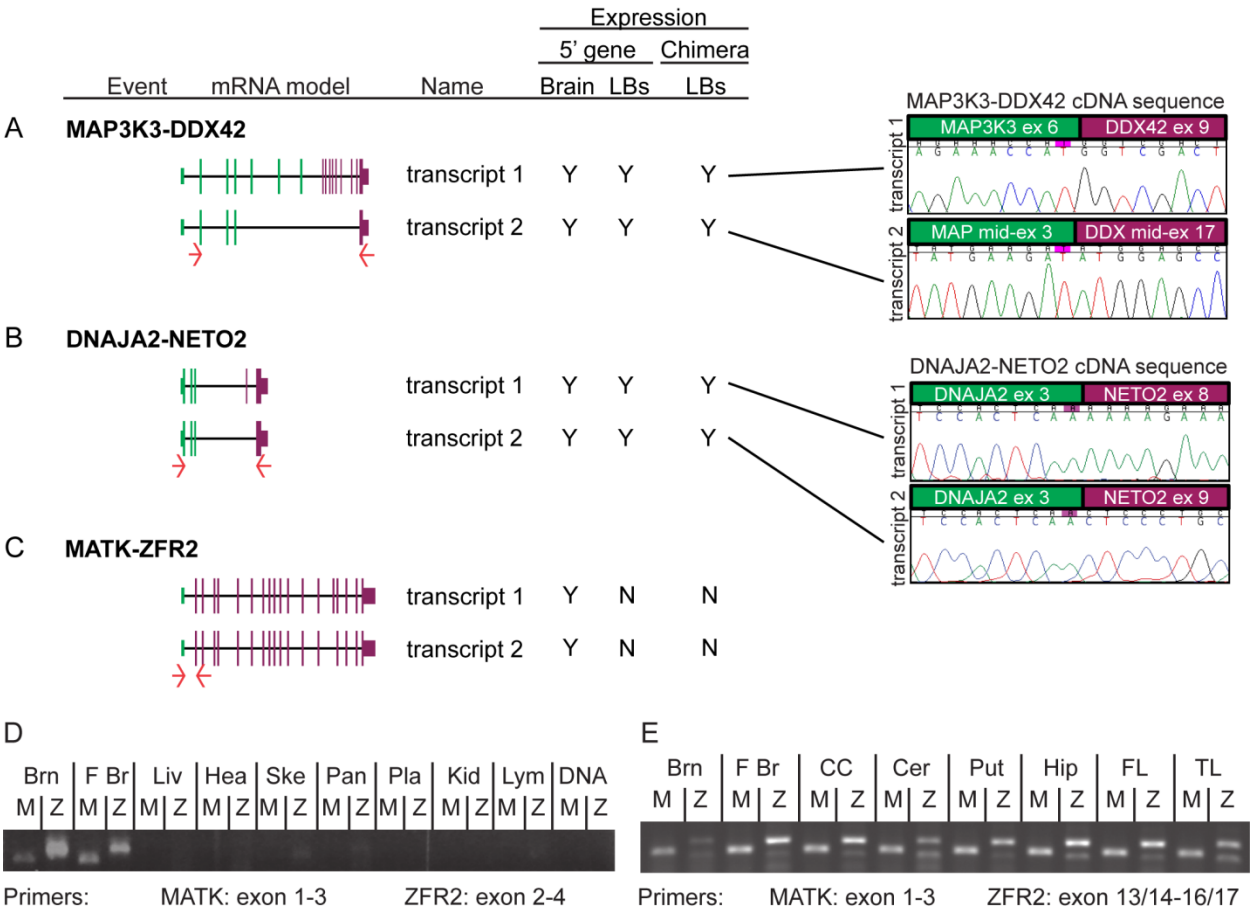

Figure 2. Characterization of parent and chimeric gene mRNA

MAP3K3-DDX42 (A) and *DNAJA2-NETO2* (B) chimeric transcripts (two each) were detected by RT-PCR in patient lymphoblasts. Red arrows indicate exons targeted by forward and reverse primers. Sanger sequence traces show chimeric junctions. Expression of 5' parent genes was also tested by RT-PCR in human brain cDNA and in patient lymphoblasts. *MATK-ZFR2* chimeric transcripts were not present in patient lymphoblasts (C). RT-PCR of cDNA from human tissue panel showed expression of parent genes *MATK* (M) and *ZFR2* (Z) only in brain and fetal brain (D). RT-PCR of cDNA from human brain subregions showed expression of *MATK* and *ZFR2* throughout the brain (E). We also detected a *ZFR2* splice variant characterized by skipping of exon 15, also expressed throughout the brain (E).

Tissue legend: Brn = Adult brain; F Br = Fetal Brain; Liv = Liver; Hea = Heart; Ske = Skeletal muscle; Pan = Pancreas; Pla = Placenta; Kid = Kidney; Lym = Lymphoblasts; DNA = genomic DNA (negative control); CC = Cerebral cortex; Cer = Cerebellum; Put = Putamen; Hip = Hippocampus; FL = Frontal lobe; TL = Temporal lobe.

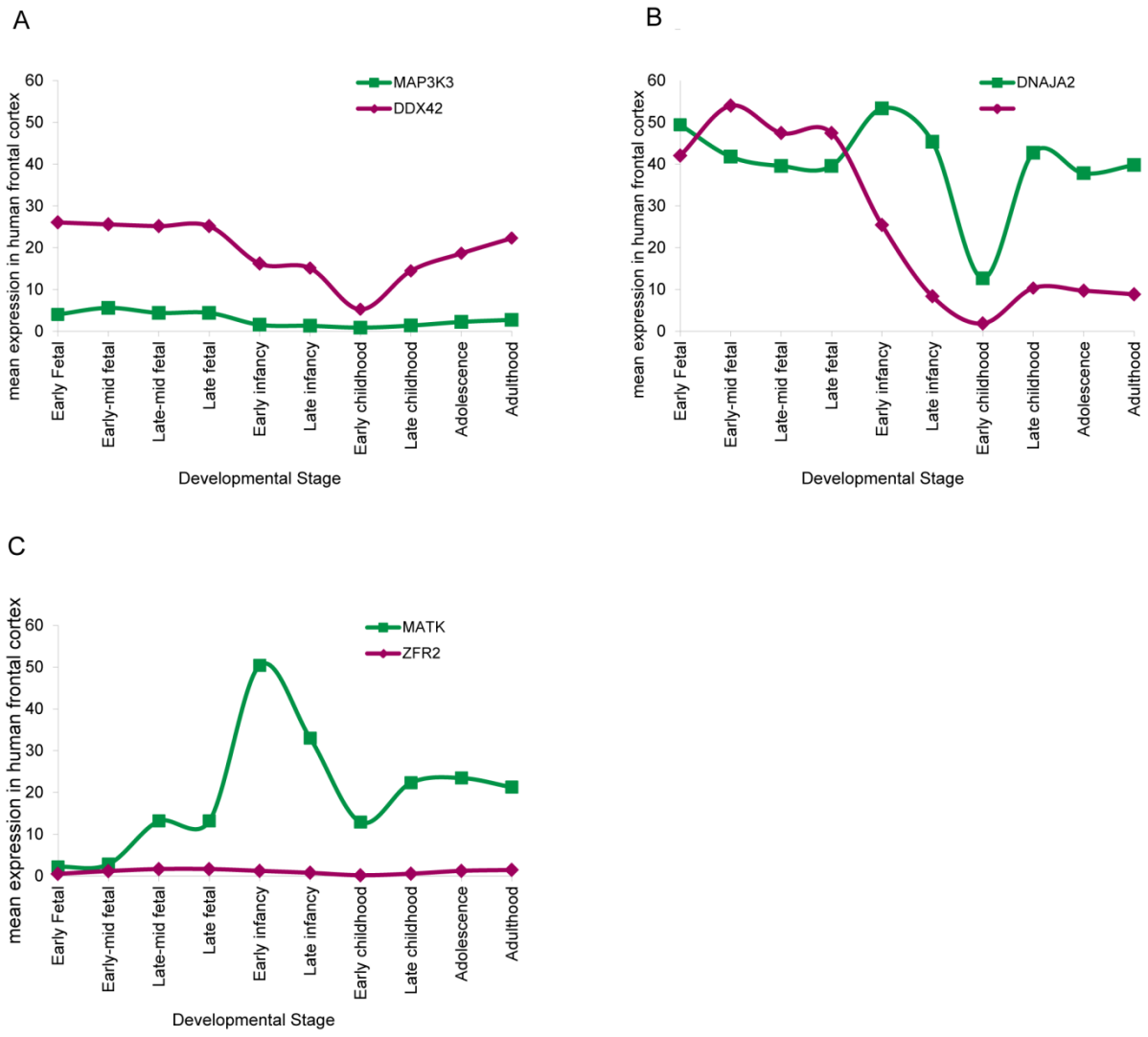

Figure 3. Normalized expression of chimera parent genes in developing human brain

Normalized expression values for each chimeric gene pair averaged from 149 human frontal cortex samples taken at different developmental ages. Data were obtained from the Allen Brain Atlas BrainSpan RNAseq dataset.

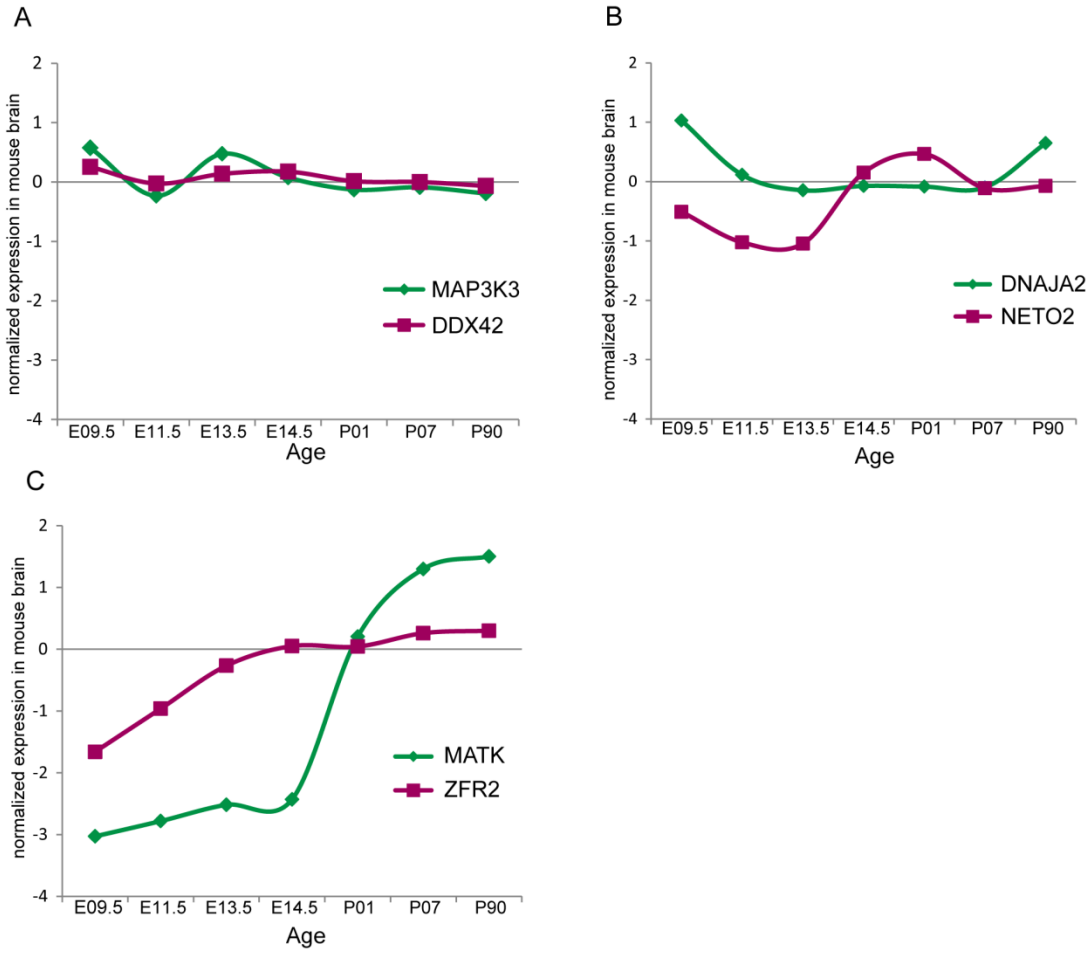

Figure 4. Normalized expression of chimera parent genes in developing mouse brain

Normalized expression values for each chimeric gene pair in mouse brain at multiple developmental timepoints. Data were obtained from two publicly available microarray datasets.

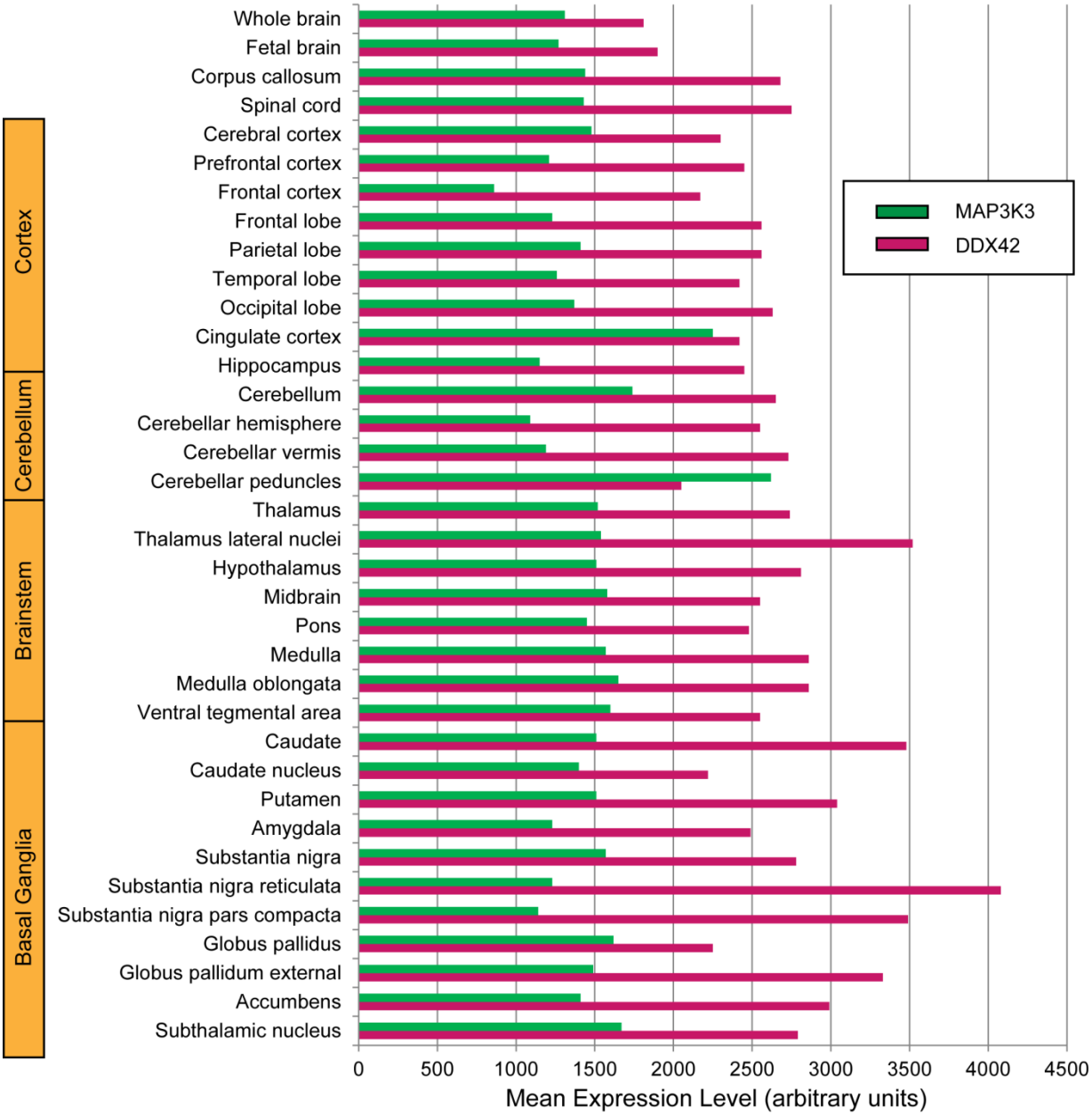

Figure 5. Expression of parent genes *MAP3K3* and *DDX42* in human brain subregions

Normalized mean expression levels are shown for *MAP3K3* and *DDX42* in human brain subregions grouped by anatomical location. Data were obtained from a publicly available dataset of microarray data curated from 1068 published studies.

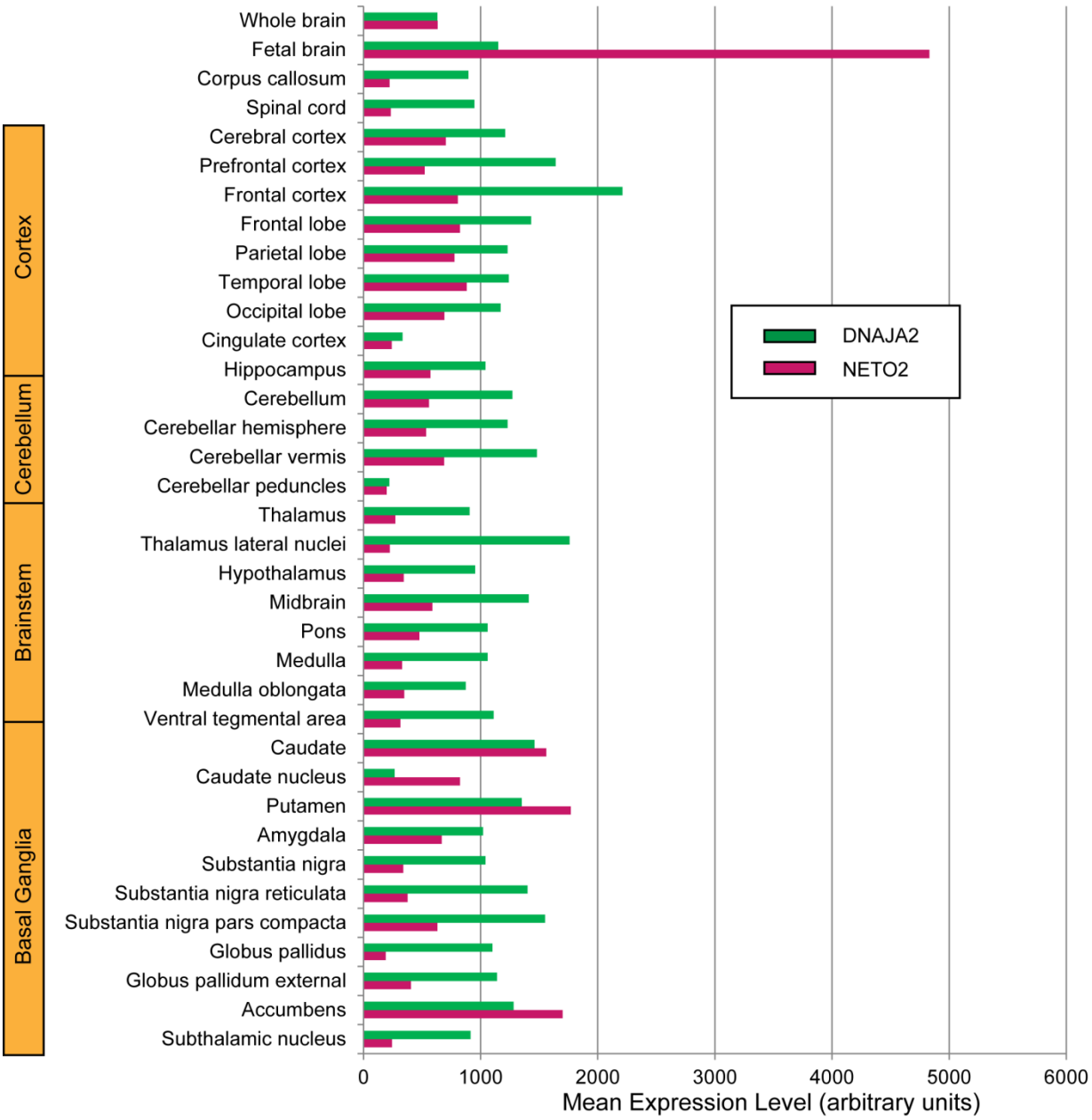

Figure 6. Expression of parent genes *DNAJA2* and *NETO2* in human brain subregions

Normalized mean expression levels are shown for *DNAJA2* and *NETO2* in human brain subregions grouped by anatomical location. Data were obtained from a publicly available dataset of microarray data curated from 1068 published studies.

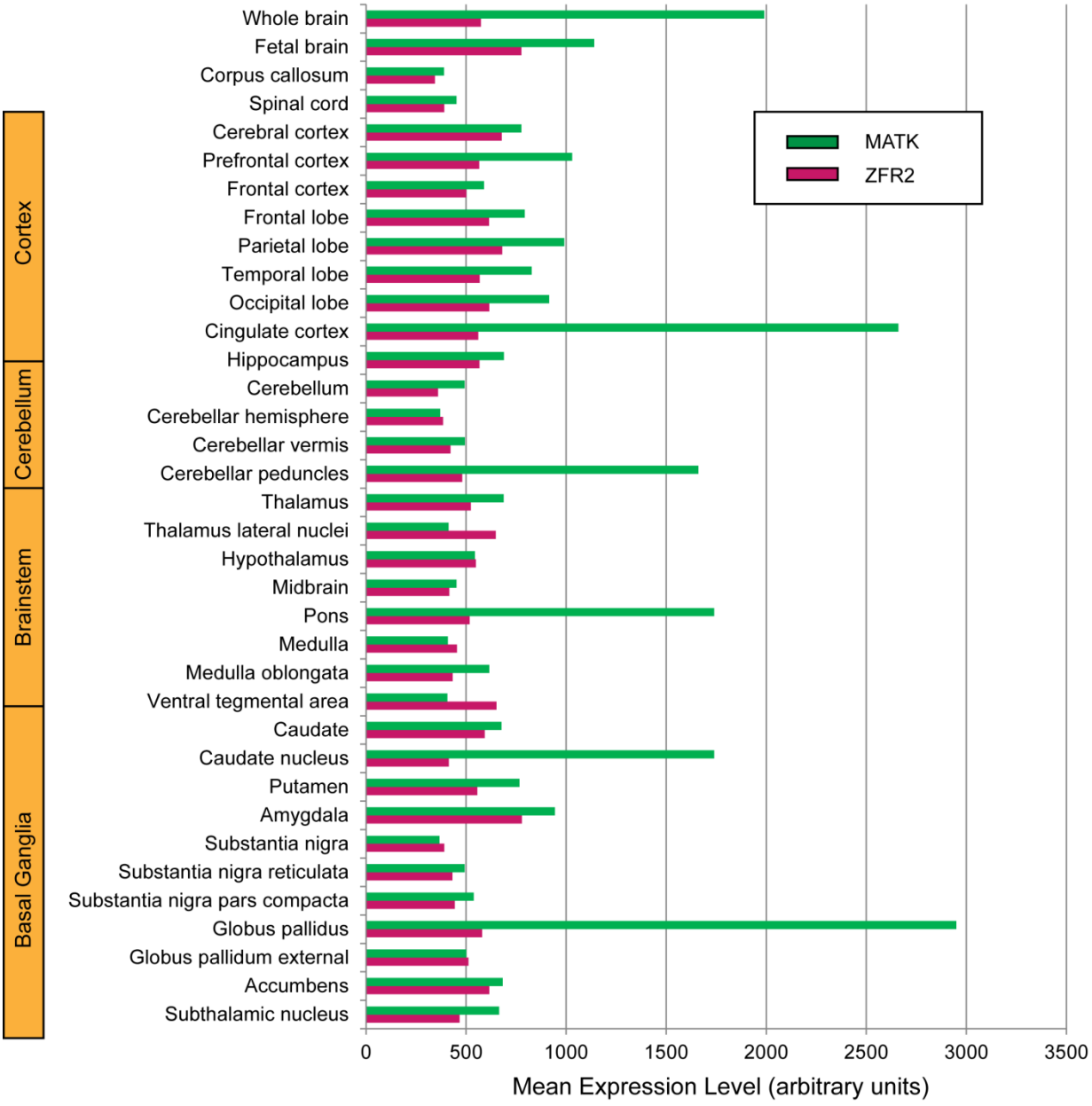

Figure 7. Expression of parent genes *MATK* and *ZFR2* in human brain subregions

Normalized mean expression levels are shown for *MATK* and *ZFR2* in human brain subregions grouped by anatomical location. Data were obtained from a publicly available dataset of microarray data curated from 1068 published studies.

