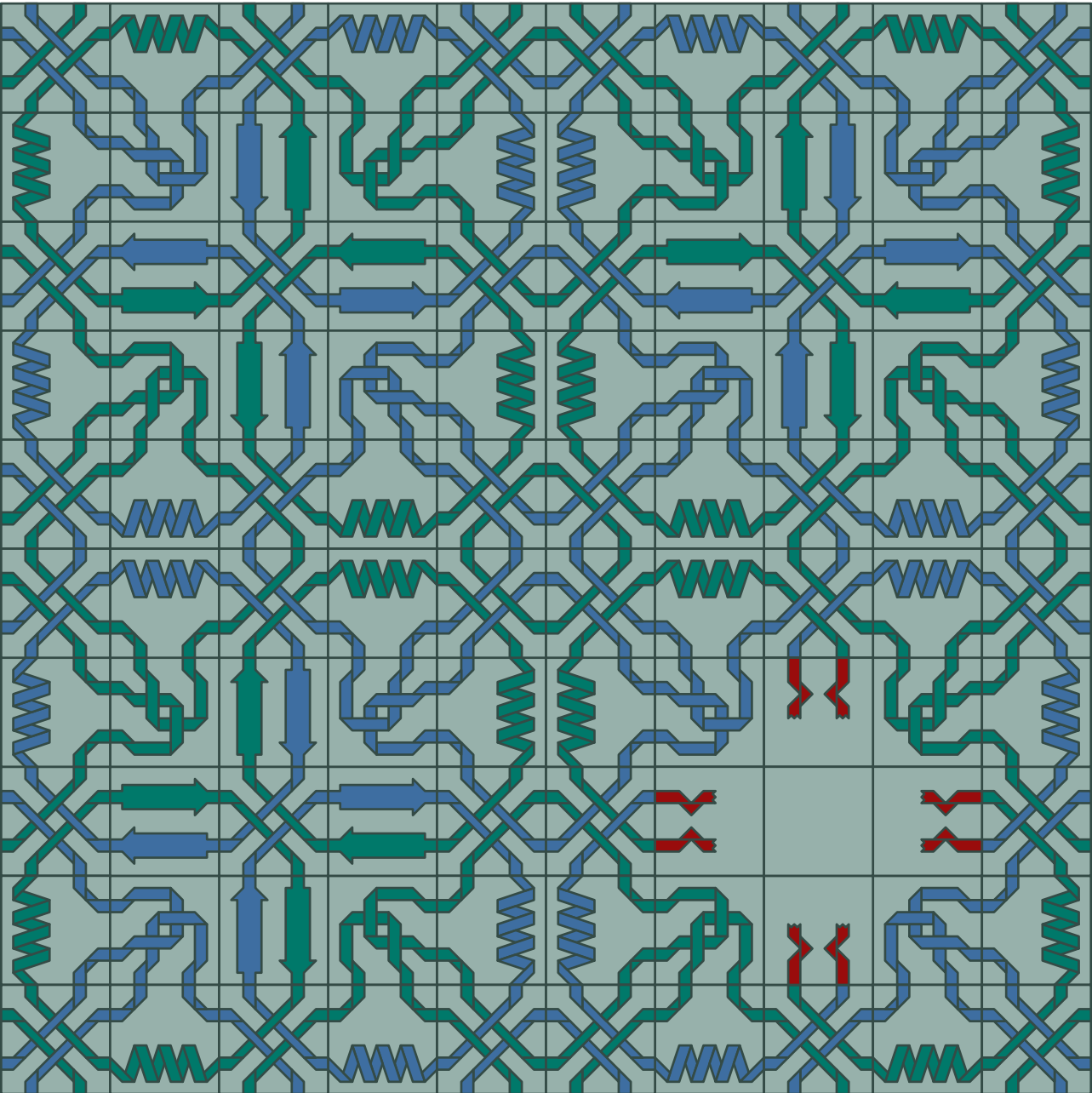


M A X
P L A
N C K

MAX PLANCK INSTITUTE
FOR PSYCHOLINGUISTICS

WWW.MPI.NL



A Network of Interacting Proteins Disrupted in Language-Related Disorders

ELLIOT KURT SOLLIS



A NETWORK OF INTERACTING PROTEINS
DISRUPTED IN LANGUAGE-RELATED DISORDERS

© 2019, Elliot Sollis

ISBN: 978-94-92910-04-2

Cover illustration by Elliot Sollis. Inspired by protein secondary structure and the *azulejos* of Lisbon.

Printed and bound by Ipskamp Drukkers b.v.

A Network of Interacting Proteins Disrupted in Language-Related Disorders

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op donderdag 21 november 2019
om 12.30 uur precies

door

Elliot Kurt Sollis

geboren op 4 december 1989
te Perth, Australië

Promotor:

Prof. dr. S.E. Fisher

Copromotor:

Dr. P. Derizioti (MPI)

Manuscriptcommissie:

Prof. dr. ir. J.H.L.M. van Bokhoven

Prof. dr. J.M.J. Kremer

Dr. D.F. Newbury (Oxford Brookes University, Verenigd Koninkrijk)

A Network of Interacting Proteins Disrupted in Language-Related Disorders

Doctoral Thesis

to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus prof. dr. J.H.J.M. van Krieken,
according to the decision of the Council of Deans
to be defended in public on Thursday 21 November 2019
at 12.30 hours

by

Elliot Kurt Sollis

born on 4 December 1989

in Perth, Australia

Supervisor:

Prof. dr. S.E. Fisher

Co-supervisor:

Dr. P. Derizioti (MPI)

Doctoral Thesis Committee:

Prof. dr. ir. J.H.L.M. van Bokhoven

Prof. dr. J.M.J. Kremer

Dr. D.F. Newbury (Oxford Brookes University, United Kingdom)

CONTENTS

1	INTRODUCTION	9
1.1	Language and disorder.....	9
1.2	FOXP2 haploinsufficiency disorder	9
1.3	The search for new language-related genes.....	10
1.4	Obstacles to discovery	13
1.5	Casting a wider net	14
1.6	Protein interaction networks in language-related disorders.....	14
1.7	Proteomics techniques discussed in this thesis.....	15
1.8	Aims of this thesis	25
2	IDENTIFICATION AND FUNCTIONAL CHARACTERISATION OF <i>DE NOVO</i> <i>FOXP1</i> VARIANTS PROVIDES NOVEL INSIGHTS INTO THE AETIOLOGY OF NEURODEVELOPMENTAL DISORDER	27
2.1	Introduction	29
2.2	Methods	30
2.3	Results	33
2.4	Discussion.....	53
3	EQUIVALENT MISSENSE VARIANT IN THE <i>FOXP2</i> AND <i>FOXP1</i> TRANSCRIPTION FACTORS CAUSES DISTINCT NEURODEVELOPMENTAL DISORDERS	59
3.1	Introduction	61
3.2	Methods	63
3.3	Results	66
3.4	Discussion.....	80
4	FOXP PROTEINS INTERACT WITH THE CHROMATIN REMODELLING FACTOR GATAD2B	85
4.1	Introduction	87
4.2	Methods	91
4.3	Results	93
4.4	Discussion.....	109
5	MASS SPECTROMETRY-BASED SCREENING FOR TBR1-INTERACTING PROTEINS....	115
5.1	Introduction	117
5.2	Methods	118
5.3	Results	122
5.4	Discussion.....	147
5.5	Appendix	151

6	TBR1 INTERACTS WITH TRANSCRIPTIONAL REGULATORS DISRUPTED IN ASD/ID-RELATED DISORDERS	157
6.1	Introduction	159
6.2	Methods	164
6.3	Results	166
6.4	Discussion.....	213
7	GENERAL DISCUSSION	219
7.1	Background.....	219
7.2	Interactions among language-related proteins	219
7.3	An emerging network of transcriptional regulators in neurodevelopment....	221
7.4	Characterising mutations in language-related disorders.....	224
7.5	The importance of functional characterisation	226
7.6	Conclusions and future directions	228
	REFERENCES	231
	SAMENVATTING.....	249
	SUMMARY	255
	CURRICULUM VITAE	261
	PUBLICATIONS	261
	ACKNOWLEDGEMENTS	265
	MPI SERIES IN PSYCHOLINGUISTICS	269

1 INTRODUCTION

1.1 LANGUAGE AND DISORDER

Language is a complex and uniquely human ability that enables communication and cultural transmission between people. The most common modality for language is speech, but it can be readily transferred to other modalities with comparable complexity and effectiveness, as in sign languages. Given adequate exposure to linguistic input, children are able to acquire any language without conscious effort or formal tuition and are able to produce complex sentences by ~4 years of age (Dosman et al., 2012). In contrast, other primates – including our closest relatives, chimpanzees and bonobos – do not reach this level of linguistic competence even when raised in a human environment and given extensive training (Graham et al., 2015). The apparent uniqueness and innateness of language capacities in humans suggest that genetic factors may underlie the neurodevelopmental processes that make language possible. Indeed, genetic contributions to language are supported by a) the tendency for language disorders to run in families and b) the higher concordance rate of language traits and disorders in monozygotic vs dizygotic twins (Stromswold, 2006; Newbury et al., 2010). Language disorders therefore represent a valuable starting point to understand the molecular mechanisms that are fundamental to the human language capacity.

Several distinct disorders have been described in which speech/language impairment is the core phenotype (Graham and Fisher, 2015). Specific language impairment (SLI, also now known as developmental language disorder, DLD) is a delay or impairment in language acquisition in the absence of other explanatory physical, neurological or environmental causes (e.g. oral malformations, hearing loss, intellectual disability (ID), lack of linguistic input during development). Childhood apraxia of speech (CAS), also known as developmental verbal dyspraxia, is characterised by impaired acquisition and coordination of the fine orofacial movements required for speech, leading to frequent speech errors, sometimes accompanied by other deficits in expressive and receptive language. Stuttering is an impairment of speech fluency, including involuntary pauses, and the repetition or extension of sounds and syllables. Developmental dyslexia, also known as reading disorder, is a difficulty in learning to read or spell that cannot be explained by general cognitive impairment, visual problems or poor learning opportunities. From a research perspective, identifying genetic variants that disrupt speech and language can help us to uncover the molecular pathways that enable these unique abilities.

1.2 *FOXP2* HAPLOINSUFFICIENCY DISORDER

At the beginning of this PhD project, only one gene had been directly linked to a distinct disorder that disproportionately affects speech/language. *FOXP2* (OMIM 605317) was identified through linkage analysis of a multigenerational UK pedigree (KE) with an autosomal dominant form of severe speech and language disorder (Fisher et al., 1998)

(OMIM 602081). Subsequent DNA sequencing identified a point mutation leading to an arginine-to-histidine substitution at residue 553 of the *FOXP2* protein in the affected family members (Lai et al., 2001). The presence of a balanced chromosomal translocation also disrupting *FOXP2* in an unrelated individual (CS) with a very similar phenotype had also been instrumental in pinpointing this locus (Lai et al., 2001). Since then, at least eight additional independent pathogenic variants have been identified that disrupt the *FOXP2* coding region, including nonsense, frameshift and missense variants and intragenic deletions (MacDermot et al., 2005; Turner et al., 2013; Reuter et al., 2017), as well as a number of larger 7q31 deletions and reciprocal translocations that affect *FOXP2* in addition to other genes (Feuk et al., 2006; Shriberg et al., 2006; Zeesman et al., 2006; Lennon et al., 2007; Tomblin et al., 2009; Palka et al., 2012; Rice et al., 2012; Zilina et al., 2012). Heterozygous disruption of *FOXP2* is sufficient to cause the disorder, suggesting haploinsufficiency as the major pathogenic mechanism.

FOXP2 is one of 50 human transcription factors (TFs) in the FOX protein family, which share a conserved forkhead box (FOX) domain required for DNA-binding and transcriptional regulation (Li et al., 2004; Vernes et al., 2006). It also contains leucine zipper and zinc finger motifs important for interactions with other proteins, and a glutamine-rich region of uncertain function. All known pathogenic *FOXP2* variants either damage the FOX domain or truncate the protein before the FOX domain, thereby preventing transcriptional regulation by damaging nuclear localisation and/or DNA-binding (Vernes et al., 2006; Estruch et al., 2016a).

The core phenotype associated with *FOXP2* variants is CAS, with typically delayed and highly unintelligible speech during childhood (Morgan et al., 2017). Other findings include oral motor dyspraxia (impairment in planning and sequencing of non-verbal oral movements); dysarthria (difficult or abnormal articulation of phonemes); moderate-to-severe receptive and expressive language deficits, with expressive language typically more severely affected; and reading and spelling impairments. Most affected individuals have normal gross and fine motor skills. In addition to low verbal IQ, non-verbal IQ may also be lower than average, though usually not sufficiently impaired to constitute ID (Morgan et al., 2017). However, while the link between *FOXP2* and CAS is solid, *FOXP2* mutations explain a small proportion of CAS cases: 1/49 cases, or ~2%, in one study (MacDermot et al., 2005); and they do not account for other forms of speech/language disorder, such as SLI, dyslexia and stuttering (Morgan et al., 2017).

1.3 THE SEARCH FOR NEW LANGUAGE-RELATED GENES

After the discovery of *FOXP2*, a number of studies sought to identify new genes involved in speech/language disorders. Traditional linkage and targeted association studies have detected several candidate genes that may be associated with stuttering, developmental dyslexia and specific language impairment (Table 1.1). However, many of these associations have not been replicated, and most of the genes remain tentative candidates (Deriziotis and Fisher, 2017).

Table 1.1. Representative selection of speech/language-related disorder candidate genes identified through traditional linkage and targeted association studies.

Disorder	Candidate gene	Reference
Stuttering	GNPTAB	Kang et al. (2010)
	GNPTG	Kang et al. (2010)
	NAGPA	Kang et al. (2010)
	AP4E1	Raza et al. (2015)
Dyslexia	ROBO1	Hannula-Jouppi et al. (2005)
	KIAA0319	Francks et al. (2004), Paracchini et al. (2006)
	DCDC2	Francks et al. (2004), Meng et al. (2005)
	DNAAF4	Taipale et al. (2003)
	MRPL19/GCFC2	Anthoni et al. (2007)
SLI	CNTNAP2	Vernes et al. (2008)
	CMIP	Newbury et al. (2009)
	ATP2C2	Newbury et al. (2009)
	NFXL1	Villanueva et al. (2015)

SLI = specific language impairment.

Since then, researchers have adopted larger-scale techniques to uncover more of the genetic basis for speech/language. Two complementary strategies have been employed: the first treats language-related phenotypes as complex traits influenced by multiple common variants with small effect sizes, while the second approach searches for rare high-penetrance variants responsible for monogenic speech/language disorders (Deriziotis and Fisher, 2017).

1.3.1 SPEECH AND LANGUAGE DISORDERS AS COMPLEX TRAITS

Investigations of speech and language as complex traits have relied on genome-wide association studies (GWAS) to search for the effects of common genetic variants on speech/language-related phenotypes. So far, these studies have revealed few significant associations (Graham and Fisher, 2015; Deriziotis and Fisher, 2017). In many cases, they have been hampered by limited sample sizes relative to the expected small effects of the putative contributing variants (Deriziotis and Fisher, 2017). Case-control studies on speech/language disorders, with case numbers in the low hundreds, have been particularly underpowered to identify significant associations (Eicher et al., 2013; Kornilov et al., 2016). Studies of speech/language abilities as quantitative measurements have achieved larger sample sizes (up to ~10,000 individuals), but even then, only two genome-wide significant associations have been identified: a locus at 3p22.2 near the *SCN11A* gene, associated with social communication ability (St Pourcain et al., 2014b); and another at 3p12.3 near the *ROBO2* gene, associated with expressive vocabulary in early infancy (St Pourcain et al., 2014a). Another major challenge when studying language-related impairments has been the difficulty of diagnosing this highly varied group of disorders. In some studies, increasing the sample size has come at the cost of making cohorts more phenotypically heterogeneous (Eicher et al., 2013; Gialluisi et al., 2014). The field is likely to benefit from ongoing efforts to combine existing data from multiple cohorts in larger meta-analyses with increased power

(Deriziotis and Fisher, 2017). As a comparison, GWAS investigations of another brain-related phenotype – schizophrenia – have required sample numbers in the tens of thousands in order to discover robust and replicable associations (Bergen and Petryshen, 2012), and similar numbers may be required for speech/language disorders.

1.3.2 MONOGENIC SPEECH AND LANGUAGE DISORDERS

More recently, there has been a hope that next generation sequencing methods such as whole exome sequencing (WES) and whole genome sequencing (WGS) might identify new causative mutations responsible for monogenic speech and language disorders. These techniques have been used to great effect for other developmental disorders, such as autism spectrum disorder (ASD) and ID, by focussing on rare *de novo* variants in severe, sporadic cases (Vissers et al., 2010; O’Roak et al., 2011).

The first WES study in patients with CAS identified heterozygous variants in *FOXP1*, *CNTNAP2*, *CNTNAP1*, *ATP13A4*, *KIAA0319* and *SETX* (Worthey et al., 2013). However, the study has been criticised (Deriziotis and Fisher, 2017) for concentrating on a set of known candidate genes in the interpretation of the results, potentially overlooking deleterious variants in novel genes. Furthermore, the study did not include sequencing of parental DNA, making it impossible to determine whether the reported variants were *de novo*, nor did it include experimental validation to assess the functional effects of the variants, many of which are likely to be benign.

A more recent study sequenced the whole genomes of 19 unrelated individuals diagnosed with CAS (Eising et al., 2018). *De novo* mutations were identified in *CHD3*, *SETD1A* and *WDR5*, in patients where DNA from unaffected parents was available. In other patients, novel loss-of-function variants were detected in multiple members of a shared module of co-expressed regulatory genes in early developing brain tissue (*KAT6A*, *SETBP1*, *ZFHX4*, *TNRC6B*, *MKL2*). Interestingly, *CHD3* has now been implicated in 34 other cases worldwide, with a variable neurodevelopmental disorder in which speech/language impairment is one of the phenotypic features (Snijders Blok et al., 2018). Overall, the study supported a prominent role for rare high-penetrance disruptive variants in the aetiology of CAS.

WES of 43 unrelated individuals affected by severe SLI identified potentially pathogenic inherited and *de novo* variants in several genes previously implicated in language-related disorders (including *ERCC1*, *GRIN2A*, *SRPX2*) as well as novel candidates (*OXRI*, *SCN9A*, *KMT2D*) (Chen et al., 2017). In a broader analysis of all variants across the SLI cohort, the authors found that variants predicted to be deleterious were over-represented in gene classes relating to ciliary function, microtubule transport and cytoskeletal organisation, highlighting candidate pathways for further investigation.

Cytogenetic methods have identified candidate genes disrupted by rare microdeletions and translocations in patients with speech/language disorders in the absence of other significant clinical findings. These candidates include *DOCK4* in dyslexia (Pagnamenta et al., 2010), *ERCC1* in CAS (Thevenon et al., 2013) and *SEMA6D* in a developmental language

disorder (Ercan-Sencicek et al., 2012). A *de novo* deletion involving *ATP2C2*, identified in an individual with SLI (Smith et al., 2015), provided support for the previously-identified association with that gene (Table 1.1; Newbury et al., 2009). Rare copy number variations (CNVs) are probably most relevant in severe speech and language disorders. Studies have found little overall correlation between rare CNV burden and either dyslexia (Girirajan et al., 2011; Gialluisi et al., 2016) or SLI (Simpson et al., 2015), although common CNVs likely contribute to SLI risk under a polygenic model (Simpson et al., 2015).

Further research will determine whether rare variants in the genes identified so far in sequencing and cytogenetic studies are major causes of language disorders. There may be additional monogenic disorders with predominantly linguistic phenotypes that remain undiscovered. Overall, however, progress has been relatively slow in identifying novel speech/language-related genes.

1.4 OBSTACLES TO DISCOVERY

While it is likely that a number of different genes play important roles in the development of a brain with the capacity to acquire and use language, mutations in any one of these genes will not necessarily lead to a primary speech and language disorder.

One factor that can obstruct gene discovery is redundancy. A particular gene may be involved in speech/language-related pathways, but a mutation in that gene will not necessarily result in a recognizable phenotype if another gene with a similar function can compensate for the loss. Mutations in this type of gene might contribute to disorder only when they occur in combination with one of more additional mutations at a different locus. The WES study of SLI described above (Chen et al., 2017) identified four probands with multiple rare coding variants in different genes (e.g. a novel missense variant in the developmental language disorder candidate gene *SEMA6D*, as well as a rare stop-gain in *SYNPR*), which might represent this kind of “multiple-hit” mechanism.

Another scenario is pleiotropy, where a speech/language-related gene may also be involved in other aspects of development, such that a mutation in that gene can result in broader neurological deficits, and even impact other organ systems. In certain disorders, the most severe symptoms may be justifiably prioritised in diagnosis, treatment and research, so that concomitant speech/language deficits remain under-reported and poorly understood. In extreme cases, a gene influencing speech/language might also be essential for one or more core biological processes, such that mutations in the gene are incompatible with life. Furthermore, the diagnosis of language disorders often relies on exclusionary criteria (e.g. a diagnosis of SLI requires the absence of general cognitive deficits or hearing difficulties), so that pleiotropic disorders may be specifically excluded from studies of speech/language impairment.

Given these complications, a narrow focus on primary speech/language disorders may only uncover a small proportion of the key genes that influence speech/language.

1.5 CASTING A WIDER NET

Two complementary strategies can be employed in order to overcome the obstacles above and help provide more insight into the genetic bases of speech and language. Firstly, established speech/language-related genes can serve as a starting point to investigate larger molecular networks, branching out to include genes that have not necessarily been linked to a primary speech/language disorder, but which may contribute to pathways shared with those genes that have. For example, chromatin immunoprecipitation (ChIP) and expression studies have identified hundreds of FOXP2 target genes, revealing candidate pathways for involvement in the neurobiology of speech/language, including differentiation, axon guidance, neurite outgrowth, synaptic plasticity, and cell motility and migration (Spiteri et al., 2007; Vernes et al., 2007, 2011).

The second approach is to consider phenotypically broader disorders that include additional clinical features beyond those that are purely linguistic. In many cases, speech/language delays co-occur with ID or ASD (Tomblin, 2011). For example, *de novo* mutations in *TBR1* cause an ASD phenotype with prominent language difficulties (Deriziotis et al., 2014b), while mutations in *SRPX2*, *GRIN2A* and *GRIN2B* have been reported to cause disorders with variable presentation of aphasia, speech dyspraxia, epilepsy and ID (Roll et al., 2006; O’Roak et al., 2011; Lemke et al., 2013; Lesca et al., 2013). Relevant disorders might also involve non-neurological symptoms, as illustrated by the rare *BCL11A* mutations that have been reported in patients with CAS, language delay, mild ID and global developmental delay, as well as the persistence of foetal haemoglobin after birth (Peter et al., 2014; Dias et al., 2016).

1.6 PROTEIN INTERACTION NETWORKS IN LANGUAGE-RELATED DISORDERS

The two approaches described above have been successfully combined in studies investigating the interaction partners of the FOXP2 transcription factor. So far, only a small number of FOXP2 interaction partners have been identified, but several are notably implicated in neurodevelopmental disorders that include speech/language deficits.

FOXP2 forms homodimers, as well as heterodimers with the paralogues FOXP1 and FOXP4 (Li et al., 2004, 2015a; Deriziotis et al., 2014a). Indeed, FOXP dimerisation is necessary for DNA-binding and transcription factor activity (Li et al., 2004). The FOXP2-FOXP1 interaction may be significant in regions of the brain where both are expressed, including parts of the striatum, thalamus and a small subset of cortical neurons (Ferland et al., 2003; Hisaoka et al., 2010). Interestingly, FOXP1 and FOXP2 have been shown to regulate some of the same downstream targets, including those with roles in language, such as *CNTNAP2*, suggesting a potential co-regulatory role for this interaction (Vernes et al., 2008; O’Roak et al., 2011).

The high degree of similarity between *FOXP2* and *FOXP1* initially drew attention to *FOXP1* as a potential candidate for speech and language disorders. However, an early candidate gene study investigating 49 patients with CAS did not find any causative *FOXP1* mutations (Vernes et al., 2009). *De novo* *FOXP1* mutations were later identified in WES

studies of patients with neurodevelopmental disorders and it became clear that the *FOXP1*-related phenotype is broader and more severe than the *FOXP2*-related phenotype, involving ID, global developmental delay and sometimes autistic features (Le Fevre et al., 2013; Sollis et al., 2016, 2017). However, speech and language delay are consistent core features, with expressive language affected more severely than receptive language, as in individuals with *FOXP2* variants (Morgan et al., 2017)

FOXP2 also interacts with the brain-specific TBR1 transcription factor (Deriziotis et al., 2014b). Both proteins are expressed in deep layers of the cortex, and in the amygdala and cerebellum (Hevner et al., 2001; Ferland et al., 2003; Fink et al., 2006; Remedios et al., 2007). TBR1 mutations have been identified in individuals with ASD and language delay (O’Roak et al., 2011; Deriziotis et al., 2014b). Mutations in either TBR1 or FOXP2 disrupt the interaction, suggesting molecular links between these two disorders (Deriziotis et al., 2014b).

Evidence from protein-protein interaction assays and from phenotypic investigations can therefore be combined to identify candidate proteins likely to be involved in the molecular networks underlying speech/language.

1.7 PROTEOMICS TECHNIQUES DISCUSSED IN THIS THESIS

Several different methods are available to investigate protein-protein interactions *in vitro*. These can be separated into methods used primarily to identify novel interaction partners, and those that have mainly been used to validate putative interactions. A brief explanation of several techniques is given here, together with some of the relative advantages and disadvantages of each.

1.7.1 IDENTIFYING NEW INTERACTIONS

1.7.1.1 *Yeast two-hybrid screening*

Yeast two-hybrid (Y2H) screening is a molecular biology technique used to identify physical interactions (i.e. binding) between two proteins in live yeast (*Saccharomyces cerevisiae*) cells. The procedure makes use of a TF that binds to a particular regulatory sequence to activate transcription of a downstream reporter gene, which produces an observable phenotype (Suter et al., 2008). For example, the active reporter gene may allow biosynthesis of an essential nutrient without which the yeast cells cannot survive. The TF is expressed in yeast cells as two separate fragments, one containing the DNA-binding domain (DBD) and the other containing the activating domain (AD). A protein of interest (known as the bait) is fused to the DBD, and a library of possible interactors (preys) are fused to the AD. A successful interaction between the bait protein and a prey protein reconnects the DBD and AD to form a functional TF, thereby driving activation of the reporter gene (Fig 1.1).

One strength of the Y2H assay is that it is relatively cost-effective and straightforward to perform, without the need for expensive or specialised equipment. It also allows simultaneous high-throughput screening of many potential interaction partners. On the other

hand, Y2H screens can be vulnerable to both false positive and false negative identifications (Suter et al., 2008). False positives can occur due to the overexpression of both baits and preys above physiological levels, as well as abnormal localisation of these proteins in cellular compartments where they are not naturally expressed, leading to non-specific interactions (Suter et al., 2008). There are also examples of false positives caused by interaction between the prey protein and the DBD fusion protein, rather than with the bait protein itself (Brückner et al., 2009). False negatives can arise if the interaction is hindered by the DBD/AD fusion protein (often at the N-terminus of the test protein), which can disrupt protein folding and may block important binding regions (Suter et al., 2008). Other problems stem from the intracellular conditions of the assay. For example, because the Y2H assay relies on DNA-binding within the nucleus, it may fail to detect interactions between non-nuclear proteins (Brückner et al., 2009). Furthermore, interactions can be missed by Y2H if they rely on mammalian-specific post-translational modifications not present in yeast cells (Suter et al., 2008). The non-human cellular environment can be beneficial, however, as it makes the detection of direct pair-wise interactions more confident, because other human proteins that might facilitate indirect interactions are absent from the experiment (Galletta and Rusan, 2015).

In this thesis, I refer to a Y2H assay performed under the direction of Prof. Svante Pääbo at the Max Planck Institute for Evolutionary Anthropology, which has been reported in the supplementary material of a previous publication (Estruch et al., 2016b). Human FOXP2 was used as the bait, against a library of human foetal brain proteins. The assay identified 34 proteins that were represented by at least one clone, and 6 that were represented by two or more. The top hit, with 7 positive clones, was PIAS1, a small ubiquitin-like modifier (SUMO) ligase, which was later confirmed to interact with, and promote SUMOylation of FOXP2 (Estruch et al., 2016b). A number of chromatin-remodelling factors were also identified in the assay, including CHD3, GATAD2B and KANSL1, which are followed up in Chapter 4 of this thesis.