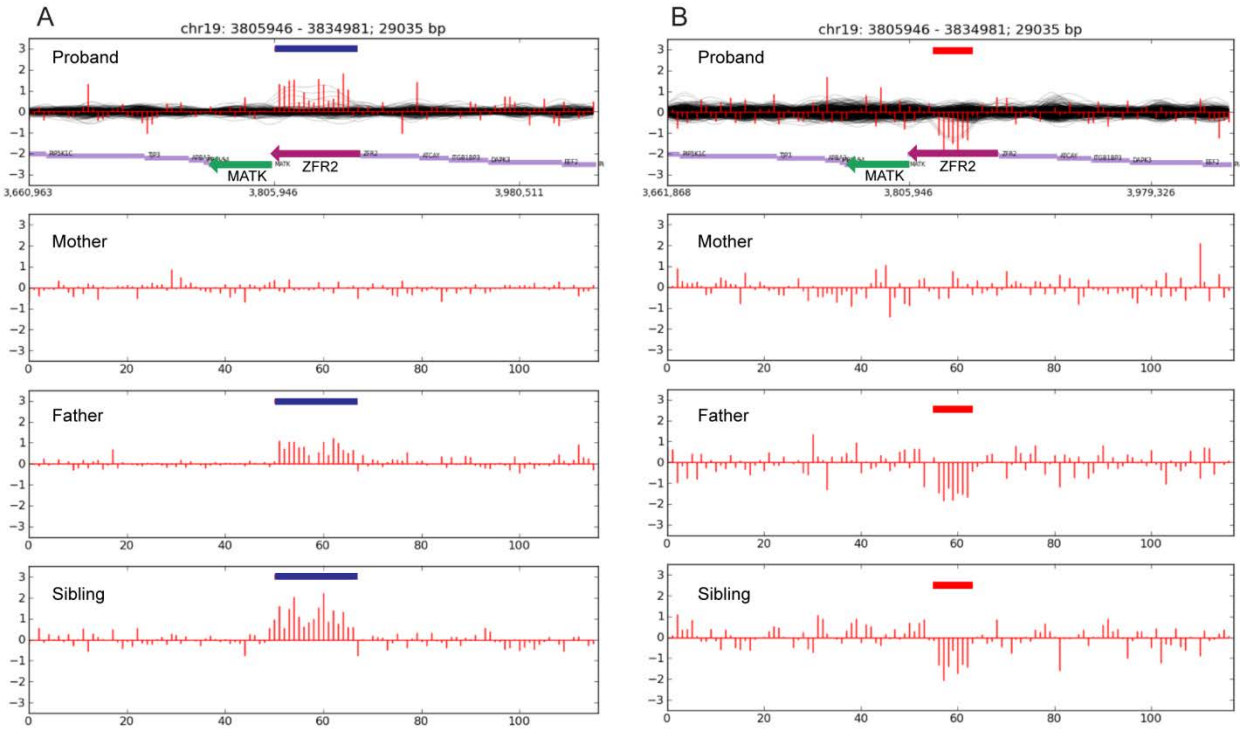

Figure 8. Copy number variants in *ZFR2* in simplex autism families

Rare CNVs derived from exome data in simplex autism quads (proband, unaffected sibling, mother and father) using CoNIFER. x-axis shows genomic coordinates, y-axis shows normalized Z-scores derived from exome read-depth. Targeted genes and transcription direction are indicated, as well as predicted location of duplication (blue bar) and deletion (red bar). We detected one potential MATK-ZFR2 chimera (A) and one intragenic deletion in ZFR2 (B). Both CNVs were inherited from apparently healthy father and transmitted to unaffected sibling.

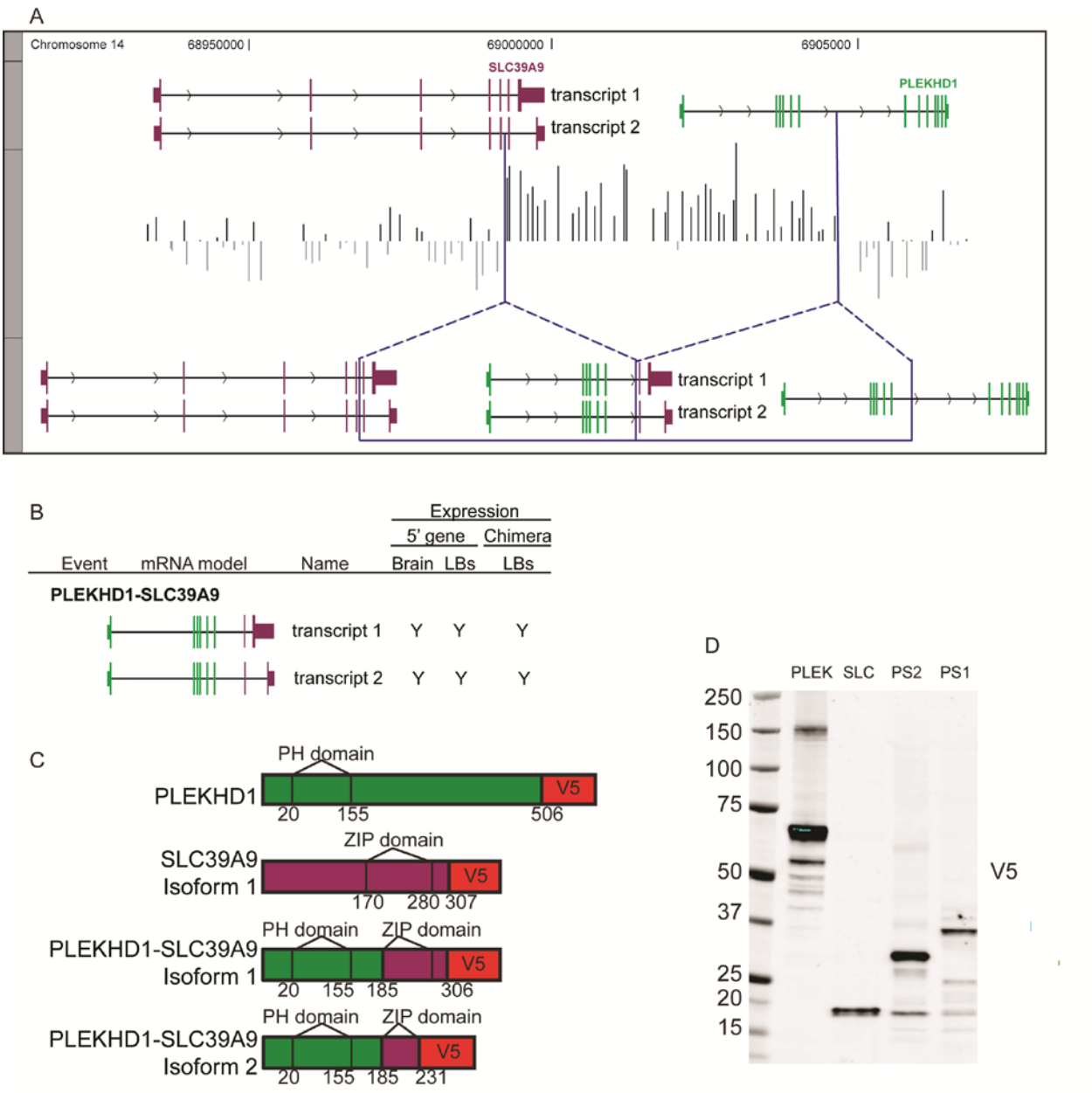

Figure 9. Features of *PLEKHD1-SLC39A9* chimeric gene present in one case and multiple controls

aCGH data shows duplication (blue dashed lines) resulting in the formation of *PLEKHD1-SLC39A9* chimeric gene. Indicated on the figure: genomic coordinates, impacted genes with direction of transcription (including two *SLC39A9* transcripts involving distinct 3'-UTRs), histograms of aCGH hybridization Z-scores, and chimeric genes resulting from *PLEKHD1* fusion with each *SLC39A9* transcript (A). Both transcripts were detected in patient lymphoblast cDNA by RT-PCR. Expression of 5' parent gene PLEKHD1 was also tested by RT-PCR in lymphoblast and brain cDNA. (B) C-terminal V5 epitope-tagged mammalian expression constructs for parent genes and two chimeric isoforms. Known functional domains are indicated. (C) Each construct was transfected into HEK293 cells, and protein expression ascertained by western blots stained for V5 and actin loading control (D) Ladder is shown along with molecular weights in kD.

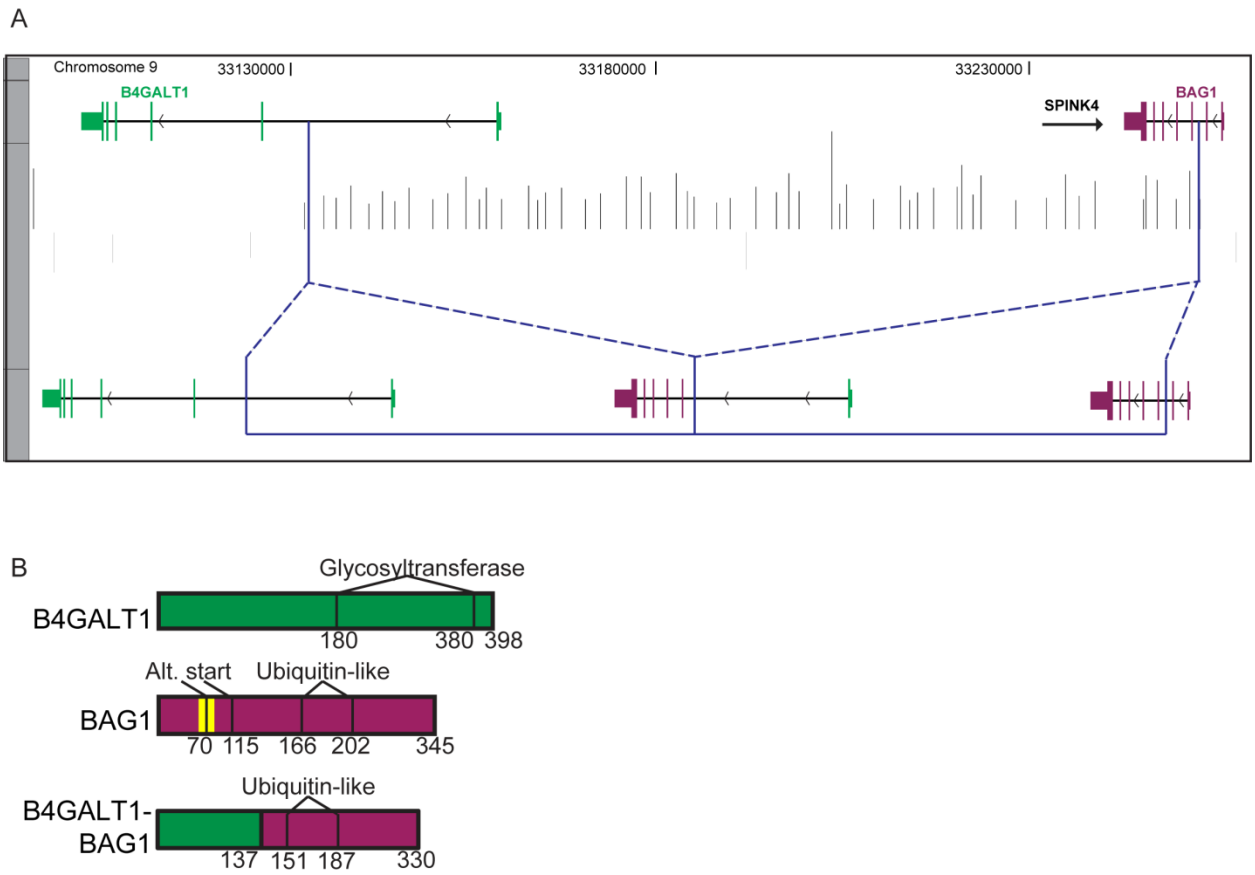

Figure 10. Features of *B4GALT1-BAG1* chimeric gene present in multiple controls

aCGH data shows duplication (blue dashed lines) resulting in the formation of *B4GALT1-BAG1* chimeric gene. Indicated on the figure: genomic coordinates, impacted genes with direction of transcription, histograms of aCGH hybridization Z-scores, and resulting chimeric gene (A). Known protein domains of parent genes are indicated along with domains present in predicted fusion protein. (B)

Chapter 3: Functional evaluation of fusion proteins

3.1 Introduction

In Chapter 2 I presented genomic evidence that rare germline chimeras, as a group, contribute to schizophrenia. This may be because chimeras are especially prone to producing drastic functional changes, by such mechanisms as shuffling of protein functional domains, recombination of regulatory regions and alterations in subcellular localization signals. In this chapter, I functionally characterize at the molecular and cellular levels three chimeric genes unique to individuals with schizophrenia in our series. To accomplish this, I generated epitope-tagged expression vectors for each chimera and its parent genes. Then, in mammalian cells, I compared expression levels and subcellular localization for each chimera-parent set. I also compared chimera and parent activity in gene-specific functional assays when gene functions were known. When possible, I focused functional analysis on pathways and cellular mechanisms potentially pertinent to neurodevelopmental processes.

For some genes involved in these chimeras, little or no functional information was available. For these, I used bioinformatic tools including analyses of sequence homology, protein structure, developmental co-regulation and protein-protein interactions to establish potential functional roles.

My goal with these studies is twofold. First, to determine whether these newly arisen germline chimeras share the often dramatic functional features of somatic chimeras and ancient chimeras targeted by selection. Second, to identify genes and pathways potentially relevant to schizophrenia.

3.2 Results

For those events with detectable transcript in lymphoblasts, I tested whether fusion proteins were present in patient lymphoblasts by western blotting. No bands of sizes other than those observed in control samples were observed for the MAP3K3-DDX42 fusion, and only a very faint band of the predicted fusion size was observed for the DNAJA2-NETO2 fusion. This suggests that in lymphoblasts, these fusion proteins are present at very low levels or not at all. However, their expression in brain may be different, and they may be able to exert an effect even at low levels of expression. Therefore, to test whether the predicted proteins are able to be stably synthesized, and to characterize their subcellular localization and function, I cloned chimeric transcripts and full-length parent genes into epitope-tagged mammalian expression vectors.

3.2.1 Functional characterization of MAP3K3-DDX42 fusion proteins

The *MAP3K3-DDX42* fusion is unique in our series in several critical ways. First, it is the only chimera resulting from a deletion, so the resulting fusion protein is acting against a backdrop of only a single functional copy of its parent genes. Second, it is the only event to impact more genes than the two involved in the fusion; in this case, three interstitial genes are wholly deleted (*LIMD2*, *STRADA*, *CCDC47*), one of which – *STRADA* - is known to regulate dendritic and axonal outgrowth [80; 81]. And finally, the chimera contains substantial coding sequence from both constituent genes, such that the fusion protein could impact two separate pathways. For the present study, I focused on function of fusion proteins relative to their 5' parent gene, *MAP3K3*.

MAP3K3-DDX42 fusion proteins were 75.1 kD and 12.9 kD (Fig. 11A and B). MAP3K3 is 70.9 kD and DDX42 is 103.0 kD. In transfected HEK293 cells, localizations of both fusion isoforms differed from those of their parent proteins (Fig. 12). MAP3K3 was predominantly cytoplasmic and DDX42 was predominantly nuclear, consistent with previously reported localizations (e.g. [82; 83]). By contrast, localization of transfected MAP3K3-DDX42 isoforms 1 and 2 was highly variable: preferentially or exclusively in the nucleus in some cells, in the cytoplasm in others, and in still others, diffused equally throughout the cell. A similar pattern was observed for MAP3K3-DDX42 isoform 1 and parents in transfected mouse cortical neurons (Fig. 13). Because of low transfection efficiency in primary neurons, I could not image enough cells to similarly quantify localization. However, based on visualization of ~10 cells, MAP3K3-DDX42 appears to be more frequently localized to the cytoplasm than in HEK293 cells.

The canonical role of *MAP3K3* is activation of the mitogen-activated protein kinase (MAPK) Extracellular signal-related kinase 5 (ERK5) [84]. In response to extracellular signals, MAP3K3 and its downstream kinase MEK5 are activated by heterodimerization and phosphorylation, and in turn phosphorylate ERK5 [85; 86]. Dimerization is mediated by N-terminal PB1 (Phox Bem 1p) domains [87] and phosphorylation by C-terminal kinase domains [84]. Expression of a truncated MAP3K3 PB1 domain or a kinase-dead MAP3K3 results in dominant negative inhibition of ERK5 activation [86-88]. MAP3K3-DDX42 isoform 1 includes an intact PB1 domain but has the DDX42 helicase domain in place of its kinase domain (Fig. 11A), and may therefore act in a similar fashion. I sought to determine whether this fusion protein binds to MEK5. Immunoprecipitation of transfected HEK293 cells indicated that MAP3K3 and MAP3K3-DDX42 isoform 1 bind endogenous phosphorylated MEK5, whereas DDX42 and MAP3K3-DDX42 isoform 2 do not (Fig. 14). This suggests that MAP3K3-DDX42 isoform 1

likely acts as a dominant negative inhibitor of ERK5 signaling.

MAP3K3 also has a non-canonical role in activation of three other MAPK cascades: p38 MAPK [82; 89; 90], JNK [89; 90], and ERK1/2 [91; 92]. Its activity in these pathways is not thought to involve heterodimerization at the PB1 domain [87; 93]. I compared the activation of each of these pathways in cells transfected with MAP3K3 or MAP3K3-DDX42 fusions alone, or co-transfected with MAP3K3 and fusions. MAP3K3, but not MAP3K3-DDX42 fusions, activated ERK1/2 and p38, as predicted by the lack of a kinase domain in fusion proteins (Fig. 15 A-D). These data also suggest that both fusions may slightly suppress activation of ERK1 and p38. JNK was not activated by MAP3K3 or the fusion proteins, but there was evidence that JNK activation was suppressed by the fusion proteins (Fig. 15 A and E).

3.2.2 Functional characterization of DNAJA2-NETO2 fusion proteins

DNAJA2-NETO2 fusion proteins were 14.7 kD and 14.3 kD (Fig. 16). DNAJA2 is 45.7 kD. Observed expression level of isoform 1 was lower than that of isoform 2 and of full length DNAJA2. This may result from nonsense mediated decay of *DNAJA2-NETO2* transcript variant 1 as its premature termination codon arises >100 nucleotides 5' of the nearest exon boundary [94]. DNAJA2-NETO2 fusion localization was also altered in HEK293 cells. The parent protein DNAJA2 was excluded from the nucleus as has been previously reported [95] whereas fusion proteins were consistently expressed throughout the nucleus and cytoplasm (Fig. 17).

Two *in vitro* demonstrations of DNAJA2 function have been previously published. One group showed that DNAJA2 enhances folding of beta-adrenergic receptor, which is the G-protein coupled receptor responsible for cyclic AMP (cAMP) synthesis. Cells co-transfected with

DNAJA2 and beta-adrenergic receptor produced significantly more cAMP when stimulated with isoproterenol than cells transfected with beta-adrenergic receptor alone [96]. Another group showed that DNAJA2 significantly reduced luciferase aggregation following heat shock, as measured by the proportion of luciferase in the insoluble fraction vs. the soluble fraction of cell lysate [97]. I repeated both of these assays using the protocols provided by the authors, however I did not see an effect of DNAJA2 in either case (not shown). Accordingly, I was unable to use these assays to test functional changes resulting from DNAJA2-NETO2 fusions.

DNAJA2 is a co-chaperone of Heat-Shock Protein 70 (HSP70), thought to play a general role in folding of neural G-protein coupled receptors [96]. Little is known, however, about its specific targets. We used *in silico* tools to predict likely targets of DNAJA2 in the brain. First, we used a database of published protein-protein interactions to find all proteins shown to interact with DNAJA2; a total of 792 genes were found to interact with DNAJA2. In order to identify interacting proteins most likely to be relevant to schizophrenia, we searched that list for genes developmentally co-expressed with DNAJA2 in the second trimester in the dorsolateral prefrontal cortex. This yielded 52 genes that physically interact with and are co-expressed with DNAJA2 (Table 3). The protein classes most highly represented (14 genes each) were the "Heat shock protein and chaperone" family, consistent with DNAJA2 functionality within that pathway, and, intriguingly, nucleic acid-binding proteins, including both DNA- and RNA-binding proteins (Fig. 18). This list also included a number of genes with roles in functions known to be disrupted in schizophrenia. These included *DYNC1LI1*, a dynein subunit gene involved in neuronal development. Engineered mutations in this gene in mice result in altered function of the developing cortex and increased anxiety behaviors [98]. *ARF4* regulates dendritic spine development in the dentate gyrus region of the hippocampus, and mice with mutations in

this gene showed deficits in memory-related pattern separation tasks [99]. *DLD* encodes a mitochondrial dehydrogenase with a direct role in adult neurogenesis and neuroprotection [100; 101]. The diversity of genes on this list points to the potential for broad-ranging cellular impacts potentially resulting from alteration of *DNAJA2* function.

Gene	R at DFC	Chaperone	Nucleic Acid Binding
YWHAE	0.993	Y	
DYNC1LI1	0.965		
CCT2	0.954	Y	
DNAJB6	0.953	Y	
PTGES3	0.95	Y	
COPB1	0.949		
SNX4	0.947		
CFL2	0.942		
MATR3	0.941		Y
SAMHD1	0.936		
POLR2B	0.934		Y
ACTR3	0.93		
ATG3	0.929		
CSE1L	0.92		
DLD	0.917		
TCP1	0.917		
SKP1	0.916		Y
CCT6A	0.911	Y	
PFDN1	0.911		
ARF4	0.907		
HSPD1	0.906	Y	
TTC27	0.903		
SSBP1	0.899		
XPO1	0.897		
PSMC6	0.896		
SUCLA2	0.895		
LRPPRC	0.894		Y
DNAJA1	0.89	Y	
NDUFA5	0.89		
VBP1	0.885	Y	
MSH2	0.88		Y
HSPA9	0.874	Y	
GTF2B	0.873		
PPP1CC	0.869		
HDAC2	0.865		Y
HNRNPH2	0.861		Y
DDX1	0.859		Y
CCT8	0.856	Y	
DHX15	0.854		Y
XRCC5	0.85		Y
CCT4	0.844	Y	
RPS3A	0.831		Y
PSMC2	0.829		
SERBP1	0.827		Y
TPM3	0.825		
WDR48	0.823		
SMARCA5	0.82		Y
C1orf25	0.815		Y
DNAJB11	0.809	Y	
HSPA4	0.808	Y	
GMPS	0.806		
HSP90AA1	0.803	Y	

Table 3. DNAJA2 interacting partners co-expressed in developing prefrontal cortex

Table lists 52 genes that both physically interact with DNAJA2 and whose expression is correlated with that of DNAJA2 in the dorsolateral prefrontal cortex during the second trimester. Genes that are members of Chaperone and Nucleic acid binding are indicated.

3.2.3 Functional characterization of MATK-ZFR2 fusion proteins

MATK-ZFR2 fusion proteins were 97.3 kD and 97.8 kD (Fig. 19). Both include a 5'-UTR of *MATK* and all but the first exon of *ZFR2*, and as a result, lack a natural start codon. Observed molecular weights of the fusion proteins were consistent with translation initiation occurring at the first Methionine codon in *ZFR2* exon 2, Met41, which is in the correct reading frame and has a strong Kozak sequence (GGG**ATG**G) [102]. Expression of the novel ZFR2 isoform 2 was higher than expression of ZFR2 isoform 1 (Fig. 19B). The same was true for their respective fusions. Isoform 2 lacks exon 15 and the resulting frameshift leads to a longer protein with 198 novel amino acids (Fig. 20A). This novel C-terminus contains no putative conserved domains and has no homologs by BLAST-p search. Instead, the exon-skipping and subsequent frameshift disrupt the highly conserved dimerization zinc finger (DZF) domain and alter the conserved secondary structure (Fig. 20B). These differences in secondary structure may contribute to functional divergence between the two isoforms. Maintenance of two functional open reading frames of nearly 600 base pairs requires a high degree of sequence constraint. Because 3 out of every 64 codons are stop codons, I estimate that the probability of maintaining an open reading frame of 198 codons by chance is 0.000074. We sequenced exons 16-19 of chimpanzee ZFR2 and found that a frameshift of the homologous region would also produce a novel protein product of 198 amino acids, whereas an analogous frameshift in mouse or rat Zfr2 would result in an earlier stop, after 77 codons (Fig. 20C).

Because MATK and ZFR2 are expressed only in brain, we evaluated localization in transfected mouse cortical neurons. Localization of MATK-ZFR2 chimeras did not appear to differ from their full-length ZFR2 counterparts. Exogenously expressed ZFR2 isoform 1 and

MATK-ZFR2 isoform 1 localized predominantly to the nucleus (Fig. 21A and B). In many neurons, they were present exclusively in the nucleus, but in some they were expressed diffusely throughout the nucleus, cell soma, and processes. In contrast, ZFR2 isoform 2 and MATK-ZFR2 isoform 2 were almost entirely excluded from the nucleus, and showed a granular staining pattern in all cell compartments where they were present (Fig. 21C and D). Within the dendrite, ZFR2 isoform 2-containing granules localized preferentially to branch sites (Fig. 22).

Sequence comparison of ZFR2 isoform 1 and its well-characterized homolog ZFR revealed potentially functionally important differences between ZFR, ZFR2 isoform 1, and ZFR2 isoform 2 (Fig. 23). Nuclear localization of ZFR2 isoform 1 may be directed by a nuclear localization signal (NLS) that is disrupted by the frameshift in ZFR2 isoform 2. By homology, it appears that ZFR2 has one putative NLS at amino acids 913-933, consisting of two clusters of basic amino acids separated by 11 spacer amino acids, consistent with the structure of a typical bipartite NLS [103]. This NLS is within the region disrupted by the frameshift, which may explain why isoform 1 but not isoform 2 localizes to the nucleus (Figs. 19A, 23). Interestingly, ZFR also has a second, more N-terminal NLS, which is entirely absent from ZFR2.

3.3 Discussion

3.3.1 Functional characterization of chimeric genes

MAP3K3-DDX42

Our results suggest that MAP3K3-DDX42 isoform 1 is likely to act as a dominant negative inhibitor of ERK5 signaling by binding to activated MEK5 (Fig. 24A). The ERK5 signaling cascade is critically important in neuronal differentiation and proliferation, neuroprotection, and adult neurogenesis [104-106] and may be misregulated in neuropsychiatric illnesses including autism and major depressive disorder [107; 108]. Our findings also suggest that both MAP3K3-DDX42 isoforms may have slight inhibitory effects on JNK, p38, and ERK1 signaling, all of which function in the brain, in processes including synaptic plasticity and inflammation [109-111]. These effects may be more pronounced against a genetic background haploid at this locus, as in the patient's cells.

DDX42, the 3' parent gene of MAP3K3-DDX42, is a brain-expressed RNA helicase thought to mediate response to viral infection of the central nervous system (CNS) [112]. Recently, DDX42 was shown to modulate the effect of ASPP2, known primarily as a major apoptosis inducer [83]. DDX42 interferes with the induction of apoptosis by ASPP2, and excludes ASPP2 from the nucleus. Intriguingly, a new, brain-specific function for ASPP2 has recently emerged: control of polarity and proliferation of neural progenitor cells during CNS development [113]. ASPP2 targeting to the apical junctional complex in these neural progenitors is thought to be critical for these functions. MAP3K3-DDX42 isoform 1 includes the DDX42 residues that bind

ASPP2, but the protein is mislocalized. DDX42 was predominantly nuclear whereas the subcellular localization of fusion isoform 1 is highly variable. Aberrant cytoplasmic localization of the fusion protein, including the C-terminus of DDX42, may allow it to access ASPP2 at the apical junctional complexes, potentially interfering with its critical role in neural progenitor proliferation (Fig. 24B).

Our evaluation of MAP3K3-DDX42 structure and function reveals unique features of chimeras compared to other mutation types. Inclusion of functional domains from two proteins allows the aberrant protein products to be involved in multiple pathways. Furthermore, one fusion gene may act both by dominant negative and gain-of-function mechanisms, due to variable localization. In a subset of cells, fusions are present in the same compartment as each parent, allowing them to carry out predicted dominant negative functions like inhibition of MAPK pathway signaling. However, in other cells, fusions are mislocalized relative to parent proteins, enabling a potential gain-of-function such as aberrant binding to ASPP2.

DNAJA2-NETO2

DNAJA2-NETO2 is an out-of-frame chimera, and as such NETO2 does not contribute any protein coding sequence to the fusion protein. However, the contribution of a naturally occurring 3'-end to the mRNA may improve ability of the RNA processing machinery to recognize and maintain this aberrant mRNA, compared with deletions or duplications that result in a simple truncation. 3'-UTRs have been shown to have profound, sequence-dependent regulatory effects on their respective genes, via microRNA binding, polyadenylation, translational control, and mRNA localization (for recent review see [114]). NETO2 is involved in

glutamate signaling in the brain [115; 116], so its 3' end should stabilize rather than destabilize the chimeric mRNA in brain tissue.

DNAJA2 is a brain-expressed member of the J-protein family, which are co-chaperones of Heat Shock Protein 70 (HSP70) proteins [95]. It has been shown to promote degradation of the cardiac hERG potassium channel [117] and to enhance HSP70-mediated refolding of neural G-proteins [96]. These roles are likely mediated by two distinct but interrelated functions of DNAJA proteins. First, they activate folding by HSP70 via ATP hydrolysis (Fig. 25A) [118]. Second, they homodimerize, forming clamp-like structures that bind target proteins for delivery to HSP70 [119; 120]. The first function is shared by all J-proteins, and is mediated by the N-terminal J-domain. The latter function is specific to DNAJA proteins, and is mediated by central and C-terminal domains. DNAJA2-NETO2 fusion proteins include only the J-domain. Similar mutations in other J-domain family proteins have been studied. One pertinent study showed that expression of an isolated J-domain in mammalian cells completely blocks luciferase refolding by endogenous HSP70 and impairs cell growth [121]. This dominant negative effect may result from inefficient ATPase activity of the isolated J-domain, which therefore prevents binding and activation by full-length J-proteins and other chaperones (Fig. 25B). Alternatively, toxicity may result from overactivation of HSP70s by free J-domains, trapping potentially important target proteins within the refolding machinery. DNAJA2-NETO2 fusion proteins may have an analogous dominant negative function, preventing proper folding of key substrates. Furthermore, the observed mislocalization of DNAJA2-NETO2 proteins to the cytoplasm may also contribute to a deleterious effect via a gain-of-function mechanism.

MATK-ZFR2

ZFR2 is a previously uncharacterized gene, similar by sequence homology and predicted protein structure to *ZFR*, a member of the double-stranded RNA binding family[122]. ZFR interacts with the brain-expressed RNA binding protein Staufen2, and together they mediate transport of ribonuclear particles (RNPs) into neurites [123]. The ZFR2 region homologous to the Staufen2 binding residues of ZFR is quite divergent, suggesting that ZFR and ZFR2 may have different binding partners. Mice have orthologs to both *ZFR* and *ZFR2*; loss of Zfr is embryonic lethal [124; 125]. C. elegans have a single gene orthologous to *ZFR* and *ZFR2*, Y95B8A.7, which when targeted by RNAi leads to defects in axon guidance [126].

Localization of ZFR2 isoform 1 is similar to endogenous expression of ZFR in neuronal cells: present at high levels in the nucleus and low levels in the cytoplasm. By contrast, the newly characterized ZFR2 isoform 2 was excluded from the nucleus, and showed preferential localization of granules to dendritic branch sites. This staining pattern is similar to that of other neuronal RNA-binding proteins, such as Staufen and FMRP in *Drosophila* neurons [127], and Zipcode binding protein 1 (ZBP1), which regulates the formation of dendritic arbors in hippocampal neurons [128]. Interestingly, ZBP1 was enriched at branch points of developing but not mature neurons, suggesting that its role is in establishment rather than maintenance of branching. There is considerable evidence that ZFR2 isoform 2 is functional. It is ubiquitously expressed in human brain regions, is stable when exogenously expressed, and localizes to dendritic branching sites similar to other members of the mRNA granule transport system. If so, the frameshift differentiating ZFR2 isoforms 1 and 2 may have acted as an evolutionary mechanism for the emergence of a novel protein in the primate brain. Frameshifting has been proposed and demonstrated to be a source of new proteins that can be targeted by selection, but

generally in the evolutionarily permissive context of a prior gene duplication [129]. To our knowledge, this is the first description of a frameshift resulting in translation of such extensively overlapping open reading frames.

Based on the evidence presented above, ZFR2 isoform 1 may play a role in nucleo-cytoplasmic shuttling of RNPs, like its homolog ZFR. ZFR2 isoform 2 may transport RNPs containing mRNA coding for proteins involved in branching and target them to appropriate sites in the dendrite, where their translation regulates branching (Fig. 26A). The fusion of *ZFR2* to the 5'-UTR and putative regulatory regions of *MATK* may result in overexpression of two isoforms of ZFR2 in the brain of the individual carrying this mutation at critical developmental time points. The result may be disruption of dendritic arborization, either positively (Fig. 26B) or negatively.

3.4 Materials and Methods

Expression Constructs

Full length *MAP3K3*, *DDX42*, *DNAJA2*, *MAP3K3-DDX42 transcripts 1 and 2*, *andDNAJA2-NETO2 transcripts 1 and 2* were amplified from patient lymphoblast cDNA using Phusion polymerase (New England Biolabs, Ipswich MA). M-D chimeras and parents were amplified with a reverse primer mutating the natural stop codon and PCR products were directly cloned into TOPO-TA pcDNA3.1 C-terminal V5/His tagged mammalian expression vector (Life Technologies, Carlsbad CA). D-N chimeras and parents were first cloned into a TOPO-TA sequencing vector (Life Technologies, Carlsbad CA), then moved into an N-terminal V5-tagged pcDNA3 expression vector (custom vector gift of Wendy Roeb) using the Gateway system (Life Technologies, Carlsbad CA). *ZFR2* transcripts 1 and 2 and *MATK-ZFR2* transcripts 1 and 2constructs were assembled using the Gibson method [130]. In brief, *MATK*, *ZFR2* and pcDNA3.1 C-terminal V5/His were PCR amplified using primers with overlapping overhangs of ~20 base pairs, denatured, annealed, and ligated together. *MATK* fragment was amplified from genomic DNA. *ZFR2* fragments were amplified from human brain cDNA (Clontech Laboratories Inc, Mountain View CA). We were unable to amplify full length *ZFR2*, so we performed nested PCR to amplify two halves of *ZFR2* separately. All constructs were fully Sanger sequenced and free of non-synonomous PCR errors.

Cell culture and transfections

Human Embryonic Kidney 293 cells (ATCC, Manassas VA) were grown in DMEM

supplemented with 10% Fetal Bovine Serum and 1% Penicillin/Streptomycin. Cortical neurons from neonatal (P0) C57BL/6 mice of either sex mice, were plated at a density of ~25,000-50,000/well onto a bed of confluent wild-type astrocytes on poly-D lysine and collagen treated coverslips in 12- well plates. Neurons were cultured in MEM with B27 and 10% horse serum. To prevent proliferation of astrocytes, mitotic inhibitor (5 μM 5-fluoro-2'-deoxyuridine and 12.5 μM uridine) was added to the neuronal cultures at 2 days in vitro (DIV). Cells were grown for a total of 10-12 days. Experiments were performed according to the guidelines for the care and use of animals approved by the Institutional Animal Care and Use Committee at the University of Washington. All transfections were performed with Lipofectamine 2000 (Life Technologies, Carlsbad CA) according to manufacturer instructions. Cortical cells were co-transfected after 9 or 10 DIV with GFP pCMV (0.6ug) and experimental construct or empty vector (2.4ug). Media was changed after 4 hours, and transfections proceeded for 24 hours.

Immunocytochemistry

HEK293 cells or cultured neurons were fixed with 4% paraformaldehyde for 20 minutes at room temperature. Neurons were then permeablized with 1xPBS/0.5% Triton x100 for 10 minutes at room temperature, and then blocked in 5% normal goat serum (NGS)-0.25%Triton-X/PBS for 1 hour at 4°C. Cultured cells were incubated overnight at 4°C with Mouse monoclonal anti-V5 primary antibody (1:200, Life Technologies, Carlsbad CA)**.** Next, sections were rinsed 4 times in PBS and incubated in Alexa 568 (red)-conjugated goat secondary antibodies (1:400, Life Technologies, Carlsbad CA) for 1-2 hours at room temperature. Sections were washed 3 times in PBS, mounted on slides using ProLong Gold Antifade media containing

DAPI (Life Technologies, Carlsbad CA).

Western Blotting and Immunoprecipitation

HEK293 cells or patient lymphoblast cell lines were lysed in RIPA buffer (50mM Tris-HCL (pH 7.4), 1% NP-40, 0.25% sodium deoxycholate, 150 mM NaCl,1mM EDTA) with cOmplete mini protease inhibitor cocktail tablets (Roche Diagnostics Corporation, Indianapolis IN) and Halt Phosphatase Inhibitor Single-Use Cocktail (Thermo Fisher Scientific, Rockford IL) and quantitated by the Bradford method. We separated 10-50ug total protein by SDS/PAGE, transferred to Immobilon-FL membrane (Millipore, Billerica MA), and incubated overnight with primary antibody and for one hour with IRDye conjugated secondary antibodies (LI-COR Biosciences, Lincoln NE), followed by analysis with an Odyssey infrared imaging system (LI-COR Biosciences, Lincoln NE). For analysis of MAPK activation, blots were probed first with phospho-specific antibodies, stripped with NewBlot™ stripping buffer (LI-COR Biosciences, Lincoln NE), and reprobed with antibody to total JNK, p38 or actin. Primary antibodies used were: mouse anti-V5 (1:5000, Life Technologies, Carlsbad CA), rabbit anti-β-actin (1:1000, Santa Cruz Biotechnology, Santa Cruz CA), mouse anti-β-actin (1:10000, Sigma-Aldrich, St. Louis MO), rabbit anti-MEK5 [EP648Y] (1:10000, Abcam, Cambridge MA), rabbit anti-phospho-MEK5 [sc-135702] (1:500, Santa Cruz Biotechnology, Santa Cruz CA), rabbit anti-p-ERK1/2 [9101] (1:1000, Cell Signaling Technology, Danvers MA), mouse anti-phospho-JNK [sc-6254] (1:500, Santa Cruz Biotechnology, Santa Cruz CA), rabbit anti-phospho-p38 MAPK [3D7] (1:1000, Cell Signaling Technology, Danvers MA), rabbit anti-JNK [56G8] (1:1000, Cell Signaling Technology, Danvers MA), rabbit anti-p38α [sc-728] (1:500, Santa Cruz

Biotechnology, Santa Cruz CA) .

For immunoprecipitation, transfected HEK293 cells were lysed with NP-40 buffer (1% NP-40, 0.15M NaCl, 0.01M Sodium phosphate) with same protease and phosphatase inhibitors and quantitated as above. 50ul protein G Dynabeads (Life Technologies, Carlsbad CA) were incubated with mouse anti-V5 (1:100, Life Technologies, Carlsbad CA) for 15 minutes at room temperature. Conjugated beads were incubated with 400ug lysate for 90 minutes at 4°C. Beads were rinsed 3x in NP-40 buffer, transferred to a new tube, and eluted in 30ul LDS sample loading buffer with 2.5% 2-mercaptoethanol.

Microscopy

Immunofluorescence images were acquired on a Zeiss 510 META confocal microscope using a 40x or 63x oil immersion objectives. Gain and offset were adjusted for each image to optimize dynamic range. For localization in HEK293 cells, we measured fluorescence intensity of DAPI and transfected protein along a 30 um line drawn through the center of each transfected cell in a 40x image using Plot Profile function in imageJ. Dendritic localization was analyzed in images taken with a 63x objective and 3-4x zoom. Images were acquired in z-series (0.4 um steps) then rendered using maximum-intensity projections and analyzed in imageJ. Profiles of GFP and Alexa 568 fluorescence were made using the Plot Profile function for a rectangular selection around the longest straight segment of a dendrite. Branching sites were indicated manually based on GFP imaging of dendrites. To correct for variation in background staining intensity and transfection efficiency, fluorescence values were normalized and mean intensities at and between branching sites calculated. To correct for increased fluorescence at branch sites

due to widening of dendrites, we divided normalized mean Alexa 568 fluorescence intensity by normalized mean GFP fluorescence intensity. Data presented are from 3 independent experiments.

DNAJA2 co-expression analysis

Physical interactions from the Homo sapiens database were collected from GeneMANIA 2011-08-03 release which consisted of 150 studies [131]. Co-expression based network analysis used existing data-sets in BrainSpan Atlas of Developing Human Brain [60]. For interacting genes, we calculated expression correlation coefficients with DNAJA2 in human dorsolateral prefrontal cortex in the 2nd trimester. Co-expression was defined as correlation coefficient $|R| >= 0.8$.

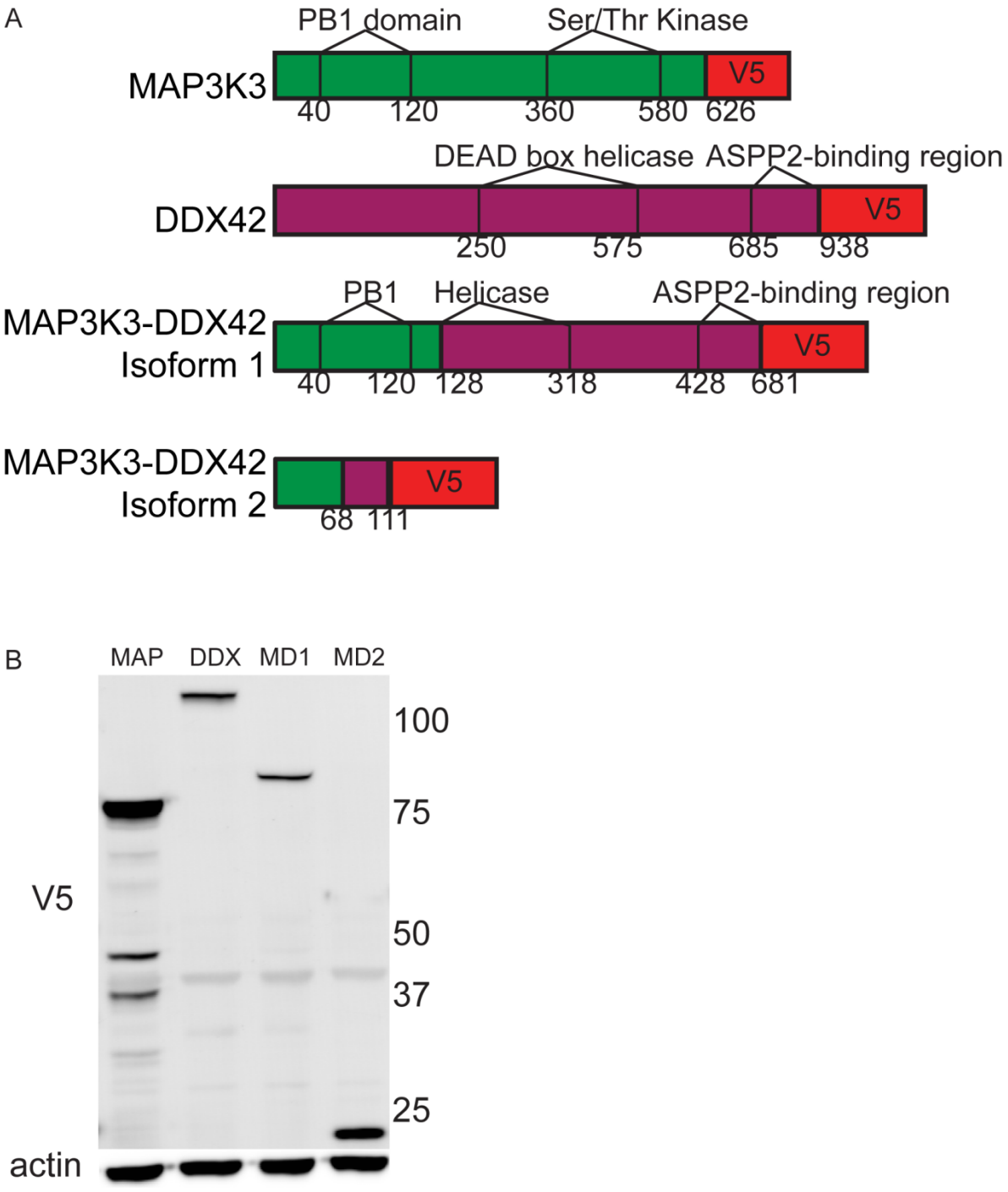

Figure 11. Structure and expression of MAP3K3-DDX42 fusion proteins

C-terminal V5 epitope-tagged mammalian expression constructs for parent genes and two chimeric isoforms. Known functional domains are indicated. (A) Each construct was transfected into HEK293 cells, and protein expression ascertained by western blots stained for V5 and actin loading control (B). Ladder is shown along with molecular weights in kD.