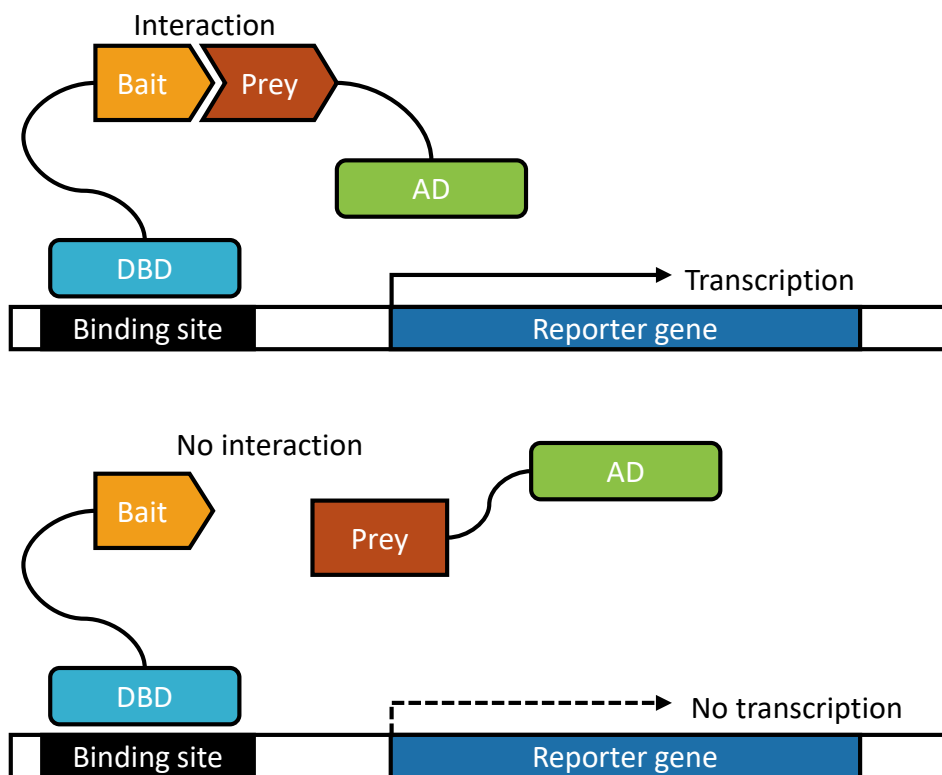


Figure 1.1. Schematic of a typical yeast-two-hybrid assay for protein-protein interactions. An interaction between the bait and prey proteins allows the DNA-binding domain (DBD) and activating domain (AD) to form a functional transcription factor, facilitating transcription of a downstream reporter gene. If the proteins do not interact, the reporter gene is not transcribed.

1.7.1.2 Affinity purification-mass spectrometry

Another major strategy for identifying novel protein-protein interactions is to use affinity purification (AP), followed by mass spectrometry (MS). In this method, a protein of interest is expressed in an appropriate cellular system, and then immunoprecipitated from the lysed cells together with its interacting proteins (Gingras et al., 2007) (Fig 1.2). This can be performed with an antibody specific to the protein, or alternatively, the assay can be generalised to any protein of interest by using an epitope tag, such as FLAG (Gingras et al., 2007). The purified proteins can then be separated (e.g. by SDS-PAGE) and lysed, usually using trypsin, to produce peptide fragments for mass spectrometry analysis (Gingras et al., 2007). MS analysis for this application typically involves peptide separation by liquid chromatography, followed by two MS scans (Gingras et al., 2007). The first scan measures the mass/charge ratio (m/z) of the complete peptide, which is then fragmented and measured again to produce a tandem mass spectrum. These spectra can then be searched against a database to identify the proteins that were present in the original sample. The resulting list contains the tagged protein-of-interest and its interaction partners.

The AP-MS approach is rapid, sensitive and comprehensive, allowing the detection of thousands of peptides in a single run (Gingras et al., 2007). MS can also be adapted to quantify peptide abundance, which may prove beneficial in order to prioritise candidates for further study. Importantly, protein detection is unbiased and does not rely on the *a priori* selection of a library of candidate proteins, as is the case for Y2H. As another advantage over Y2H, AP-MS can be performed in almost any cell or tissue type, without the reliance on yeast cells. This is beneficial for investigating human protein-protein interactions, as it enables the use of human cell lines or even cultured cells/tissues, with much more physiologically relevant conditions, including the presence of mammalian-specific post-translational modifications that may influence interactions (Gingras et al., 2007).

On the other hand, AP-MS can be vulnerable to both false positive and false negative results. Contaminants can be introduced at multiple stages. For example, the immunoprecipitation step can non-specifically pull down highly abundant proteins (e.g. actin), chaperones that bind to unfolded polypeptides and proteins that can interact with the antibodies and beads used for purification (Gingras et al., 2007). Overexpression of the protein-of-interest can also allow non-physiological interactions to form (Gingras et al., 2007). Appropriate controls and filtering steps are necessary to remove common contaminants from the final analysis. The MS stage can also introduce experimental artefacts, such as the trypsin used for fragmentation and human keratins present in the laboratory, although these can be identified and removed effectively during peptide identification. Conversely, true interactions can be missed if the AP conditions (e.g. washing steps) are too stringent to preserve the interaction, or if there is a large imbalance in the relative abundance of the bait and its interactors, especially if the bait is overexpressed. For these reasons, AP-MS is good at detecting stable interactions with more abundant proteins but less effective than other methods at detecting weak or transient interactions (Gingras et

al., 2007). False negative results can arise if the epitope tag used for AP-MS blocks important binding sites.

Recently, an AP-MS approach was successfully employed to screen for FOXP1/2/4-interacting proteins, and subsequent functional validation led to the identification of seven novel FOXP-interacting transcription factors (*NR2F1*, *NR2F2*, *SATB1*, *SATB2*, *SOX5*, *YY1* and *ZMYM2*) (Estruch et al., 2018). I use a similar methodology in Chapters 5 and 6 of this thesis, to investigate novel interaction partners of the TBR1 transcription factor.

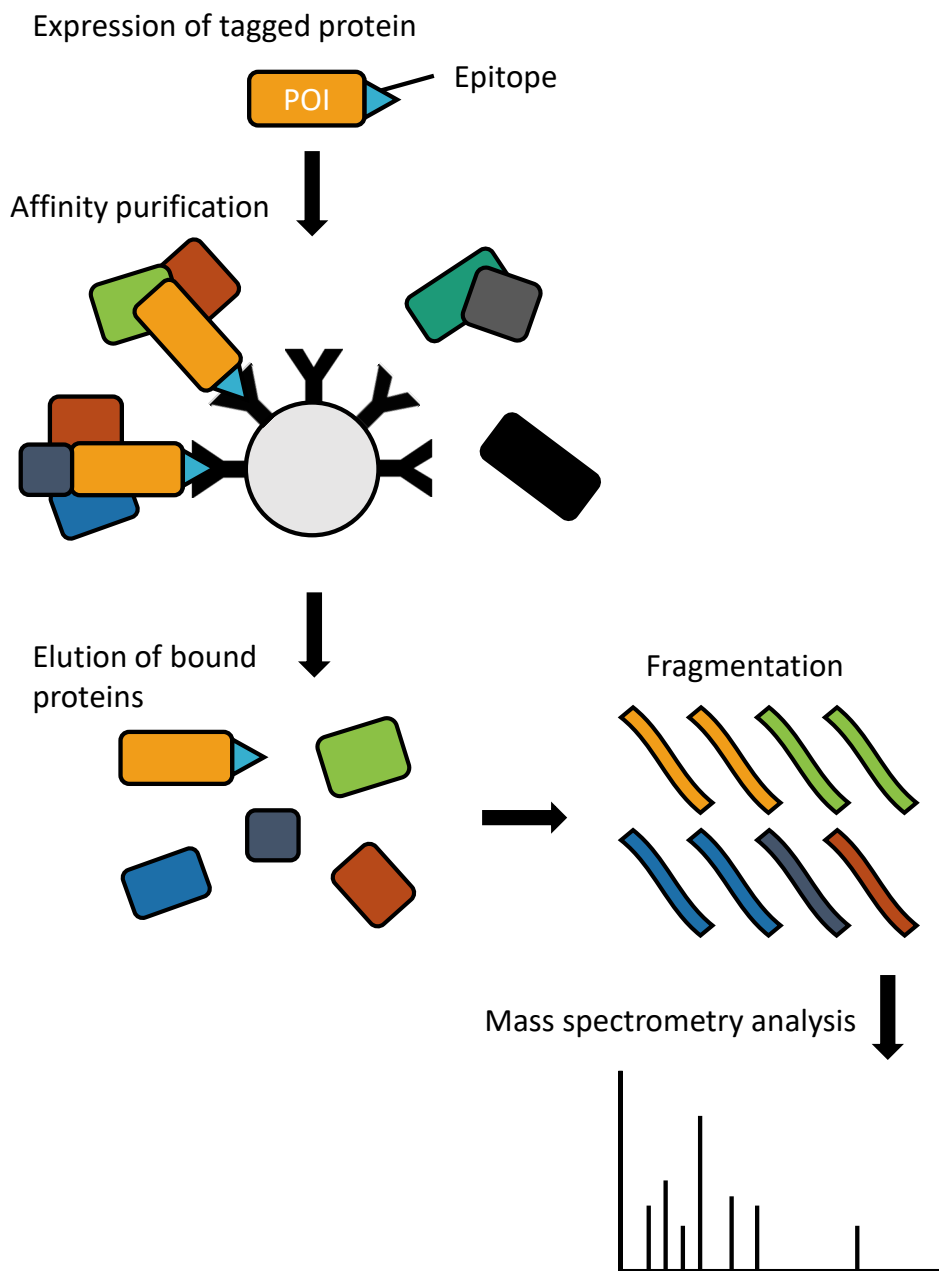


Figure 1.2. The affinity purification-mass spectrometry procedure as used in this thesis. The protein of interest (POI) is expressed in cells with an epitope tag. The protein is then captured, along with any bound proteins, using an antibody against the epitope, while non-interacting proteins remain in solution and are washed away. The bound proteins are then eluted, separated and fragmented (e.g. using trypsin) and then identified by mass spectrometry.

1.7.2 VALIDATING PUTATIVE INTERACTIONS

1.7.2.1 *Co-immunoprecipitation*

Co-immunoprecipitation (co-IP) assays are used to test for the physical association of two or more proteins in a sample (Fig 1.3). Proteins of interest can be over-expressed by transient transfection, or endogenous protein-protein interactions can be tested. The cells are then lysed, and an antibody is used to bind and isolate one of the proteins of interest together with any bound proteins. Western blotting is then employed to detect whether the putative interactor has been co-immunoprecipitated. Co-IP with over-expressed proteins has the advantage that epitope tags can be added to the proteins of interest, enabling the use of the same general detection antibody across experiments (Klenova et al., 2002). The high level of protein expression means that even weak or unstable protein-protein interactions can be detected, but also increases the detection of spurious interactions, as the proteins are expressed at much higher than physiological levels (Klenova et al., 2002). Co-IP with endogenous proteins is less vulnerable to false positives but requires the optimisation of specific antibodies for each protein tested (Klenova et al., 2002).

Co-IP is a valuable method for testing small numbers of interactions but is difficult to scale up for more high-throughput experiments. It can also require considerable hands-on time and is often difficult to optimise for certain interactions (Klenova et al., 2002). Other disadvantages are that cell lysis can destroy the specific intracellular conditions that may be required to maintain certain protein-protein interactions, and weak or transient interactions can be disrupted during washing steps (Klenova et al., 2002). Co-IP assays were not utilised in this thesis. However, co-IPs are discussed in relation to other studies reported in the literature, which have informed the experiments performed here, as well as the interpretation of results.

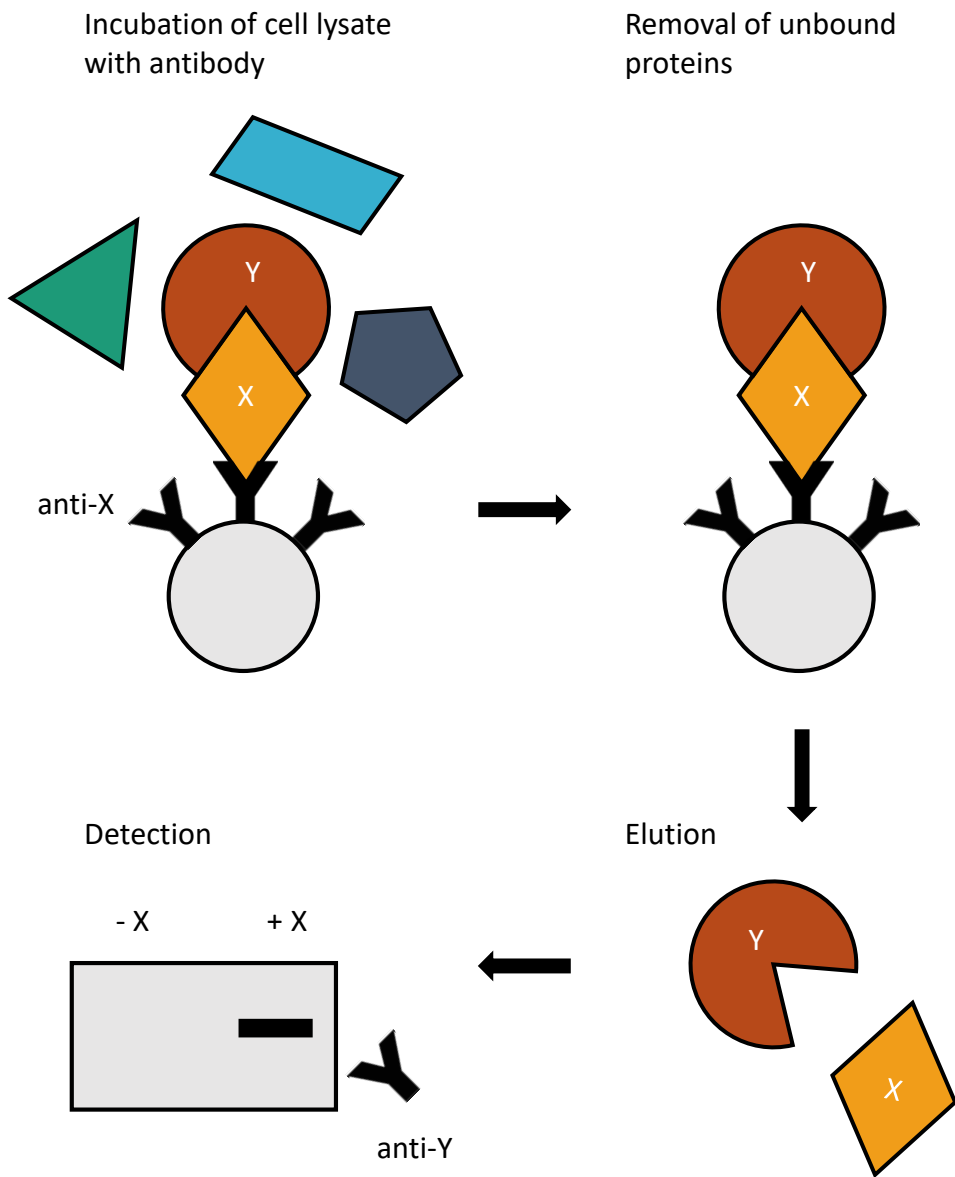


Figure 1.3. Schematic of a typical co-immunoprecipitation experiment. Two proteins of interest (X and Y) are expressed in a cell. The cells are lysed and incubated with an antibody against protein X. Protein X remains bound to the antibody, along with any interacting proteins, while unbound proteins are removed by washing. Proteins are eluted and analysed by Western blot. If protein Y can be detected in an elution containing protein X, then the interaction is validated.

1.7.2.2 Bioluminescence resonance energy transfer

The bioluminescence resonance energy transfer (BRET) assay is a method for monitoring protein-protein interactions in live cells (Pfleger and Eidne, 2006; Deriziotis et al., 2014a). BRET occurs naturally in marine organisms such as the sea pansy (*Renilla reniformis*) and involves the non-radiative (dipole-dipole) transfer of energy generated by a 'donor' luciferase enzyme to an 'acceptor' fluorescent protein, after oxidation of a luciferase substrate such as coelenterazine. This energy transfer occurs when the emission spectrum of the luciferase reaction overlaps with the excitation spectrum of the fluorescent acceptor. The most common donor-acceptor combination for molecular biology applications is *Renilla* luciferase (Rluc) and yellow fluorescent protein (YFP).

In order to test protein-protein interactions, expression constructs are transfected into live cells, to produce one protein of interest fused to Rluc, and a putative interactor fused to YFP (Fig 1.4). Luminescence measurements are taken after the addition of the luciferase substrate. If there is no interaction, the luciferase reaction proceeds independently, producing a luminescent signal with a peak at 480 nm. If the two proteins interact, however, Rluc and YFP may be brought into close enough proximity (< 10 nm) for energy transfer to occur, generating a YFP emission peak at 527 nm. A positive interaction can therefore be measured as a shift in the ratio of blue-to-green luminescence.

The BRET assay is relatively cost- and time-efficient. It can be scaled up to measure multiple interactions at the same time in 96-well plates. Because interactions are monitored directly in live cells, problems associated with cell lysis can be avoided, and weak/transient interactions that might otherwise be disrupted during washing/sample preparation can be successfully detected. It also has a relatively low false positive rate, because the BRET tags have to be so close together to generate a signal. BRET assays can also be adapted to map binding sites, and to assess the effects of mutations and/or post-translational modifications (Deriziotis et al., 2014a). False positives can occur however, if the proteins of interest are over-expressed to excessive levels, which can enable non-specific interactions (Deriziotis et al., 2014a). The main disadvantage of BRET is that it can be vulnerable to false negatives, which can occur if the Rluc- and YFP-tags obstruct key binding domains, or if other factors such as protein length or conformation place the tags too far away from one another to allow energy transfer.

BRET assays have been used successfully to validate several interactions involving speech/language-related proteins, as well as to characterise how these interactions are affected by changes in the protein sequence, e.g. to characterise putative causative mutations or to test hypotheses about binding sites and other functionally important residues (Deriziotis et al., 2014a, 2014b; Lozano et al., 2015; Estruch et al., 2016a, 2016b, 2018). BRET assays are an important part of the work described in this thesis, specifically in Chapters 2, 3, 4 and 6.

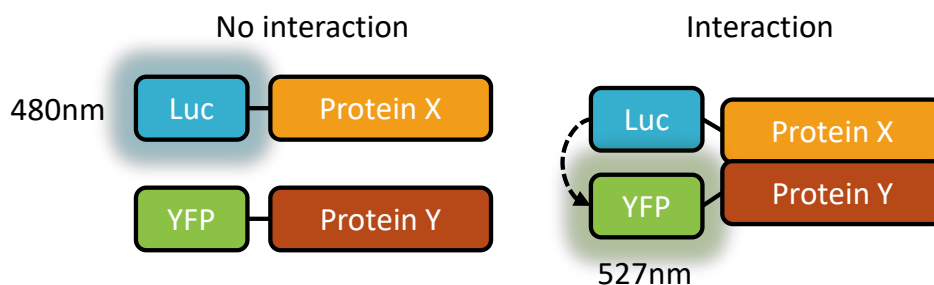


Figure 1.4. Bioluminescence resonance energy transfer (BRET) assay as used in this thesis. Proteins of interest (X and Y) are fused to either a luciferase (Luc) or yellow fluorescent protein (YFP) tag. In the absence of an interaction, the luciferase reaction proceeds independently and only blue light at a wavelength of 480 nm is emitted. When Proteins X and Y interact, the luciferase and YFP tags become close enough for some energy to be transferred from luciferase to YFP, which emits green light at 527 nm. A shift in the observed ratio between the two wavelengths provides evidence of an interaction.

1.8 AIMS OF THIS THESIS

The major aims of this thesis are to: 1) identify and characterise novel protein-protein interactions involving known language-related proteins, in order to further explore the molecular networks involved in the speech/language phenotype; and 2) investigate the consequences of mutations affecting these proteins, at both the phenotypic and molecular level, including their effects on interactions with other proteins in the emerging speech/language networks.

In **Chapter 2** of this thesis, I investigate *FOXP1*, a close paralogue of *FOXP2* with variants causing a syndrome characterised by global developmental delay, ID and speech/language delay. I report three novel pathogenic variants in *FOXP1* and use *in vitro* functional genomic methods to determine their effects on protein expression, localisation, transcription factor activity and protein-protein interactions in human cell models. In particular, I show that *FOXP1* variants disrupt FOXP1 dimerisation, and that a subset of these variants may also have deleterious consequences for the functions of wild-type FOXP1, as well as FOXP2. I suggest that the disruption of this interaction may contribute to the language impairment phenotype in patients with *FOXP1* mutations.

In **Chapter 3**, I focus on one particular pathogenic variant in *FOXP1*, which has a direct equivalent in *FOXP2* that is a well-known cause of speech and language disorder. I compare these two variants at the molecular and phenotypic level and draw attention to the functional differences between the closely related FOXP1 and FOXP2 proteins.

In **Chapter 4**, I investigate potential interactions between the FOXP transcription factors and three candidate proteins identified in a yeast-two-hybrid assay: GATAD2B, CHD3 and KANSL1. These proteins are all subunits of chromatin-remodelling complexes and are implicated in neurodevelopmental disorders involving speech and language impairment. I confirm the interaction of FOXP1/2/4 with GATAD2B, but not with CHD3 or KANSL1. I go on to refine the location of the GATAD2B-binding site in the FOXP protein sequence and to investigate the effects of pathogenic *GATAD2B* and *FOXP1/2* variants on these interactions.

In **Chapters 5 and 6**, I focus on TBR1, a brain-specific transcription factor involved in ASD and ID, which also interacts with the FOXP1/2/4 proteins (Deriziotis et al., 2014b). In **Chapter 5**, I conduct an AP-MS screen to expand the known interaction network of TBR1 and identify a set of 248 candidate TBR1-interactors. I show that the TBR1 interactome is enriched for ASD/ID-related proteins, and for proteins involved in transcriptional regulation and chromatin remodelling. In **Chapter 6**, I validate six novel TBR1-interactors identified in the AP-MS screen, including multiple transcriptional regulators. Strikingly, several TBR1 interaction partners also interact with FOXP1/2/4, strengthening the molecular links between these important transcription factors. I also investigate how these interactions are affected by variants identified in patients with neurodevelopmental disorders.

Finally in **Chapter 7**, I summarise the contributions of this thesis to the study of language and genetics and consider further avenues of research that could build on this work in the future.

2 IDENTIFICATION AND FUNCTIONAL CHARACTERISATION OF *DE NOVO FOXP1* VARIANTS PROVIDES NOVEL INSIGHTS INTO THE AETIOLOGY OF NEURODEVELOPMENTAL DISORDER

PUBLISHED AS:

Sollis E, Graham SA, Vino A, Froehlich H, Vreeburg M, Dimitropoulou D, Gilissen C, Pfundt R, Rappold GA, Brunner HG, Deriziotis P, Fisher SE. 2016. Identification and functional characterisation of *de novo FOXP1* variants provides novel insights into the aetiology of neurodevelopmental disorder. *Hum Mol Genet* 25:546–557.

ABSTRACT

De novo disruptions of the neural transcription factor FOXP1 are a recently discovered, rare cause of sporadic intellectual disability (ID). We report three new cases of FOXP1-related disorder identified through clinical whole-exome sequencing. Detailed phenotypic assessment confirmed that global developmental delay, autistic features, speech/language deficits, hypotonia and mild dysmorphic features are core features of the disorder. We expand the phenotypic spectrum to include sensory integration disorder and hypertelorism. Notably, the aetiological variants in these cases include two missense variants within the DNA-binding domain of FOXP1. Only one such variant has been reported previously. The third patient carries a stop-gain variant. We performed functional characterisation of the three missense variants alongside our stop-gain and two previously described truncating/frameshift variants. All variants severely disrupted multiple aspects of protein function. Strikingly, the missense variants had similarly severe effects on protein function as the truncating/frameshift variants. Our findings indicate that a loss of transcriptional repression activity of FOXP1 underlies the neurodevelopmental phenotype in FOXP1-related disorder. Interestingly, the three novel variants retained the ability to interact with wild-type FOXP1, suggesting these variants could exert a dominant-negative effect by interfering with the normal FOXP1 protein. These variants also retained the ability to interact with FOXP2, a paralogous transcription factor disrupted in rare cases of speech and language disorder. Thus, speech/language deficits in these individuals might be worsened through deleterious effects on FOXP2 function. Our findings highlight that *de novo FOXP1* variants are a cause of sporadic ID and emphasise the importance of this transcription factor in neurodevelopment.

2.1 INTRODUCTION

FOXP1 (forkhead-box protein P1; OMIM 605515) is a member of the forkhead box family of transcription factors that is crucial for embryonic development (Wang et al., 2004a). Heterozygous disruptions of the *FOXP1* gene result in global developmental delay and intellectual disability (ID) (OMIM 613670). The first cases of *FOXP1*-related disorder were individuals carrying *de novo* deletions of *FOXP1* and neighbouring genes (Pariani et al., 2009). Since then, additional cases have been found with either truncating/frameshift variants or deletions encompassing only *FOXP1*, confirming that disruption of one copy of this gene results in disorder (Carr et al., 2010; Hamdan et al., 2010; Horn et al., 2010; O’Roak et al., 2011; Bacon and Rappold, 2012; Le Fevre et al., 2013; Lozano et al., 2015). All reported aetiological *FOXP1* variants to date have occurred *de novo*, consistent with the widespread observation of *de novo* protein-disrupting variants in cases of severe sporadic neurodevelopmental disorder (Veltman and Brunner, 2012).

Comparison of the phenotypes of patients with *FOXP1* disruptions has led to the delineation of a syndrome in which mild-to-moderate ID is frequently accompanied by features of autism spectrum disorder (ASD) (Hamdan et al., 2010; Horn et al., 2010; O’Roak et al., 2011; Le Fevre et al., 2013; Lozano et al., 2015). Some individuals present with additional behavioural problems such as obsessions and compulsions, aggression and hyperactivity (Hamdan et al., 2010; O’Roak et al., 2011; Lozano et al., 2015). Macrocephaly and abnormal facial features have been reported in certain cases (Le Fevre et al., 2013; Lozano et al., 2015). Probands carrying *FOXP1* variants also show speech and language impairments, which range from moderate to severe and affect expressive language to a greater degree than receptive language (Carr et al., 2010; Hamdan et al., 2010; Horn et al., 2010; Le Fevre et al., 2013; Lozano et al., 2015). The speech deficits in these patients may be in part related to orofacial motor dysfunction, which has been noted in several cases (Horn et al., 2010; Le Fevre et al., 2013).

The presence of speech and language impairments in this emerging *FOXP1*-related disorder is of particular interest because *FOXP1* is the closest paralogous gene to *FOXP2* (OMIM 605317). The *FOXP2* gene is disrupted in a rare form of speech and language disorder that is characterised by developmental verbal dyspraxia (DVD) [also known as childhood apraxia of speech (CAS)] as well as deficits in expressive and receptive language affecting spoken and written domains (OMIM 602081) (Lai et al., 2001). FOXP1 and FOXP2 heterodimerise *via* a leucine zipper domain (Ferland et al., 2003) and are co-expressed in several brain regions, with the potential to co-regulate downstream targets, including those involved in language, such as *CNTNAP2* (Ferland et al., 2003; Li et al., 2004; O’Roak et al., 2011).

In mouse and human foetal brain, *FOXP1* expression is observed in regions including the striatum, cerebral cortex (layers 3-5), hippocampus and thalamus (Ferland et al., 2003; Teramitsu et al., 2004). Mice with global *Foxp1* deletion die around embryonic day E14.5 (Wang et al., 2004a). Mice with selective deletion of *Foxp1* in the brain are viable but display gross malformations of the striatum that develop postnatally, along with electrophysiological

abnormalities in the hippocampus (Bacon et al., 2015). These animals also exhibit learning and memory deficits and reduced social interests (Bacon et al., 2015).

Here, we report the identification of three novel *de novo* *FOXP1* variants in individuals with ID through clinical whole-exome sequencing, including two missense variants in the FOX DNA-binding domain. We present detailed phenotypic information on the affected probands, extending the phenotype of *FOXP1*-related disorder. We report functional characterisation of the three novel variants reported here, alongside three previously detected *de novo* aetiological variants (Hamdan et al., 2010; O’Roak et al., 2011; Srivastava et al., 2014). Our functional investigations include the first analyses of missense *FOXP1* variants. We demonstrate deleterious effects on subcellular localisation, transcriptional repression and protein-protein interactions for all six aetiological variants, shedding light on the pathological mechanisms underlying *FOXP1*-related disorder.

2.2 METHODS

2.2.1 EXOME SEQUENCING

For patient 1, whole-exome sequencing was performed by GeneDx (MD, USA). For patients 2 and 3, routine diagnostic exome sequencing was carried out as previously described (Neveling et al., 2013). All variants were validated by Sanger sequencing in probands and parents and were found to be *de novo*. Informed oral consent was obtained for the use of the data and photographs according to relevant institutional and national guidelines and regulations.

Variants are described throughout this article according to the following reference sequences: *FOXP1* transcript NM_032682.5, FOXP1 protein NP_116071.2. Variants have been submitted to the NCBI ClinVar database (see URL¹).

2.2.2 PROTEIN MODELLING OF *FOXP1* VARIANTS

The *de novo* FOXP1 variants (p.R465G, p.R514C and p.W534R) were modelled using the YASARA structural-simulation software (see URL²). The protein model for human FOXP2 bound to NFAT and DNA (RSCB Protein Data Bank 2AS5) served as a structural template. The FoldX plugin for YASARA was used to calculate $\Delta\Delta G$ values.

2.2.3 DNA CONSTRUCTS

Wild-type FOXP1 and FOXP2 were amplified by PCR and cloned into pCR2.1-TOPO (Invitrogen) as described (Deriziotis et al., 2014b). FOXP1 variant constructs were generated using the QuikChange II Site-Directed Mutagenesis Kit (Agilent) (primers in Table 2.1).

¹ ClinVar: <http://www.ncbi.nlm.nih.gov/clinvar> (see accession numbers SCV000246199, SCV000246200 and SCV000246201)

² YASARA: <http://www.yasara.org/>

The c.1017_1018insT (p.A339Sfs*4) variant construct was synthesised by GenScript USA Inc. as described previously (O’Roak et al., 2011). FOXP cDNAs were subcloned using BamHI/XbaI restriction sites into pLuc (Deriziotis et al., 2014a), pYFP (Deriziotis et al., 2014a) and a modified pmCherry-C1 expression vector (Clontech). All constructs were verified by Sanger sequencing.

Table 2.1. Primers used to generate *FOXP1* variants by site directed mutagenesis.

Variant	Forward primer (5’-3’)	Reverse primer (5’-3’)
c.1393A>G (p.R465G)	ATATGTAAATGGTGGTCCAAC TTCTGCGTTCCTTATAAAATTCT TGGTT	AACCAAGAATTTTATAAGAACGC AGAAGTTGGACCACCATTACATA T
c.1540C>T (p.R514C)	TGAAGACTAAGATTATGACAC ACTGCATTCTTCCACGTG	CACGTGGAAGAATGCAGTGTGTC ATAATCTTAGTCTCA
c.1317C>G (p.Y439*)	GTTGTATTTGTCTGATTACCG CCTGCGGATGGG	CCCATCCGCAGGCGGTAATCAGA CAAATACAAC
c.1573C>T (p.R525*)	CCCTTAAACGTTTTCTACTCAC ACAAAACACTTGTGAAGAC	GTCTTCACAAGTGTTTTGTGTGAG TAGAAAACGTTAAAGGG
c.1600T>C (p.W534R)	ACTCATCCACTGTCCGTACT GCCCCTTAACGTT	AACGTTAAAGGGGCAGTACGGAC AGTGGATGAAGT
c.226_228dup (p.Q76dup)	CAGCAGCAACAGCAGCAGCA GCAAGTTAGTGGATTAATAA	TTTTAATCCACTAACTTGCTGCTG CTGCTGTTGCTGCTG
c.320T>C (p.I107T)	GCTATGATGACACCTCAAGTT ACCACTCCCCAGCAA	TTGCTGGGGAGTGGTAACCTGAGG TGTCATCATAGC
c.643C>G (p.P215A)	GGCAGCCTGCCCTTGCCCTTC AACC	GGTTGAAGGGCAAGGGCAGGCTG CC
c.1709A>G (p.N570S)	CGCCTACTGCACACCTCTCAG TGCAGCTTTAC	GTAAAGCTGCACTGAGAGGTGTG CAGTAGGCG
c.1790A>C (p.N597T)	CCCACTCTGGGCACCTTAGCC AGCGCA	TGCGCTGGCTAAGGTGCCAGAGT GGG

2.2.4 CELL CULTURE AND TRANSFECTION

HEK293 cells were cultured in DMEM supplemented with 10% foetal bovine serum (both from Invitrogen). Transfections were performed using GeneJuice, according to manufacturer’s instructions (Merck-Millipore).

2.2.5 WESTERN BLOTTING

Cells were transfected with equimolar concentrations of FOXP1 expression plasmids and cultured for 24 h. Whole-cell lysates were extracted by treatment with lysis buffer (100 mM Tris pH 7.5, 150 mM NaCl, 10 mM EDTA, 0.2% Triton X-100, 1% PMSF, protease inhibitor cocktail; all from Sigma-Aldrich) for 10 min at 4°C, before centrifuging at 10,000 x g for 30 min at 4°C to remove cell debris. Proteins were resolved on 4-15% Tris-Glycine gels and transferred onto polyvinylidene fluoride membranes (both Bio-Rad). Blots were probed with mouse anti-EGFP (for pYFP constructs; 1:8,000; Clontech) overnight at 4°C, followed by incubation with HRP-conjugated goat anti-mouse IgG for 45 min at room temperature

(1:2,000; Bio-Rad). Proteins were visualised using Novex ECL Chemiluminescent Substrate Reagent Kit (Invitrogen) and the ChemiDoc XRS+ System (Bio-Rad).

2.2.6 PROTEIN EXPRESSION ANALYSIS

Cells were transfected in clear-bottomed 96-well plates in triplicate with equimolar concentrations of FOXP1 expression plasmids, together with a modified pmCherry-C1 plasmid to normalise for transfection efficiency. YFP and mCherry fluorescence intensities were measured in live cells 42 h post-transfection in a TECAN M200PRO microplate reader at 37°C and 5% CO₂. For each well, the background-subtracted YFP intensity was divided by the background-subtracted mCherry intensity. Triplicate conditions were averaged.

2.2.7 FLUORESCENT MICROSCOPY

Cells were seeded onto coverslips coated with poly-L-lysine (Sigma-Aldrich), and were fixed 24 h post transfection using 4% paraformaldehyde (Electron Microscopy Sciences) for 10 min at room temperature. YFP and mCherry fusion proteins were visualised by direct fluorescence. Nuclei were visualised with Hoechst 33342 (Invitrogen). Fluorescence images were obtained using a LSM510 confocal microscope with LSM Image Software (Zeiss).

2.2.8 LUCIFERASE REPORTER ASSAYS

Cells were seeded in 24-well plates and transfected with 45 ng of firefly luciferase reporter construct (pGL3-prom; Promega), 5 ng of *Renilla* luciferase normalisation control (pRL-TK; Promega) and 200 ng FOXP1 expression construct (WT or mutant in pYFP) or empty vector (pYFP; control). Firefly luciferase and *Renilla* luciferase activities were measured 48 h post transfection in a TECAN F200PRO microplate reader using the Dual-Luciferase Reporter Assay system (Promega). The statistical significance of the luciferase reporter assays was analysed using a one-way analysis of variance and a Tukey's post hoc test.

2.2.9 BRET ASSAYS

BRET assays were performed as previously described (Deriziotis et al., 2014a, 2014b). In summary, cells were transfected with pairs of *Renilla* luciferase and YFP fusion proteins in 96-well plates. *Renilla* luciferase and YFP fused to a C-terminal nuclear localisation signal were used as control proteins. EnduRen luciferase substrate (Promega) was added to cells 36-48 h after transfection at a final concentration of 60 μM and incubated for 4 hours. Emission measurements were taken with a TECAN F200PRO microplate reader using the Blue1 and Green1 filters, and corrected BRET ratios were calculated as follows: $[\text{Green1}_{(\text{experimental condition})} / \text{Blue1}_{(\text{experimental condition})}] - [\text{Green1}_{(\text{control condition})} / \text{Blue1}_{(\text{control condition})}]$. YFP fluorescence was then measured separately, with excitation at 505nm and emission at 545, to quantify expression of the YFP-fusion proteins. *Renilla* luciferase and YFP fused to a C-terminal nuclear localisation signal were used as control proteins.

2.3 RESULTS

2.3.1 CLINICAL DESCRIPTION OF PATIENTS

Patient 1 is an 11-year-old boy from the US born to healthy non-consanguineous parents (Table 2.2, Fig 2.1). At the age of 6 months the patient was noted to have hypotonia and at the age of 8 months he was diagnosed with global developmental delay and macrocephaly. At the age of 12 months he was noted by MRI to have mild, diffuse periventricular leukomalacia. He had delays in all motor milestones; he walked at 21 months and has fine motor problems. In addition, he presented with speech delays and has speech impairment, including articulation deficits. He has been diagnosed with mild ID; an IQ test revealed a wide range in his IQ score (between 50 and 80 in the different tasks). He displays a number of autistic features such as stereotypic behaviour, obsessive-compulsive tendencies, ADHD, and sensory processing disorder, but does not fulfil criteria for classical autism. Anxiety-related behaviours are well-managed by Prozac. During childhood, he presented with recurrent infections of the skin and ear. Other findings were a large forehead and short stature. Biochemical tests (plasma amino acids and urine organic acids) were normal. Genetic investigations with normal results included chromosome analysis, chromosome microarray, fragile X syndrome screening and *PTEN* sequencing.

Patient 2 is a 7-year-old Dutch boy born to healthy non-consanguineous parents (Table 2.2, Fig 2.1). His family history includes a maternal uncle diagnosed with a pervasive developmental disorder. The patient presented with global developmental delay, ID and speech and language problems. Delays in cognitive development (total IQ = 53 at the age of 5 years) and mild delays in motor development (walked at the age of 17 months) were noted. Currently, he can speak in sentences, but with poor pronunciation/articulation. The child neurologist described an apraxia of the tongue. When playing, he constantly talks and makes noises. He has also been diagnosed with a sensory integration disorder. In addition, his behaviour can be demanding and impulsive; he has ADHD and received medication (Ritalin). He shows many features of pervasive developmental disorder such as stereotypic behaviour, obsessive-compulsive tendencies and a great need for structure in daily life but does not fulfil criteria for classical autism. He has mild dysmorphic features including hypertelorism, small down slanted eyes, a short nose and mild retrognathia. Height, weight and head circumference measurements are normal for his age. Other findings include strabismus and enuresis. Investigations with normal results included screens for fragile X and Angelman syndromes, SNP array testing, *MECP2* and *TCF4* sequencing and metabolic tests. A screen for deletions/duplications of the poly-alanine stretch in exon 2 of *ARX* was negative.

Patient 3 is a 15-year-old Dutch girl born to healthy non-consanguineous parents (Table 2.2, Fig 2.1). The pregnancy was complicated by gestational diabetes. She was born at 38 weeks *via* C-section and had a birth weight of 4,160 grams. During her first months of life, she developed nystagmus and was prescribed glasses because of hypermetropia (+5 dioptre). The patient showed developmental delays; she walked at the age of 21 months and has been

diagnosed with severe ID. At 3 years her BSID-II was equivalent to 18 months and at 5 years her development was equivalent to about 2.3 years (SON-R; non-verbal intelligence test). Her language skills at 12 years were equivalent to 3 years. Currently, she mostly lacks speech and can say a few words with poor articulation. In addition, the patient has behavioural problems: autistic features, anxiety, aggression, obsessive-compulsive behaviour and screaming. Mild dysmorphic features include hypertelorism, prominent forehead and a broad tip of the nose. Menarche was at the age of 12 years. Her height and head circumference are normal (-1SD and +1SD respectively in growth charts for Dutch children). Investigations with normal findings included karyotyping, fragile X screening, fluorescence *in situ* hybridisation, SNP arrays and metabolic tests.

2.3.2 CLINICAL WHOLE-EXOME SEQUENCING

Clinical whole-exome sequencing revealed *de novo* heterozygous *FOXP1* variants in all three patients (Fig 2.1, Table 2.3). Patient 1 carries a missense *FOXP1* variant p.R465G which affects the first residue of the FOX DNA-binding domain. Patient 2 carries a missense *FOXP1* variant p.R514C within the DNA-recognition helix of the FOX DNA-binding domain. Patient 3 carries a stop-gain *FOXP1* variant p.Y439*, which truncates the protein between the leucine zipper dimerisation domain and FOX DNA-binding domain.

Table 2.2. Phenotypic description of patients with *de novo* *FOXP1* variants.

Study (proband)	This study (1)	This study (2)	This study (3)	This study (B)	O’Roak	Srivastava (41)	Lozano
Variant	p.R465G	p.R514C	p.Y439*	p.R525*	p.A339Sfs*4	p.W534R	p.V423His*37
Age	11 y	7 y	15 y	9 y 11 m	9 y 5 m	2 y	14 y
Sex	male	male	female	male	male	ND	female
ID	mild	mild/moderate	severe	moderate	severe	+	moderate
ASD	autistic traits	PDD-NOS	PDD-NOS	+	+	ND	+
ADHD	+	+	-	ND	ND	ND	+
Language impairment	+	+	+	+	+	ND	+
Regression	ND	-	-	ND	+	ND	ND
Anxiety	+	ND	+	ND	ND	ND	ND
Obsessions/compulsions	+	+	+	+	ND	ND	+
Aggression	ND	ND	+	+	ND	ND	+
Motor delay	+	+	+	+	ND	ND	+
Hypotonia	+	+	+	ND	ND	+	+
Macrocephaly	+	-	-	ND	ND	+	+
Other	leukomalacia, sensory disorder	strabismus, enuresis, sensory disorder, hypertelorism	nystagmus, hypertelorism	jejunal/ileal atresia	seizures	-	-

Variants are annotated based on protein reference sequence NP_116071.2. Whole-gene deletions are not included here - see Le Fevre et al. (2013) for a summary of those cases. ND = no data; PDD-NOS = pervasive developmental disorder not otherwise specified.

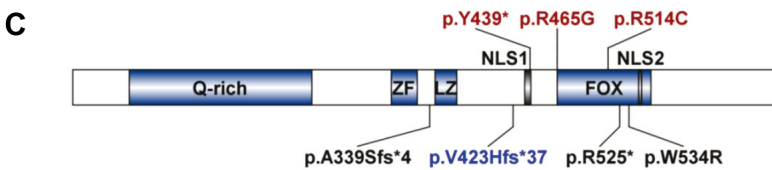
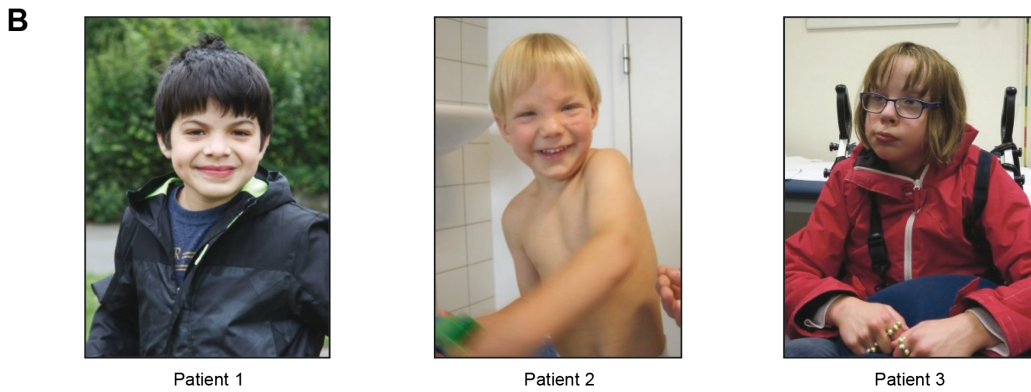
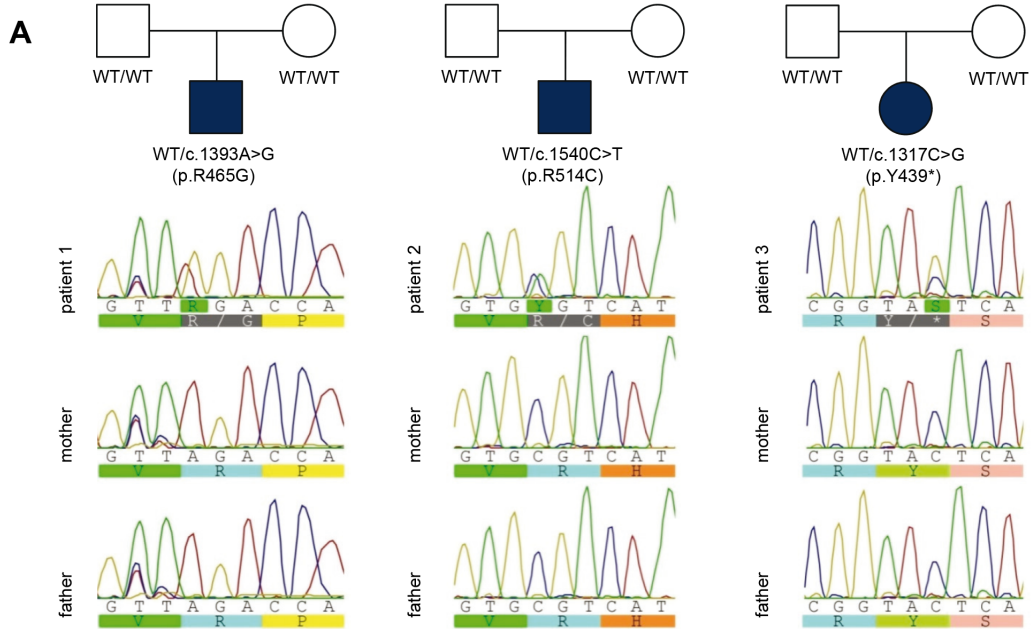


Figure 2.1 (opposite page). Identification of *de novo* FOXP1 variants in three patients with neurodevelopmental disorder. (A) Pedigrees of the three patients with Sanger traces of genomic DNA from the probands and their unaffected parents. (FOXP1 transcript accession number = NM_032682.5; FOXP1 protein accession number = NP_116071.2) (B) Photographs of patients 1, 2 and 3 at eleven, three and twelve years of age, respectively. (C) Schematic representation of the FOXP1 protein indicating *de novo* changes found in individuals with neurodevelopmental disorder. The three variants identified in this study are shown in red. The p.V423Hfs*37 variant characterised previously (Lozano et al., 2015) is shown in blue and three additional previously reported variants are shown in black. The major domains in FOXP1 are indicated: a glutamine-rich region (Q-rich), zinc finger (ZF), leucine zipper (LZ) and forkhead box (FOX) domains, and two nuclear localisation signals (NLS1 and NLS2).

Table 2.3. *De novo* FOXP1 variants in patients with neurodevelopmental disorder.

Study (proband)	Genomic coordinates (hg19)	Transcript variant	Type	Protein variant	PolyPhen2
This study (1)	chr3:71026829	c.1393A>G	missense	p.R465G	probably damaging
This study (2)	chr3:71021818	c.1540C>T	missense	p.R514C	probably damaging
This study (3)	chr3:71027010	c.1317C>G	nonsense	p.Y439*	N/A
Hamdan (B)	chr3:71021785	c.1573C>T	nonsense	p.R525*	N/A
O’Roak	chr3:71050170 71050171	c.1017_1018insT	frameshift	p.A339Sfs*4	N/A
Srivastava (41)	chr3:71021758	c.1600T>C	missense	p.W534R	probably damaging
Lozano*	chr3:71027059 71027060	c.1267_1268del	frameshift	p.V423Hfs*37	N/A

Variants are annotated based on the following reference sequences: transcript NM_032682.5, protein NP_116071.2. *Functional characterisation for the p.V423Hfs*37 described in the original publication (Lozano et al., 2015). N/A = not available (PolyPhen2 predictions only for missense variants).

2.3.3 FUNCTIONAL CHARACTERISATION OF AETIOLOGICAL *FOXP1* VARIANTS

Prior observations of *de novo* *FOXP1* deletions in sporadic ID strongly suggest that haploinsufficiency of this gene is the main pathogenic mechanism in *FOXP1*-related disorder (Carr et al., 2010; Hamdan et al., 2010; Horn et al., 2010; O’Roak et al., 2011; Le Fevre et al., 2013; Lozano et al., 2015). In addition, *de novo* truncating and frameshift *FOXP1* variants have been reported previously in individuals with ID, and most likely result in haploinsufficiency via nonsense mediated decay (NMD) of the variant transcript. Indeed, this has been demonstrated for the p.A339Sfs*4 and the p.V423Hfs*37 variants (O’Roak et al., 2011; Lozano et al., 2015). (In the case of the p.Y439* and p.R525* variants, no patient material was available to confirm NMD of the altered transcript.)

We hypothesised that the *de novo* missense variants in *FOXP1* in patients 1 and 2 would also result in a loss of protein function. Only one other missense variant in *FOXP1* (p.W534R) has been reported to date, and, to our knowledge, the effect on function has not been investigated (Table 2.3) (Srivastava et al., 2014). Therefore, we sought to characterise the effects of all three currently known missense variants in *FOXP1* on protein function (Fig 2.1C). For comparison, we performed parallel characterisation on three aetiological truncating and frameshift variants, including the p.Y439* variant found in patient 3 and two previously reported variants (p.R525* and p.A339Sfs*4) (Fig 2.1C; Table 2.3) (Hamdan et al., 2010; O’Roak et al., 2011). The p.R525* variant truncates the protein within the FOX domain and abolishes transcriptional repression activity (Hamdan et al., 2010). The p.A339Sfs*4 variant truncates the protein between the zinc finger and leucine zipper domains and results in mislocalisation of the protein within the cell (O’Roak et al., 2011). We recently reported an additional aetiological frameshift variant in *FOXP1* (p.V423Hfs*37) (Fig 2.1C; Table 2.3) (Lozano et al., 2015). Because functional analyses for this variant have been reported elsewhere (Lozano et al., 2015), we did not assess the biological consequences of the p.V423Hfs*37 variant here.

2.3.4 EFFECTS OF *FOXP1* VARIANTS ON PROTEIN EXPRESSION

Wild-type (WT) *FOXP1* and the six aetiological *FOXP1* variants were expressed as fusions with YFP and produced proteins at the expected molecular weights (Fig 2.2A). Western blotting suggested that the p.A339Sfs*4 variant was expressed at a substantially higher level than WT *FOXP1*. We quantified expression levels of all variants in live cells based on fluorescence intensity, confirming the increased expression level of the p.A339Sfs*4 variant (Fig 2.2B). The dramatically increased expression of the p.A339Sfs*4 variant, but not of the p.Y439* variant, indicates that the region between residues 339 and 439 may play a role in the regulation of *FOXP1* protein levels. This region contains the leucine zipper dimerisation motif as well as a serine/threonine-rich region that may be subject to post-translational modification.

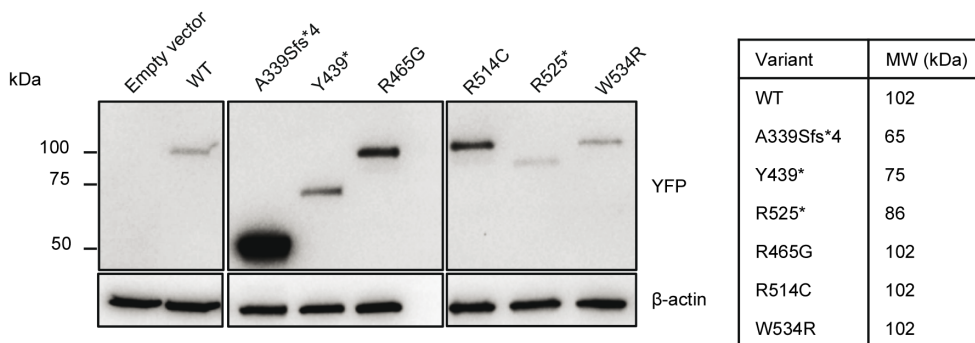
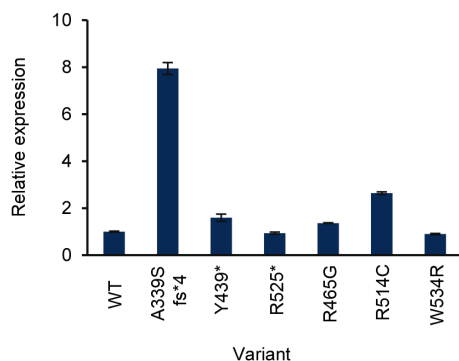
A**B**

Figure 2.2. Effects of *FOXP1* variants on protein expression. (A) Immunoblot of whole-cell lysates of cells expressing YFP-tagged *FOXP1* variants probed with anti-EGFP antibody. Blot was stripped and re-probed with anti- β -actin antibody to confirm equal loading. The predicted molecular weights of the YFP-*FOXP1* fusion proteins are indicated on the right-hand side. (B) Relative expression of *FOXP1* protein variants in live cells as assessed by YFP fluorescence (average of three experiments \pm S.D.).

2.3.5 SUBCELLULAR LOCALISATION OF FOXP1 VARIANTS

Direct fluorescence imaging of YFP-tagged FOXP1 variants showed that the WT protein localised to the nucleus and was excluded from nucleoli (Fig 2.3), as reported previously (Lozano et al., 2015). Strikingly, all six FOXP1 variants showed aberrant protein localisation, suggesting impaired function (Fig 2.3). The p.A339Sfs*4 variant displayed a diffuse distribution in the nucleus and cytoplasm, consistent with previous observations (O’Roak et al., 2011). This variant lacks both nuclear localisation signals (NLSs) (Fig 2.1C) but is small enough to passively diffuse into the nucleus. The p.Y439* variant formed large cytoplasmic aggregates and was entirely absent from cell nuclei, similar to the recently reported p.V423Hfs*37 variant, which is truncated at a similar position in the protein (Lozano et al., 2015). The p.R525* variant was also excluded from the nucleus, despite retaining one of the two NLSs, and formed large aggregates, suggesting misfolding of the aberrant protein.

Notably, all three missense variants displayed dramatically disturbed localisation patterns, despite having intact NLSs (Fig 2.3, 2.1C). The p.R465G and p.R514C variants formed cytoplasmic and nuclear aggregates, whereas the p.W534R variant was observed exclusively in cytoplasmic aggregates. The more severe effects of the p.W534R variant on protein localisation may be due to dramatic destabilisation of the FOX domain, as suggested by molecular modelling (p.W534R $\Delta\Delta G$ +5.35 kcal/mol; p.R465G $\Delta\Delta G$ +0.83 kcal/mol; p.R514C $\Delta\Delta G$ +0.74 kcal/mol).

2.3.6 TRANSCRIPTIONAL REPRESSION BY FOXP1 VARIANTS

To assess the effect of variants on the ability of FOXP1 to regulate transcription, we performed luciferase reporter assays using the SV40 promoter (Fig 2.4) (Vernes et al., 2006; Hamdan et al., 2010; Lozano et al., 2015). As reported previously, WT FOXP1 repressed luciferase activity ($P < 0.001$). All six FOXP1 variants showed significant ($P < 0.001$) loss of repressive activity, similar to the p.V423Hfs*37 variant reported previously (Lozano et al., 2015), suggesting that they would not be able to regulate transcription of target genes.

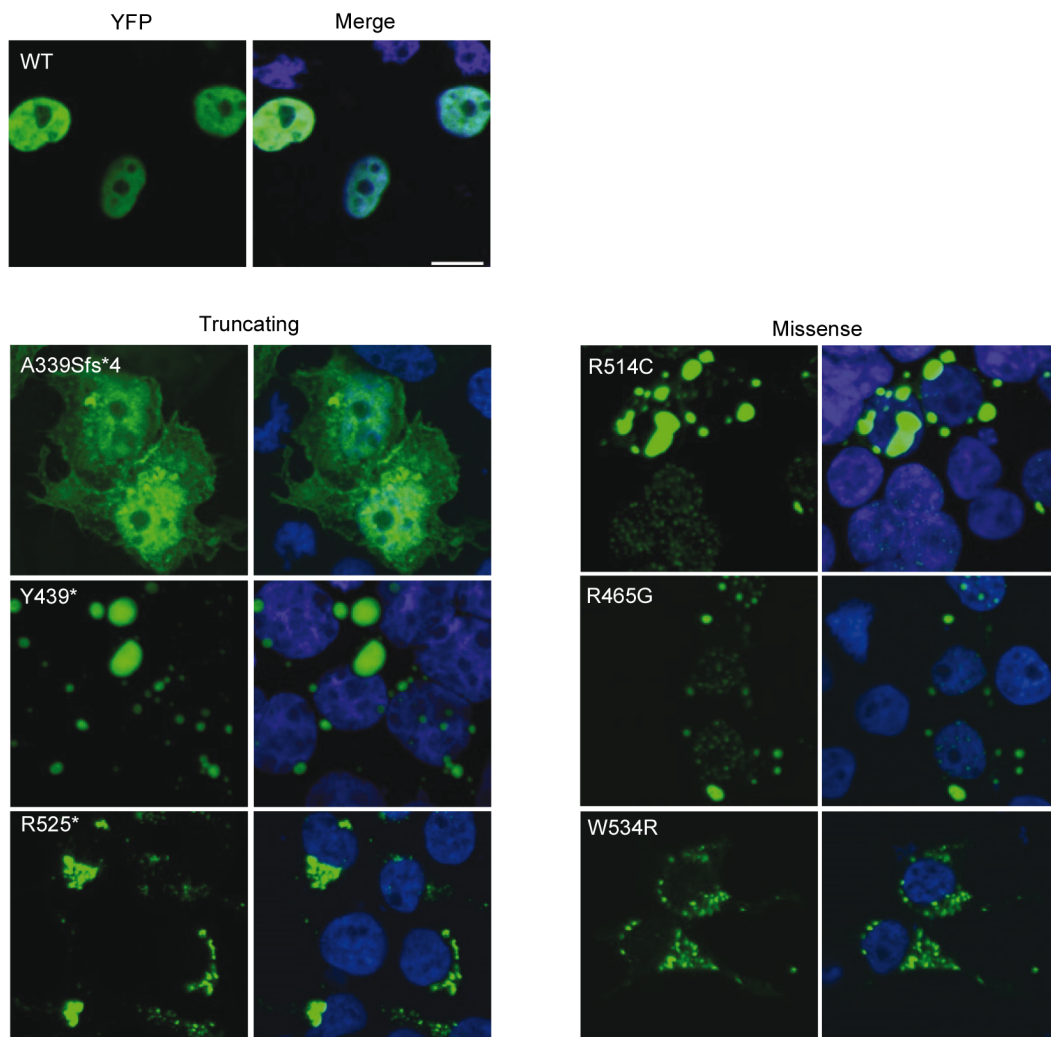


Figure 2.3. *FOXP1* variants severely disrupt cellular localisation. Fluorescence imaging of cells expressing YFP-tagged FOXP1 variants (green). Nuclei were stained with Hoechst 33342 (blue). Scale bar, 10 μ m.