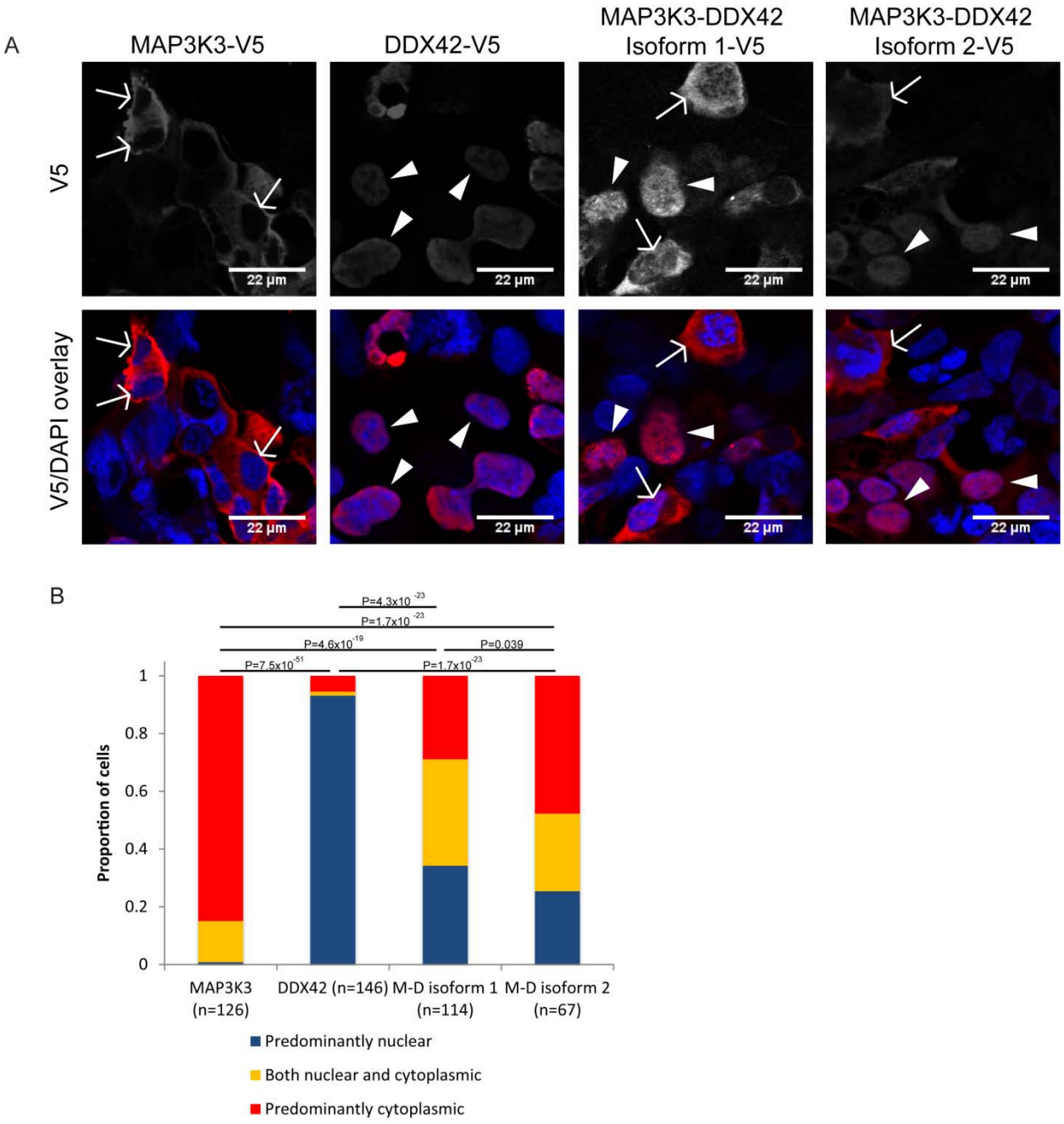

Figure 12. Altered localization of MAP3K3-DDX42 fusion proteins in HEK293 cells

V5 immunostaining of transfected HEK293 cells shows cytoplasmic localization (arrows) of MAP3K3 and nuclear localization (triangles) of DDX42. In contrast, the localization of both M-D fusion proteins is highly variable, ranging from exclusively nuclear (triangles) to exclusively cytoplasmic (arrows). (A) To quantify this effect, we binned cells into three categories according to V5 staining: predominantly nuclear, both nuclear and cytoplasmic, or predominantly cytoplasmic. Frequency of cells in each category is shown for both parent genes and both fusion proteins (B). We show P-values from chi-squared comparisons of distributions between each group.

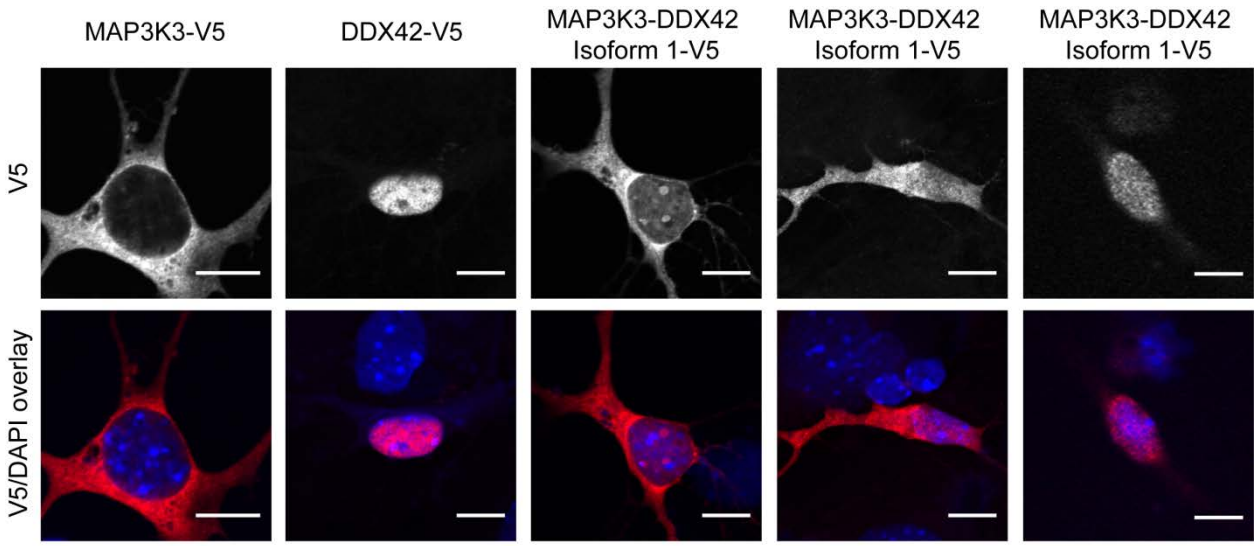

Figure 13. Altered localization of MAP3K3-DDX42 fusion proteins in cultured cortical neurons

V5 immunostaining of transfected mouse cultured cortical neurons shows cytoplasmic localization of MAP3K3 and nuclear localization of DDX42. We show representative images of cells transfected with predominantly cytoplasmic, nuclear and cytoplasmic, and predominantly nuclear localization of MAP3K3-DDX42 isoform 1. Scale bars represent 10 µm.

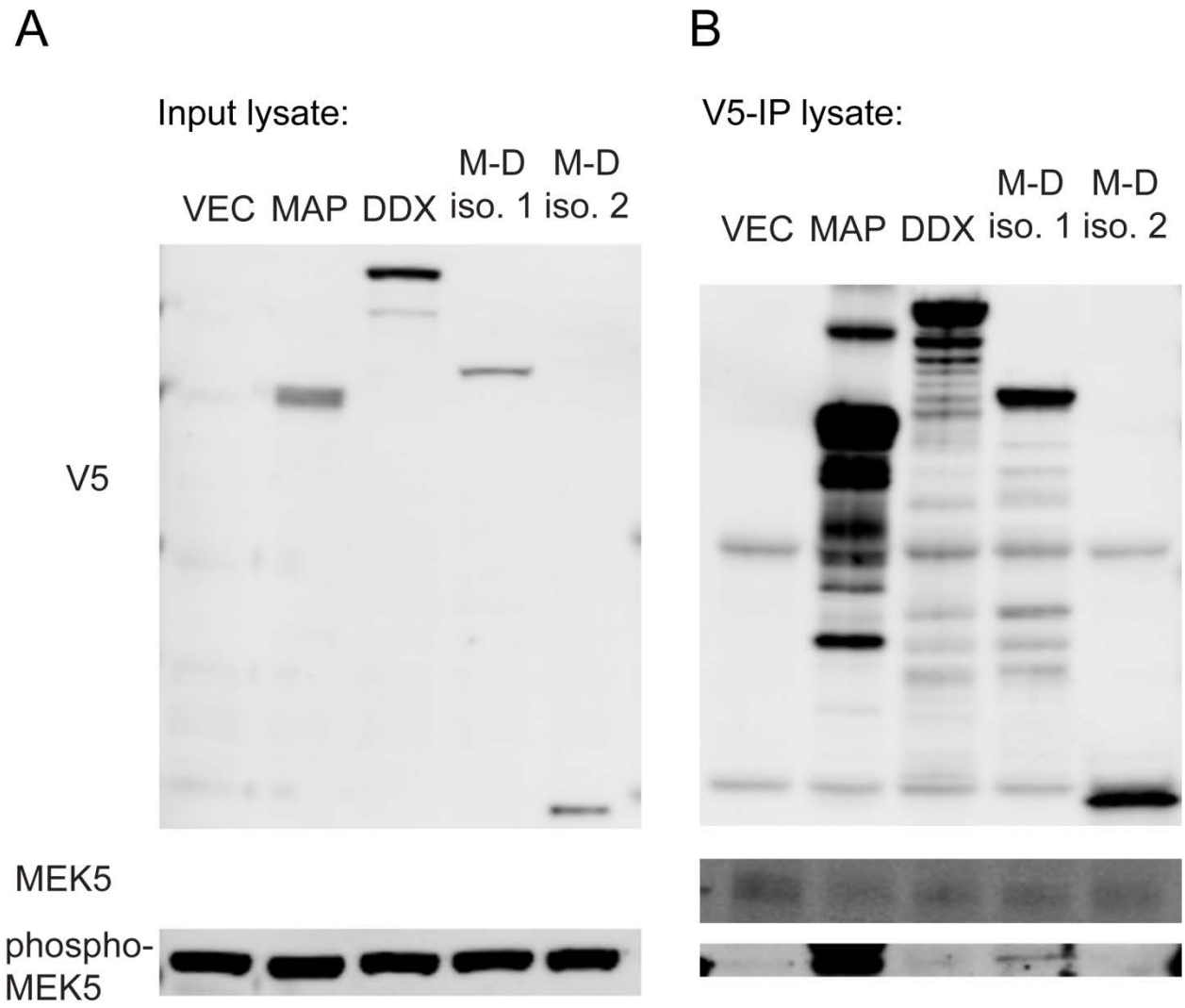

Figure 14. MAP3K3-DDX42 isoform 1 binds phosphorylate MEK5

Lysates from HEK293 cells transfected with empty vector or MAP3K3-DDX42 parent gene or chimeric constructs were immunoprecipitated with a V5 antibody. Western blots were run for input lysates (A) and immunoprecipitated lysates (B) and stained with antibodies to V5, MEK5 and phospho-MEK5. After immunoprecipitation, we see phospho-MEK5 bands only in the MAP3K3 and MAP3K3-DDX42 isoform 1 transfected cells, suggesting that both bind phospho-MEK5. None of the constructs appeared to bind unphosphorylated MEK5.

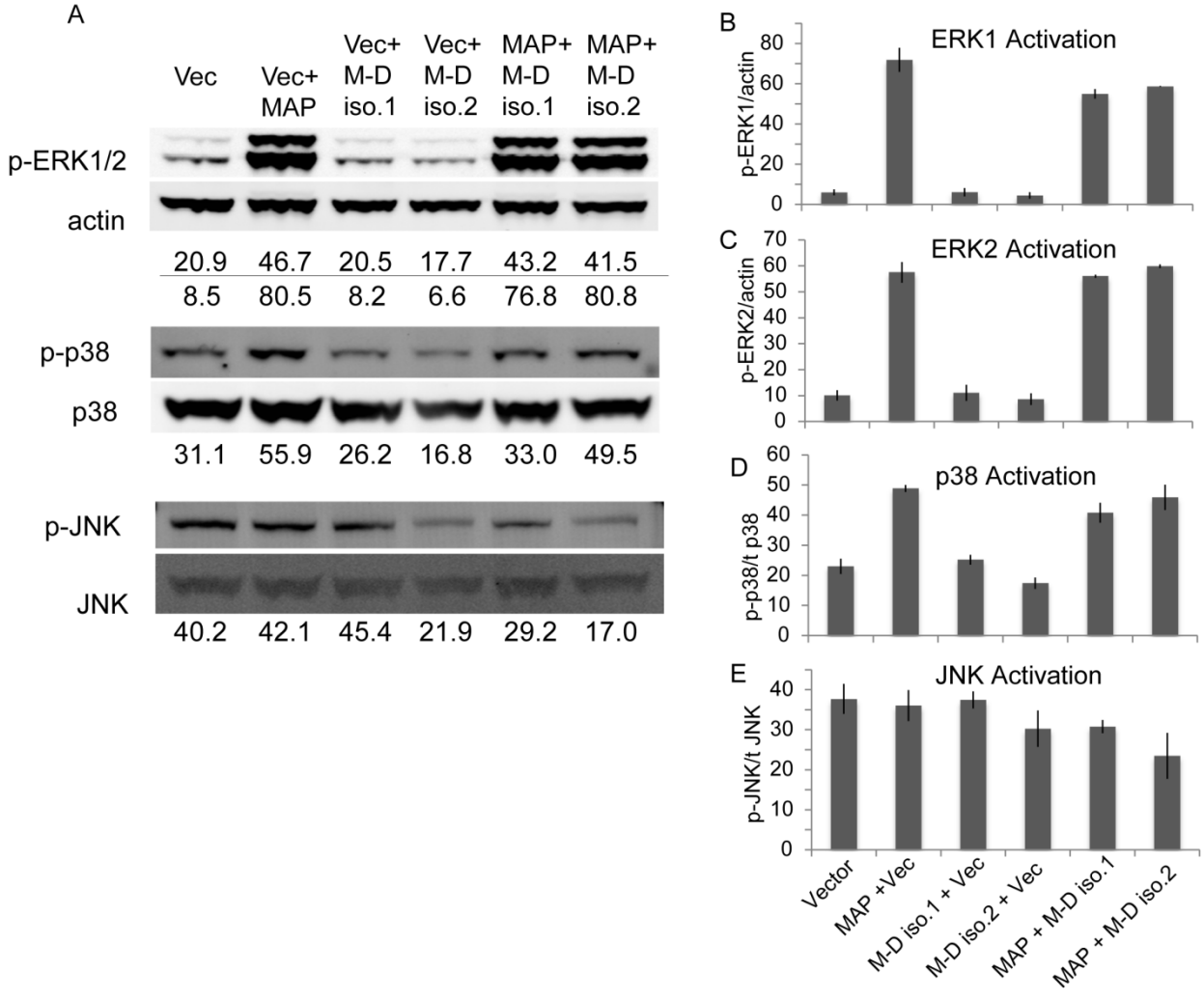

Figure 15. Effect of MAP3K3-DDX42 on signaling by MAP kinases

HEK293 cells were transfected with combinations of MAP3K3 and MAP3K3-DDX42 constructs to observe their effect on activation of MAP kinases. Western blots were probed with antibodies specific to phosphorylated ERK1/2, p38 and JNK, followed by antibodies to actin, total p38 and total JNK. Representative blots and optical density of each band normalized to appropriate loading control are shown in (A). Mean normalized densitometry values from three replicate experiments are shown for ERK1 (B), ERK2 (C), p38 (D) and JNK (E). Error bars represent S.E.M.

Transfection conditions were: empty pcDNA vector (VEC), MAP3K3 (MAP + VEC), MAP3K3-DDX42 isoform 1 (VEC+M-D iso.1), MAP3K3-DDX42 isoform 1 (VEC+M-D iso.2), or co-transfected with MAP3K3 and either M-D long (MAP + M-D iso.1) or M-D short (MAP + M-D iso.2). Empty vector was included to control for total amount of transfected DNA.

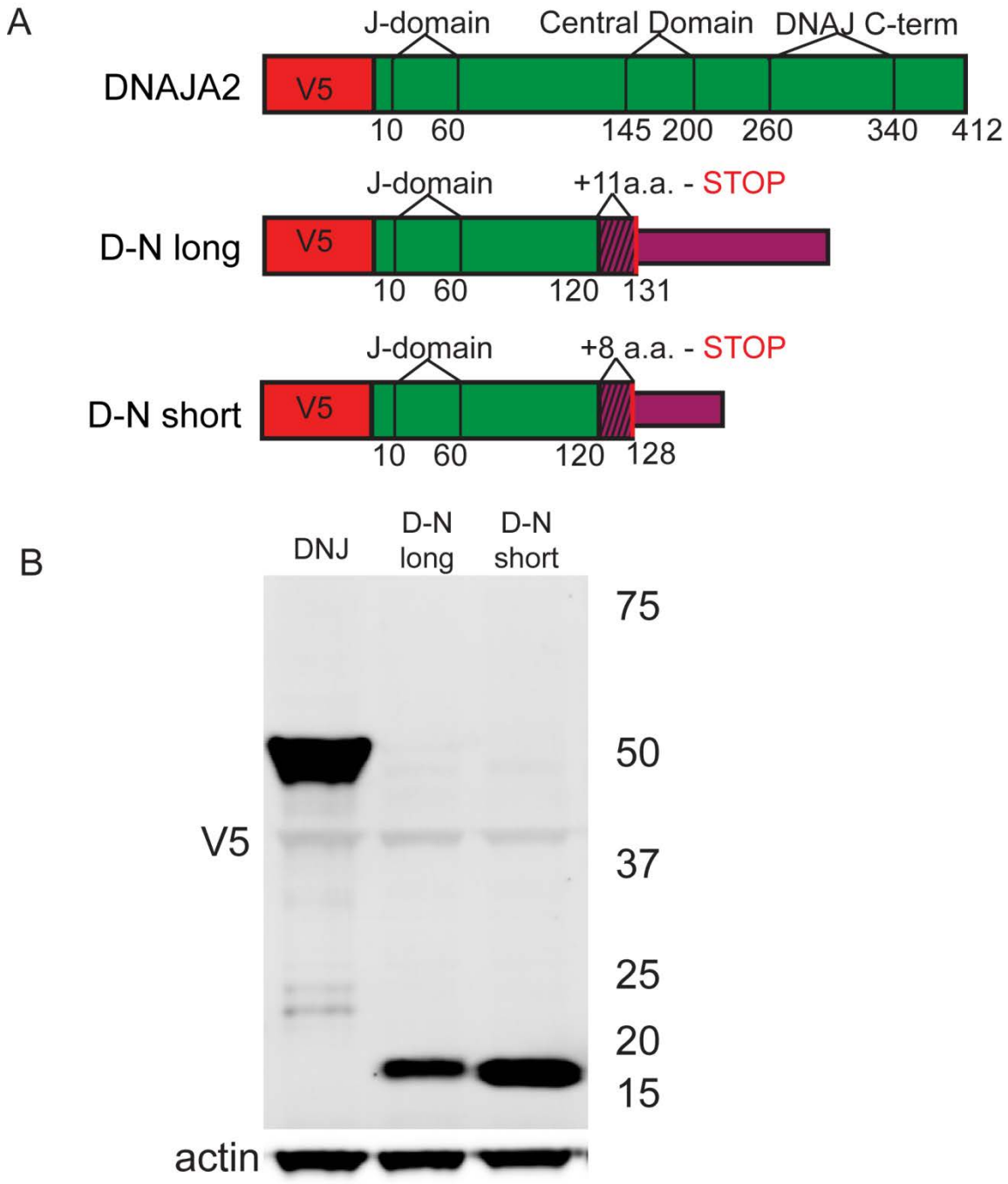

Figure 16. Structure and expression of DNAJA2-NETO2 fusion proteins

N-terminal V5 epitope-tagged mammalian expression constructs for DNAJA2 and two chimeric isoforms. Known functional domains are indicated. (A) Each construct was transfected into HEK293 cells, and protein expression ascertained by western blots stained for V5 and actin loading control (B). Ladder is shown along with molecular weights in kD.