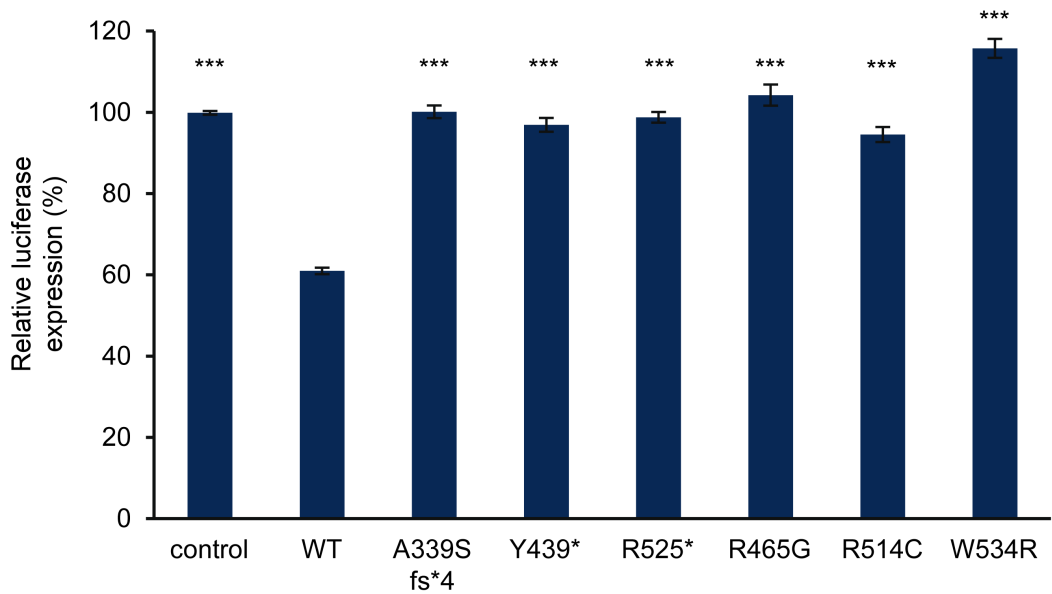


Figure 2.4. *FOXPI* variants abolish transcriptional repression. Luciferase reporter assays using the SV40 promoter. The mean \pm S.E.M. of three independent experiments is shown. Values are expressed relative to the control. All variants were significantly different to WT FOXPI (***) ($P < 0.001$).

2.3.7 PROTEIN INTERACTIONS OF FOXP1 VARIANTS

To examine the effects of the FOXP1 variants on self-association, and on interaction with WT FOXP1 and FOXP2 proteins, we employed the bioluminescence resonance energy transfer (BRET) assay, which monitors protein-protein interactions in live cells (Deriziotis et al., 2014a, 2014b; Lozano et al., 2015).

The p.Y439* variant retained the ability to interact with WT FOXP1 and FOXP2 and to self-associate (Fig 2.5A, 2.6A, 2.7), similar to the recently reported p.V423Hfs*37 variant (Lozano et al., 2015). Strikingly, co-transfection of the p.Y439* variant with WT FOXP1 or FOXP2 resulted in translocation of the WT proteins into cytoplasmic aggregates (Fig 2.5B, 2.6B). The other two truncating variants, p.A339Sfs*4 and p.R525*, showed a complete loss of interaction with WT FOXP1 and FOXP2, and were also unable to self-associate (Fig 2.5A, 2.6A, 2.7). Consistent with this loss of interaction, the localisation of WT FOXP1 and FOXP2 was not perturbed by co-transfection with these variants (Fig 2.5B, 2.6B). In the case of the p.A339Sfs*4 variant, the loss of interaction is expected due to the lack of the leucine zipper motif. However, the p.R525* variant retains the leucine zipper and could therefore be expected to interact with itself and WT FOXP proteins. Furthermore, the loss of interaction with WT proteins is unlikely to be due to mislocalisation of the variant since the variant also fails to self-associate (Fig 2.5, 2.6, 2.7). The absence of interactions may be due to misfolding of this variant.

Of the three missense variants, the p.R465G and p.R514C variants retained the ability to interact with WT FOXP1 and FOXP2 and to self-associate (Fig 2.5A, 2.6A, 2.7). Accordingly, when co-expressed with WT FOXP1 or FOXP2 proteins, the p.R465G and p.R514C variants led to mislocalisation of the WT proteins in nuclear aggregates (Fig 2.5B, 2.6B). These variants may therefore sequester WT protein and exert a dominant negative effect in patient cells. In contrast, the p.W534R variant showed loss of interaction with WT FOXP1 and FOXP2 and a reduced ability to self-associate (Fig 2.5A, 2.6A, 2.7). Furthermore, co-expression of p.W534R with FOXP1 or FOXP2 did not affect localisation of the WT proteins (Fig 2.5B, 2.6B). The loss of interactions resulting from the p.W534R variant may be due to protein misfolding.

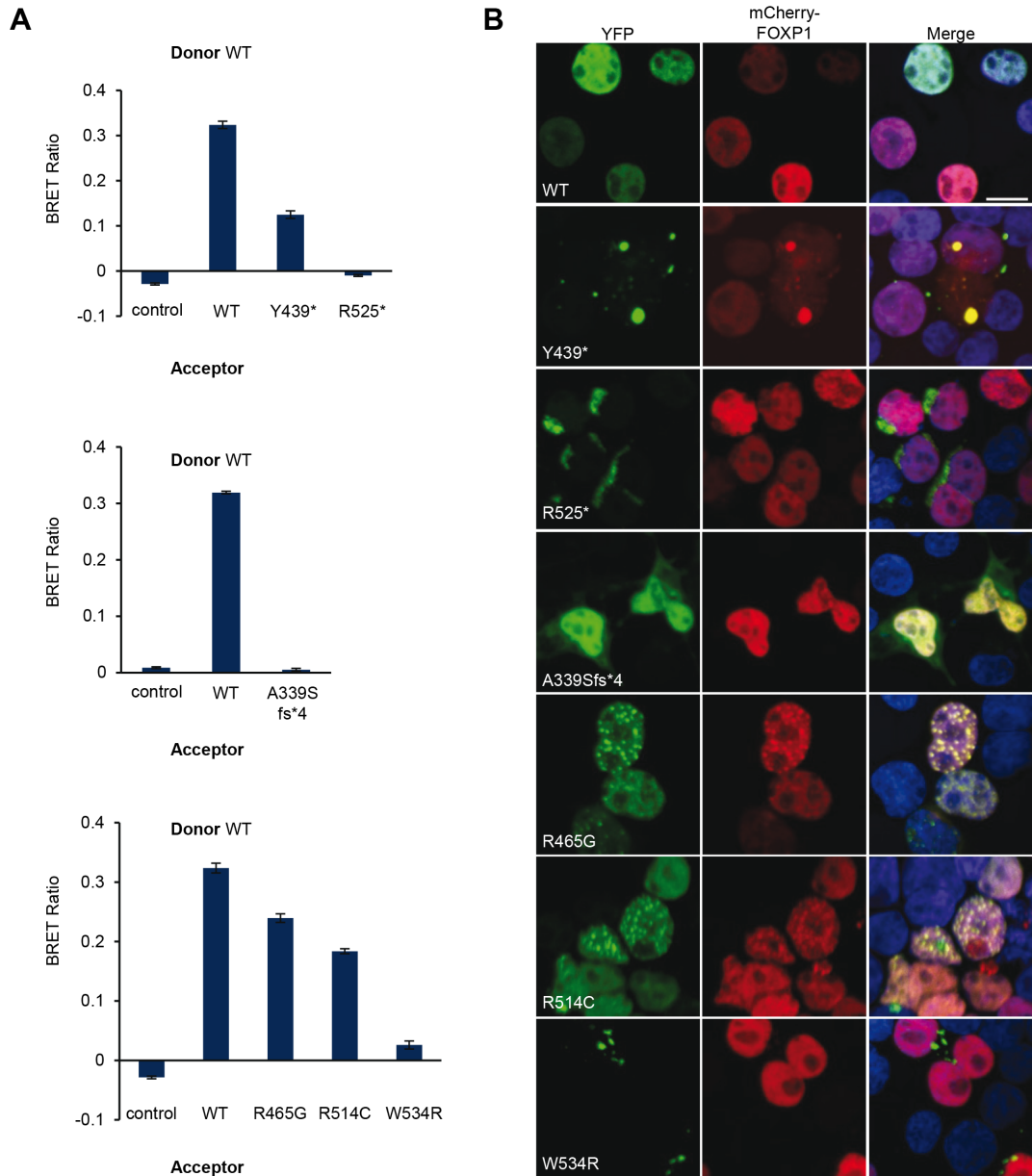
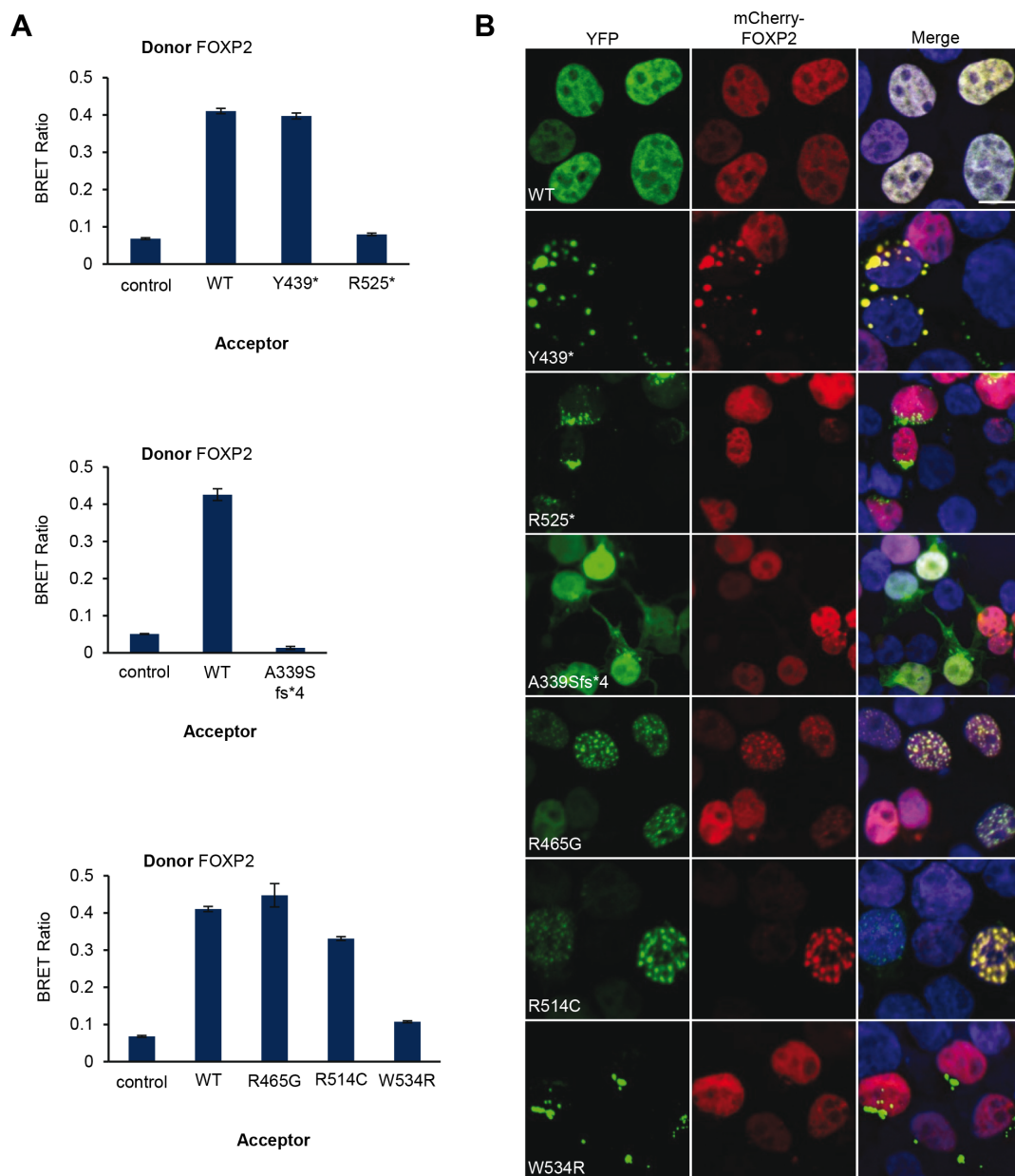


Figure 2.5. *FOXP1* variants disrupt interactions with WT *FOXP1*. (A) BRET assays for interaction between WT *FOXP1* and *FOXP1* variants. Bars represent the corrected mean BRET ratios \pm S.D. of one experiment performed in triplicate. (B) Fluorescence imaging of cells co-transfected with *FOXP1* variants and WT *FOXP1*. *FOXP1* variants fused to YFP are shown in green (left panel) and WT *FOXP1* fused to mCherry is shown in red (middle panel). Nuclei were visualised using Hoechst 33342 (blue). Scale bar, 10 μ m.



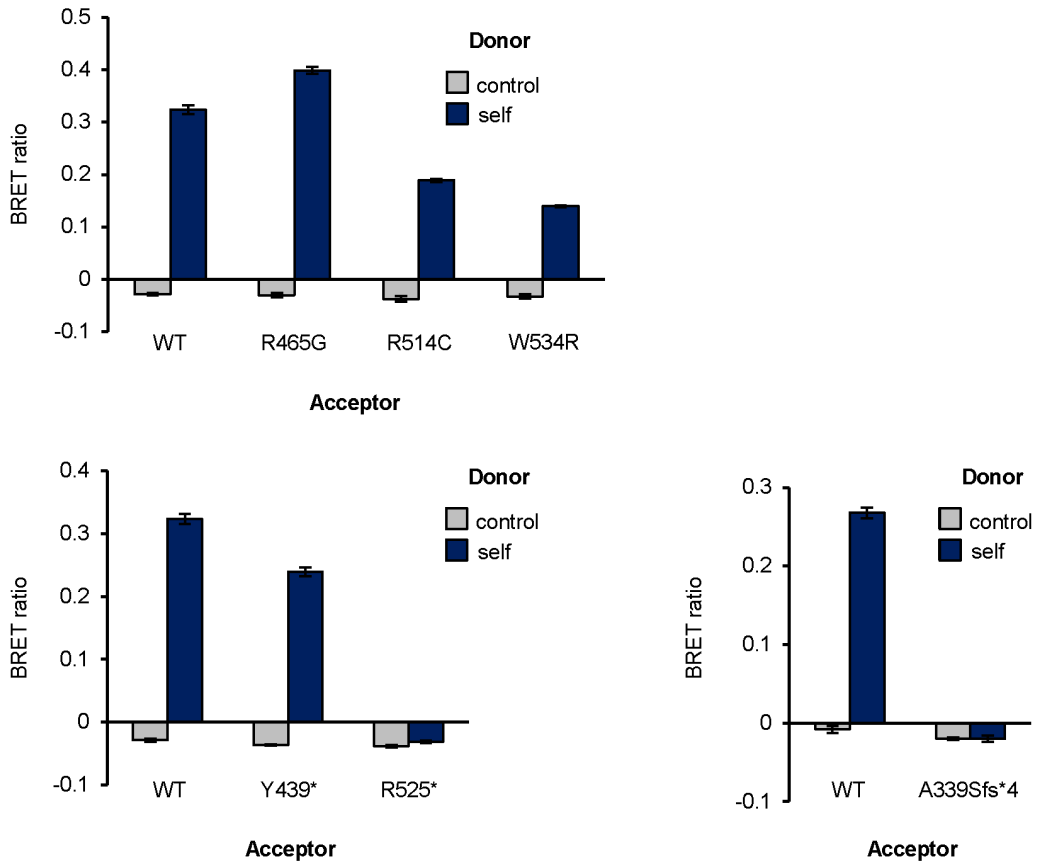


Figure 2.7. Effects of aetioloical FOXP1 variants on self-association. BRET assays for self-association of FOXP1 variants. Bars represent the corrected mean BRET ratios \pm S.D. of one experiment performed in triplicate.

2.3.8 *FOXP1* AS A RISK FACTOR IN ID AND CAS

Given the striking effects of the *de novo* variants identified in individuals with ID on FOXP1 protein function, we also assessed the effects of five additional *FOXP1* variants observed in probands with ID or CAS (Table 2.4, Fig 2.8A). Three of these variants - p.Q76dup, p.P215A, p.N570S - are known to have been inherited from unaffected parents, and the p.P215A and p.N570S variants have subsequently been found to occur at low frequency in the general population (Table 2.4) (Vernes et al., 2009; Hamdan et al., 2010; Horn et al., 2010). These variants may represent risk factors for ID, or cause a less severe form of *FOXP1*-related disorder that manifests primarily as a speech impairment. In this case they might have effects on protein function that are milder in comparison to the effects of *de novo* variants. The other two variants – p.I107T and p.N597T - are of unknown inheritance status, but have not been observed in the general population, such that it is possible that they may be causal *de novo* variants in the probands (Horn et al., 2010; Worthey et al., 2013). The p.I107T variant was found in a proband with CAS and was attributed a potentially causal role on the basis of the relationship of FOXP1 to FOXP2 (Worthey et al., 2013).

We found that all five variants were expressed and showed normal nuclear localisation and transcriptional repression activity (Fig 2.8B, 2.8C, 2.9A, 2.9B). Furthermore, all variants were able to self-associate and to interact with WT FOXP1 and FOXP2 (Fig 2.9C, 2.10, 2.11). Therefore, our data do not support the contribution of these variants to the disorders in these patients. The lack of functional effects of the inherited variants suggests that the involvement of FOXP1 in neurodevelopmental disorder may be limited to highly penetrant *de novo* variants. Of particular note, the lack of effects of the p.I107T and p.N597T variants on protein function emphasises the importance of performing functional characterisation of missense variants observed in known disorder genes before attributing a causal role, especially when the inheritance status is unknown.

Table 2.4. Variants of unknown significance in *FOXPI*.

Genomic coordinates (hg19)	dbSNP	MAF	Transcript	Effect on protein	Protein	PolyPhen2	Phenotypes	Inheritance
chr3:71161740_71161741	N/A	0	c.226_228dup	in-frame insertion	p.Q76dup	N/A	ID ¹	inherited
chr3:71096114	rs146606219	0.002	c.643C>G	missense	p.P215A	probably damaging	ID ^{1,2} CAS ³ controls ^{1,2,3}	inherited
chr3:71019900	rs140161845	0.002	c.1709A>G	missense	p.N570S	benign	ID ¹ controls ²	inherited
chr3:71015140	N/A	0	c.1790A>C	missense	p.N597T	benign	ID ¹	unknown
chr3:71102887	N/A	0	c.320T>C	missense	p.I107T	possibly damaging	CAS ⁴	unknown

Variants are annotated based on the following reference sequences: transcript NM_032682.5, protein NP_116071.2. Phenotype references: ¹ Horn et al. (2010), ² Hamdan et al. (2010), ³ Vermees et al. (2009), ⁴ Worthey et al. (2013). MAF = minor allele frequency in controls, from NHBLI Exome Sequencing Project (n = 13006). N/A = not available.

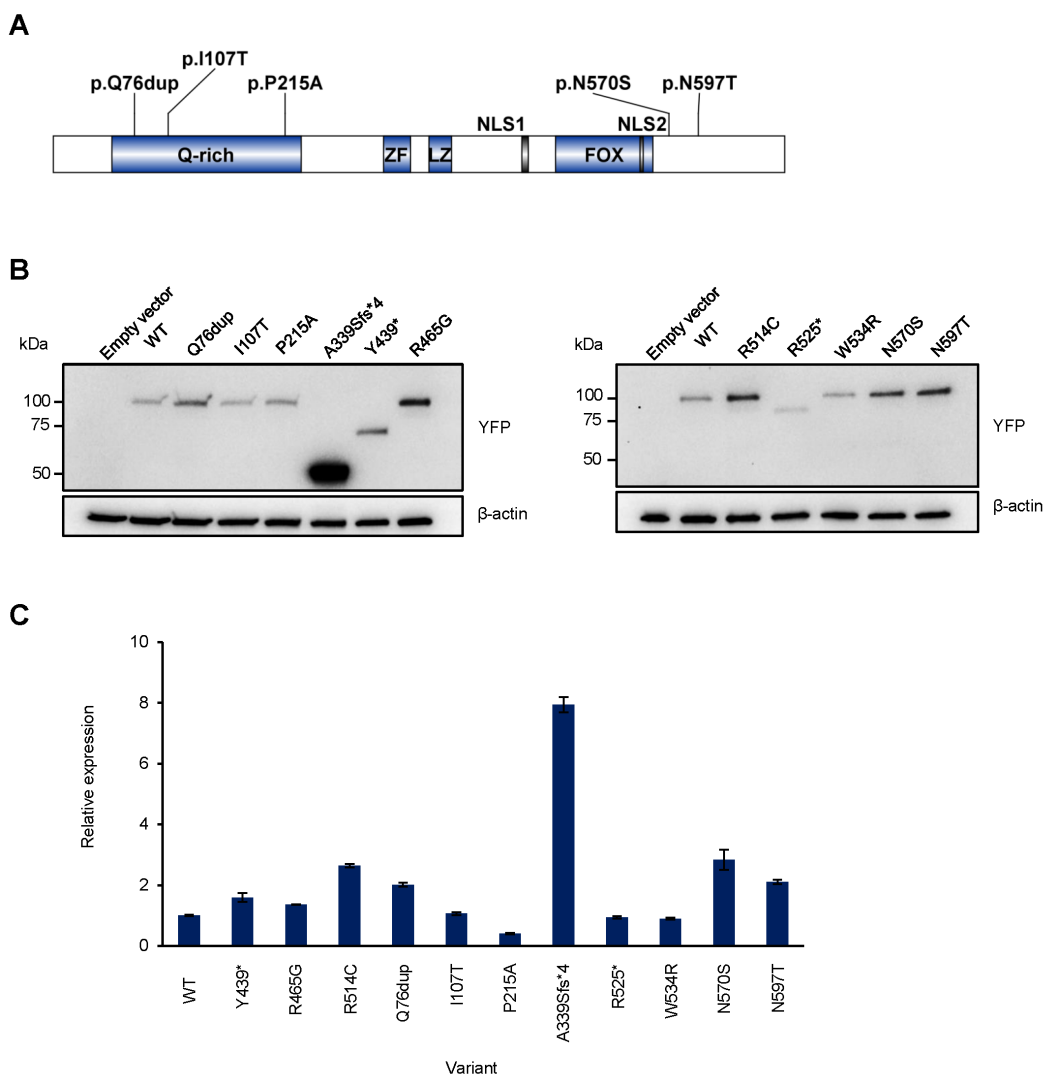


Figure 2.8. FOXP1 variants of unknown significance in neurodevelopmental disorder. (A) Schematic representation of the FOXP1 protein indicating variants of unknown significance found in cases of ID or CAS. The major domains in FOXP1 are indicated: a glutamine-rich region (Q-rich), zinc finger (ZF), leucine zipper (LZ) and forkhead box (FOX) domains, and two nuclear localisation signals (NLS1 and NLS2). (B) Immunoblot of whole-cell lysates of cells expressing FOXP1 variants probed with anti-EGFP antibody. Blots were stripped and re-probed with anti- β -actin antibody to confirm equal loading. (C) Relative expression of FOXP1 protein variants in live cells as assessed by YFP fluorescence (average of three experiments \pm S.D.).

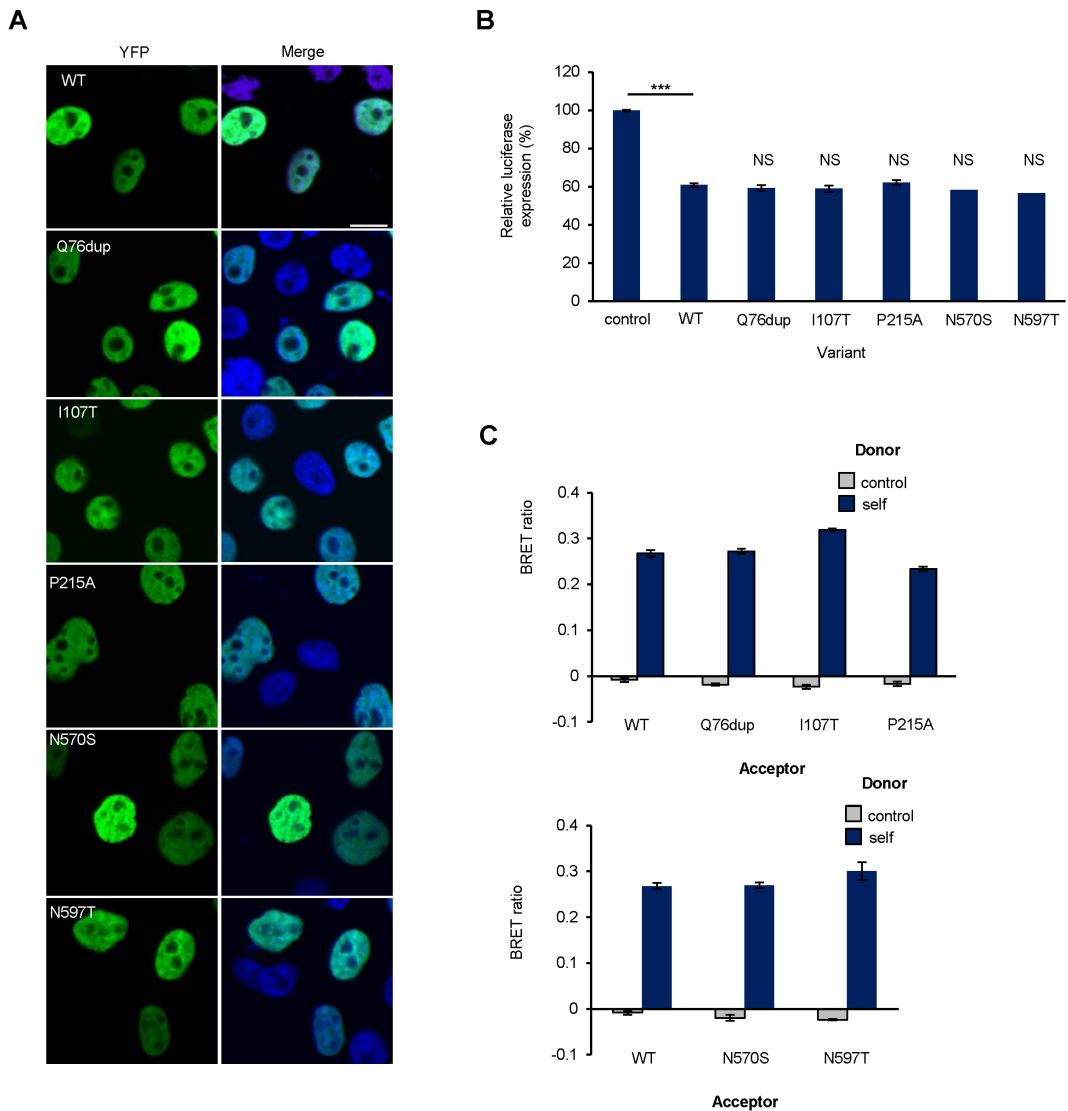


Figure 2.9. Functional characterisation of FOXP1 variants of unknown significance. (A) Fluorescence imaging of cells expressing YFP-tagged FOXP1 variants (green). Nuclei were stained with Hoechst 33342 (blue). Scale bar, 10 μ m. **(B)** Luciferase reporter assays using the SV40 promoter. The mean \pm S.E.M. of three independent experiments is shown. Values are expressed relative to the control. WT FOXP1 was significantly different to the control (***) P <0.001), but not to the FOXP1 variants. **(C)** BRET assays for self-association of FOXP1 variants. Bars represent the corrected mean BRET ratios \pm S.D. of one experiment performed in triplicate.

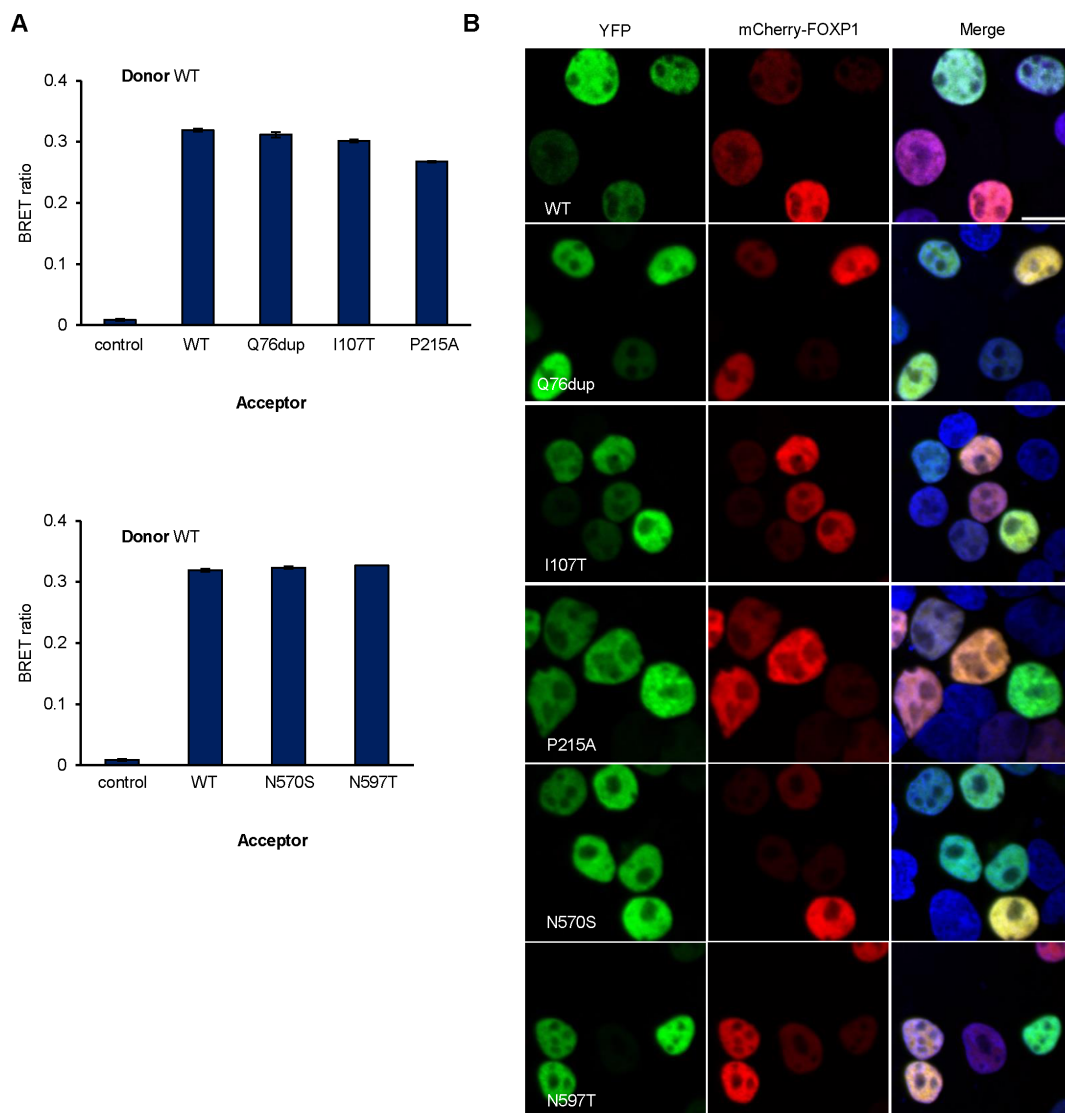


Figure 2.10. Effects of FOXP1 variants of unknown significance on interaction with WT FOXP1. (A) BRET assays for interaction between WT FOXP1 and FOXP1 variants. Bars represent the corrected mean BRET ratios \pm S.D. of one experiment performed in triplicate. (B) Fluorescence imaging of cells co-transfected with WT FOXP1 and FOXP1 variants. FOXP1 variants fused to YFP are shown in green (left panel) and WT FOXP1 fused to mCherry is shown in red (middle panel). Nuclei were visualised using Hoechst 33342 (blue). Scale bar, 10 μ m.