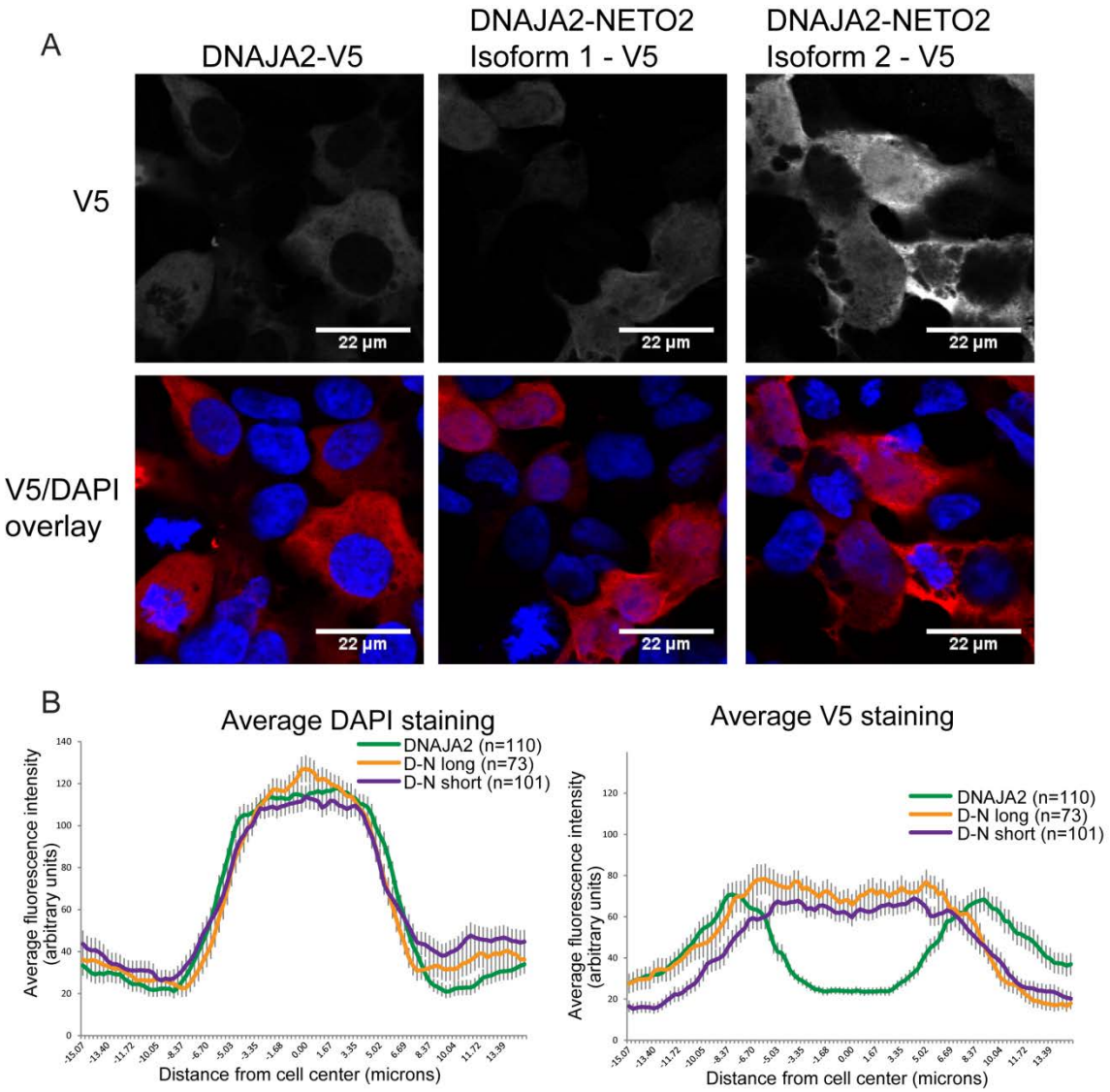

Figure 17. Altered localization of DNAJA2-NETO2 fusion proteins in HEK293 cells

V5 immunostaining of transfected HEK293 cells shows nuclear exclusion of DNAJA2, compared with increased nuclear expression of both DNAJA2-NETO2 chimeras (A). We quantified localization by measuring fluorescence intensity of either DAPI (B) or V5 (C) along a 30µm line drawn centered on the nucleus. Plots show the average fluorescence intensity of indicated number of cells taken from three independent experiments; error bars represent S.E.M.

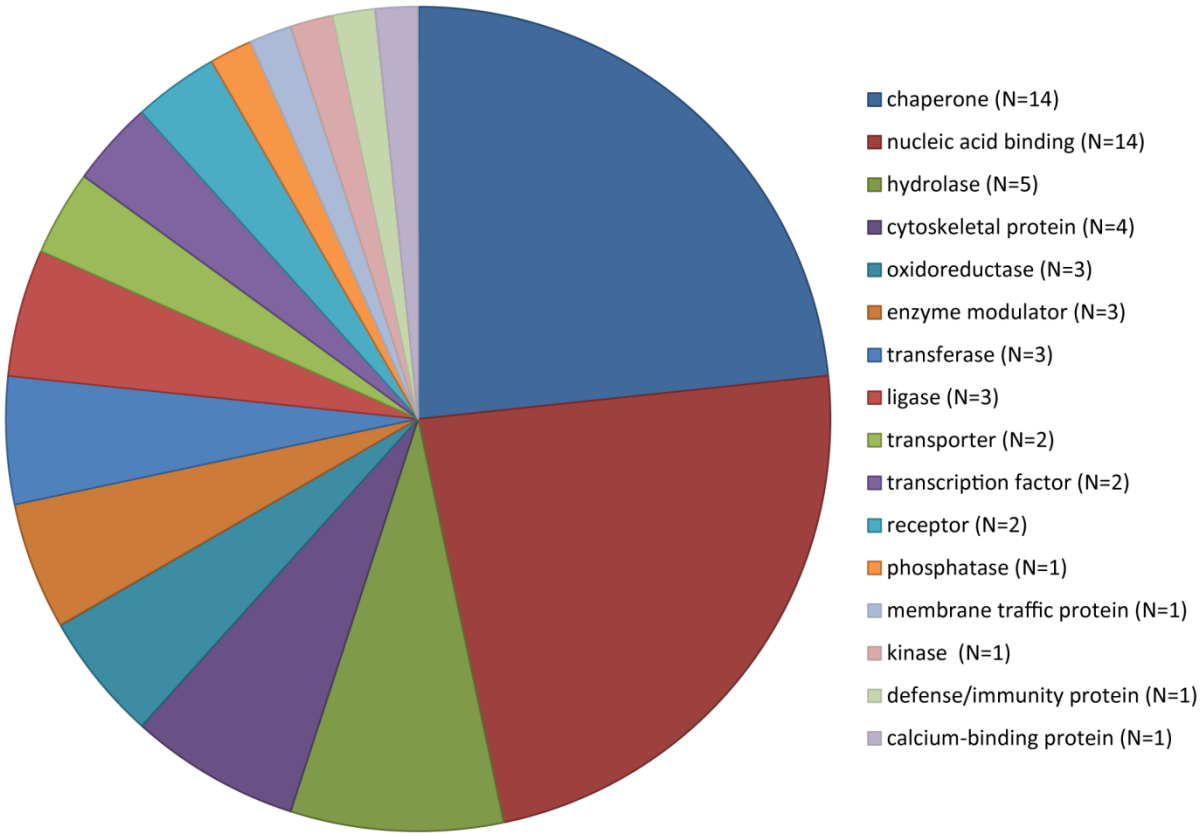

Figure 18. Functional classes of DNAJA2 interacting partners co-expressed in developing prefrontal cortex

Using publicly available databases, we established a list of 52 genes that both physically interact with DNAJA2 and whose expression is correlated with that of DNAJA2 in the dorsolateral prefrontal cortex during the second trimester. Chart shows proportion of those genes in indicated protein classes based on annotation in Panther.

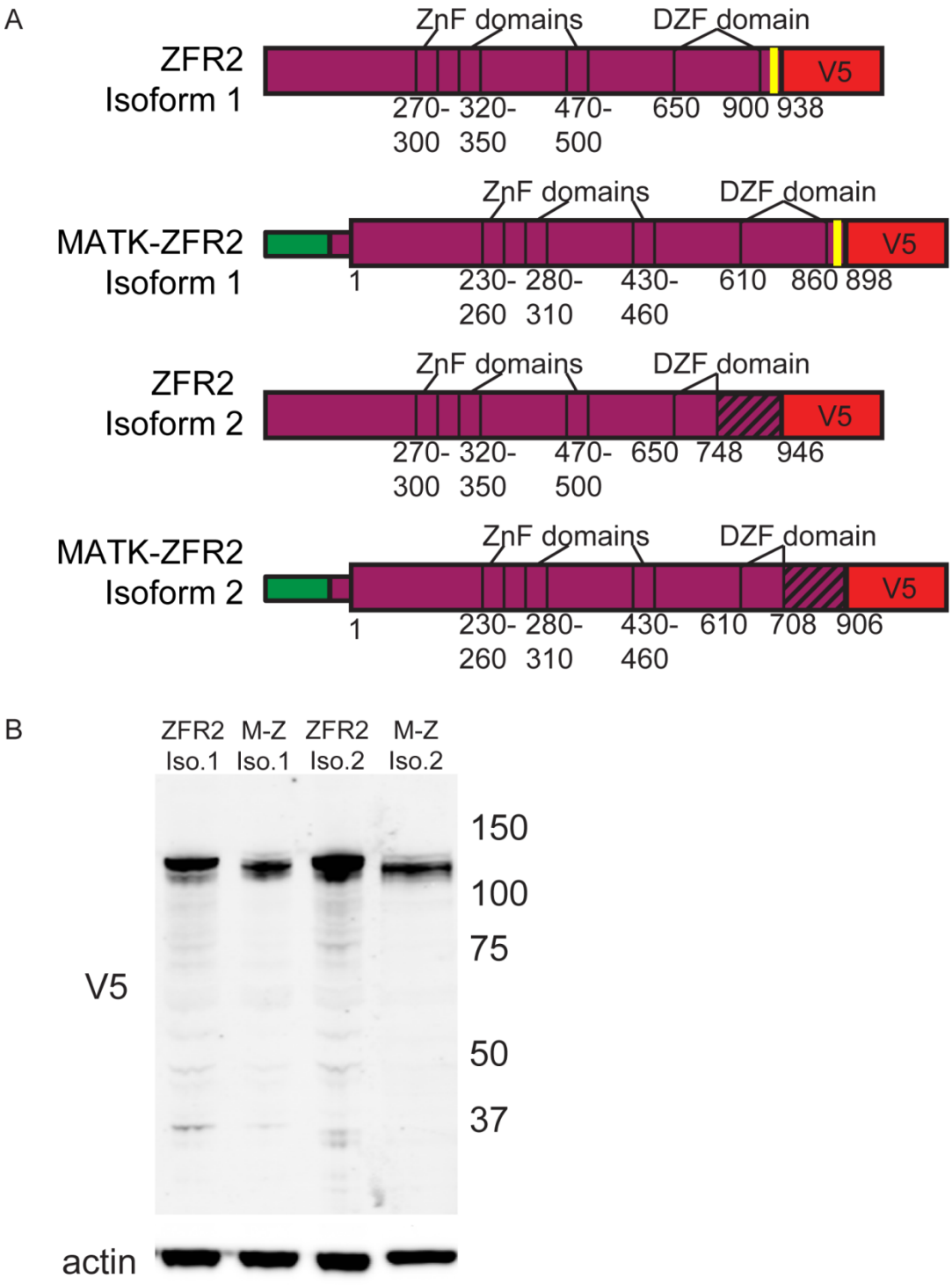

Figure 19. Structure and expression of MATK-ZFR2 fusion proteins

C-terminal V5 epitope-tagged mammalian expression constructs for both isoforms of ZFR2 and both respective chimeric isoforms. Known functional domains are indicated (A). Each construct was transfected into HEK293 cells, and protein expression ascertained by western blots stained for V5 and actin loading control (B). Ladder is shown along with molecular weights in kD.

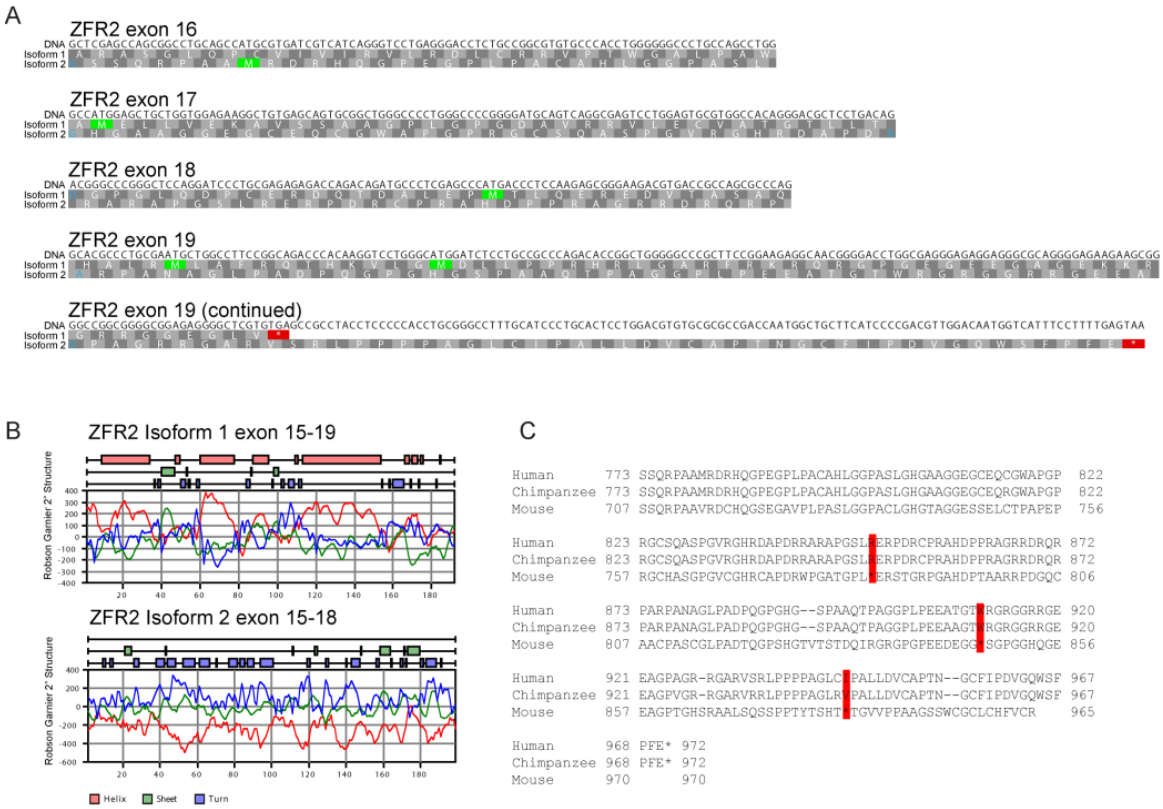

Figure 20. ZFR2 isoform 2 frameshifted C-terminus

Skipping of ZFR2 exon 15 results in a frameshift but not in a premature stop. The two open reading frames in ZFR2 exons 16-19 are shown in (A). We compared the predicted secondary structures of the differing C-termini and showed dramatic differences, particularly in helical content (B). We then compared equivalent frameshifted translations of exons 16-19 of human, chimpanzee, and mouse *ZFR2*. Red highlighted residues are positions where the frameshift leads to a premature stop in mouse but not in human or chimpanzee.

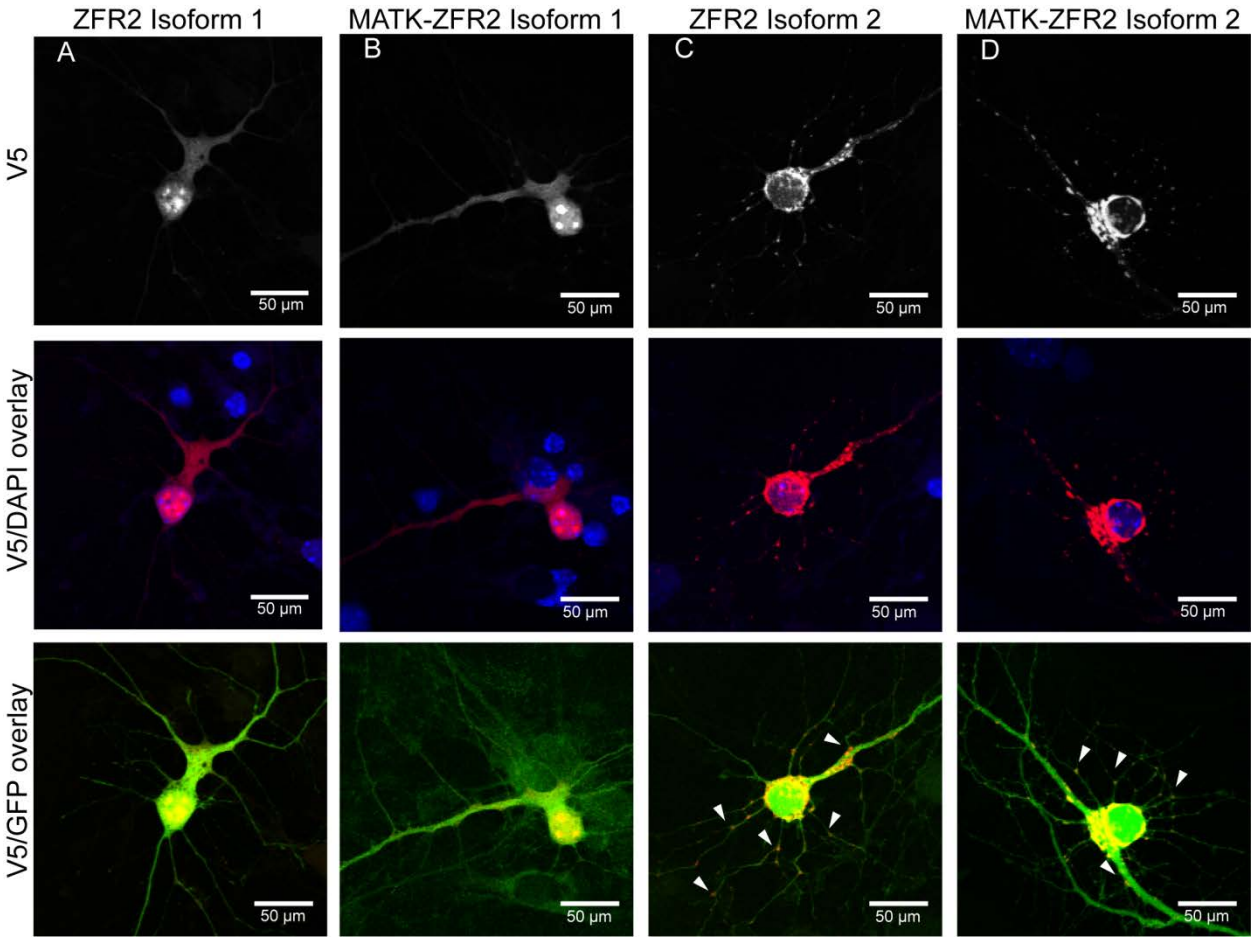

Figure 21. Localization of ZFR2 isoforms and MATK-ZFR2 fusion proteins in cultured cortical neurons

V5 immunostaining of transfected mouse cultured cortical neurons shows diffuse expression of ZFR2 isoform 1 (A) and MATK-ZFR2 isoform 1 (B) throughout the nucleus and cytoplasm, with brighter puncta in the nucleus not co-localized with nucleoli. ZFR2 isoform 2 (C) and MATK-ZFR2 isoform 2 (D) are largely excluded from the nucleus. In dendrites they have a granular staining pattern, with many granules present at branch sites (triangles).

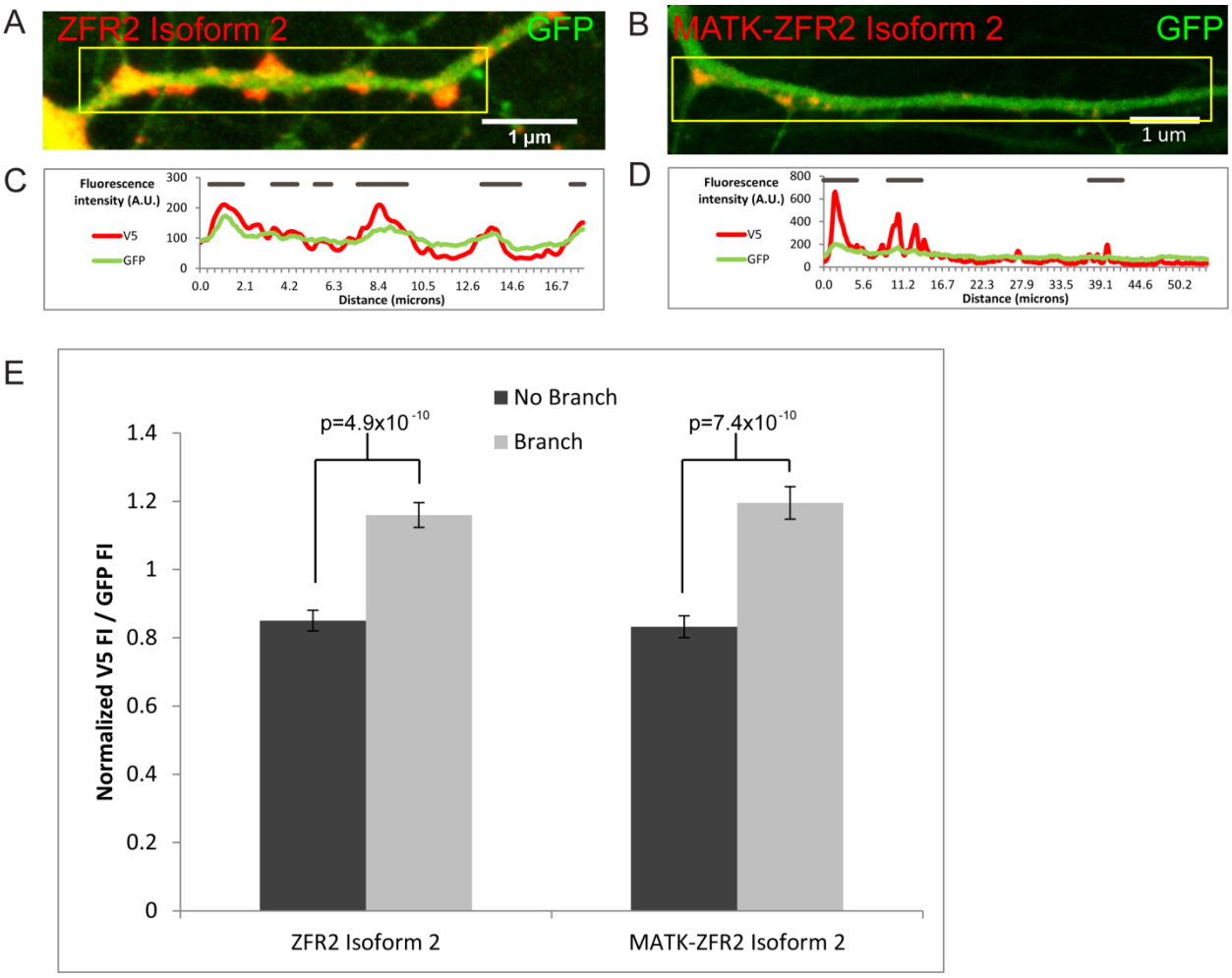

Figure 22. ZFR2 isoform 2 and MATK-ZFR2 isoform 2 localize preferentially to dendritic branch sites

Representative micrographs of cultured mouse cortical cells transfected with ZFR2 isoform 2 (A) or MATK-ZFR2 isoform 2 (B). Dendritic morphology was visualized by cotransfected GFP. (C) and (D) show fluorescence profiles of boxed area. Horizontal gray bars indicate manually annotated branch sites. (E) shows mean intensities of V5 fluorescence along non-branching and branching segments of dendrites, divided by mean GFP fluorescence at those segments to correct for increased dendrite diameter at branch sites. P-values represent significance by 2-tailed T-test. Error bars represent SEM.