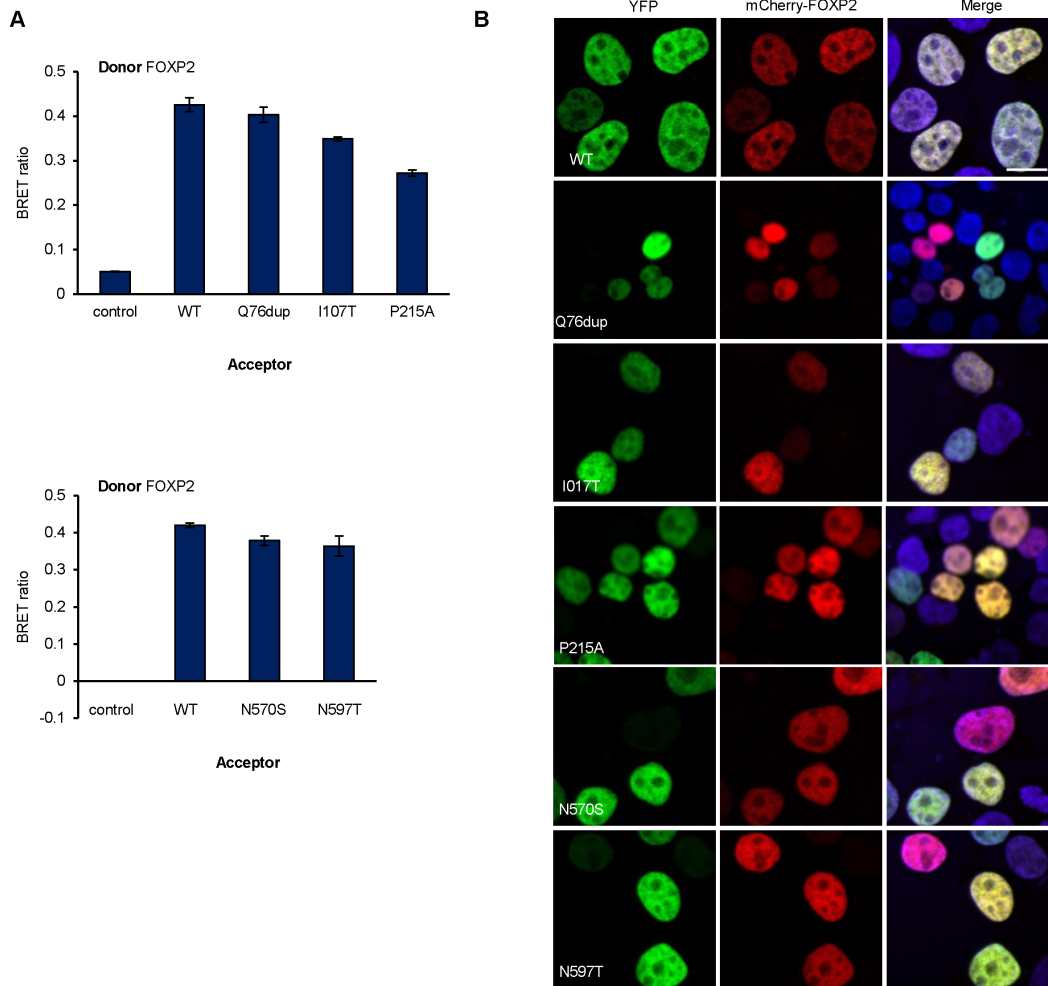


Figure 2.11. Effects of FOXP1 variants of unknown significance on interaction with WT FOXP2. (A) BRET assays for interaction between WT FOXP2 and FOXP1 variant proteins. Bars represent the corrected mean BRET ratios \pm S.D. of one experiment performed in triplicate. (B) Fluorescence imaging of cells co-transfected with WT FOXP2 and FOXP1 variants. FOXP1 variants fused to YFP are shown in green (left panel) and WT FOXP2 fused to mCherry is shown in red (middle panel). Nuclei were visualised using Hoechst 33342 (blue). Scale bar, 10 μ m.

2.4 DISCUSSION

Here we report three novel *de novo* *FOXP1* variants, including two missense changes, detected by clinical whole-exome sequencing in patients presenting with ID. We performed functional characterisation of these variants, together with three previously reported *de novo* variants, and found that all six severely disrupted multiple aspects of protein function (Table 2.5). Notably, missense variants had similarly deleterious effects on protein activity as truncating and frameshift variants, confirming that these variants have a causal role in disorder in these patients. The loss of function demonstrated by molecular screens for missense, nonsense and frameshift variants reported here, coupled with the observation of whole gene deletions of *FOXP1* in patients with ID (Pariani et al., 2009; Carr et al., 2010; Horn et al., 2010), indicates that haploinsufficiency is the main pathogenic mechanism in individuals with *FOXP1*-related disorder. However, the retention of dimerisation capacity of some of the variants suggests that the abnormal protein could exert a dominant negative effect by preventing WT *FOXP1* and *FOXP2* from binding to DNA and modulating transcription.

Table 2.5. Summary of functional characterisation of aetiological *FOXP1* variants.

Variant	Expression	Localisation	Repression	Interactions	
				FOXP1	FOXP2
p.R465G	similar to WT	N+C*	-		
p.R514C	similar to WT	N+C*	-		
p.Y439*	similar to WT	C*	-		
p.A339Sfs*4	> WT	N+C	-		
p.R525*	similar to WT	C*	-		
p.W534R	similar to WT	C*	-		

WT = wild-type. Localisation: N = nuclear, C = cytoplasmic, N+C = combination of nuclear and cytoplasmic localisation, *forms aggregates. Minus sign indicates loss of repressive function. Interactions: green = interaction, red = no interaction.

The three patients in our study share phenotypic characteristics, including developmental delay, mild to severe ID, autistic features, speech/language impairment (including articulation and pronunciation deficits), hypotonia, and obsessive-compulsive tendencies (Table 2.2). These traits are consistent with those previously described for individuals with whole gene deletions or protein-disrupting *FOXP1* variants (Carr et al., 2010; Hamdan et al., 2010; Horn et al., 2010; O’Roak et al., 2011; Le Fevre et al., 2013; Lozano et al., 2015). We observed additional physical features such as hypertelorism, and neurological traits such as sensory integration disorder, which have not been previously described, thereby extending the phenotype associated with *FOXP1*-related disorder. Interestingly, patient 2 was noted to have apraxia of the tongue in addition to articulation and pronunciation problems. Deficits in learning and executing the fine motor sequences of the mouth, lips, tongue and larynx

required for speech are seen in cases with disruptions of *FOXP2*, which is a paralogue of *FOXP1* (Lai et al., 2001). Oromotor dysfunction has also been previously reported in individuals with *FOXP1* deletions in the absence of dyspraxic speech (Horn et al., 2010; Le Fevre et al., 2013). Assessment of speech skills is warranted in cases of *FOXP1* disruption to ascertain if problems with orofacial praxis are a common phenotypic trait in *FOXP1*-related disorder.

The loss of transcriptional repression activity we observed for missense variants in our assays is consistent with the mechanism of DNA-binding by FOX transcription factors. The arginine at position 514, which is mutated in patient 2, is conserved in 49/50 human FOX proteins (Fig 2.12). No structure has been reported for FOXP1 bound to DNA, but the equivalent residue in FOXP2 (R553) is in the DNA-recognition helix and makes van der Waals contacts plus a water-mediated hydrogen bond to DNA bases and is therefore important for sequence-specific DNA binding (Stroud et al., 2006). The tryptophan at position 534 in FOXP1, which is mutated in a patient previously described (Srivastava et al., 2014), is conserved in all 50 human FOX proteins (Fig 2.12). The side chain of the equivalent residue in FOXP2 (W573) is in a β -strand in the FOX domain and makes a hydrogen bond to the DNA backbone (Stroud et al., 2006). The arginine at position 465 in FOXP1, which is mutated in patient 1, is conserved in all human FOXP subfamily proteins (Fig 2.12). Furthermore, all human FOX proteins (with the exception of the FOXO subfamily) have a positively charged residue at this position (Fig 2.12). The equivalent residue in FOXP2 (R504) lies at the N-terminus of the FOX domain and is important for DNA binding since it makes hydrogen bonds to the DNA backbone via its side chain (Stroud et al., 2006).

Strikingly, missense variants affecting residues equivalent to FOXP1 p.R514 have been reported in five FOX-related monogenic disorders. A missense variant affecting the equivalent residue in FOXP2 (R553) is responsible for CAS and expressive and receptive language deficits in a multigenerational pedigree (KE family) (Lai et al., 2001). In cellular assays, this variant (p.R553H) has been shown to cause mislocalisation of the protein, which forms nuclear and cytoplasmic aggregates (Vernes et al., 2006). The effects of the FOXP1 p.R514C variant identified here on protein localisation are slightly more severe than the effects of the FOXP2 p.R553H variant, possibly as a consequence of the substitution with cysteine rather than histidine. The FOXP2 p.R553H variant also resulted in loss of DNA-binding and transcriptional repression activity (Vernes et al., 2006), consistent with our observations on the FOXP1 p.R514C variant. Missense variants affecting residues homologous to FOXP1 p.R514 have also been reported in lymphedema-distichiasis syndrome (OMIM 153400) (FOXC2 p.R121H, heterozygous), Bamforth-Lazarus syndrome (OMIM 241850) (FOXE1 p.R102C, homozygous), alveolar capillary dysplasia with misalignment of pulmonary veins (OMIM 265380) (FOXF1 p.R97H, heterozygous) and blepharophimosis, ptosis and epicanthus inversus (OMIM 110100) (FOXL2 p.R103C, heterozygous) (Brice et al., 2002; Baris et al., 2006; Beysen et al., 2008; Sen et al., 2013).

A variant affecting the residue homologous to FOXP1 W534 has been reported in FOXC1 (p.W152G) in a case of aniridia and congenital glaucoma (OMIM 601090) and had

a severe impact on protein function (Ito et al., 2009). Variants affecting residues homologous to R465 have not been reported. However, while experiments were being performed for the current study, an additional missense variant in FOXP1 (p.P466L) was reported in a child with global developmental delay (Deciphering Developmental Disorders Study, 2015). The relevant residue is adjacent to the R465 residue affected by the missense variant in patient 1. Missense variants affecting the amino acid equivalent to P466 have been reported in FOXF1 (p.P49S; p.P49Q) in alveolar capillary dysplasia with misalignment of pulmonary veins (OMIM 265380) (Sen et al., 2013). In sum, the *FOXP1* variants identified in our patients mirror disease-causing variants across multiple FOX proteins, highlighting the key role of these amino acids in DNA binding by these transcription factors. This comparison of pathogenic variants in FOX family proteins underscores the value of considering paralogues and 3D structure together with functional studies in evaluating the aetiological contribution of novel variants arising from next-generation sequencing.

The identification of *FOXP1* disruptions in cases of ID revealed the critical role of this transcription factor in neurodevelopment. However, the precise functions of this gene in brain development remain unclear. In particular, neural target genes regulated by FOXP1 remain to be identified. *Foxp1* is expressed in the striatum, cerebral cortex (layers 3-5), hippocampus and thalamus (Ferland et al., 2003). Investigations of the consequences of loss of this protein in the mouse brain were initially precluded because global deletion of both copies of *Foxp1* results in embryonic lethality due to cardiac defects (Wang et al., 2004a). Recently, mice have been engineered that lack *Foxp1* only in the brain (Bacon et al., 2015). These animals display an imbalance of excitatory to inhibitory input in hippocampal neurons, but the most prominent effects are in the striatum. Following largely normal embryonic brain development, the mice develop enlargement of the ventral region of the striatum with reduction in the dorsal volume and increased ventricular volume during the first three post-natal weeks (Bacon et al., 2015). Moreover, striatal neurons display altered morphology. In addition to morphological and electrophysiological abnormalities, mice lacking *Foxp1* in the brain have behavioural alterations with potential relevance for the neurological phenotype of human individuals carrying *FOXP1* disruptions (Bacon et al., 2015). Specifically, these animals display hyperactivity, increased repetitive behaviours, impaired short-term memory, reduced social interests and elevated anxiety.

We have shown that several *FOXP1* variants found in patients with ID and language impairment result in aberrant interactions with FOXP2. FOXP1 is co-expressed with FOXP2 in regions of the brain, including the striatum and a small subset of cortical neurons (Ferland et al., 2003; Hisaoka et al., 2010). Striatal dysfunction in mice lacking *Foxp1* in the brain is particularly intriguing because of alterations affecting this region in mice with one functional copy of *Foxp2* (Groszer et al., 2008; French et al., 2012). These animals show impairments in motor-skill learning which may relate to electrophysiological anomalies in the striatum, including impaired long-term depression in corticostriatal synapses (Groszer et al., 2008; French et al., 2012). Future studies aimed at identifying downstream targets of the FOXP1 protein in the striatum, in the presence or absence of FOXP2 protein, will be crucial in

furthering our understanding of ways that these transcription factors work together to regulate gene expression during neurodevelopment.

While *FOXP1* disruptions reported to date in humans are all heterozygous, no abnormalities have yet been described in mice with heterozygous deletion of *Foxp1* either globally or in a brain-restricted manner (Wang et al., 2004a; Bacon et al., 2015). However, there has been no detailed analysis of brain anatomy or electrophysiology in mice with heterozygous *Foxp1* deletion. It is challenging to find appropriate phenotypes to assay in mouse models of human neurodevelopmental disorders because the surface phenotypes in the two species may differ while still reflecting a common neurobiological deficit. In the case of mice with heterozygous disruptions of *Foxp2* mirroring the genotype of humans with speech and language disorder, detailed characterisation was required to uncover behavioural and electrophysiological deficits (French and Fisher, 2014). Furthermore, these deficits relate to motor circuitry and motor-skill learning behaviour and not to vocalisation, which is not an appropriate proxy for human speech (French and Fisher, 2014). In future, it will be of interest to conduct detailed characterisation of mice with heterozygous *Foxp1* disruption to try to identify deficits which might model the underlying molecular mechanism of neurodevelopmental disorder seen in patients.

In summary, we report three new patients presenting with sporadic ID and carrying *FOXP1* disruptions. Our clinical assessment of these patients contributes to the delineation of the core phenotype of *FOXP1*-related disorder and has identified novel features which may help identify future cases. Furthermore, functional characterisation of the *FOXP1* variants found in these patients has both confirmed the diagnosis of *FOXP1*-related disorder and shed light on the molecular mechanisms that underlie this condition. Future work should focus on elucidating the role of this transcription factor in brain development.

Figure 2.12 (opposite page). Protein sequence alignment of FOX domains from human FOX transcription factors. *FOXP1* variants arising from *de novo* missense mutations, including the ones reported in this study, are indicated in red, whereas as the variant resulting from a *de novo* nonsense mutation is indicated in blue. UniProt accession numbers: *FOXP1* (Q9H334), *FOXP2* (O15409), *FOXP3* (Q9BZS1), *FOXP4* (Q8IVH2), *FOXO1* (Q12778), *FOXO3* (Q43523), *FOXO4* (P98177), *FOXO6* (A8MYZ6), *FOXM1* (Q08050), *FOXF1* (Q12946), *FOXF2* (Q12947), *FOXQ1* (Q9C009), *FOXC1* (Q12948), *FOXC2* (Q99958), *FOXS1* (O43638), *FOXA1* (P55317), *FOXA2* (Q9Y261), *FOXA3* (P55318), *FOXB1* (Q99853), *FOXB2* (Q5VYV0), *FOXD4L1* (Q9NU39), *FOXD4L2* (Q6VB85), *FOXD4L3* (Q6VB84), *FOXD4L4* (Q8WXT5), *FOXD4L5* (Q5VV16), *FOXD4L6* (Q3SYB3), *FOXD1* (Q16676), *FOXD2* (O60548), *FOXD3* (Q9UJU5), *FOXD4* (Q12950), *FOX E1* (O00358), *FOX E3* (Q13461), *FOXG1* (P55316), *FOXI1* (Q12951), *FOXI2* (Q6ZQN5), *FOXI3* (A8MTJ6), *FOX L1* (Q12952), *FOX L2* (P58012), *FOXJ1* (Q92949), *FOXJ2* (Q9P0K8), *FOXJ3* (Q9UPW0), *FOXK1* (P85037), *FOXK2* (Q01167), *FOXH1* (O75593), *FOXN1* (O15353), *FOXN2* (P32314), *FOXN3* (O00409), *FOXN4* (Q96NZ1), *FOXR1* (Q6PIV2), *FOXR2* (Q6PIQ5).

FOX1	R465G	R525*	R514C	W534R
FOX2				
FOX3				
FOX4				
FOX01				
FOX02				
FOX03				
FOX04				
FOX05				
FOX06				
FOX01				
FOX2				
FOX3				
FOX4				
FOX01				
FOX02				
FOX03				
FOX04				
FOX05				
FOX06				
FOX01				
FOX2				
FOX3				
FOX4				
FOX01				
FOX02				
FOX03				
FOX04				
FOX05				
FOX06				
FOX01				
FOX2				
FOX3				
FOX4				
FOX01				
FOX02				
FOX03				
FOX04				
FOX05				
FOX06				
FOX01				
FOX2				
FOX3				
FOX4				
FOX01				
FOX02				
FOX03				
FOX04				
FOX05				
FOX06				
FOX01				
FOX2				
FOX3				
FOX4				
FOX01				
FOX02				
FOX03				
FOX04				
FOX05				
FOX06				
FOX01				
FOX2				
FOX3				
FOX4				
FOX01				
FOX02				
FOX03				
FOX04				
FOX05				
FOX06				
FOX01				
FOX2				
FOX3				
FOX4				
FOX01				
FOX02				
FOX03				
FOX04				
FOX05				
FOX06				
FOX01				
FOX2				
FOX3				
FOX4				
FOX01				
FOX02				
FOX03				
FOX04				
FOX05				
FOX06				
FOX01				
FOX2				
FOX3				
FOX4				
FOX01				
FOX02				
FOX03				
FOX04				
FOX05				
FOX06				
FOX01				
FOX2				
FOX3				
FOX4				
FOX01				
FOX02				
FOX03				
FOX04				
FOX05				
FOX06				
FOX01				
FOX2				
FOX3				
FOX4				
FOX01				
FOX02				
FOX03				
FOX04				
FOX05				
FOX06				
FOX01				
FOX2				
FOX3				
FOX4				
FOX01				
FOX02				
FOX03				
FOX04				
FOX05				
FOX06				
FOX01				
FOX2				
FOX3				
FOX4				
FOX01				
FOX02				
FOX03				
FOX04				
FOX05				
FOX06				
FOX01				
FOX2				
FOX3				
FOX4				
FOX01				
FOX02				
FOX03				
FOX04				
FOX05				
FOX06				

3 EQUIVALENT MISSENSE VARIANT IN THE *FOXP2* AND *FOXP1* TRANSCRIPTION FACTORS CAUSES DISTINCT NEURODEVELOPMENTAL DISORDERS

PUBLISHED AS:

Sollis E, Deriziotis P, Saitsu H, Miyake N, Matsumoto N, Hoffer MJV, Ruivenkamp CAL, Alders M, Okamoto N, Bijlsma EK, Plomp AS, Fisher SE. 2017. Equivalent missense variant in the *FOXP2* and *FOXP1* transcription factors causes distinct neurodevelopmental disorders. *Hum Mutat* 38:1542–1554.

ABSTRACT

The closely related paralogues *FOXP2* and *FOXP1* encode transcription factors with shared functions in the development of many tissues, including the brain. However, while mutations in *FOXP2* lead to a speech/language disorder characterised by childhood apraxia of speech (CAS), the clinical profile of *FOXP1* variants includes a broader neurodevelopmental phenotype with global developmental delay, intellectual disability, and speech/language impairment. Using clinical whole-exome sequencing, we report an identical *de novo* missense *FOXP1* variant identified in three unrelated patients. The variant, p.R514H, is located in the forkhead-box DNA-binding domain and is equivalent to the well-studied p.R553H *FOXP2* variant that co-segregates with CAS in a large UK family. We present here for the first time a direct comparison of the molecular and clinical consequences of the same mutation affecting the equivalent residue in *FOXP1* and *FOXP2*. Detailed functional characterisation of the two variants in cell model systems revealed very similar molecular consequences, including aberrant subcellular localisation, disruption of transcription factor activity, and deleterious effects on protein interactions. Nonetheless, clinical manifestations were broader and more severe in the three cases carrying the p.R514H *FOXP1* variant than in individuals with the p.R553H variant related to CAS, highlighting divergent roles of *FOXP2* and *FOXP1* in neurodevelopment.

3.1 INTRODUCTION

The *FOXP2* (OMIM 605317; NM_014491.3; NP_055306.1) and *FOXP1* (OMIM 605515; NM_032682.5; NP_116071.2) genes are very closely related paralogues with important roles in embryonic development, including in the brain (Wang et al., 2004a; Shu et al., 2007; French and Fisher, 2014; Bacon et al., 2015). They encode transcription factors of the forkhead-box (FOX) family and display a high degree of similarity at the amino acid level (total protein: 64% identity, 82% similarity; FOX DNA-binding domain: 87% identity, 96% similarity). *FOXP2* and *FOXP1* can heterodimerise via a leucine zipper domain to regulate transcription (Li et al., 2004), and in brain regions where they are co-expressed, such as the striatum and certain cortical neurons in layers 5 and 6 (Ferland et al., 2003; Hisaoka et al., 2010), they may cooperatively regulate downstream targets (Vernes et al., 2008; O’Roak et al., 2011).

Heterozygous disruptions of *FOXP2* and *FOXP1* cause distinct neurodevelopmental phenotypes (Lai et al., 2001; Sollis et al., 2016). *FOXP2* variants cause a rare form of neurodevelopmental disorder characterised by severe speech deficits (childhood apraxia of speech; CAS) accompanied by impairments in expressive and receptive language affecting oral and written domains (OMIM 602081) (Lai et al., 2001). In contrast, mutations in *FOXP1* cause a broader neurodevelopmental syndrome involving global developmental delay, intellectual disability (ID), speech/language impairment and autistic features (OMIM 613670) (Sollis et al., 2016). These phenotypic differences are consistently evident despite a similar spectrum of causative variants in the two genes, which includes nonsense, frameshift and missense variants (Table 3.1, 3.2), as well as larger deletions (Feuk et al., 2006; Carr et al., 2010), and therefore seems to provide evidence of distinct roles for *FOXP2* and *FOXP1* in neurodevelopment. However, it is also possible that this disparity could instead be explained by the different amino acid changes so far documented in each gene (Table 3.1, 3.2). Thus far, there are no published studies comparing directly equivalent pathogenic variants in *FOXP2* and *FOXP1*.

Table 3.1. Published pathogenic variants in FOXP1.

Variant	Reference	Phenotype
Initiation loss	1	DD, ID, speech/language delay
p.P225T	2	DD
p.A339Sfs*4	3	DD, ID, speech/language delay, ASD
p.L414Dfs*46	2	DD
p.V283Pfs*11	2	DD, ID
p.V423Hfs*37	4	DD, ID, speech/language delay, ASD
p.T431Gfs*29	2	DD, ID
p.Y439*	Chapt. 2	DD, ID, speech/language delay, PDD-NOS
p.Q456*	5	behavioural problems
p.R465G	Chapt. 2	DD, ID, speech/language delay, autistic features
p.P466L	5	DD
p.R514C	Chapt. 2	DD, ID, speech/language delay, PDD-NOS, ADHD
p.R514H	This study	DD, ID, speech/language delay
p.H515D	2	DD
p.R525Q	2	DD
p.R525*	6	DD, ID, speech/language delay, ASD
p.W534R	7	DD, ID
p.F541Lfs*5	2	DD, behavioural problems

Missense, nonsense, frameshift and small indel variants are shown. Variants are annotated based on FOXP1 transcript NM_032682.5. References: ¹Song et al. (2015), ²Bekheirnia et al. (2017), ³O’Roak et al. (2011), ⁴Lozano et al. (2015), ⁵Deciphering Developmental Disorders Study (2015), ⁶Hamdan et al. (2010), ⁷Srivastava et al. (2014). Phenotypes: DD = developmental delay, ID = intellectual disability, ASD = autism spectrum disorder, PDD-NOS = pervasive developmental disorder not otherwise specified.

Table 3.2. Published pathogenic variants in FOXP2.

Variant	Reference	Phenotype
p.Q17L	1	CAS
p.Q188 Q191dup	1	CAS
p.R328*	1, 2	CAS, language impairment
p.Q390Vfs*7	3	CAS, language impairment
p.M406T	4	Language impairment, cognitive impairment
p.R478*	2	Speech problems, language impairment
p.P505L	2	Speech problems, language impairment
p.R536P	2	Language impairment
p.F538Lfs*28	2	Speech problems, language impairment
p.R553H	5	CAS, language impairment
p.R564*	2	Language impairment
p.N597H	6	CAS, language impairment

Missense, nonsense, frameshift and small indel variants are shown. Variants are annotated based on FOXP2 transcript NM_014491.3. References: ¹MacDermot et al. (2005), ²Reuter et al. (2017), ³Turner et al. (2013), ⁴Roll et al. (2010), ⁵Lai et al. (2001), ⁶Laffin et al. (2012). Phenotypes: CAS = childhood apraxia of speech.

The most rigorously studied aetiological *FOXP2* variant is an arginine-to-histidine substitution at residue 553 (p.R553H) co-segregating with CAS in multiple members of a large multigenerational UK pedigree (Lai et al., 2001). Note that other studies, such as Reuter et al. (2016), have used a different isoform (NM_148898.3; NP_683696.2) for *FOXP2* annotation, and therefore refer to the p.R553H variant as p.R578H. The affected arginine residue in *FOXP2* lies within the FOX DNA-binding domain and makes direct contact with the backbone of the target DNA to which the protein binds when acting as a transcription factor (Stroud et al., 2006). Human cell-based assays have shown that the p.R553H variant alters subcellular localisation and abolishes transcriptional repression activity (Vernes et al., 2006). Moreover, electrophoretic mobility shift assays (EMSA) have robustly demonstrated that the p.R553H variant prevents the FOX domain from binding to DNA (Vernes et al., 2006). The functional importance of R553 is further highlighted in *in vivo* studies of mice that are heterozygous for an equivalent p.R552H variant and display impaired motor skill learning, decreased synaptic plasticity and altered firing properties in corticostriatal circuits, as well as producing sequences of ultrasonic vocalisations with reduced complexity (Groszer et al., 2008; French et al., 2012; Chabout et al., 2016).

In the present study, we report for the first time an arginine-to-histidine substitution at the equivalent residue of *FOXP1* (p.R514H), the result of an identical heterozygous *de novo* variant in three unrelated probands. This provides a unique opportunity to directly compare equivalent mutations in *FOXP2* and *FOXP1*. We present thorough functional characterisation of the p.R514H *FOXP1* variant in human cellular models, assessing multiple protein characteristics and directly comparing the effects to those of the equivalent p.R553H *FOXP2* variant. We also compare the clinical profile of the three p.R514H *FOXP1* patients to the p.R553H *FOXP2* phenotype.

3.2 METHODS

3.2.1 WHOLE-EXOME SEQUENCING

For Patient 1, WES was performed as previously described (Fukai et al., 2015). In brief, approximately 3 µg DNA was sheared and used for a SureSelect Human All Exon V5 library (Agilent Technologies, Santa Clara, CA) according to the manufacturer's instructions. Samples were sequenced on a HiSeq2000 (Illumina, San Diego, CA) with 101-bp paired-end reads. Of all variants within exons or ±30 bp from exon–intron boundaries, those registered in dbSNP137 (minor allele frequency ≥0.01), the National Heart Lung and Blood Institute Exome Sequencing Project Exome Variant Server (NHLBI-ESP 6500; see URL¹) and an in-house database (exome data from 575 Japanese individuals) were removed. Variants were confirmed by Sanger sequencing using an ABI PRISM 3500xl autosequencer (Life Technologies, Carlsbad, CA).

¹ Exome Variant Server: <http://evs.gs.washington.edu/EVS/>

For Patient 2, exome sequencing was performed by GenomeScan, Leiden, the Netherlands, where exomes were enriched with the SureSelect Human All Exon V5 kit (Agilent) followed by HiSeq2500 system sequencing. The in-house sequence analysis pipeline MAGPIE (Modular GATK-Based Variant Calling Pipeline) based on read alignment using Burrows-Wheeler Alignment (BWA) (Li and Durbin, 2009) and variant calling using Genome Analysis Toolkit (GATK) (McKenna et al., 2010) was used for quality control, and to generate BAM and VCF files. Variants were annotated using Variant Effect Predictor (VEP, Ensembl). Before variant analysis and interpretation was started, intergenic and frequent variants (>5% present in Genome of the Netherlands or the 1000 Genomes Project database) were excluded. Further filtering and analysis steps were performed using a custom-made version of the Leiden Open Variation Database (LOVD) called LOVDplus. The variant was confirmed by Sanger sequencing.

For Patient 3, whole exome sequencing (WES) was performed in a trio diagnostic approach (patient and both parents). Libraries were prepared using the Kapa HTP kit (Illumina) and capture was performed using the SeqCap EZ Exome v3.0 (Roche Nimblegen, Madison, WI). Sequencing was done on an Illumina HiSeq2500 HTv4 (Illumina) with paired-end 125-bp reads. Read alignment to hg19 and variant calling were done with a pipeline based on BWA-MEM0.7 and GATK 3.3. Variant annotation and prioritising were done using Cartagenia NGS Bench (Cartagenia Inc Cambridge, MA). Only one *de novo* variant was found in a gene panel for ID (consisting of 842 genes). The variant was confirmed by Sanger sequencing.

For all patients, informed consent was obtained for the use of the data and photographs according to relevant institutional and national guidelines and regulations.

3.2.2 DNA CONSTRUCTS

WT *FOXP1/2*, *TBRI* (OMIM 604616) and *CTBP1/2* (OMIM 602618; 602619) were amplified by PCR and subcloned into pLuc, pYFP and a modified pmCherry-C1 expression vector (Clontech) as previously described (Deriziotis et al., 2014a, 2014b; Estruch et al., 2016a). Variants were generated using the QuikChange II Site-Directed Mutagenesis Kit (Agilent) using the following primers: *FOXP1* p.R514H, sense 5'-ACGTGGAAGAATGCAGTGCATCATAATCTTAGTCTTCAC-3' and antisense 5'-GTGAAGACTAAGATTATGATGCACTGCATTCTTCCACGT-3'; *FOXP2* p.R553H, sense 5'-CTTGGAAGAATGCAGTACATCATAATCTTAGCCTGCAC-3' and antisense 5'-GTGCAGGCTAAGATTATGATGTACTGCATTCTTCCAAG-3'. All constructs were verified by Sanger sequencing. *FOXP* DNA variants are numbered according to the cDNA reference sequences NM_032682.5 (*FOXP1*) and NM_014491.3 (*FOXP2*), where +1 is the A of the ATG translation initiation codon. The initiation codon is codon 1.

3.2.3 CELL CULTURE AND TRANSFECTION

HEK293 cells (ATCC® CRL-1573™) were cultured in DMEM supplemented with 10% foetal bovine serum (both Invitrogen). Transfections were performed using GeneJuice, according to manufacturer's instructions (Merck-Millipore).

3.2.4 WESTERN BLOTTING

Cells were transfected with equimolar concentrations of WT or variant *FOXP1* expression plasmids and cultured for 24 h. Whole-cell lysates were extracted by treatment with lysis buffer (100mM Tris pH 7.5, 150mM NaCl, 10mM EDTA, 0.2% Triton X-100, 1% PMSF, protease inhibitor cocktail; all from Sigma-Aldrich) for 10 min at 4°C, before centrifuging at 10,000 x g for 30 min at 4°C to remove cell debris. Proteins were resolved on a 4-15% Tris-Glycine gel and transferred onto a polyvinylidene fluoride membrane (both Bio-Rad). Blots were probed with mouse anti-EGFP (for pYFP constructs; 1:8,000; Clontech) and mouse anti-β-actin (as loading control; 1:10,000; Sigma) overnight at 4°C, followed by incubation with HRP-conjugated goat anti-mouse IgG for 60 min at room temperature (1:2,000; Bio-Rad). Proteins were visualised using Novex ECL Chemiluminescent Substrate Reagent Kit (Invitrogen) and the ChemiDoc XRS+ System (Bio-Rad).

3.2.5 FLUORESCENCE MICROSCOPY

Cells were seeded onto coverslips coated with poly-L-lysine (Sigma-Aldrich) and were fixed 24 h post-transfection using 4% paraformaldehyde (Electron Microscopy Sciences) for 10 min at room temperature. YFP and mCherry fusion proteins were visualised by direct fluorescence. HisV5-tagged proteins were visualised by immunofluorescence, using anti-V5 primary antibody (SV5-Pk1, GeneTex; 1:500) and donkey anti-mouse Alexa 488 secondary antibody (Invitrogen). Nuclei were visualised with Hoechst 33342 (Invitrogen). Fluorescence images were obtained using a Zeiss Axio Imager M2 upright microscope.

3.2.6 LUCIFERASE REPORTER ASSAYS

Cells were seeded in 24-well plates and transfected with 45 ng of firefly luciferase reporter construct (pGL3-prom; Promega), 5 ng of *Renilla* luciferase normalisation control (pRL-TK; Promega) and 200ng *FOXP1* expression construct (WT or variant in pYFP) or empty vector (pYFP; control). Cells were lysed in 24-well plates with 1X Passive Lysis Buffer (Promega) 48 h post-transfection and transferred to opaque white 96-well plates for luminescence measurements. Firefly luciferase and *Renilla* luciferase activities were measured in a TECAN F200PRO microplate reader with injectors using the Dual-Luciferase Reporter Assay system (Promega). Briefly, luminescence due to firefly and then *Renilla* luciferase activity was measured for 10 seconds after addition of Luciferase Assay Reagent II and Stop & Glo Reagent, respectively. Each transfection was performed in triplicate and the experiment was conducted three times. The statistical significance of the luciferase

reporter assays was analysed using a one-way analysis of variance and a Tukey's *post hoc* test.

3.2.7 BRET ASSAYS

BRET assays were performed as previously described (Deriziotis et al., 2014a, 2014b). In summary, cells were transfected with pairs of *Renilla* luciferase and YFP-fusion proteins in 96-well plates. *Renilla* luciferase and YFP fused to a C-terminal nuclear localisation signal were used as control proteins. EnduRen luciferase substrate (Promega) was added to cells 48 h after transfection at a final concentration of 60µM and incubated for 4 h. Emission measurements were taken with a TECAN F200PRO microplate reader using the Blue1 and Green1 filters and corrected BRET ratios were calculated as follows: $[\text{Green1}_{(\text{experimental condition})} / \text{Blue1}_{(\text{experimental condition})} - \text{Green1}_{(\text{control condition})} / \text{Blue1}_{(\text{control condition})}]$. YFP fluorescence was then measured separately, with excitation at 485 nm and emission at 535 nm, to quantify expression of the YFP-fusion proteins. The statistical significance of the BRET assays was analysed using independent two-sample *t*-tests.

3.3 RESULTS

3.3.1 CLINICAL DESCRIPTION OF PATIENTS

Patient 1, a 2-year-old girl, is the third child of healthy and non-consanguineous Japanese parents with no family history of neurological disease (Fig 3.1A, Table 3.3). She was born at 36 weeks' gestation. Her birth weight was 2,380 g, length was 47.5 cm, and head circumference was 33 cm. She was hypotonic and her developmental milestones were delayed. At 19 months old, she could crawl and stand with support. Her physical growth was also disturbed. At 21 months old, she had a height of 76.2 cm (-2.1 SD), weight of 7.6 kg (-2.7 SD) and head circumference of 45.0 cm (-1.1 SD). At 2 years her language perception was poor and she did not speak any meaningful words. She exhibited signs of severe ID. During her infantile period, she experienced febrile seizures on three occasions. She also exhibited visual problems, including esotropia and hypermetropia. Biochemical examinations (blood cell count and blood smear, renal and liver function, uric acid, albumin, serum electrolytes, lactate, pyruvate, ammonia, amino acids, blood gases, thyroid function and serum transferrin analysis) were normal. Genetic tests with normal results included karyotyping and array CGH. Electroencephalography and brain magnetic resonance imaging (MRI) showed no significant abnormalities.

Patient 2 is a 3-year-old Dutch boy born to healthy non-consanguineous parents (Fig 3.1A, Table 3.3). The father has a healthy daughter from a previous relationship. The mother has 3 male cousins (in one sibship) with developmental delay early in childhood, with catch-up later on. Family history is otherwise normal. He was born at 36 weeks' gestation. His Apgar scores were 5/7/10 after 1/5/10 minutes respectively, with signs of foetal distress due to a nuchal cord wrapped multiple times around the neck. His birth weight was 2,500 g.

Developmental delay was noted at the age of 12 months. He started to walk independently at 26 months. At the age of 2 years and 5 months he did not have a pincer grip. Growth parameters were within the normal range. His speech was severely delayed and he used less than 10 single words, although comprehension was reported as good. On follow up at the age of 3 years and 6 months, he had developed more speech and was able to speak in sentences with 2-3 words. A recent IQ test (WIPPSIII-NL) showed a total IQ score of 60. He is described as a friendly boy, with a tendency to repetitive behaviour, but without major behavioural problems. Parents noticed that he is highly sensitive to temperature and to certain textures. He was born with undescended testes for which he had orchidopexy, a small umbilical hernia, and a sacral dimple. Other findings include a broad forehead, hypertelorism, short palpebral fissures with mild down slant at a younger age, and recurrent otitis media for which he received grommets. He has received treatment for strabismus and has been diagnosed with cerebral visual impairment. SNP-array showed 2 small copy number variations that were inherited from the unaffected mother.

Patient 3 is an 8-year-old Dutch boy born to healthy non-consanguineous parents (Fig 3.1A, Table 3.3). Growth delay was detected by ultrasound at 19 weeks. He was born at 39 weeks' gestation and had good Apgar scores. He spent 24 hours in an incubator with a little extra oxygen. Transient hypoglycaemia was also noted. His birth weight was 2,710 g. A large head circumference prompted brain echography, which revealed no abnormalities. At the age of 3 days, he received phototherapy for hyperbilirubinemia. He exhibited many uncontrolled movements and could walk unsupported just before his 3rd birthday. His speech is profoundly delayed, with no speech at 7 years and 11 months. A postnatal hearing test was normal. Contact was good during the neonatal period, but parents found him less alert and less interested in his surroundings than other babies. He has had severe sleeping problems (awakening at night followed by staying awake for a long time, sleeping in the daytime), which was managed with melatonin. He did not have severe behavioural problems but moved all day and showed occasional hand biting and head banging. He was born with undescended testes, underdevelopment of the scrotum and a small penis. He received orchidopexy for one testis; the other was not found. At 7 years 11 months, he had a height of 124.5 cm (-1.5 SD), weight of 26 kg (+1 SD) and head circumference of 57.7 cm (>+2.5 SD). Other findings include prominent forehead, widow's peak, curly hair (not familial), low and posteriorly-rotated ears, low nasal bridge, mildly anteverted nares, telecanthus, epicanthus, thin upper lip, wide internipple distance and foetal pads on the fingers. He has strabismus and has been diagnosed with amblyopia and hypermetropia. Cytogenetic investigation revealed a normal male karyotype. FISH with subtelomeric probes, fragile X screening and metabolic investigations all returned normal results. DNA-analysis of Noonan syndrome (like) genes (*PTPN11*, *SOS1*, *KRAS*, *RAF1*, *BRAF*, *MAP2K1*, *MAP2K2*, *HRAS*), as well as *PTEN*, identified no mutations. Array CGH also showed no abnormalities. Brain MRI showed mild widening of the extracerebral space.

3.3.2 *DE NOVO* MISSENSE *FOXP1* VARIANT IDENTIFIED IN ALL THREE CASES BY CLINICAL WHOLE-EXOME SEQUENCING

Clinical whole-exome sequencing was performed with DNA from the three probands and their unaffected parents to identify putative pathogenic variants. Using this method, we identified an identical heterozygous *de novo* missense *FOXP1* variant present in all three unrelated probands (Fig 3.1B; Table 3.3). Patient 1 carried an additional compound heterozygous variant in *PEX10* (OMIM 602859). Variants in *PEX10* cause peroxisome biogenesis disorder, characterised by hepatic and renal abnormalities and ID (OMIM 614870). Normal kidney and liver function in Patient 1 rule out a contribution towards the observed phenotype. No additional *de novo* variants were identified in the other two patients. Based on the available data it was not possible to determine for any patient whether the *de novo* mutation had arisen on the paternal or maternal copy of the gene. The variant (NM_032682.5: c.1541G>A) was validated as *de novo* by Sanger sequencing (Fig 3.1B) and has been submitted to the NCBI ClinVar database (see URL¹).

The *FOXP1* variant detected here results in an arginine to histidine substitution (p.R514H) within the DNA-recognition helix of the FOX DNA binding domain (Fig 3.1C). The severity of the mutation was assessed using PolyPhen-2 (v2.2.2r398; see URL²; Adzhubei et al., 2010) and found to be Probably Damaging, with a score of 0.999 (sensitivity 0.09, specificity 0.99). Cellular assays have demonstrated that the equivalent change in *FOXP2* (p.R553H) found in cases of CAS results in abnormal localisation, loss of DNA binding and transcriptional repression activity (Vernes et al., 2006). To investigate whether the p.R514H variant in *FOXP1* results in disruption of protein function and to enable comparisons to the p.R553H variant in *FOXP2*, we performed detailed functional characterisation of the two variants in parallel.

Figure 3.1 (opposite page). Identification of an identical *de novo* *FOXP1* variant in three unrelated patients with global developmental delay. (A) Photographs of Patients 1 (14 mo), 2 (top, 2 yr 6 mo; bottom, 3 yr 2 mo) and 3 (top, 11.5 mo; bottom, 3 yr). (B) Sanger traces of genomic DNA from the probands and their unaffected parents. (C) Schematic representation of recombinant FOXP1 and FOXP2 proteins used in our assays. Both proteins contain a glutamine-rich region (Q-rich), and zinc finger (ZnF), leucine zipper (LeuZ) and FOX DNA-binding domains. The p.R514H FOXP1 and p.R553H FOXP2 variants at equivalent positions within the FOX domain are also labelled. The following cDNA and protein reference sequences were used for annotation in this article: *FOXP1* transcript NM_032682.5 and protein NP_116071.2; *FOXP2* transcript NM_014491.3 and protein NP_055306.1.

¹ ClinVar: <http://www.ncbi.nlm.nih.gov/clinvar> (see accession number SCV000494541)

² PolyPhen2: <http://genetics.bwh.harvard.edu/pph2>

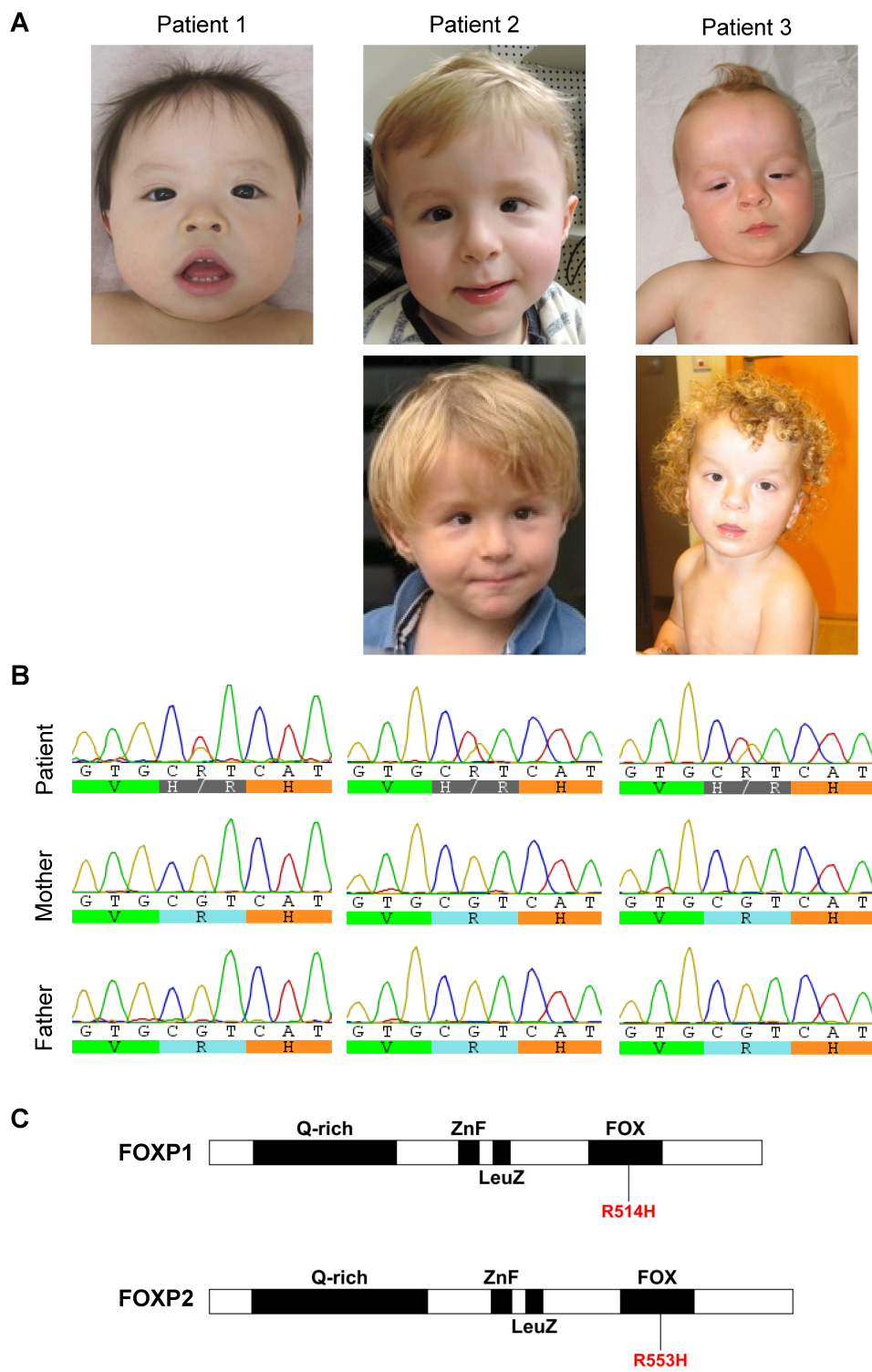


Table 3.3. Phenotypic comparison of patients with de novo variants at residue 514 of FOXP1.

	Patient 1	Patient 2	Patient 3	Chapter 2 (Patient 2)
Variant	p.R514H	p.R514H	p.R514H	p.R514C
Age	2 y	3 y 8 m	7 y 11 m	7 y
Sex	female	male	male	male
Neurodevelopmental features				
ID	moderate	mild	moderate/severe	mild/moderate
Speech/ language delay	+	+	+	+
Production more severely affected	+	+	+	ND
Autistic features	-	-	ND	+
Behavioural problems	ND	repetitive behaviour (not major)	hand biting, head banging (not major)	obsessions/compulsions, stereotypic behaviour, impulsive behaviour, ADHD
MRI	normal	ND	mild widening of extracerebral space	ND
Motor and sensory features				
Motor delay	+	+	+	+
Hypotonia	+	-	-	+
Sensory symptoms	-	sensitive to temperature/textures	-	sensory integration disorder
Visual symptoms	esotropia, hypermetropia	strabismus, cerebral visual impairment	strabismus, hypermetropia	strabismus

	Patient 1	Patient 2	Patient 3	Chapter 2 (Patient 2)
<i>Physical features</i>				
Growth delay	+	-	+	ND
Prominent forehead	+	-	+	ND
Macrocephaly	-	-	+	-
Eyes	down-slanted	short palpebral fissures, hypertelorism	telecanthus, epicanthus	small, down-slanted, hypertelorism
Short nose with broad tip	+	ND	+	+
Prominent digit pads	-	ND	+	ND
Urogenital anomalies	-	cryptorchidism	cryptorchidism, small penis	ND
Other physical features	-	sacral dimple	ears low-set and tilted back, thin upper lip, crooked little toes, curly hair (not familial)	mild retrognathism
Other medical problems	febrile seizures in infancy	recurrent otitis media	postnatal hyperbilirubinemia, severe sleep problems	enuresis

ND = no data; PDD-NOS = pervasive developmental disorder not otherwise specified; ADHD = attention deficit/hyperactivity disorder.

3.3.3 THE p.R514H FOXP1 VARIANT DISRUPTS MULTIPLE PROTEIN FUNCTIONS

FOXP variants were expressed as fusions with YFP or mCherry in HEK293 cells and produced proteins at the expected molecular weights (Fig 3.2B). Unlike wild-type (WT) FOXP1, which is diffusely expressed in the nucleus, the p.R514H variant showed a small increase in cytoplasmic expression and formed nuclear or cytoplasmic aggregates in ~30% of cells, consistent with loss of function (Fig 3.2A). Aberrant localisation of the variant protein is observed not only in YFP fusion proteins, but also in FOXP1 tagged with a smaller HisV5 epitope (Fig 3.3). Similarly, an increase in cytoplasmic expression and aggregation was observed for the variant at the equivalent residue in FOXP2 (p.R553H) (Fig 3.2A, Fig 3.3), as reported previously (Vernes et al., 2006). The crystal structure of the FOX domain from FOXP2 shows that R553 makes direct contact with the DNA backbone (Stroud et al., 2006). Accordingly, functional assays in human cells have shown that the p.R553H FOXP2 variant does not bind DNA carrying a consensus target sequence and cannot repress transcription (Vernes et al., 2006). Although there is no available crystal structure of FOXP1 bound to DNA, the FOX domains in FOXP1 and FOXP2 are 87% identical (Fig 3.4). Therefore, we hypothesised that R514 in FOXP1 may also be crucial for transcriptional regulation and used luciferase reporter assays to test this. Indeed, in our reporter assays the p.R514H FOXP1 variant completely abolished transcriptional repression activity (Fig 3.2C). Overall, these data indicate that, like R553 in FOXP2, this R514 residue is crucial for FOXP1 function.

To regulate transcription, FOXP1 forms homodimers and heterodimers with itself and other FOXP proteins, including FOXP2 (Li et al., 2004). We have previously demonstrated that *de novo* missense *FOXP1* variants located in the FOX domain may exert a dominant negative effect by interacting with and mislocalising WT FOXP proteins to nuclear and cytoplasmic aggregates (Sollis et al., 2016). Similarly, the p.R553H variant in FOXP2 has recently been shown to interact with and mislocalise WT FOXP1 and FOXP2 proteins to the cytoplasm (Estruch et al., 2016a). We used the bioluminescence resonance energy transfer (BRET) assay to monitor protein interactions in live cells and found that the p.R514H FOXP1 variant can interact with WT FOXP1 and FOXP2 proteins (Fig 3.5A, 3.5B). Moreover, consistent with prior observations for *FOXP* variants disrupting the FOX domain (Estruch et al., 2016a; Sollis et al., 2016), the p.R514H variant mislocalises WT FOXP1 and FOXP2 proteins to nuclear and cytoplasmic aggregates (Fig 3.5C, 3.5D). Overall, these findings suggest that the pathogenic mechanism in these three new patients may also involve a dominant negative effect, conferred by the p.R514H FOXP1 variant.

Few interaction partners for FOXP proteins are currently known. Thus far, FOXP1 and FOXP2 have been shown to interact with the transcriptional corepressors CTBP1 and CTBP2 (Li et al., 2004; Estruch et al., 2016a) and the ASD-related TBR1 transcription factor (Deriziotis et al., 2014b). We assessed FOXP1 interactions with these partners and found that the p.R514H variant retains the ability to interact with CTBP1 and CTBP2 and does not alter the localisation of these two proteins in co-transfection experiments (Fig 3.6). The same effect is seen with the p.R553H variant in FOXP2 (Fig 3.6) (Estruch et al., 2016a).

Interestingly, in our BRET assays neither the p.R514H FOXP1 variant, nor the p.R553H FOXP2 variant interacted with TBR1 (Fig 3.7A) (Deriziotis et al., 2014b). Previous work has shown that the FOX domain is not required for FOXP2-TBR1 interaction (Deriziotis et al., 2014b), suggesting that damage to the FOX domain does not directly account for the loss of interaction observed here. Instead, the loss of interaction may result from aberrant localisation of the FOXP variants. Indeed, although TBR1 partially colocalised with both FOXP variants when they occurred in the nucleus, TBR1 was absent from any nuclear or cytoplasmic aggregates formed by the variants (Fig 3.7B).

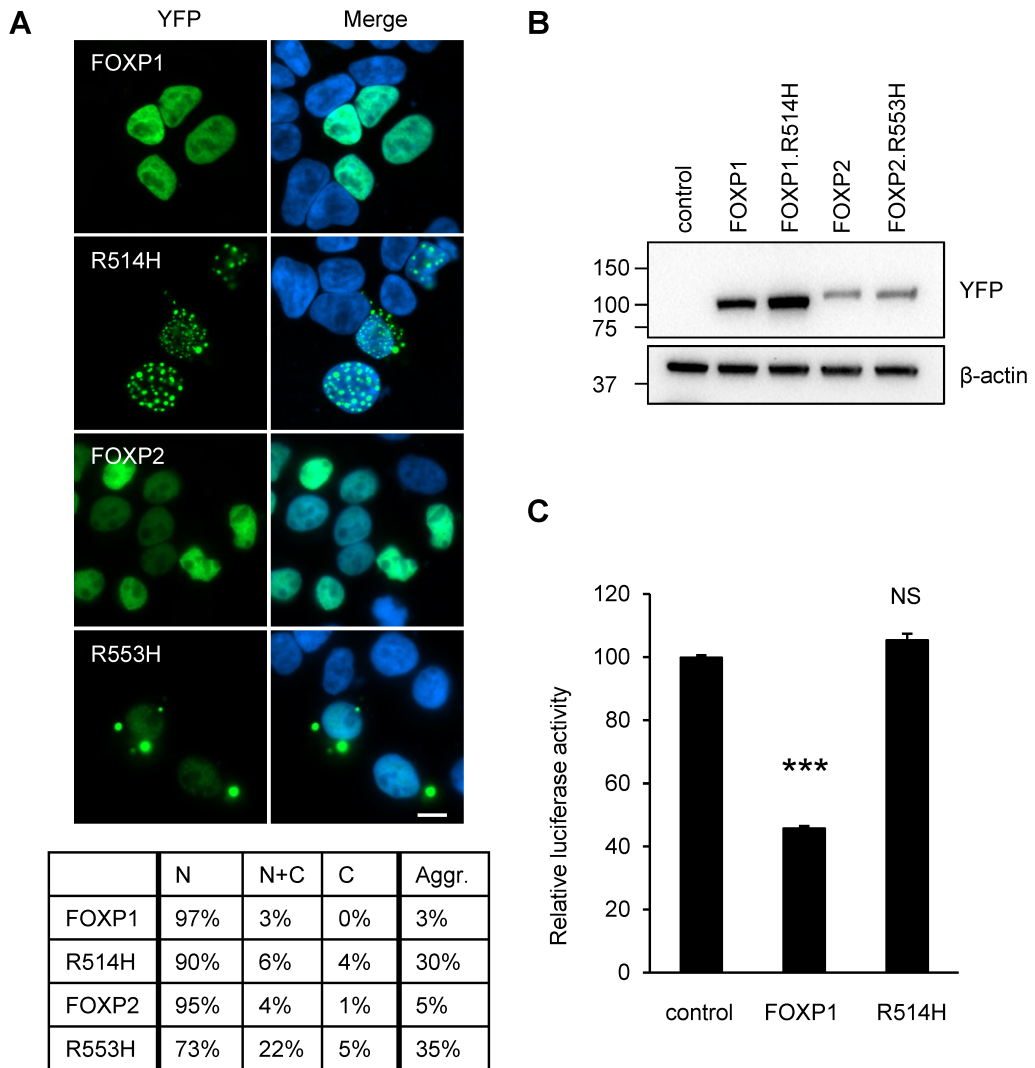


Figure 3.2. The p.R514H FOXP1 variant disrupts subcellular localisation. (A) Upper panel: Fluorescence microscopy images of HEK293 cells transfected with FOXP1/2 variants. FOXP proteins fused to YFP are shown in green. Nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μ m. Lower panel: The percentage of cells expressing each FOXP protein variant in the nucleus only (N), nucleus and cytoplasm (N+C) or cytoplasm only (C). The percentage of cells containing protein aggregates (Aggr.) are also shown. More than 400 cells were scored for each variant. (B) Immunoblotting of whole-cell lysates from HEK293 cells transfected with FOXP1 and FOXP2 variants fused to YFP. The control condition represents cells transfected with an empty pYFP plasmid. β -actin served as a loading control. (C) Luciferase reporter assays for transcriptional regulatory activity of the p.R514H variant in HEK293 cells. Values are expressed relative to the control (***) $P < 0.001$; NS, not significant). The mean \pm SEM of three independent experiments performed in triplicate is shown.