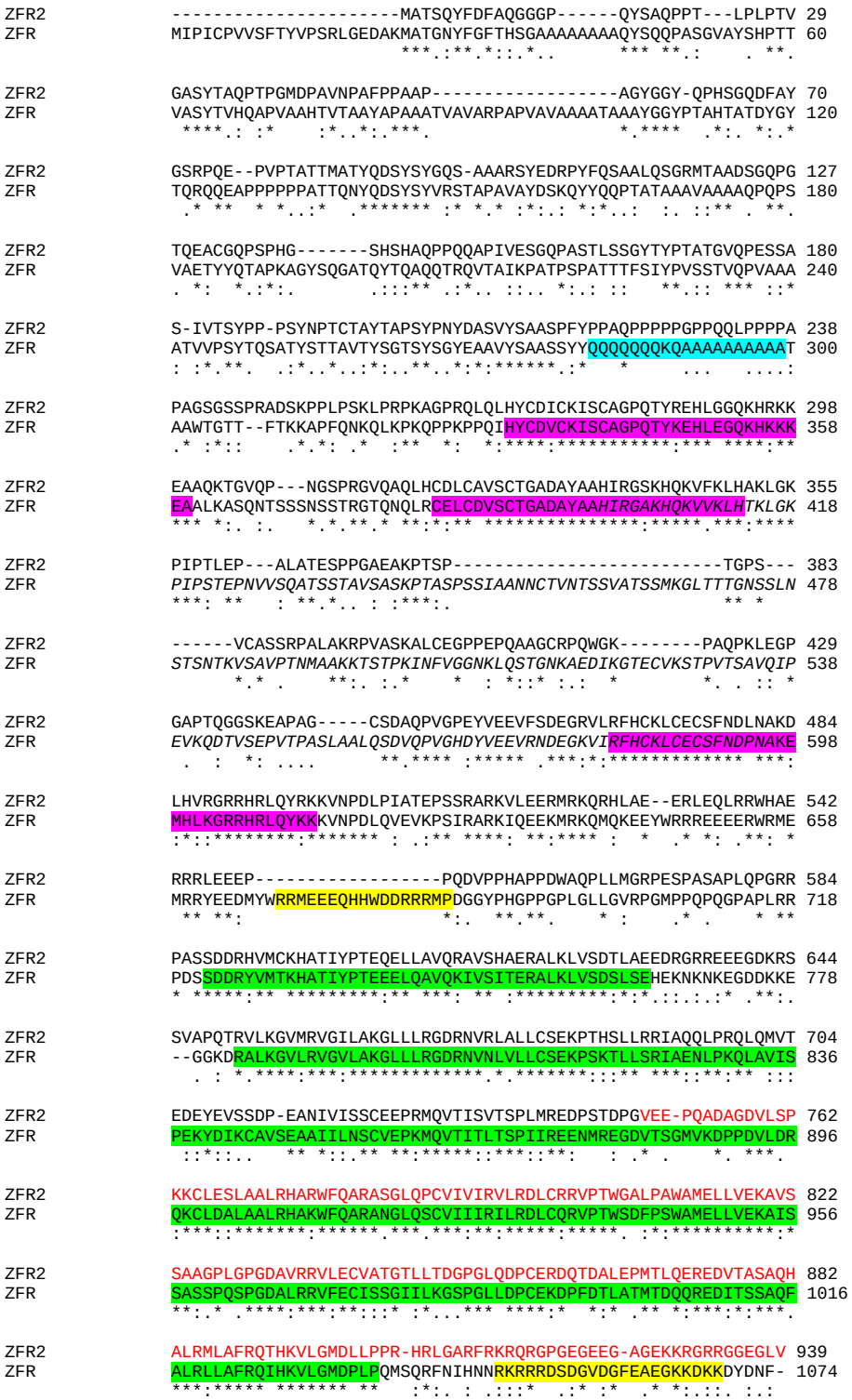

Figure 23.

Alignment of human ZFR2 isoform 1 and human ZFR2 isoform 2

Annotation Key:

C2H2 zinc fingers

DZF domain

Putative nuclear localization signal

polyQ/polyA

Stau2-binding region

Amino acids absent in ZFR2 isoform 2

* positions which have a single, fully conserved residue.

: conservation between groups of strongly similar properties - scoring > 0.5 in the Gonnet PAM 250 matrix.

. conservation between groups of weakly similar properties - scoring =< 0.5 in the Gonnet PAM 250 matrix.

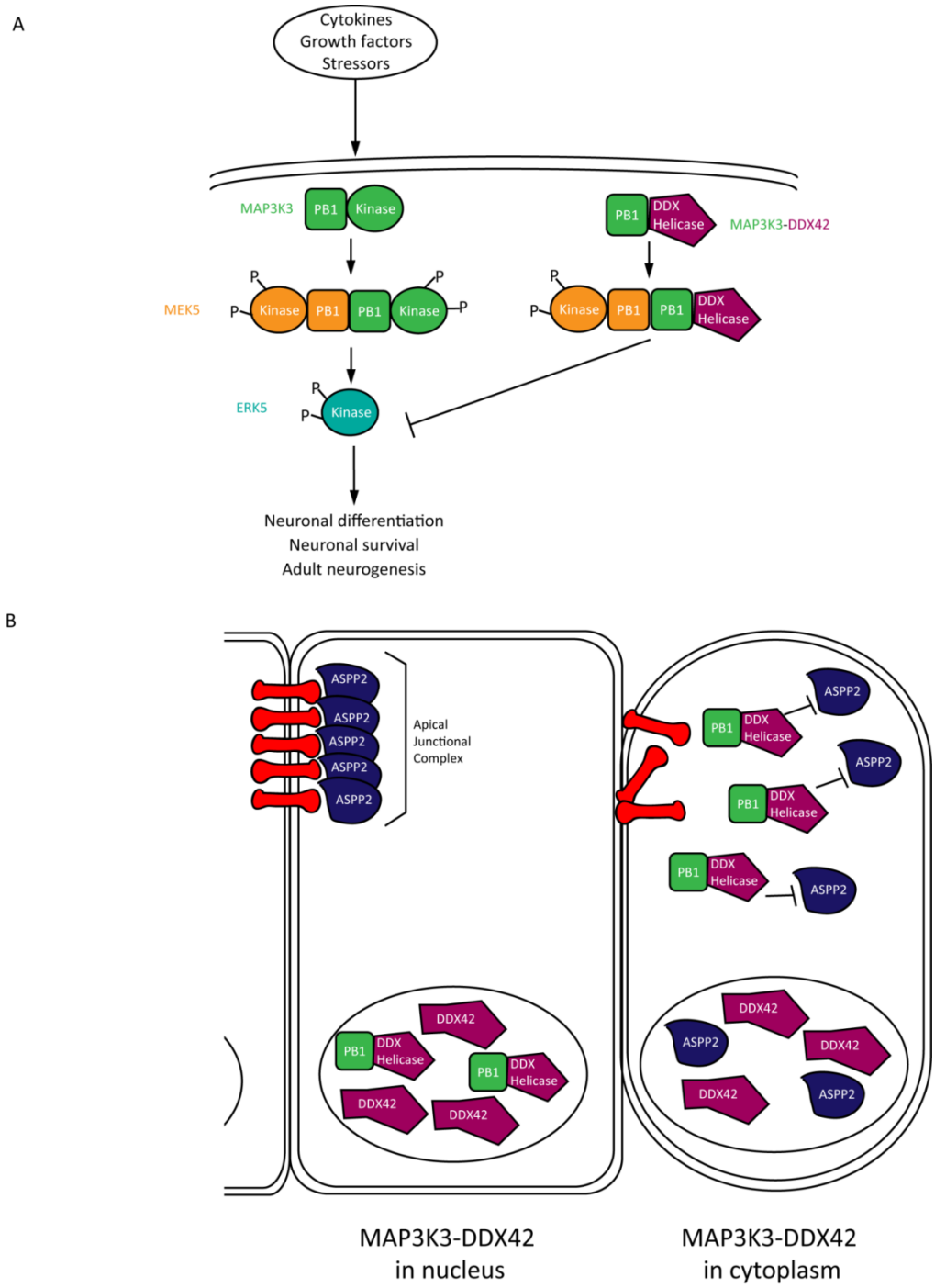

Figure 24. Predicted effects of MAP3K3-DDX42 proteins on ERK5 and ASPP2 pathways

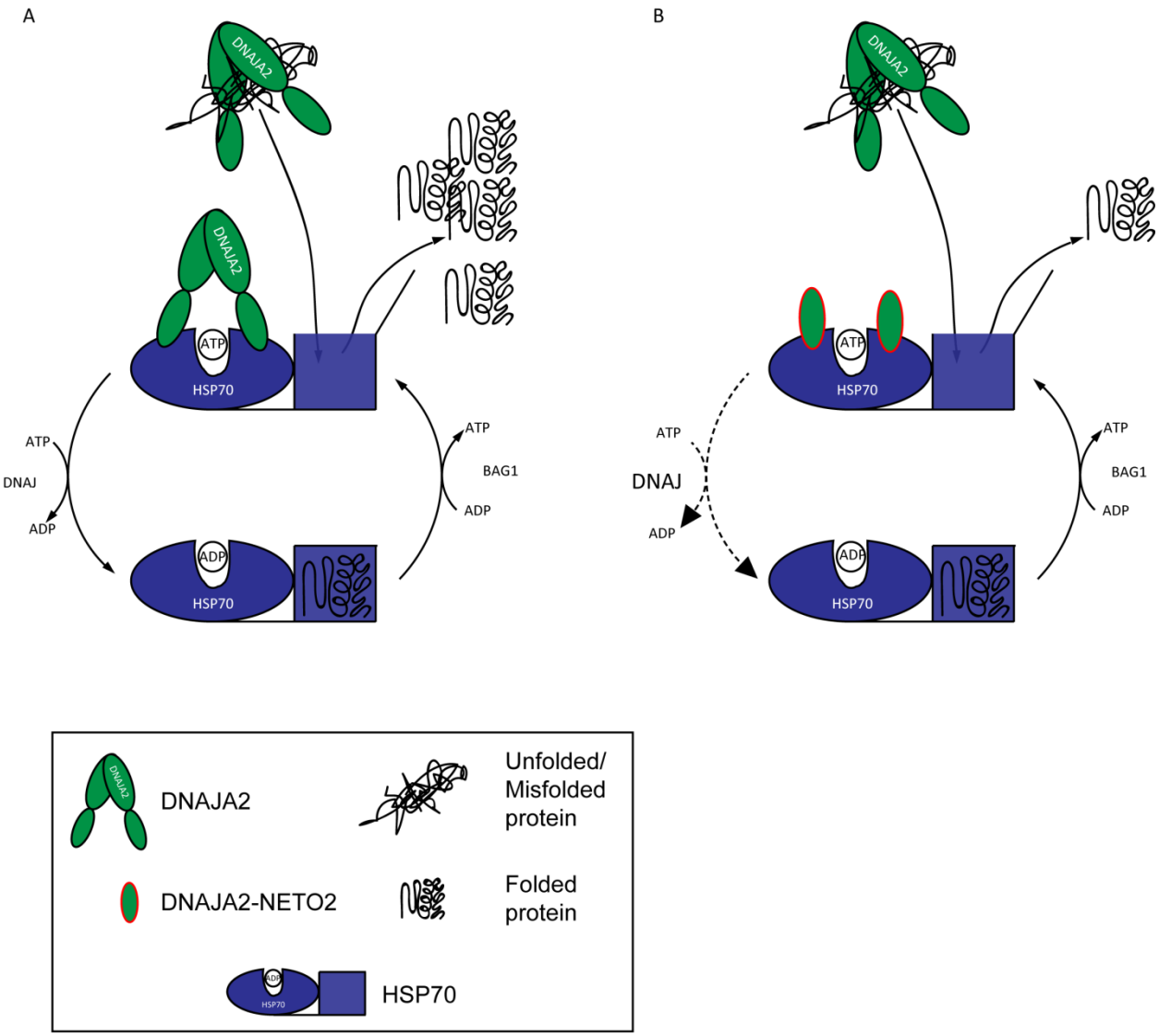

Figure 25. Predicted effects of DNAJA2-NETO2 fusion proteins on HSP70

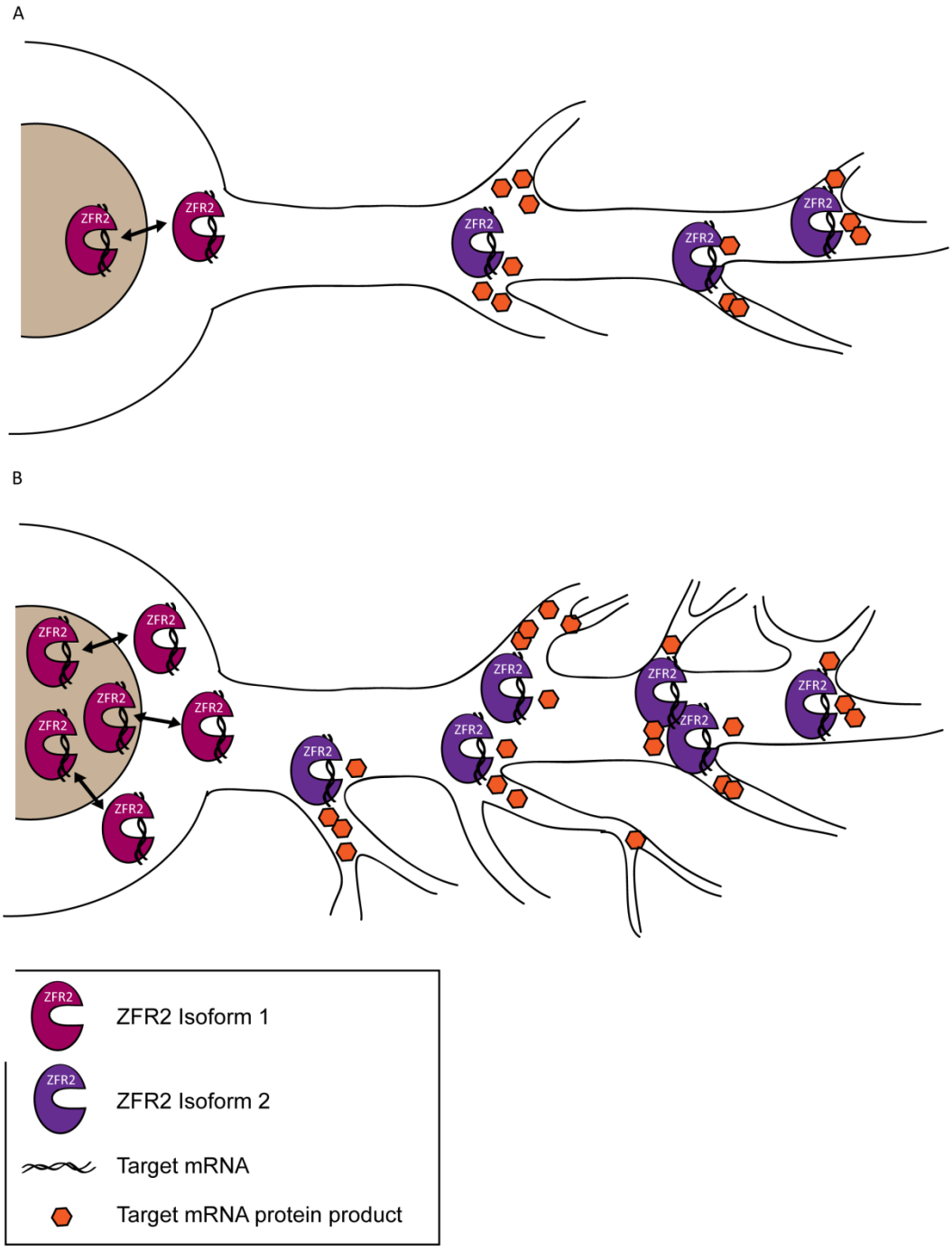

Figure 26. Predicted effects of MATK-ZFR2 proteins on mRNA transport and dendritic branching

Chapter 4: Summary and Future Work

4.1 Summary and Discussion

In the first part of this study, I detected and characterized at the genomic level CNVs leading to formation of chimeric genes in individuals with schizophrenia and in controls. I found that a small proportion of rare CNVs (~3% in this series), both deletions and duplications, result in chimera formation. Most of these (5 out of 7 tested) were transcribed into chimeric mRNA detectable in carrier lymphoblasts. Furthermore, there was an increased burden of rare, brain-expressed chimeric genes among individuals with schizophrenia. The prevalence of such events among cases evaluated by two different aCGH platforms was 6/172, or 3.5%, in contrast to zero among 390 controls. This finding supports a role for chimeric events in schizophrenia.

The second part, functional analysis of three case-specific chimeric events, showed that these chimeras differ dramatically in genomic and transcript architecture, in the functions of the genes involved, and in the predicted mode of action. I presented evidence that MAP3K3-DDX42 fusion proteins have a dominant negative impact on MAP kinase signaling. I also demonstrated that fusion proteins are dramatically mislocalized when exogenously expressed, which may contribute to an additional gain-of-function effect, potentially on establishment of neuron polarity by DDX42-interacting-protein ASPP2 Because this chimera results from a deletion, these effects may be further amplified by reduced expression of parent genes in the patient's cells. DNAJA2-NETO2 fusion results in translation of two stable truncated proteins that are mislocalized to the nucleus. Similar truncations in related proteins have been shown to act as dominant negative inhibitors of protein refolding. We identified candidate genes that may be

specific targets of DNAJA2 chaperone activity in prefrontal cortex. Decreased chaperoning of these proteins could disrupt a wide range of critical processes. MATK-ZFR2 fusion proteins, on the other hand, are nearly identical to their parent ZFR2 proteins and their subcellular localizations are unchanged by fusion to MATK. Their function is likely to be equivalent to that of ZFR2; in this case, it is most likely their regulation that is affected. Based on parent gene expression patterns, we predict that MATK-ZFR2 fusions are likely to be dramatically overexpressed during late fetal development, potentially impacting RNP shuttling and dendritic arborization.

Our analysis of the previously uncharacterized gene *ZFR2* revealed an alternately spliced variant resulting in an intriguing frameshift. Rather than resulting in a premature stop, the frameshift codes for what is essentially a new protein with 198 novel C-terminal amino acids. We provide evidence that the frameshifted protein is new in the primate lineage, and is involved in dendritic branching. If so, this represents a striking evolutionary mechanism that has not been previously observed, to our knowledge.

For private or very rare mutations, including these chimeras, a proof of causality can best be approached by functional analysis. Some genes may impact these or other critical functions under experimental conditions without contributing to schizophrenia risk. The ongoing refinement of the spectrum of experimental phenotypes resulting from mutation of schizophrenia genes will improve our ability to test causality of novel candidates. In addition, genetic criteria including recurrence, linkage within families, and *de novo* mutation as well as bioinformatic analyses of protein-protein interactions, networks, and co-expression during development also contribute to evaluation of causality. We suggest that a focus on chimerism as a mutational mechanism allows the identification of mutations that are likely to prove damaging when

evaluated functionally. These findings suggest a role for germline gene fusions in schizophrenia, implicate new candidate genes and pathways, and provide a new approach to the identification of likely pathogenic variation.

4.2 Future Directions

One limitation of this study is the array CGH platform used, which, though state-of-the-art at the time, has a relatively low resolution of ~1 probe per kb with many technical difficulties in correct event calling. Since we began this study, transformative new sequencing technologies have emerged. Exome and whole genome sequencing data in large case series can be readily exploited to detect chimeric genes, and will reveal whether the overall contribution of this class of mutation to schizophrenia is similar to that in our series. Similar analyses in other neurodevelopmental illnesses may also be revealing. Beyond their role in disease, chimeras are intriguing candidates for emerging traits in the human lineage. We focused on extremely rare, case-specific chimeras, but a population genetics approach to chimeras present in specific populations, or chimeras undergoing selection has the potential to reveal fascinating new variants.

Furthermore, this analysis was limited to germline chimerism, but somatic mutations arising in development may also produce chimeras whose expression is restricted to the CNS, or even to specific brain regions. Deep sequencing of DNA and RNA from post-mortem brain tissue of affected individuals has the potential to reveal such somatic mutations, chimeric or otherwise.

On a much broader note, one of the key challenges facing geneticists and clinicians concerned with psychiatric illness is how to incorporate these complex genetic findings into

clinical practice. A daunting possibility is a complete overhaul of diagnostic categories, with new clinical entities defined both by genotype and phenotype. While I believe that serious gaps in our current knowledge preclude a comprehensive new nosology at present, targeting future studies to specifically fill in those gaps will prove useful. These approaches may include pharmacogenetic studies of subjects with known genetic lesions and targeted resequencing of genes involved in neurodevelopmental illnesses in mentally ill probands, regardless of diagnostic status. Another approach that I think will be enormously revealing is the generation of induced pluripotent stem cells (iPSCs) from many individuals, particularly with known genetic lesions. It may be that phenotypic characterization of neurons derived from these iPSCs will define groups of mutations leading to shared cellular phenotypes.

Bibliography

1. Velligan, D.I., Alphs, L., Lancaster, S., Morlock, R., and Mintz, J. (2009). Association between changes on the Negative Symptom Assessment scale (NSA-16) and measures of functional outcome in schizophrenia. Psychiatry research 169, 97-100.
2. Milev, P., Ho, B.C., Arndt, S., and Andreasen, N.C. (2005). Predictive values of neurocognition and negative symptoms on functional outcome in schizophrenia: a longitudinal first-episode study with 7-year follow-up. The American journal of psychiatry 162, 495-506.
3. Velligan, D.I., Mahurin, R.K., Diamond, P.L., Hazleton, B.C., Eckert, S.L., and Miller, A.L. (1997). The functional significance of symptomatology and cognitive function in schizophrenia. Schizophrenia research 25, 21-31.
4. Puig, O., Penades, R., Gasto, C., Catalan, R., Torres, A., and Salamero, M. (2008). Verbal memory, negative symptomatology and prediction of psychosocial functioning in schizophrenia. Psychiatry research 158, 11-17.
5. Leucht, S., Corves, C., Arbter, D., Engel, R.R., Li, C., and Davis, J.M. (2009). Second-generation versus first-generation antipsychotic drugs for schizophrenia: a meta-analysis. Lancet 373, 31-41.
6. Murray, R.M., and Lewis, S.W. (1987). Is schizophrenia a neurodevelopmental disorder? British medical journal 295, 681-682.
7. Thaker, G.K., and Carpenter, W.T., Jr. (2001). Advances in schizophrenia. Nature medicine 7, 667-671.
8. Cardno, A.G., and Gottesman, II. (2000). Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. American journal of medical genetics 97, 12-17.
9. Murphy, K.C., and Owen, M.J. (2001). Velo-cardio-facial syndrome: a model for understanding the genetics and pathogenesis of schizophrenia. The British journal of psychiatry : the journal of mental science 179, 397-402.
10. Karayiorgou, M., Morris, M.A., Morrow, B., Shprintzen, R.J., Goldberg, R., Borrow, J., Gos, A., Nestadt, G., Wolyniec, P.S., Lasseter, V.K., et al. (1995). Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. Proceedings of the National Academy of Sciences of the United States of America 92, 7612-7616.
11. Blackwood, D.H., Fordyce, A., Walker, M.T., St Clair, D.M., Porteous, D.J., and Muir, W.J. (2001). Schizophrenia and affective disorders--cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and P300 findings in a family. American journal of human genetics 69, 428-433.
12. Millar, J.K., Wilson-Annan, J.C., Anderson, S., Christie, S., Taylor, M.S., Semple, C.A., Devon, R.S., St Clair, D.M., Muir, W.J., Blackwood, D.H., et al. (2000). Disruption of two novel genes by a translocation co-segregating with schizophrenia. Human molecular genetics 9, 1415-1423.
13. Walsh, T., McClellan, J.M., McCarthy, S.E., Addington, A.M., Pierce, S.B., Cooper, G.M., Nord, A.S., Kusenda, M., Malhotra, D., Bhandari, A., et al. (2008). Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science 320, 539-543.

14. Stefansson, H., Rujescu, D., Cichon, S., Pietilainen, O.P., Ingason, A., Steinberg, S., Fossdal, R., Sigurdsson, E., Sigmundsson, T., Buizer-Voskamp, J.E., et al. (2008). Large recurrent microdeletions associated with schizophrenia. Nature 455, 232-236.
15. Kirov, G., Grozeva, D., Norton, N., Ivanov, D., Mantripragada, K.K., Holmans, P., International Schizophrenia, C., Wellcome Trust Case Control, C., Craddock, N., Owen, M.J., et al. (2009). Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. Human molecular genetics 18, 1497-1503.
16. Malhotra, D., and Sebat, J. (2012). CNVs: harbingers of a rare variant revolution in psychiatric genetics. Cell 148, 1223-1241.
17. Xu, B., Roos, J.L., Levy, S., van Rensburg, E.J., Gogos, J.A., and Karayiorgou, M. (2008). Strong association of de novo copy number mutations with sporadic schizophrenia. Nature genetics 40, 880-885.
18. Kirov, G., Pocklington, A.J., Holmans, P., Ivanov, D., Ikeda, M., Ruderfer, D., Moran, J., Chambert, K., Toncheva, D., Georgieva, L., et al. (2012). De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Molecular psychiatry 17, 142-153.
19. Malhotra, D., McCarthy, S., Michaelson, J.J., Vacic, V., Burdick, K.E., Yoon, S., Cichon, S., Corvin, A., Gary, S., Gershon, E.S., et al. (2011). High frequencies of de novo CNVs in bipolar disorder and schizophrenia. Neuron 72, 951-963.
20. Xu, B., Roos, J.L., Dexheimer, P., Boone, B., Plummer, B., Levy, S., Gogos, J.A., and Karayiorgou, M. (2011). Exome sequencing supports a de novo mutational paradigm for schizophrenia. Nature genetics 43, 864-868.
21. Xu, B., Ionita-Laza, I., Roos, J.L., Boone, B., Woodrick, S., Sun, Y., Levy, S., Gogos, J.A., and Karayiorgou, M. (2012). De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. Nature genetics 44, 1365-1369.
22. Girirajan, S., Rosenfeld, J.A., Cooper, G.M., Antonacci, F., Siswara, P., Itsara, A., Vives, L., Walsh, T., McCarthy, S.E., Baker, C., et al. (2010). A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. Nature genetics 42, 203-209.
23. Williams, H.J., Monks, S., Murphy, K.C., Kirov, G., O'Donovan, M.C., and Owen, M.J. (2013). Schizophrenia two-hit hypothesis in velo-cardio facial syndrome. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics 162B, 177-182.
24. Williams, N.M. (2011). Molecular mechanisms in 22q11 deletion syndrome. Schizophr Bull 37, 882-889.
25. Jaspers, K. (1963). General psychopathology.([Chicago]: University of Chicago Press).
26. Toro, C.T., and Deakin, J.F. (2007). Adult neurogenesis and schizophrenia: a window on abnormal early brain development? Schizophrenia research 90, 1-14.
27. Mao, Y., Ge, X., Frank, C.L., Madison, J.M., Koehler, A.N., Doud, M.K., Tassa, C., Berry, E.M., Soda, T., Singh, K.K., et al. (2009). Disrupted in schizophrenia 1 regulates neuronal progenitor proliferation via modulation of GSK3beta/beta-catenin signaling. Cell 136, 1017-1031.
28. Duan, X., Chang, J.H., Ge, S., Faulkner, R.L., Kim, J.Y., Kitabatake, Y., Liu, X.B., Yang, C.H., Jordan, J.D., Ma, D.K., et al. (2007). Disrupted-In-Schizophrenia 1 regulates integration of newly generated neurons in the adult brain. Cell 130, 1146-1158.
29. Ghashghaei, H.T., Weber, J., Pevny, L., Schmid, R., Schwab, M.H., Lloyd, K.C., Eisenstat, D.D., Lai, C., and Anton, E.S. (2006). The role of neuregulin-ErbB4 interactions on the

proliferation and organization of cells in the subventricular zone. Proceedings of the National Academy of Sciences of the United States of America 103, 1930-1935.
30. Liu, Y., Ford, B.D., Mann, M.A., and Fischbach, G.D. (2005). Neuregulin-1 increases the proliferation of neuronal progenitors from embryonic neural stem cells. Developmental biology 283, 437-445.
31. Schwartz, T.L., Sachdeva, S., and Stahl, S.M. (2012). Glutamate neurocircuitry: theoretical underpinnings in schizophrenia. Frontiers in pharmacology 3, 195.
32. Wang, Q., Charych, E.I., Pulito, V.L., Lee, J.B., Graziane, N.M., Crozier, R.A., Revilla-Sanchez, R., Kelly, M.P., Dunlop, A.J., Murdoch, H., et al. (2011). The psychiatric disease risk factors DISC1 and TNIK interact to regulate synapse composition and function. Molecular psychiatry 16, 1006-1023.
33. Kwon, O.B., Longart, M., Vullhorst, D., Hoffman, D.A., and Buonanno, A. (2005). Neuregulin-1 reverses long-term potentiation at CA1 hippocampal synapses. The Journal of neuroscience : the official journal of the Society for Neuroscience 25, 9378-9383.
34. Li, B., Woo, R.S., Mei, L., and Malinow, R. (2007). The neuregulin-1 receptor erbB4 controls glutamatergic synapse maturation and plasticity. Neuron 54, 583-597.
35. Yang, K., Lei, G., Jackson, M.F., and Macdonald, J.F. (2010). The involvement of PACAP/VIP system in the synaptic transmission in the hippocampus. Journal of molecular neuroscience : MN 42, 319-326.
36. Ma, X., Fei, E., Fu, C., Ren, H., and Wang, G. (2011). Dysbindin-1, a schizophrenia-related protein, facilitates neurite outgrowth by promoting the transcriptional activity of p53. Molecular psychiatry 16, 1105-1116.
37. Hayashi-Takagi, A., Takaki, M., Graziane, N., Seshadri, S., Murdoch, H., Dunlop, A.J., Makino, Y., Seshadri, A.J., Ishizuka, K., Srivastava, D.P., et al. (2010). Disrupted-in-Schizophrenia 1 (DISC1) regulates spines of the glutamate synapse via Rac1. Nature neuroscience 13, 327-332.
38. Miyoshi, K., Honda, A., Baba, K., Taniguchi, M., Oono, K., Fujita, T., Kuroda, S., Katayama, T., and Tohyama, M. (2003). Disrupted-In-Schizophrenia 1, a candidate gene for schizophrenia, participates in neurite outgrowth. Molecular psychiatry 8, 685-694.
39. Vacic, V., McCarthy, S., Malhotra, D., Murray, F., Chou, H.H., Peoples, A., Makarov, V., Yoon, S., Bhandari, A., Corominas, R., et al. (2011). Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. Nature 471, 499-503.
40. Friedman, J.I., Vrijenhoek, T., Markx, S., Janssen, I.M., van der Vliet, W.A., Faas, B.H., Knoers, N.V., Cahn, W., Kahn, R.S., Edelmann, L., et al. (2008). CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy. Molecular psychiatry 13, 261-266.
41. Ishiguro, H., Koga, M., Horiuchi, Y., Noguchi, E., Morikawa, M., Suzuki, Y., Arai, M., Niizato, K., Iritani, S., Itokawa, M., et al. (2010). Supportive evidence for reduced expression of GNB1L in schizophrenia. Schizophr Bull 36, 756-765.
42. Mata, I., Perez-Iglesias, R., Roiz-Santianez, R., Tordesillas-Gutierrez, D., Gonzalez-Mandly, A., Berja, A., Vazquez-Barquero, J.L., and Crespo-Facorro, B. (2010). Additive effect of NRG1 and DISC1 genes on lateral ventricle enlargement in first episode schizophrenia. NeuroImage 53, 1016-1022.
43. Ranz, J.M., and Parsch, J. (2012). Newly evolved genes: Moving from comparative genomics to functional studies in model systems: How important is genetic novelty for species adaptation and diversification? Bioessays 34, 477-483.

44. Viale, A., Courseaux, A., Presse, F., Ortola, C., Breton, C., Jordan, D., and Nahon, J.L. (2000). Structure and expression of the variant melanin-concentrating hormone genes: only PMCHL1 is transcribed in the developing human brain and encodes a putative protein. Molecular biology and evolution 17, 1626-1640.
45. Schmieder, S., Darre-Toulemonde, F., Arguel, M.J., Delerue-Audegond, A., Christen, R., and Nahon, J.L. (2008). Primate-specific spliced PMCHL RNAs are non-protein coding in human and macaque tissues. BMC evolutionary biology 8, 330.
46. Newman, J.C., Bailey, A.D., Fan, H.Y., Pavelitz, T., and Weiner, A.M. (2008). An abundant evolutionarily conserved CSB-PiggyBac fusion protein expressed in Cockayne syndrome. PLoS genetics 4, e1000031.
47. Bailey, A.D., Gray, L.T., Pavelitz, T., Newman, J.C., Horibata, K., Tanaka, K., and Weiner, A.M. (2012). The conserved Cockayne syndrome B-piggyBac fusion protein (CSB-PGBD3) affects DNA repair and induces both interferon-like and innate antiviral responses in CSB-null cells. DNA repair 11, 488-501.
48. Peisajovich, S.G., Garbarino, J.E., Wei, P., and Lim, W.A. (2010). Rapid diversification of cell signaling phenotypes by modular domain recombination. Science 328, 368-372.
49. Rogers, R.L., and Hartl, D.L. (2012). Chimeric genes as a source of rapid evolution in Drosophila melanogaster. Molecular biology and evolution 29, 517-529.
50. Rogers, R.L., Bedford, T., and Hartl, D.L. (2009). Formation and longevity of chimeric and duplicate genes in Drosophila melanogaster. Genetics 181, 313-322.
51. Heisterkamp, N., Stam, K., Groffen, J., de Klein, A., and Grosveld, G. (1985). Structural organization of the bcr gene and its role in the Ph' translocation. Nature 315, 758-761.
52. Annala, M.J., Parker, B.C., Zhang, W., and Nykter, M. (2013). Fusion genes and their discovery using high throughput sequencing. Cancer letters.
53. Edwards, P.A. (2010). Fusion genes and chromosome translocations in the common epithelial cancers. The Journal of pathology 220, 244-254.
54. Zhou, X., Chen, Q., Schaukowitch, K., Kelsoe, J.R., and Geyer, M.A. (2010). Insoluble DISC1-Boymaw fusion proteins generated by DISC1 translocation. Molecular psychiatry 15, 669-672.
55. Eykelenboom, J.E., Briggs, G.J., Bradshaw, N.J., Soares, D.C., Ogawa, F., Christie, S., Malavasi, E.L., Makedonopoulou, P., Mackie, S., Malloy, M.P., et al. (2012). A t(1;11) translocation linked to schizophrenia and affective disorders gives rise to aberrant chimeric DISC1 transcripts that encode structurally altered, deleterious mitochondrial proteins. Human molecular genetics 21, 3374-3386.
56. Holt, R., Sykes, N.H., Conceicao, I.C., Cazier, J.B., Anney, R.J., Oliveira, G., Gallagher, L., Vicente, A., Monaco, A.P., and Pagnamenta, A.T. (2012). CNVs leading to fusion transcripts in individuals with autism spectrum disorder. European journal of human genetics : EJHG 20, 1141-1147.
57. Buizer-Voskamp, J.E., Blauw, H.M., Boks, M.P., van Eijk, K.R., Veldink, J.H., Hennekam, E.A., Vorstman, J.A., Mulder, F., Tiemeier, H., Uitterlinden, A.G., et al. (2013). Increased paternal age and the influence on burden of genomic copy number variation in the general population. Human genetics 132, 443-450.
58. Nord, A.S., Roeb, W., Dickel, D.E., Walsh, T., Kusenda, M., O'Connor, K.L., Malhotra, D., McCarthy, S.E., Stray, S.M., Taylor, S.M., et al. (2011). Reduced transcript expression of genes affected by inherited and de novo CNVs in autism. European journal of human genetics : EJHG 19, 727-731.

59. Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., and Lee, C. (2004). Detection of large-scale variation in the human genome. Nature genetics 36, 949-951.
60. (2011). BrainSpan: Atlas of the Developing Human Brain. In. (
61. Chaignat, E., Yahya-Graison, E.A., Henrichsen, C.N., Chrast, J., Schutz, F., Pradervand, S., and Reymond, A. (2011). Copy number variation modifies expression time courses. Genome research 21, 106-113.
62. Hartl, D., Irmler, M., Romer, I., Mader, M.T., Mao, L., Zabel, C., de Angelis, M.H., Beckers, J., and Klose, J. (2008). Transcriptome and proteome analysis of early embryonic mouse brain development. Proteomics 8, 1257-1265.
63. Kupershmidt, I., Su, Q.J., Grewal, A., Sundaresh, S., Halperin, I., Flynn, J., Shekar, M., Wang, H., Park, J., Cui, W., et al. (2010). Ontology-based meta-analysis of global collections of high-throughput public data. PloS one 5.
64. Acharya, S., Foletta, V.C., Lee, J.W., Rayborn, M.E., Rodriguez, I.R., Young, W.S., 3rd, and Hollyfield, J.G. (2000). SPACRCAN, a novel human interphotoreceptor matrix hyaluronan-binding proteoglycan synthesized by photoreceptors and pinealocytes. The Journal of biological chemistry 275, 6945-6955.
65. Taniguchi, M., Fukunaka, A., Hagihara, M., Watanabe, K., Kamino, S., Kambe, T., Enomoto, S., and Hiromura, M. (2013). Essential role of the zinc transporter ZIP9/SLC39A9 in regulating the activations of Akt and Erk in B-cell receptor signaling pathway in DT40 cells. PloS one 8, e58022.
66. Meuillet, E.J. (2011). Novel Inhibitors of AKT: Assessment of a Different Approach Targeting the Pleckstrin Homology Domain. Current medicinal chemistry.
67. Masri, K.A., Appert, H.E., and Fukuda, M.N. (1988). Identification of the full-length coding sequence for human galactosyltransferase (beta-N-acetylglucosaminide: beta 1,4-galactosyltransferase). Biochemical and biophysical research communications 157, 657-663.
68. Lo, N.W., Shaper, J.H., Pevsner, J., and Shaper, N.L. (1998). The expanding beta 4-galactosyltransferase gene family: messages from the databanks. Glycobiology 8, 517-526.
69. Guillard, M., Morava, E., de Ruijter, J., Roscioli, T., Penzien, J., van den Heuvel, L., Willemsen, M.A., de Brouwer, A., Bodamer, O.A., Wevers, R.A., et al. (2011). B4GALT1-congenital disorders of glycosylation presents as a non-neurologic glycosylation disorder with hepatointestinal involvement. The Journal of pediatrics 159, 1041-1043 e1042.
70. Mengle-Gaw, L., McCoy-Haman, M.F., and Tiemeier, D.C. (1991). Genomic structure and expression of human beta-1,4-galactosyltransferase. Biochemical and biophysical research communications 176, 1269-1276.
71. Aveic, S., Pigazzi, M., and Basso, G. (2011). BAG1: the guardian of anti-apoptotic proteins in acute myeloid leukemia. PloS one 6, e26097.
72. Sun, N.F., Meng, Q.Y., Hu, S.Y., Tian, A.L., Wang, R.H., Liu, Z.X., and Xu, L. (2011). Correlation between the expression of the BAG-1 gene and clinicopathologic factors in colorectal cancer. Journal of cancer research and clinical oncology 137, 1419-1424.
73. Zhang, X.Y., Pfeiffer, H.K., Mellert, H.S., Stanek, T.J., Sussman, R.T., Kumari, A., Yu, D., Rigoutsos, I., Thomas-Tikhonenko, A., Seidel, H.E., et al. (2011). Inhibition of the single

downstream target BAG1 activates the latent apoptotic potential of MYC. Molecular and cellular biology 31, 5037-5045.
74. Kermer, P., Krajewska, M., Zapata, J.M., Takayama, S., Mai, J., Krajewski, S., and Reed, J.C. (2002). Bag1 is a regulator and marker of neuronal differentiation. Cell death and differentiation 9, 405-413.
75. Planchamp, V., Bermel, C., Tonges, L., Ostendorf, T., Kugler, S., Reed, J.C., Kermer, P., Bahr, M., and Lingor, P. (2008). BAG1 promotes axonal outgrowth and regeneration in vivo via Raf-1 and reduction of ROCK activity. Brain : a journal of neurology 131, 2606-2619.
76. Liman, J., Ganesan, S., Dohm, C.P., Krajewski, S., Reed, J.C., Bahr, M., Wouters, F.S., and Kermer, P. (2005). Interaction of BAG1 and Hsp70 mediates neuroprotectivity and increases chaperone activity. Molecular and cellular biology 25, 3715-3725.
77. Maeng, S., Hunsberger, J.G., Pearson, B., Yuan, P., Wang, Y., Wei, Y., McCammon, J., Schloesser, R.J., Zhou, R., Du, J., et al. (2008). BAG1 plays a critical role in regulating recovery from both manic-like and depression-like behavioral impairments. Proceedings of the National Academy of Sciences of the United States of America 105, 8766-8771.
78. Sroka, K., Voigt, A., Deeg, S., Reed, J.C., Schulz, J.B., Bahr, M., and Kermer, P. (2009). BAG1 modulates huntingtin toxicity, aggregation, degradation, and subcellular distribution. J Neurochem 111, 801-807.
79. Krumm, N., Sudmant, P.H., Ko, A., O'Roak, B.J., Malig, M., Coe, B.P., Project, N.E.S., Quinlan, A.R., Nickerson, D.A., and Eichler, E.E. (2012). Copy number variation detection and genotyping from exome sequence data. Genome research 22, 1525-1532.
80. Parker, W.E., Orlova, K.A., Parker, W.H., Birnbaum, J.F., Krymskaya, V.P., Goncharov, D.A., Baybis, M., Helfferich, J., Okochi, K., Strauss, K.A., et al. (2013). Rapamycin Prevents Seizures After Depletion of STRADA in a Rare Neurodevelopmental Disorder. Science translational medicine 5, 182ra153.
81. Orlova, K.A., Parker, W.E., Heuer, G.G., Tsai, V., Yoon, J., Baybis, M., Fenning, R.S., Strauss, K., and Crino, P.B. (2010). STRADalpha deficiency results in aberrant mTORC1 signaling during corticogenesis in humans and mice. The Journal of clinical investigation 120, 1591-1602.
82. Uhlik, M.T., Abell, A.N., Johnson, N.L., Sun, W., Cuevas, B.D., Lobel-Rice, K.E., Horne, E.A., Dell'Acqua, M.L., and Johnson, G.L. (2003). Rac-MEKK3-MKK3 scaffolding for p38 MAPK activation during hyperosmotic shock. Nature cell biology 5, 1104-1110.
83. Uhlmann-Schiffler, H., Kiermayer, S., and Stahl, H. (2009). The DEAD box protein Ddx42p modulates the function of ASPP2, a stimulator of apoptosis. Oncogene 28, 2065-2073.
84. Chao, T.H., Hayashi, M., Tapping, R.I., Kato, Y., and Lee, J.D. (1999). MEKK3 directly regulates MEK5 activity as part of the big mitogen-activated protein kinase 1 (BMK1) signaling pathway. The Journal of biological chemistry 274, 36035-36038.
85. Nakamura, K., Uhlik, M.T., Johnson, N.L., Hahn, K.M., and Johnson, G.L. (2006). PB1 domain-dependent signaling complex is required for extracellular signal-regulated kinase 5 activation. Molecular and cellular biology 26, 2065-2079.
86. Xu, B.E., Stippec, S., Lenertz, L., Lee, B.H., Zhang, W., Lee, Y.K., and Cobb, M.H. (2004). WNK1 activates ERK5 by an MEKK2/3-dependent mechanism. The Journal of biological chemistry 279, 7826-7831.