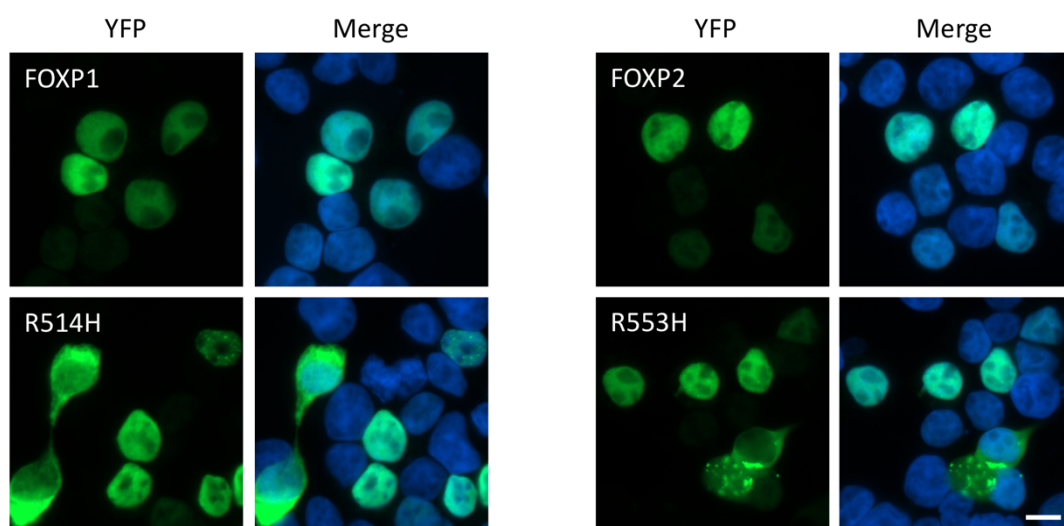


Figure 3.3. Aberrant subcellular localisation of the p.R514H FOXP1 and p.R553H FOXP2 variants in small epitope-tagged proteins. Immunofluorescence staining of HEK293 cells transfected with HisV5-tagged FOXP variants (green). Nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μm.

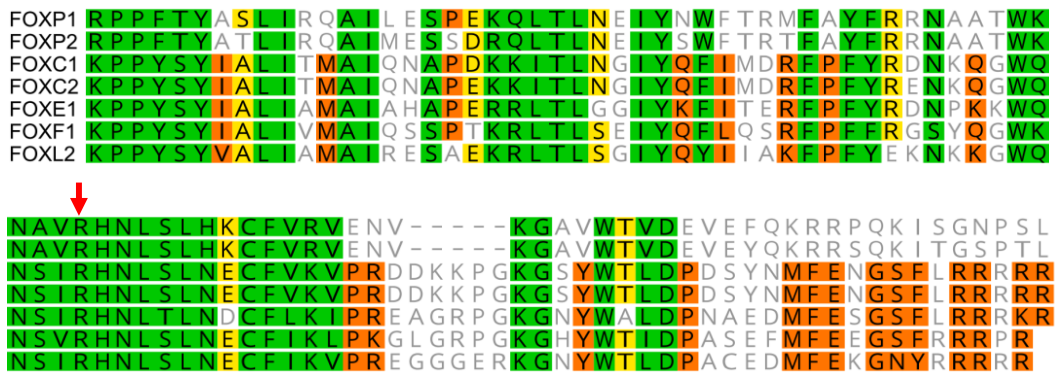


Figure 3.4. Protein sequence alignment of FOX domains from human FOX transcription factors. Red arrow indicates the arginine residue affected by the p.R553H FOXP2 and p.R514H FOXP1 variants. Pathogenic variants have been reported at the equivalent arginine residue in each of the FOX transcription factors shown here. UniProt accession numbers: FOXP1 (Q9H334), FOXP2 (O15409), FOXC1 (Q12948), FOXC2 (Q99958), FOXE1 (O00358), FOXF1 (Q12946), FOXL2 (P58012).

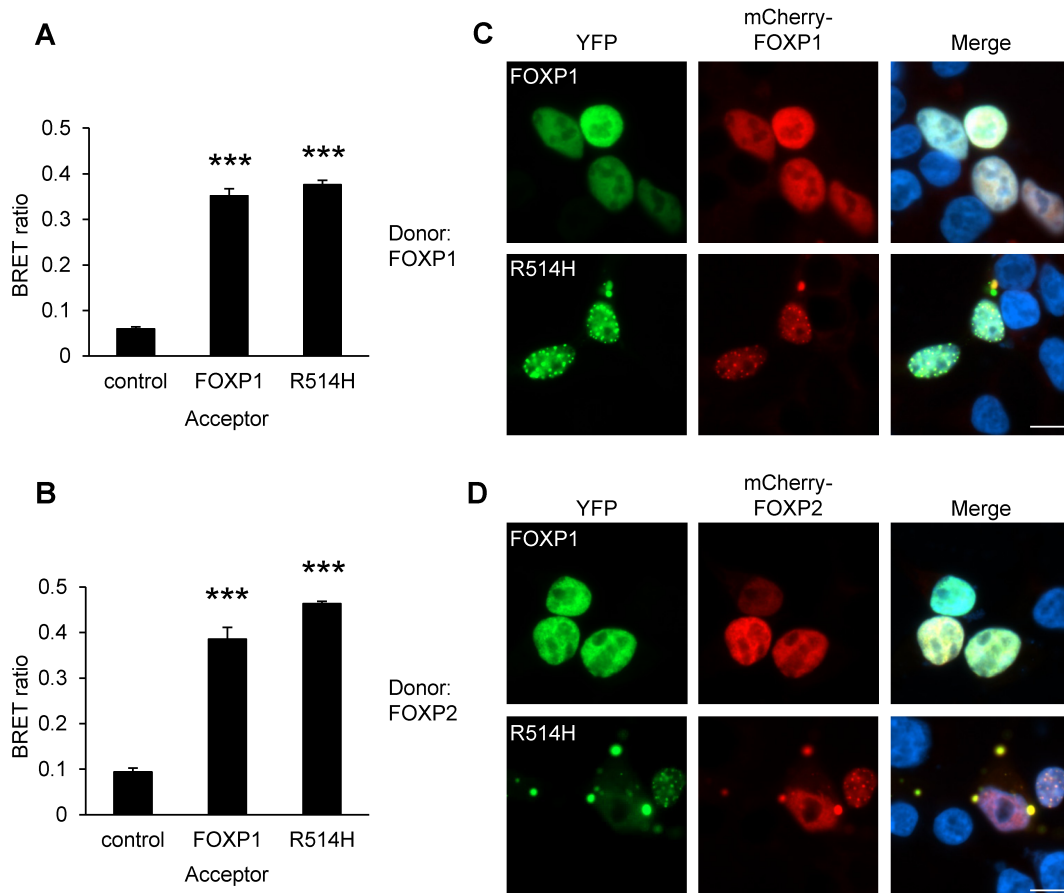


Figure 3.5. The p.R514H FOXP1 variant translocates WT FOXP proteins to nuclear and cytoplasmic aggregates. (A and B) BRET assays for interaction between the p.R514H FOXP1 variant and WT FOXP1 or FOXP2. Bars represent the corrected mean BRET ratios \pm SD of one experiment performed in triplicate. Asterisks indicate significant differences compared to control ($***P < 0.001$, independent two-sample *t*-test). (C) Fluorescence microscopy images of HEK293 cells co-transfected with WT FOXP1 (fused to mCherry, red) and either WT FOXP1 or p.R514H FOXP1 (fused to YFP, green). (D) Fluorescence microscopy images of HEK293 cells co-transfected with WT FOXP2 (fused to mCherry, red) and either WT FOXP1 or p.R514H FOXP1 (fused to YFP, green). Nuclei were stained with Hoechst 33342 (blue). Scale bars = 10 μ m.

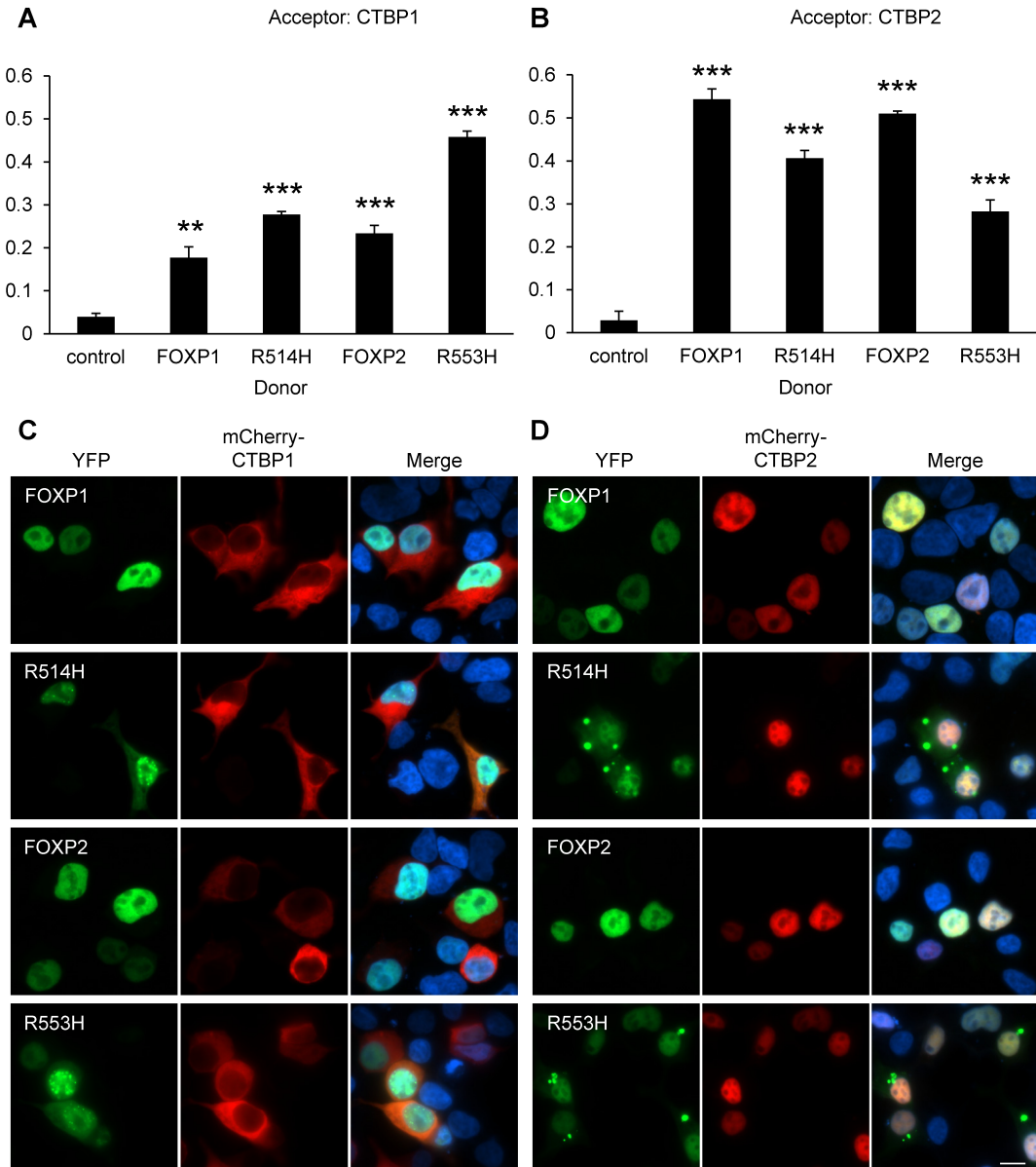


Figure 3.6. The p.R514H FOXP1 variant maintains interactions with CTBP1/2. BRET assays for interaction between the p.R514H FOXP1 variant and (A) CTBP1 or (B) CTBP2. The p.R553H FOXP2 variant is included for comparison. Bars represent the corrected mean BRET ratios \pm SD of one experiment performed in triplicate. Asterisks indicate significant differences compared to control (** $P < 0.01$, *** $P < 0.001$, independent two-sample t -test). (C) Fluorescence microscopy images of HEK293 cells co-transfected with CTBP1 (fused to mCherry, red) and FOXP1/2 variants (fused to YFP, green). (D) Fluorescence microscopy images of HEK293 cells co-transfected with CTBP2 (fused to mCherry, red) and FOXP1/2 variants (fused to YFP, green). Nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μ m.

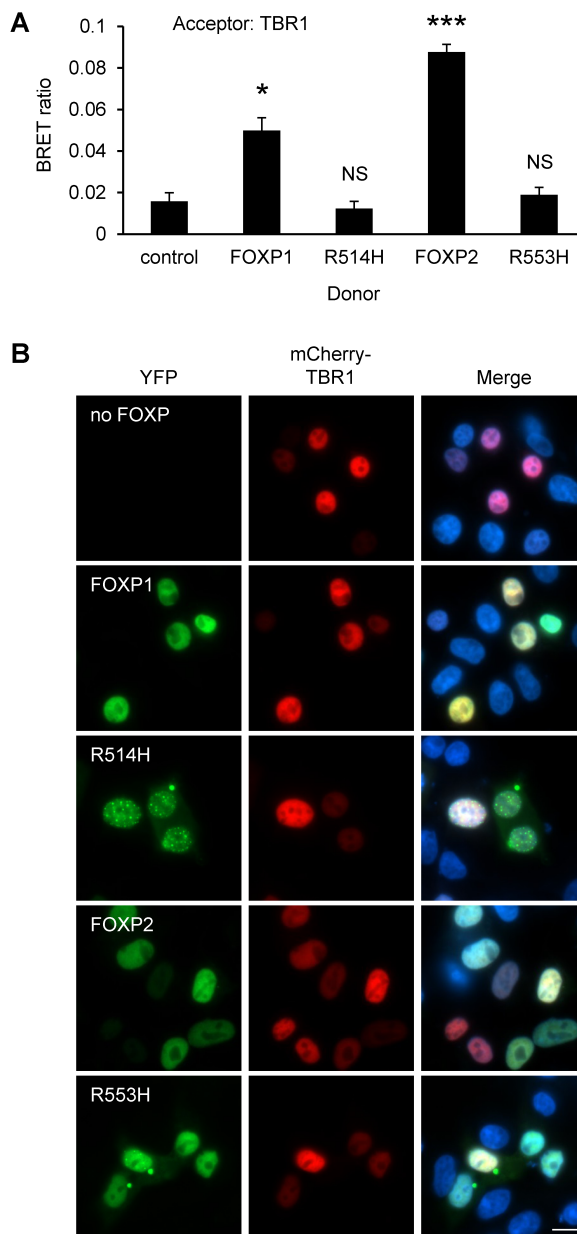


Figure 3.7. The p.R514H FOXP1 variant abolishes the interaction between FOXP1 and TBR1.

(A) BRET assay for interaction between the p.R514H FOXP1 variant and TBR1. The p.R553H FOXP2 variant is included for comparison. Bars represent the corrected mean BRET ratios \pm SD of one experiment performed in triplicate. Asterisks indicate significant differences compared to control ($*P < 0.05$, $***P < 0.001$, independent two-sample *t*-test). NS, not significant. (B) Fluorescence microscopy images of HEK293 cells co-transfected with TBR1 (fused to mCherry, red) and WT FOXP1, p.R514H FOXP1, WT FOXP2 or p.R553H FOXP2 (fused to YFP, green). Nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μ m.

3.4 DISCUSSION

Our parallel functional characterisation of the *de novo* p.R514H variant detected in three new patients and the pathogenic p.R553H FOXP2 variant previously found co-segregating with CAS has revealed similar effects on protein function. Both variants lead to aberrant subcellular localisation and loss of transcriptional repression activity, and they exert the same effect on protein interactions with TBR1, CTBP1 and CTBP2. Notably, the two variants dimerise with WT FOXP1 and FOXP2 and translocate these proteins into nuclear and cytoplasmic aggregates. While the existence of *FOXP* whole-gene deletions and truncating/frameshift variants points to haploinsufficiency as the key pathogenic mechanism, our findings suggest an additional dominant-negative effect, whereby the FOXP1/2 variant prevents the WT protein from binding DNA and regulating transcription.

Despite having very similar effects at the protein level, the matching arginine-to-histidine substitutions in FOXP1 and FOXP2 cause distinct neurodevelopmental phenotypes. The p.R514H FOXP1 variant results in broader and more severe effects on general cognition, motor development and behaviour, while the effects of the p.R553H FOXP2 variant are largely confined to speech, language processing and orofacial motor function, with milder consequences for other aspects of cognition and development (Vargha-Khadem et al., 1998; Lai et al., 2001). These observations may be partly explained by the different expression patterns of FOXP1 and FOXP2 in the brain. Although both proteins are found in the striatum, the hippocampus expresses only FOXP1, and the Purkinje cells of the cerebellum express only FOXP2 (Ferland et al., 2003). In the cortex, expression is largely non-overlapping, with FOXP1 detected in layers 3-5 and FOXP2 detected mainly in layer 6 and in restricted regions of layer 5 (Ferland et al., 2003; Hisaoka et al., 2010). In addition, FOXP1 and FOXP2 may be expressed in distinct neuronal subpopulations within the same brain regions, and even in cells where the two proteins are co-expressed, they may have distinct functions arising from differences in their downstream targets and/or interaction partners. *In vitro* studies suggest that certain genes important for nervous system development, including *NEUROD1* and *EFNB3*, might be differentially regulated by different combinations of FOXP1/2/4 homo- and hetero-dimers (Sin et al., 2015). Furthermore, an RNA-sequencing study comparing downstream targets of *Foxp1* and *Foxp2* in mouse striatum found that only 12% of putative *Foxp1* target genes were also targets of *Foxp2* (Araujo et al., 2015). Differences between the two proteins have also been noted on assays of protein interaction; for example, the SUMO-protein ligase PIAS3 interacts with FOXP2 but not FOXP1 in live-cell assays (Estruch et al., 2016b). Future work comparing the interactome of FOXP1 and FOXP2 in appropriate models may further our understanding of the pathogenic mechanisms underlying the two distinct neurodevelopmental disorders.

In this study, we demonstrated that the p.R514H FOXP1 variant, like the p.R553H FOXP2 variant, prevents interaction with TBR1. This may have *in vivo* relevance in neurons that co-express FOXP1 and TBR1, including in the hippocampus (Ferland et al., 2003; Cipriani et al., 2016) and a small subset of cortical neurons (Hisaoka et al., 2010). These regions differ from those that co-express FOXP2 and TBR1, which include cortical layer 6

and amygdala (Ferland et al., 2003; Remedios et al., 2007). Other regions, such as the deep cerebellar nuclei, may express all three proteins (Ferland et al., 2003; Fink et al., 2006). Region-dependent consequences of impaired interaction with TBR1 *in vivo* may be another reason for the distinct phenotypic effects of the same mutation in *FOXP1* versus *FOXP2*.

The p.R514H *FOXP1* variant can also be compared to the previously reported p.R514C variant affecting the same residue (Table 3.3). At the phenotypic level, p.R514C and p.R514H *FOXP1* variants lead to clinical features broadly typical of *FOXP1*-related disorder, although only the patient carrying the p.R514C variant displayed autistic features and behavioural problems (Sollis et al., 2016). Interestingly, Patient 2 in our study exhibited heightened sensitivity to temperature and textures, which may align with the sensory processing disorder previously reported in the proband carrying the p.R514C variant (Sollis et al., 2016). Strabismus and other visual problems were identified in all three p.R514H cases described here and in the prior p.R514C case (Table 3.3) (Sollis et al., 2016). To our knowledge, strabismus has been reported in two other cases of *FOXP1*-related disorder (Bekheirnia et al., 2017), whereas hypermetropia (Pariani et al., 2009) and hyperopic astigmatism (Bekheirnia et al., 2017) have each been reported once. It is important to note that visual problems are not a common feature of *FOXP1*-related disorder (Le Fevre et al., 2013) and in at least two cases, normal vision was explicitly noted (Le Fevre et al., 2013; Song et al., 2015). Further studies may determine whether visual symptoms are a common and under-reported consequence of *FOXP1* variants or restricted to a subset of mutations.

We also note urogenital abnormalities including bilateral cryptorchidism, small penis, and an underdeveloped scrotum in the two male patients carrying the p.R514H variant. A recent study identifying eight novel heterozygous *de novo* *FOXP1* variants found that while all patients had neurodevelopmental phenotypes consistent with *FOXP1*-related disorder, 6/8 also exhibited urogenital defects, including undescended testes and congenital abnormalities of the kidney and urinary tract (CAKUT) (Bekheirnia et al., 2017). The range of mutations reported by Bekheirnia et al. includes frameshift and missense variants within the FOX domain and is consistent with variants previously reported in *FOXP1*-related disorder (Sollis et al., 2016). It is therefore possible that urogenital abnormalities may simply represent a variable or underdiagnosed feature of *FOXP1*-related disorder.

The present study focused on the major isoform of *FOXP1* to characterise the molecular effects of the p.R514H variant. It may be noted that seven additional isoforms have been reported, resulting from alternative splicing (The UniProt Consortium, 2017; see URL¹), however their physiological relevance remains relatively unclear. All but two (isoforms 5 and 8) retain the R514 residue and may therefore be affected by the p.R514H variant. Interestingly, isoform 8, which is specifically expressed in embryonic stem cells and displays distinct DNA-binding properties (Gabut et al., 2011), would not be disrupted by the variant, as it contains an alternative forkhead domain that does not include the mutated residue.

¹ UniProt entry for *FOXP1*: <https://www.uniprot.org/uniprot/Q9H334>

Eight different missense variants have now been identified at homologous sites in various FOX genes (Kawase et al., 2001; Lai et al., 2001; Brice et al., 2002; Baris et al., 2006; Beysen et al., 2008; Sen et al., 2013; Sollis et al., 2016) (Table 3.4; Fig 3.4), and in the case of arginine 514 of FOXP1, this has occurred in at least 4 independent *de novo* events. It is therefore tempting to speculate that this site may be particularly prone to mutation. In each case, the variant is a C>T or G>A transition within a CpG dinucleotide sequence. These sequences are underrepresented in the genome, because 5-methylcytosine (5-mC) undergoes spontaneous deamination leading to a C>T transition (or G>A on the complementary strand). Methylated CpG sites are therefore mutational hotspots (Pfeifer, 2006), and spontaneous deamination may account for the recurrence of mutations at this position. The fact that this CpG sequence has been maintained in the human population is consistent with the view that a change at this site is highly deleterious, as shown in our functional assays and in the patient phenotypes.

Table 3.4. Pathogenic variants in FOX proteins at arginine residues equivalent to FOXP1-R514.

Gene	FOXP1	FOXP1	FOXP2	FOXC1
Study	This study	Sollis, 2016	Lai, 2001	Kawase, 2001
Disorder	Global developmental delay and intellectual disability	Global developmental delay and intellectual disability	Speech and language disorder	Axenfeld-Rieger syndrome
OMIM	613670	613670	602081	602482
Genomic location (h38)	chr3:70,972,666	chr3:70,972,667	chr7:114,662,075	chr6:1,610,825
gDNA	g.C>T	g.G>A	g.G>A	g.G>A
Strand	-	-	+	+
cDNA	NM_032682.5 c.1541G>A	NM_032682.5 c.1540C>T	NM_014491.3 c.1658G>A	NM_001453.2 c.380G>A
Protein	p.R514H	p.R514C	p.R553H	p.R127H

Gene	FOXC2	FOXE1	FOXF1	FOXL2
Study	Brice, 2002	Baris, 2006	Sen, 2013	Beysen, 2008
Disorder	Lymphedema-distichiasis syndrome	Bamforth-Lazarus syndrome	Alveolar capillary dysplasia with misalignment of pulmonary veins	Blepharophimosis, ptosis and epicanthus inversus
OMIM	153400	241850	265380	110100
Genomic location (h38)	chr16:86,567,697	chr9:97,854,218	chr16:86,510,859	chr3:138,946,416
gDNA	g.G>A	g.C>T	g.G>A	g.G>A
Strand	+	+	+	-
cDNA	NM_005251.2 c.362G>A	NM_004473.3 c.304C>T	NM_001451.2 c.290G>A	NM_023067.3 c.307C>T
Protein	p.R121H	p.R102C	p.R97H	p.R103C

The current study is the first, to our knowledge, to compare equivalent variants in *FOXP2* and *FOXP1*. However, similar functional analyses have been performed to compare two homologous variants in *FOXC1* (OMIM 601090; p.R127H) and *FOXC2* (OMIM 602402; p.R121H). Both of these variants disrupted the normal nuclear localisation of the protein and abolished DNA-binding (Saleem et al., 2003; Berry et al., 2005). Furthermore, while both *FOXC1* and *FOXC2* could act as transcriptional activators, the variants abolished or significantly reduced transcriptional activation (Saleem et al., 2003; Berry et al., 2005). The identification of very similar molecular effects for the equivalent variants in *FOXP1/2* provides further evidence for a conserved role for this residue across the FOX transcription factor family.

In summary, we have identified a novel *de novo* missense variant in *FOXP1* that is identical to the most well-studied aetiological variant in *FOXP2*. Functional characterisation revealed clear similarities between these equivalent mutations in terms of their impact on protein function. On the other hand, the phenotypic profiles of the two mutations are highly distinct, supporting divergent roles for *FOXP2* and *FOXP1* in neurodevelopment.

4 FOXP PROTEINS INTERACT WITH THE CHROMATIN REMODELLING FACTOR GATAD2B

ABSTRACT

Mutations in the FOXP transcription factors lead to language-related and other neurodevelopmental disorders, probably through the dysregulation of target gene expression. The precise mechanisms of FOXP-mediated transcriptional regulation are not well understood but are likely to include the binding and recruitment of chromatin modifying factors to FOXP target loci. Indeed, a previous yeast two-hybrid screen for FOXP2-interactors identified several chromatin-remodelling factors, including three (GATAD2B, CHD3 and KANSL1) that have been implicated in neurodevelopmental disorders involving speech and language deficits. Here, the three candidates were characterised *in vitro* and validation was attempted for putative interactions with FOXP1/2/4. Bioluminescence resonance energy transfer (BRET) assays confirmed interactions between FOXP1/2/4 and GATAD2B. The GATAD2B-binding site on FOXP2 was mapped to residues 260-422 and included the leucine zipper domain, which must be intact for the interaction to occur, while the CR2 domain of GATAD2B was found to be essential for FOXP-binding. These results corroborate and extend the findings of previous investigations in the orthologous mouse proteins. Further experiments showed that pathogenic variants in *FOXP1*, *FOXP2* and *GATAD2B* can disrupt these interactions, potentially explaining a number of phenotypic overlaps between the disorders related to each gene. No interactions were observed with either CHD3 or KANSL1, however both genes remain important candidates for speech and language disorders in their own right.

4.1 INTRODUCTION

The FOXP transcription factors (TFs) are highly expressed in the brain and play an important role in neural development. Mutations in *FOXP2* (OMIM 605317) cause a speech and language disorder characterised by childhood apraxia of speech (CAS) and multiple linguistic deficits (Lai et al., 2001) (OMIM 602081), while *FOXP1* (OMIM 605515) mutations lead to a syndrome characterised by moderate to severe intellectual disability (ID) and speech impairment, sometimes accompanied by autistic features (Le Fevre et al., 2013; Sollis et al., 2016) (OMIM 613670). In order to regulate transcription, the FOXP1/2 proteins, as well as a third paralogue, FOXP4, bind to specific DNA loci via a forkhead box (FOX) domain. While earlier work defined these TFs as transcriptional repressors (Wang et al., 2003; Vernes et al., 2006), more recent studies in cell lines and mouse brain tissue suggest that FOXP1/2 are able to both up- and down-regulate the expression of target genes (Spiteri et al., 2007; Vernes et al., 2007; Devanna et al., 2014; Araujo et al., 2015).

It is likely that the regulatory activity of FOXP TFs relies on interactions with other proteins, including chromatin remodelling factors. These factors adjust the interactions between DNA and DNA-packaging proteins in the nucleus, thereby increasing or decreasing gene expression by controlling access of the transcriptional machinery to specific loci. Interestingly, chromatin remodelling factors have essential roles throughout neural development, and mutations in at least 28 chromatin remodelling-related genes have been implicated in neurological disorders such as ID and autism spectrum disorders (ASD), which are often associated with severe speech deficits (Ronan et al., 2013). Interactions between FOXP proteins and chromatin remodelling factors may therefore provide a mechanism for FOXP-mediated repression and/or activation and may help to explain their involvement in neurodevelopmental disorders with overlapping phenotypes.

A yeast two-hybrid (Y2H) screen was previously performed in a human foetal brain library using the full-length FOXP2 protein as bait, in order to identify candidate interaction partners. The results of this screen were previously published in an abridged form (Estruch et al., 2016b), and are presented here in full (Table 4.1). Among the putative FOXP2-interaction partners were multiple chromatin remodelling proteins, including several that have been implicated in neurodevelopmental disorders.

These candidates included two proteins – CHD3 (OMIM 602120) and GATAD2B (OMIM 614998) – that are members of the nucleosome remodelling and histone deacetylation (NuRD) complex, a large macromolecular complex with at least two enzymatic subunits (Lai and Wade, 2011) (Fig 4.1). CHD3 functions as an ATP-dependent helicase within the NuRD complex and is one of three alternative subunits (with CHD4 and CHD5) responsible for the nucleosome remodelling function of the complex. GATAD2B (also known as p66 β) and its homologue GATAD2A (p66 α) are structural subunits that directly associate with histone tails and interact with the MDB2/3 subunit to assist targeting of the NuRD complex to specific loci (Brackertz et al., 2006). Histone deacetylation activity has led to the traditional classification of the NuRD complex as a transcriptional corepressor complex, although recent studies have also found the NuRD complex at active enhancers

and promoters, suggesting it may have more nuanced effects on gene expression (Basta and Rauchman, 2015).