87. Nakamura, K., and Johnson, G.L. (2003). PB1 domains of MEKK2 and MEKK3 interact with the MEK5 PB1 domain for activation of the ERK5 pathway. The Journal of biological chemistry 278, 36989-36992.
88. Sun, W., Kesavan, K., Schaefer, B.C., Garrington, T.P., Ware, M., Johnson, N.L., Gelfand, E.W., and Johnson, G.L. (2001). MEKK2 associates with the adapter protein Lad/RIBP and regulates the MEK5-BMK1/ERK5 pathway. The Journal of biological chemistry 276, 5093-5100.
89. Garner, A.P., Weston, C.R., Todd, D.E., Balmanno, K., and Cook, S.J. (2002). Delta MEKK3:ER* activation induces a p38 alpha/beta 2-dependent cell cycle arrest at the G2 checkpoint. Oncogene 21, 8089-8104.
90. Huang, Q., Yang, J., Lin, Y., Walker, C., Cheng, J., Liu, Z.G., and Su, B. (2004). Differential regulation of interleukin 1 receptor and Toll-like receptor signaling by MEKK3. Nature immunology 5, 98-103.
91. Wang, X., Zhang, F., Chen, F., Liu, D., Zheng, Y., Zhang, Y., Dong, C., and Su, B. (2011). MEKK3 regulates IFN-gamma production in T cells through the Rac1/2-dependent MAPK cascades. Journal of immunology 186, 5791-5800.
92. Kim, K., Duramad, O., Qin, X.F., and Su, B. (2007). MEKK3 is essential for lipopolysaccharide-induced interleukin-6 and granulocyte-macrophage colony-stimulating factor production in macrophages. Immunology 120, 242-250.
93. Bonvin, C., Guillon, A., van Bemmelen, M.X., Gerwins, P., Johnson, G.L., and Widmann, C. (2002). Role of the amino-terminal domains of MEKKs in the activation of NF kappa B and MAPK pathways and in the regulation of cell proliferation and apoptosis. Cellular signalling 14, 123-131.
94. Isken, O., and Maquat, L.E. (2007). Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. Genes & development 21, 1833-1856.
95. Terada, K., and Mori, M. (2000). Human DnaJ homologs dj2 and dj3, and bag-1 are positive cochaperones of hsc70. The Journal of biological chemistry 275, 24728-24734.
96. Rosales-Hernandez, A., Beck, K.E., Zhao, X., Braun, A.P., and Braun, J.E. (2009). RDJ2 (DNAJA2) chaperones neural G protein signaling pathways. Cell stress & chaperones 14, 71-82.
97. Hageman, J., van Waarde, M.A., Zylicz, A., Walerych, D., and Kampinga, H.H. (2011). The diverse members of the mammalian HSP70 machine show distinct chaperone-like activities. The Biochemical journal 435, 127-142.
98. Banks, G.T., Haas, M.A., Line, S., Shepherd, H.L., Alqatari, M., Stewart, S., Rishal, I., Philpott, A., Kalmar, B., Kuta, A., et al. (2011). Behavioral and other phenotypes in a cytoplasmic Dynein light intermediate chain 1 mutant mouse. The Journal of neuroscience : the official journal of the Society for Neuroscience 31, 5483-5494.
99. Jain, S., Yoon, S.Y., Zhu, L., Brodbeck, J., Dai, J., Walker, D., and Huang, Y. (2012). Arf4 determines dentate gyrus-mediated pattern separation by regulating dendritic spine development. PloS one 7, e46340.
100. Klivenyi, P., Starkov, A.A., Calingasan, N.Y., Gardian, G., Browne, S.E., Yang, L., Bubber, P., Gibson, G.E., Patel, M.S., and Beal, M.F. (2004). Mice deficient in dihydrolipoamide dehydrogenase show increased vulnerability to MPTP, malonate and 3-nitropropionic acid neurotoxicity. J Neurochem 88, 1352-1360.
101. Calingasan, N.Y., Ho, D.J., Wille, E.J., Campagna, M.V., Ruan, J., Dumont, M., Yang, L., Shi, Q., Gibson, G.E., and Beal, M.F. (2008). Influence of mitochondrial enzyme

deficiency on adult neurogenesis in mouse models of neurodegenerative diseases. Neuroscience 153, 986-996.
102. Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic acids research 15, 8125-8148.
103. Dingwall, C., Robbins, J., Dilworth, S.M., Roberts, B., and Richardson, W.D. (1988). The nucleoplasmin nuclear location sequence is larger and more complex than that of SV-40 large T antigen. The Journal of cell biology 107, 841-849.
104. Pan, Y.W., Zou, J., Wang, W., Sakagami, H., Garelick, M.G., Abel, G., Kuo, C.T., Storm, D.R., and Xia, Z. (2012). Inducible and conditional deletion of extracellular signal-regulated kinase 5 disrupts adult hippocampal neurogenesis. The Journal of biological chemistry.
105. Li, T., Pan, Y.W., Wang, W., Abel, G., Zou, J., Xu, L., Storm, D.R., and Xia, Z. (2013). Targeted Deletion of the ERK5 MAP Kinase Impairs Neuronal Differentiation, Migration, and Survival during Adult Neurogenesis in the Olfactory Bulb. PloS one 8, e61948.
106. Cavanaugh, J.E., Jaumotte, J.D., Lakoski, J.M., and Zigmond, M.J. (2006). Neuroprotective role of ERK1/2 and ERK5 in a dopaminergic cell line under basal conditions and in response to oxidative stress. Journal of neuroscience research 84, 1367-1375.
107. Dwivedi, Y., Rizavi, H.S., Teppen, T., Sasaki, N., Chen, H., Zhang, H., Roberts, R.C., Conley, R.R., and Pandey, G.N. (2007). Aberrant extracellular signal-regulated kinase (ERK) 5 signaling in hippocampus of suicide subjects. Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology 32, 2338-2350.
108. Yang, K., Sheikh, A.M., Malik, M., Wen, G., Zou, H., Brown, W.T., and Li, X. (2011). Upregulation of Ras/Raf/ERK1/2 signaling and ERK5 in the brain of autistic subjects. Genes, brain, and behavior 10, 834-843.
109. Mingorance-Le Meur, A. (2006). JNK gives axons a second chance. The Journal of neuroscience : the official journal of the Society for Neuroscience 26, 12104-12105.
110. Tong, L., Prieto, G.A., Kramar, E.A., Smith, E.D., Cribbs, D.H., Lynch, G., and Cotman, C.W. (2012). Brain-derived neurotrophic factor-dependent synaptic plasticity is suppressed by interleukin-1beta via p38 mitogen-activated protein kinase. The Journal of neuroscience : the official journal of the Society for Neuroscience 32, 17714-17724.
111. Sweatt, J.D. (2001). The neuronal MAP kinase cascade: a biochemical signal integration system subserving synaptic plasticity and memory. J Neurochem 76, 1-10.
112. Lin, C.W., Cheng, C.W., Yang, T.C., Li, S.W., Cheng, M.H., Wan, L., Lin, Y.J., Lai, C.H., Lin, W.Y., and Kao, M.C. (2008). Interferon antagonist function of Japanese encephalitis virus NS4A and its interaction with DEAD-box RNA helicase DDX42. Virus research 137, 49-55.
113. Sottocornola, R., Royer, C., Vives, V., Tordella, L., Zhong, S., Wang, Y., Ratnayaka, I., Shipman, M., Cheung, A., Gaston-Massuet, C., et al. (2010). ASPP2 binds Par-3 and controls the polarity and proliferation of neural progenitors during CNS development. Developmental cell 19, 126-137.
114. Jia, J., Yao, P., Arif, A., and Fox, P.L. (2013). Regulation and dysregulation of 3'UTR-mediated translational control. Current opinion in genetics & development 23, 29-34.

115. Tang, M., Ivakine, E., Mahadevan, V., Salter, M.W., and McInnes, R.R. (2012). Neto2 interacts with the scaffolding protein GRIP and regulates synaptic abundance of kainate receptors. PloS one 7, e51433.
116. Ivakine, E.A., Acton, B.A., Mahadevan, V., Ormond, J., Tang, M., Pressey, J.C., Huang, M.Y., Ng, D., Delpire, E., Salter, M.W., et al. (2013). Neto2 is a KCC2 interacting protein required for neuronal Cl- regulation in hippocampal neurons. Proceedings of the National Academy of Sciences of the United States of America 110, 3561-3566.
117. Walker, V.E., Wong, M.J., Atanasiu, R., Hantouche, C., Young, J.C., and Shrier, A. (2010). Hsp40 chaperones promote degradation of the HERG potassium channel. The Journal of biological chemistry 285, 3319-3329.
118. Horne, B.E., Li, T., Genevaux, P., Georgopoulos, C., and Landry, S.J. (2010). The Hsp40 J-domain stimulates Hsp70 when tethered by the client to the ATPase domain. The Journal of biological chemistry 285, 21679-21688.
119. Vos, M.J., Hageman, J., Carra, S., and Kampinga, H.H. (2008). Structural and functional diversities between members of the human HSPB, HSPH, HSPA, and DNAJ chaperone families. Biochemistry 47, 7001-7011.
120. Stirling, P.C., Bakhoum, S.F., Feigl, A.B., and Leroux, M.R. (2006). Convergent evolution of clamp-like binding sites in diverse chaperones. Nature structural & molecular biology 13, 865-870.
121. Michels, A.A., Kanon, B., Bensaude, O., and Kampinga, H.H. (1999). Heat shock protein (Hsp) 40 mutants inhibit Hsp70 in mammalian cells. The Journal of biological chemistry 274, 36757-36763.
122. Kleines, M., Gartner, A., Ritter, K., and Schaade, L. (2001). Cloning and expression of the human single copy homologue of the mouse zinc finger protein zfr. Gene 275, 157-162.
123. Elvira, G., Massie, B., and DesGroseillers, L. (2006). The zinc-finger protein ZFR is critical for Staufen 2 isoform specific nucleocytoplasmic shuttling in neurons. J Neurochem 96, 105-117.
124. Meagher, M.J., Schumacher, J.M., Lee, K., Holdcraft, R.W., Edelhoff, S., Disteche, C., and Braun, R.E. (1999). Identification of ZFR, an ancient and highly conserved murine chromosome-associated zinc finger protein. Gene 228, 197-211.
125. Meagher, M.J., and Braun, R.E. (2001). Requirement for the murine zinc finger protein ZFR in perigastrulation growth and survival. Molecular and cellular biology 21, 2880-2890.
126. Schmitz, C., Kinge, P., and Hutter, H. (2007). Axon guidance genes identified in a large-scale RNAi screen using the RNAi-hypersensitive Caenorhabditis elegans strain nre-1(hd20) lin-15b(hd126). Proceedings of the National Academy of Sciences of the United States of America 104, 834-839.
127. Barbee, S.A., Estes, P.S., Cziko, A.M., Hillebrand, J., Luedeman, R.A., Coller, J.M., Johnson, N., Howlett, I.C., Geng, C., Ueda, R., et al. (2006). Staufen- and FMRP-containing neuronal RNPs are structurally and functionally related to somatic P bodies. Neuron 52, 997-1009.
128. Perycz, M., Urbanska, A.S., Krawczyk, P.S., Parobczak, K., and Jaworski, J. (2011). Zipcode binding protein 1 regulates the development of dendritic arbors in hippocampal neurons. The Journal of neuroscience : the official journal of the Society for Neuroscience 31, 5271-5285.

129. Raes, J., and Van de Peer, Y. (2005). Functional divergence of proteins through frameshift mutations. Trends in genetics : TIG 21, 428-431.
130. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6, 343-345.
131. Mostafavi, S., Ray, D., Warde-Farley, D., Grouios, C., and Morris, Q. (2008). GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. Genome biology 9 Suppl 1, S4.

Appendix: Primer sequences

Name	Primer sequence	Purpose	Target
MATK_F	TTGGCATCCTGGCTGAGACG	MATK expression	cDNA
MATK_R	TGGAGAAGCAACACCCCTCTC	MATK expression	
ZFR2_ex4_F	GGTAGGTGTATCCCGAGGACAAG	ZFR2 expression	cDNA
ZFR2_ex2_R	ACAACCCACTCCTGGGATGG	ZFR2 expression	
ZFR2-MATK_F	GGTAGGTGTATCCCGAGGACAAG	MATK-ZFR2 chimera expression	patient cDNA
ZFR2-MATK_R	AAGCAGTTGGTTAGCGAAGGG	MATK-ZFR2 chimera expression	
ZFR2_ex13-14_F	CCCGGCAGCTCCAGATGG	ZFR2 expression, alternate splicing of exon 15	cDNA
ZFR2_ex16-17_R	CATGGCCCAGGCTGGC	ZFR2 expression, alternate splicing of exon 15	
Dup16_bkpt_F	CACAAACATGCCATCATATTGAGGGCTGGG	DNAJA2-NETO2 breakpoint primers	patient gDNA
Dup16_bkpt_R	ATGACAGATACGGAGAGCAAGGTCTTCGGG	DNAJA2-NETO2 breakpoint primers	
Dup19_bkpt_F	TTAGATGAGTCAGGGTTCCAGCAGAGAACG	MATK-ZFR2 breakpoint primers	patient gDNA
Dup19_bkpt_R	GGGTTGCTAGGCGACATCGGTGAGTGTAGG	MATK-ZFR2 breakpoint primers	
Dup14_bkpt_F	AGAGTTCCTGTTGCTCCACATCCTTGCCAG	PLEKHD1-SLC39A9 breakpoint primers	patient gDNA
Dup14_bkpt_R	GTTGACATCTGGGTAAAATGCGGAGGGTTG	PLEKHD1-SLC39A9 breakpoint primers	
Del17_bkpt_F	GCCTCCTTGACAGACATTTTCGTTCCTCAG	MAP3K3-DDX42 breakpoint primers	patient gDNA
Del17_bkpt_R	CTGCATCGAGAAGACCTTGAACAATCCCAG	MAP3K3-DDX42 breakpoint primers	
MAP-DDX_cl_F	ATGGACGAACAGGAGGCATTGAACT	cloning of MAP3K3-DDX42	patient cDNA
MAP-DDX_cl_R	CGAACTGTCCCATCGACTTTTCTTG	cloning of MAP3K3-DDX42	
DDX42_WT_cl_F	GGCACCATGAACTGGAATAA	cloning of full-length DDX42	patient cDNA
DDX42_WT_cl_R	TTTAGCACATCCCCTCGAAC	cloning of full-length DDX42	
MAP3K3_WT_cl_F	CACCATGGACGAACAGGA	cloning of full-length MAP3K3	patient cDNA
MAP3K3_WT_cl_R	GGCCGTGAGAGCTGAGTACA	cloning of full-length MAP3K3	
DNA-NET_cl_F	ATGGCTAACGTGGCTGACACG	cloning of DNAJA2-NETO2	patient cDNA
DNA-NET_cl_R	TGAGAAGTCAATGGATATGGATGCT	cloning of DNAJA2-NETO2	
DNAJA2_WT_cl_F	ATGGCTAACGTGGCTGAC	cloning of full-length DNAJA2	patient cDNA
DNAJA2_WT_cl_R	TTTGCAGAGTTGACTGATGG	cloning of full-length DNAJA2	
PLEK-SLC_cl_F	ATGTTCACGTCCAAGTCCAACTCGG	cloning of PLEKHD1-SLC39A9	patient cDNA
PLEK-SLC_cl_R	GTAATGCTGGTGTCCTACTGACAGG	cloning of PLEKHD1-SLC39A9	

PLEK_WT_cl_F	GATGTTCACGTCCAAGTCCA	cloning of full-length PLEKHD1	patient cDNA
PLEK_WT_cl_R	AGGAGCGCCCATCTCTTT	cloning of full-length PLEKHD1	
SLC_WT_cl_F	AAAGGAGGGCAGAATGGA	cloning of full-length SLC39A9	patient cDNA
SLC_WT_cl_R	CATTGAATGCTGGTGTCC	cloning of full-length SLC39A9	
ZFR2_nest1_F	AAGATGGCGACGAGTCAGTATTTC	outer nesting primers for first half of ZFR2	brain cDNA
ZFR2_nest1_R	GCTGAACACCTCCTCCACATATTC	outer nesting primers for first half of ZFR2	
ZFR2_nest2_F	TGGGGGAAACCAGCCCAACC	outer nesting primers for second half of ZFR2	brain cDNA
ZFR2_nest2_R	GAGGTAGGCGGCTAACACGAG	outer nesting primers for second half of ZFR2	
MATK-ZFR2_gibs_F2	cgggttgctaggcgacatcgCGCCCAGCCTCCGACC	MATK-ZFR2 isoform 1 Gibson assembly	ZFR2_nest1 PCR product
MATK-ZFR2_gibs_R2	GAAGCGAAGCACTCGCCCTTCGTCGCTGAA CACCTCCTCCACATATTC	MATK-ZFR2 isoform 1 Gibson assembly	
MATK-ZFR2_gibs_F4	CTCGTGTTAGCCGCCTACCTCAAGGGCAATT CTGCAGATATCCAG	MATK-ZFR2 isoform 1 Gibson assembly	miniprepped vector
MATK-ZFR2_gibs_R4	accaactgcttgattttgctAAGGGCAATTCCACCACA CTGG	MATK-ZFR2 isoform 1 Gibson assembly	
MATK-ZFR2_gibs_F3	GAATATGTGGAGGAGGTGTTCAGCG	ZFR2, MATK-ZFR2 isoform 1 Gibson assembly	ZFR2_nest2 PCR product
MATK-ZFR2_gibs_R3	CTGGATATCTGCAGAATTGCCCTTGAGGTAG GCGGCTAACACGAG	ZFR2, MATK-ZFR2 isoform 1 Gibson assembly	
MATK-ZFR2_gibs_F1	CCAGTGTGGTGGAATTGCCCTTagcaaaatcaag cagttggttagcg	MATK-ZFR2 isoform 1 Gibson assembly	gDNA
MATK-ZFR2_gibs_R1	GAAGGGTCGGAGGCTGGGCGcgatgtcgcctagc aacccg	MATK-ZFR2 isoform 1 Gibson assembly	
ZFR2_gibs_F1	AGTGTGGTGGAATTGCCCTTATGGCGACGAG TCAGTATTTCGAC	ZFR2 isoform 1 and 2 Gibson assembly	ZFR2_nest1 PCR product
ZFR2_gibs_R1	GAAGCGAAGCACTCGCCCTTCGTCGCTGAA CACCTCCTCCACATATTC	ZFR2 isoform 1 and 2 Gibson assembly	
ZFR2_nest3_F	GCAACTCAGCCTCCCAAAGTG	outer nesting primers for C-terminus of ZFR2 isoform 2	brain cDNA
ZFR2_nest3_R	ATCTCCTGCCGCCCAGACACCG	outer nesting primers for C-terminus of ZFR2 isoform 2	
ZFR2_iso2_gibs_F2	CTCGTGTTAGCCGCCTACCT	ZFR2 isoform 1 and 2 Gibson assembly	ZFR2_nest3 PCR product
ZFR2_iso2_gibs_R2	atatctgcagaattgcccttTGACTCAAAAGGAAATGA CCA	ZFR2 isoform 1 and 2 Gibson assembly	
ZFR2_iso2_gibs_F1	AGTGTGGTGGAATTGCCCTTATGGCGACGAG TCAGTATTTCGAC	ZFR2 isoform 2 Gibson assembly	miniprepped ZFR2 isoform2

ZFR2_iso2_gibs_R1	GAGGTAGGCGGCTAACACGAG	ZFR2 isoform 2 Gibson assembly	
ZFR2_iso2_gibs_F3	GGTCATTTCCTTTTGAGTAAAAGGGCAATTCTGCAGATATCCAG	ZFR2 isoform 2 Gibson assembly	miniprepped ZFR2 isoform2
ZFR2_iso2_gibs_R3	GTCGAAATACTGACTCGTCGCCATAAGGGCAATTCCACCACACTGG	ZFR2 isoform 2 Gibson assembly	

Curriculum Vitae

Education	**MD** University of Washington	expected 2015 Seattle, WA
	PhD, Genome Sciences, University of Washington Dissertation: *Formation of Chimeric Genes by Copy Number Variation as a Mutational Mechanism in Schizophrenia*	2013 Seattle, WA
	B.S., Neurobiology, University of Washington	2003 Seattle, WA
	B.A., Comparative Literature, University of Washington	2003 Seattle, WA
Research	**Research Fellow** National Institute of Mental Health, Lab of Genetics PI: Michael Brownstein, MD/PhD	Bethesda, MD 6/01 – 8/01
	Research Assistant University of Washington, Department of Psychology PI: Lee Osterhout, PhD	Seattle, WA 6/02 – 9/03
	Research Technician Center for Disease Control PI: Robert Downing, PhD	Entebbe, Uganda 10/03 – 03/04
	Research Technician Fred Hutchinson Cancer Research Center PIs: Steve DeRosa, MD; Julie McElrath, MD/PhD	Seattle, WA 08/04 – 07/05
	Rotation Student University of Washington, Pathology Department PI: Robert Hevner, MD/PhD	Seattle, WA 07/05 – 09/05
	Rotation Student University of Washington, Neurology Department PI: Gwenn Garden, MD/PhD	Seattle, WA 07/06 - 09/06
	Doctoral Student University of Washington, Genome Sciences Department PIs: Mary-Claire King, PhD; Jon McClellan, MD	Seattle, WA 08/07 – present

Publications

Tom Walsh, Jon M. McClellan, Shane E. McCarthy, Anjene M. Addington, Sarah B. Pierce, Greg M. Cooper, Alex S. Nord, Mary Kusenda, Dheeraj Malhotra, Abishek Bhandari, Sunday M. Stray, Caitlin F. Rippey, Patricia Roccanova, Vlad Makarov, B. Lakshmi, Robert L. Findling, Linmarie Sikich, Thomas Stromberg, Barry Merriman, Nitin Gogtay, Philip Butler, Kristen Eckstrand, Laila Noory, Peter Gochman, Robert Long, Zugen Chen, Sean Davis, Carl Baker, Evan E. Eichler, Paul S. Meltzer, Stanley F. Nelson, Andrew B. Singleton, Ming K. Lee, Judith L. Rapoport, Mary-Claire King, Jonathan Sebat. Rare Structural Variants Disrupt Multiple Genes in Neurodevelopmental Pathways in Schizophrenia, **Science**, 2008 Apr 25;320(5875):539-43.

Hodge RD, Kowalczyk TD, Wolf SA, Encinas JM, Rippey C, Enikolopov G, Kempermann G, Hevner RF. Intermediate Progenitors in Adult Hippocampal Neurogenesis: Tbr2 Expression and Coordinate Regulation of Neuronal Output, **Journal of Neuroscience**, 2008 Apr 2; 28(14):3707-3717.

Grants & Awards

NIMH summer research fellowship	2001
Mary Gates Fellowship for Undergraduate Research	2002-03
Achievement Reward for College Scientists Foundation award	2008-10
NIMH F30 recipient	2008-12

Foreign Language

French (Fluent), Spanish, Arabic (Conversational)