Interestingly, *CHD3* variants have recently been reported in multiple patients with ID and/or language difficulties including CAS (Eising et al., 2018; Snijders Blok et al., 2018). Other members of the CHD (chromodomain) family of proteins have been implicated in ID-related syndromes such as epileptic encephalopathy (*CHD2*) (Carvill et al., 2013), Sifrim-Hitz-Weiss syndrome (*CHD4*) (Sifrim et al., 2016) and CHARGE syndrome (*CHD7*) (Visser et al., 2004); and *CHD2*, *CHD7* and *CHD8* have been recognised as candidate genes for ASD (Neale et al., 2012; O’Roak et al., 2012b, 2012a; Bernier et al., 2014). Heterozygous loss-of-function mutations in *GATAD2B* cause a distinctive syndrome involving ID, severe speech delay and neonatal hypotonia (OMIM 615074) (de Ligt et al., 2012; Willemsen et al., 2013; Luo et al., 2017), echoing major features of both *FOXP1*- and *FOXP2*-related disorders (Lai et al., 2001; Sollis et al., 2016, 2017; Morgan et al., 2017). Notably, *Gatad2b* was detected in an independent Y2H screen for Foxp-interacting proteins in adult mouse lung, and the authors subsequently confirmed its interaction with mouse *Foxp1/2* by co-immunoprecipitation (Chokas et al., 2010). In the present study, I demonstrate that the orthologous human proteins bind in a similar manner and examine for the first time whether the interactions are affected by pathogenic *GATAD2B* and *FOXP* variants. The results presented here do not provide evidence of a direct *FOXP-CHD3* interaction but suggest that an indirect association could be facilitated via *GATAD2B*.

In addition to members of the NuRD complex, the Y2H screen of human foetal brain also identified *KANSL1* (OMIM 612452) as a potential interactor of *FOXP2* (Table 4.1). *KANSL1* is a component of the NSL complex, which has histone acetyltransferase (HAT) activity, and is therefore associated with transcriptionally active genes (Feller et al., 2012). Mutations in *KANSL1* cause Koolen-de Vries Syndrome (OMIM 610443), which involves ID and speech/language impairment (Koolen et al., 2012; Zollino et al., 2012; Morgan et al., 2018). No further evidence was found for an interaction between *FOXP*s and *KANSL1*, but the significant phenotypic connections warrant further attention.

Table 4.1. Putative FOXP2-interactors in human foetal brain Y2H screen.

Gene	Preys	Disorder (OMIM #)	Complex	Function
PIAS1	7	-	-	-
FXD6	4	-	-	-
CHD3	4	Snijders Blok-Campeau syndrome (618205)	NuRD	Chromatin remodelling
NREP	3	-	-	-
FKBP1A	2	-	-	-
NRGN	2	-	-	-
BRD7	1	-	SWI/SNF BRM-BRG1	Histone modification
C20orf96	1	-	-	-
CLSTN3	1	-	-	-
CTNNB1	1	Mental retardation, autosomal dominant 19 (615075); Exudative vitreoretinopathy 7 (617572)	-	-
DDB2	1	Xeroderma pigmentosum group E (278740)	N/A	Histone ubiquitination
EFHD1	1	-	-	-
GRM5	1	-	-	-
IL3RA	1	-	-	-
LENG8	1	-	-	-
MT2A	1	-	-	-
FAM89B	1	-	-	-
NMT1	1	-	-	-
SIGMAR1	1	Amyotrophic lateral sclerosis 16, juvenile (614373); Spinal muscular atrophy, distal, autosomal recessive 2 (605726)	-	-
GATAD2B	1	Mental retardation, autosomal dominant 18 (615074)	NuRD	Histone modification read
PHC1	1	Microcephaly 11, primary, autosomal recessive (615414)	PRC1	Polycomb group (PcG) protein
PRKAB1	1	-	N/A	Histone phosphorylation
PTPRK	1	-	-	-
RBFOX2	1	-	-	-
SNAPC3	1	-	-	-
TIAF1	1	-	-	-
TUBB	1	Cortical dysplasia, complex, with other brain malformations 6 (615771); Symmetric circumferential skin creases, congenital 1 (156610)	-	-
ZFH4	1	Ptosis, congenital (178300)	-	-
KANSL1	1	Koolen-De Vries syndrome (610443)	NSL	Histone acetylation
ARL17A	1	-	-	-

Gene	Preys	Disorder (OMIM #)	Complex	Function
CDC42BPA	1	-	-	-
KIAA1211	1	-	-	-
BRPF3	1	-	MOZ/MORF	Histone acetylation
C5orf42	1	Joubert syndrome 17 (614615); Orofaciodigital syndrome VI (277170)	-	-

Known Mendelian disorders are listed with OMIM identifiers. Protein complex information and chromatin-modifying functions are taken from the EpiFactors Database (Medvedeva et al., 2015)

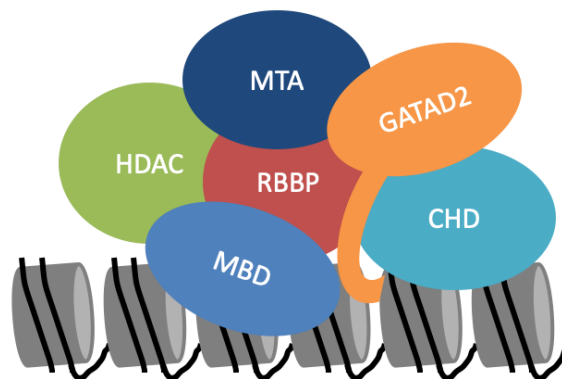


Figure 4.1. Schematic of the NuRD complex, including the putative FOXP-interacting proteins CHD3 and GATAD2B. The CHD3/4/5 subunit is responsible for the nucleosome remodelling activity of the complex, while HDAC1/2 catalyses histone deacetylation. GATAD2A/B interacts with histone tails and with the MBD2/3 subunit, which is responsible for methylated DNA-binding. Other subunits include MTA1/2/3 (involved in transcription factor-binding) and RBBP4/7 (involved in histone-binding).

4.2 METHODS

4.2.1 DNA CONSTRUCTS

WT (wild-type) *FOXP1* (isoform a; NM_032682.5; NP_116071.2), *FOXP2* (isoform I; NM_014491.3; NP_055306.1) and *FOXP4* (isoform 1; NM_001012426.1; NP_001012426.1), as well as the *FOXP2* p.R328* variant and synthetic FOXP2 truncations, were amplified by PCR and cloned into pCR2.1-TOPO (Invitrogen) as previously described (Deriziotis et al., 2014b; Estruch et al., 2016a; Sollis et al., 2016). The *FOXP1* A339Sfs*4 variant was synthesised by GenScript (O’Roak et al., 2011). The *FOXP2* p.R553H variant and the remaining *FOXP1* variants were generated using the QuikChange II Site-Directed Mutagenesis Kit (Agilent), as previously described (Lozano et al., 2015; Estruch et al., 2016a; Sollis et al., 2016, 2017). WT *GATAD2B* (NM_020699.3; NP_065750.1), *CHD3* (isoform 1; NM_001005273.2; NP_001005273.1) and *KANSL1* (isoform 1; NM_001193466.1; NP_001180395.1), as well as the *GATAD2B* p.Q470* variant, were amplified using the primers listed in Table 4.2 and cloned into pCR2.1-TOPO. The *GATAD2B* p.Q190Afs*34 and p.N195Kfs*30 variant constructs were generated by site-directed mutagenesis using the primers listed in Table 4.3. All cDNAs were then subcloned using BamHI/XbaI (FOXP2, *GATAD2B*, *CHD3*) or HindIII/XbaI (*KANSL1*) restriction sites into pLuc, pYFP and a modified pmCherry-C1 expression vector (Clontech). All constructs were verified by Sanger sequencing.

Table 4.2. Primers used to generate DNA constructs.

Construct	Forward primer (5’-3’)	Site	Reverse primer (5’-3’)	Site
GATAD2B	GGATCCTGGATAGAA TGACAGAAGATGC	BamHI	TCTAGATTATTTCTGT CCACTGATGG	XbaI
GATAD2B p.Q470*	GGATCCTGGATAGAA TGACAGAAGATGC	BamHI	TCTAGATTATAGGGCT TTCACAAATGCATTTT TC	XbaI
CHD3	GGATCCGGAAGGCG GCAGACACTGTGATC C	BamHI	GCTAGCTTAGTCGTCT ATACAGATCACCTCC	NheI
KANSL1	AAGCTTCGCTGCGAT GGCGCCCGCTCT	HindIII	TCTAGATTTATCTGTG AGTCGGGCGCT	XbaI

Table 4.3. Primers used for site-directed mutagenesis of GATAD2B variants.

VARIANT	Sense primer (5’-3’)	Antisense primer (5’-3’)
GATAD2B p. Q190Afs*34	CTGTTAAAGAACTGAGAC AGTCAGCTACAGAAAGAGA ATGT	ACATTCTCTTTCTGTAGCTGACT GTCTCAGTTTCTTTAACAG
GATAD2B p. N195Kfs*30	GTCAGCTACAGAAAGAGAA ATGTGGTCCAGAAGACT	AGTCTTCTGGACCACATTTCTCT TTCTGTAGCTGAC

4.2.2 CELL CULTURE AND TRANSFECTION

HEK293 cells were cultured in DMEM supplemented with 10% foetal bovine serum. Transfections were performed using GeneJuice, according to the manufacturers' instructions (Merck-Millipore).

4.2.3 WESTERN BLOTTING

Cells were transfected with equimolar amounts of GATAD2B or KANSL1 expression plasmids and cultured for 48 h. Whole-cell lysates were extracted by treatment with lysis buffer (100 mM Tris pH 7.5, 150mMNaCl, 10mMEDTA, 0.2% Triton X-100, 1% PMSF, protease inhibitor cocktail; all from Sigma–Aldrich) for 10 min at 4°C, before centrifuging at 10,000×g for 30 min at 4°C to remove cell debris. Proteins were resolved on 4–15% Tris–Glycine gels and transferred onto polyvinylidene fluoride membranes (both Bio-Rad). Blots were probed with mouse anti-mCherry (for pmCherry constructs; 1:2,000; Clontech) overnight at 4°C, followed by incubation with HRP-conjugated goat anti-mouse IgG for 45 min at room temperature (1:2,000; Bio-Rad). Proteins were visualised using Novex ECL Chemiluminescent Substrate Reagent Kit (Invitrogen) and the ChemiDoc XRS+ System (Bio-Rad).

4.2.4 FLUORESCENCE MICROSCOPY

Cells were seeded onto coverslips coated with poly-L-lysine (Sigma–Aldrich) and were fixed 24 h post-transfection using 4% paraformaldehyde (Electron Microscopy Sciences) for 10 min at room temperature. YFP and mCherry fusion proteins were visualised by direct fluorescence. Nuclei were visualised with Hoechst 33342 (Invitrogen). Fluorescence images were obtained using a Zeiss Axio Imager M2 upright microscope with ZEN imaging software (Zeiss).

4.2.5 BRET ASSAY

BRET assays were performed as previously described (Deriziotis et al., 2014a, 2014b). In summary, cells were transfected with pairs of *Renilla* luciferase and YFP-fusion proteins in 96-well plates. *Renilla* luciferase and YFP fused to a C-terminal nuclear localisation signal were used as control proteins. EnduRen luciferase substrate (Promega) was added to cells 36–48 h after transfection at a final concentration of 60 μM and incubated for 4 h. Emission measurements were taken with a TECAN F200PRO microplate reader using the Blue1 and Green1 filters and corrected BRET ratios were calculated as follows: $[\text{Green1}_{(\text{experimental condition})}/\text{Blue1}_{(\text{experimental condition})}] - [\text{Green1}_{(\text{control condition})}/\text{Blue1}_{(\text{control condition})}]$. YFP fluorescence was then measured separately, with excitation at 505 nm and emission at 545, to quantify expression of the YFP-fusion proteins. *Renilla* luciferase and YFP fused to a C-terminal nuclear localisation signal were used as control proteins.

4.3 RESULTS

4.3.1 GATAD2B INTERACTS WITH AND INFLUENCES LOCALISATION OF FOXP2/1/4

Chokas et al. (2010) identified *Gatad2b* (p66 β) as a potential *Foxp*-interacting protein in a Y2H screen that used a fragment of mouse *Foxp2* as bait against a murine adult lung cDNA library. Follow-up co-immunoprecipitation experiments confirmed that *Gatad2b* could interact with full-length *Foxp1* and *Foxp2*, but the article did not report direct evidence of an interaction with full-length *Foxp4*. Although the human FOXP1/2 and GATAD2B proteins are highly similar to their mouse orthologues (FOXP1 = 92.9% identity, FOXP2 = 99.4% identity, GATAD2B = 98.3% identity), it was possible that the human proteins might exhibit different interaction properties. Another Y2H screen using full-length human FOXP2 in human foetal brain independently identified GATAD2B as a putative interactor (Table 4.1) (Estruch et al., 2016b), but did not attempt to validate the findings with another method. I therefore decided to try confirming the putative interactions between human GATAD2B and human FOXP1/2/4 proteins.

Bioluminescence resonance energy transfer (BRET) assays were carried out in live HEK293 cells. These experiments clearly confirmed that GATAD2B is able to interact with FOXP1/2 and provided the first direct evidence of an interaction with FOXP4 (Fig 4.2A). Fluorescence imaging of mCherry-tagged GATAD2B in these cells revealed a speckled nuclear pattern of expression (Fig 4.2B), which is in line with previous reports (Brackertz et al., 2006). FOXP1/2 exhibited diffuse nuclear localisation when transfected alone, while FOXP4 localisation was both nuclear and cytoplasmic (Fig 4.2C). Interestingly, all three FOXP proteins formed nuclear speckles when co-transfected with GATAD2B (Fig 4.2D). These findings are consistent with the interaction shown by BRET assays, and suggests that interaction with GATAD2B alters the localisation of FOXP proteins within the nucleus.

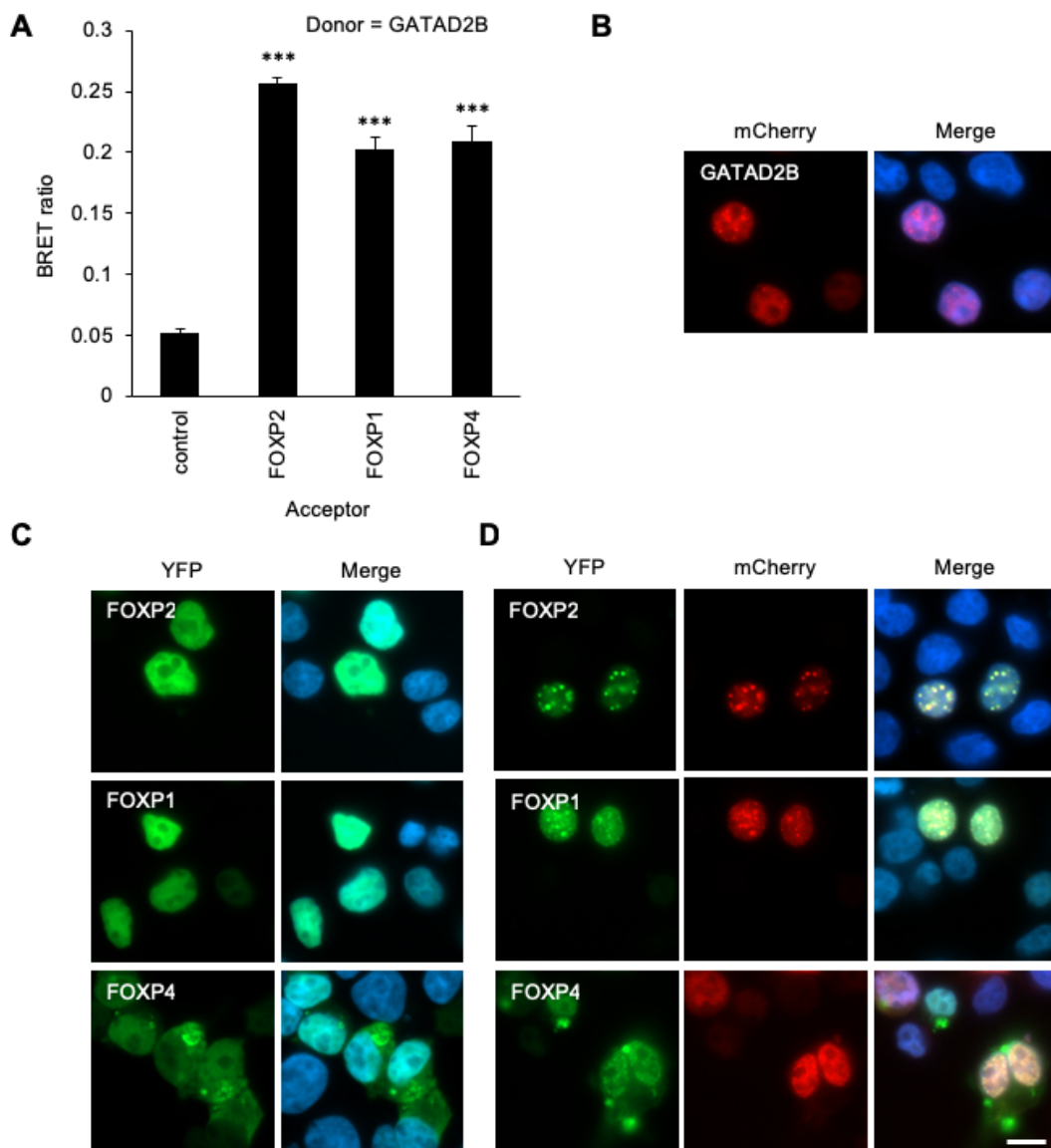


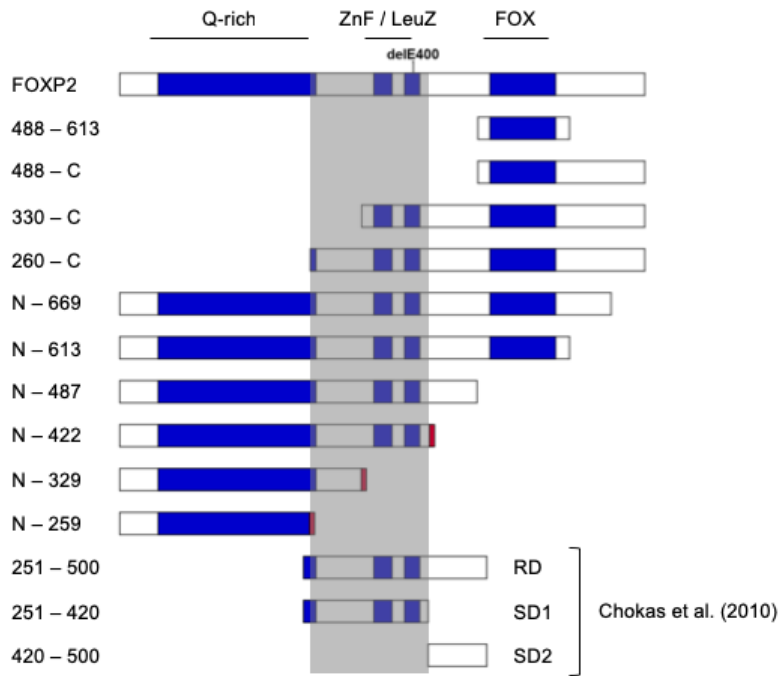
Figure 4.2. GATAD2B interacts with FOXP1/2/4. (A) BRET assays for interaction between GATAD2B and FOXP2/1/4. Bars represent the corrected mean BRET ratio \pm S.D. of one experiment performed in triplicate ($***P < 0.001$ compared to control, Student's T-test). (B-D) Fluorescence imaging of cells transfected with mCherry-tagged GATAD2B (red), YFP-tagged FOXP1/2/4 (green), or both. Nuclei are stained with Hoechst 33342 (blue). Scale bar, 10 μ m.

4.3.2 THE FOXP2 LEUCINE ZIPPER IS ESSENTIAL FOR GATAD2B-BINDING

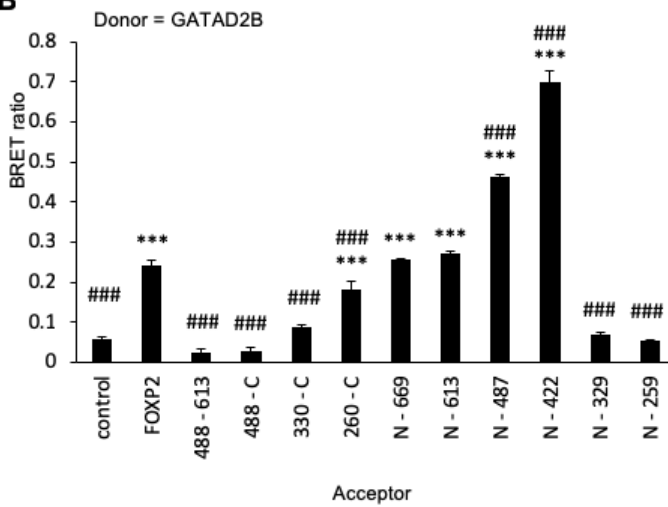
Next, I sought to identify the GATAD2B-binding region in FOXP2, using a series of synthetic FOXP2 truncations (Fig 4.3A). A previous study mapped the Gatad2b-binding site in mouse *Foxp1/2/4* to a region equivalent to hFOXP2 residues 251-500 (Chokas et al., 2010). More precise mapping experiments that were performed only in *Foxp1* showed that two smaller segments were able to independently interact with Gatad2b: SD1 (equivalent to hFOXP2 residues 251-420) and SD2 (equivalent to hFOXP2 residues 420-500). In the BRET assays, GATAD2B was unable to interact with FOXP2 fragments that lacked residues 260-422 (Fig 4.3B). This loss of interaction cannot be explained by mislocalisation of the truncated proteins, as each of the truncations has been previously shown to localise to the nucleus, either alone, or due to an NLS attached to the C-terminus (Estruch et al., 2016a). Therefore, this study has indicated that a segment very similar to the SD1 region described by Chokas et al. (2010) is sufficient for GATAD2B-binding in human FOXP2 (Fig 4.3A). On the other hand, the SD2 region does not appear to be required for the GATAD2B-FOXP2 interaction, although the existence of an additional binding site within that region cannot be ruled out. Interestingly, neither the N329* fragment nor the complementary C-terminal truncation (residue 330 - C) was able to interact, suggesting that sites on both sides of this breakpoint may be involved in GATAD2B-binding.

Notably, the identified GATAD2B-binding region in FOXP2 identified here encompasses the leucine zipper and zinc-finger domains (Fig 4.3A). The leucine zipper in particular is essential for FOXP-dimerisation (Li et al., 2004). In fact, an in-frame deletion of a single FOXP2 residue (E400) within the leucine zipper has been shown to abolish homodimerisation, despite retaining normal nuclear localisation (Deriziotis et al., 2014a). Deletion of this residue was also sufficient to prevent GATAD2B-binding (Fig 4.3C), suggesting either that GATAD2B also binds to the leucine zipper, or that GATAD2B-interaction is dependent on FOXP2-dimerisation. The high degree of sequence homology between the FOXP2 leucine zipper and that of FOXP1 (82% identity) and FOXP4 (86% identity) suggests that the same motif is likely to be important for GATAD2B-FOXP1/4 interactions.

A



B



C

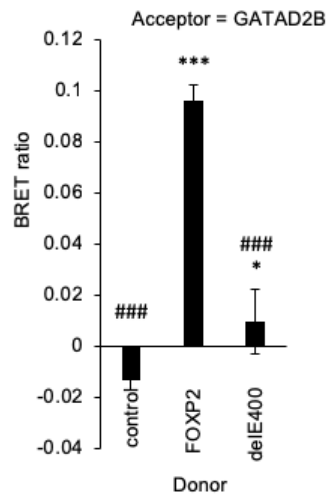


Figure 4.3 (opposite page). The FOXP leucine zipper is required for GATAD2B-interaction. (A) Schematics representing FOXP2 synthetic truncations. Regions equivalent to the mouse Foxp1 fragments used by Chokas et al. (2010) are also shown: RD = repression domain, SD1/2 = subdomain 1/2. The location of the E400 deletion is indicated on the wild-type protein. Domains are shown in blue (Q-rich = glutamine-rich region, ZnF = zinc finger, LeuZ = leucine zipper, FOX = forkhead box). Note that three truncations contain a synthetic nuclear localisation signal at the C-terminal end (red boxes). Shaded area indicates the region necessary for GATAD2B-binding (residue 260-422). (B) BRET assays for interaction between GATAD2B and synthetic FOXP2 truncations. Bars represent the corrected mean BRET ratio \pm S.D. of one experiment performed in triplicate ($***P < 0.001$ compared to control, $###P < 0.001$ compared to WT FOXP2, one-way ANOVA and *post-hoc* Tukey's test). (C) BRET assay for interaction between GATAD2B and FOXP2 p.delE400 variant. Bars represent the corrected mean BRET ratio \pm S.D. of one experiment performed in triplicate ($*P < 0.05$ and $***P < 0.001$ compared to control, $###P < 0.001$ compared to WT FOXP2, one-way ANOVA and *post-hoc* Tukey's test).

4.3.3 PATHOGENIC FOXP1/2 VARIANTS DISRUPT INTERACTION WITH GATAD2B

Mutations in *FOXP2* cause a severe speech and language disorder characterised by CAS and expressive and receptive language deficits (Lai et al., 2001; Morgan et al., 2017), while *FOXP1* variants have been identified in patients with ID and speech impairment, sometimes accompanied by autistic features (Le Fevre et al., 2013; Sollis et al., 2016, 2017). Given that disrupted protein-protein interactions are among the molecular consequences of *FOXP* mutations (Table 4.4) (Deriziotis et al., 2014b; Estruch et al., 2016a; Sollis et al., 2016, 2017), I decided to investigate the effects of these mutations on interactions with GATAD2B.

Table 4.4. Effects of FOXP variants on protein-protein interactions.

Gene	Variant	Loc.	Repr.	Interactions					
				FOXP1	FOXP2	FOXP4	GATAD2B	CTBP1/2	TBR1
FOXP2	p.R553H	N+C*	-	Green	Green	Green	Green	Green	Red
	p.R328*	N+C*	-	Red	Red	Red	Red	Orange	Red
FOXP1	p.R465G	N+C*	-	Green	Green	Grey	Green	Grey	Grey
	p.R514C	N+C*	-	Green	Green	Grey	Green	Grey	Grey
	p.R514H	N+C*	-	Green	Green	Grey	Green	Green	Red
	p.W534R	C*	-	Red	Red	Grey	Red	Grey	Grey
	p.A339Sfs*4	N+C	-	Red	Red	Grey	Red	Grey	Grey
	p.V423Hfs*37	C*	-	Green	Green	Grey	Red	Grey	Grey
	p.Y439*	C*	-	Green	Green	Grey	Red	Grey	Grey
p.R525*	C*	-	Red	Red	Grey	Red	Grey	Grey	

Localisation: N = nuclear, C = cytoplasmic, N+C = mixed localisation, *forms aggregates. Repression: minus sign = loss of repressive function. Interactions: green = interaction, red = no interaction, orange = decreased interaction, grey = no data.

Two pathogenic FOXP2 variants identified in individuals with CAS (Fig 4.4A) were expressed as YFP-fusion proteins and assessed for their effects on the FOXP2-GATAD2B interaction using BRET and co-localisation experiments. Both variant proteins – p.R553H (Lai et al., 2001) and p.R328* (MacDermot et al., 2005) – exhibited aberrant cytoplasmic localisation and aggregation (Fig 4.4B), as has been previously shown (Vernes et al., 2006; Estruch et al., 2016a; Sollis et al., 2017). The p.R553H variant protein maintained the ability to interact with GATAD2B, and the two proteins co-localised in a punctate fashion in the nucleus (Fig 4.4C, D). As GATAD2B and FOXP2 p.R553H also exhibit speckled or aggregated localisation when expressed independently, the functional significance of the observed co-localisation is not entirely clear. Nonetheless, it is possible that abnormal aggregation of the p.R553H FOXP2 variant may in turn disrupt the localisation of bound GATAD2B and adversely affect its function. The p.R328* FOXP2 variant, which lacks the

leucine zipper, abolished the interaction and did not form nuclear speckles, consistent with a loss of function mechanism for this variant (Fig 4.4C, D).

Next, eight pathogenic *FOXP1* variants (4 missense, 2 frameshift, 2 nonsense) found in patients with ID and language impairment (Sollis et al., 2016, 2017) were tested (Fig 4.5A). Three of the four *FOXP1* variants resulting from missense mutations maintained the *GATAD2B* interaction, and at least partially co-localised with *GATAD2B* in a punctate nuclear pattern (Fig 4.5B, C). As these four variants have been previously shown to form abnormal aggregates when transfected alone (Sollis et al., 2016, 2017), the observed co-localisation may reflect abnormal translocation of bound *GATAD2B* that could impact on its function. However, it is not possible to clearly distinguish this from the normal speckled expression pattern of *GATAD2B*. Conversely, the p.W534R variant, which is exclusively cytoplasmic, showed no interaction or co-localisation with *GATAD2B*. All of the *FOXP1* frameshift and nonsense variants abolished the interaction with *GATAD2B* (Fig 4.5B). The p.V423Hsf*37, p.Y439* and p.R525* variants are exclusively cytoplasmic, and therefore cannot make contact with *GATAD2B* (Fig 4.5C). The p.A339Sfs*4 variant, on the other hand, remains partially nuclear, suggesting that the lack of interaction in this case is caused by the loss of a crucial binding region (Fig 4.5C). In fact, residue A339 is positioned between the zinc finger and leucine zipper, pointing to the leucine zipper as the main site of *FOXP*-*GATAD2B* interaction, as previously observed for *FOXP2*. The lack of interaction/co-localisation between the *FOXP1* p.W534R and truncating variants and *GATAD2B* supports a loss of function mechanism for these variants.

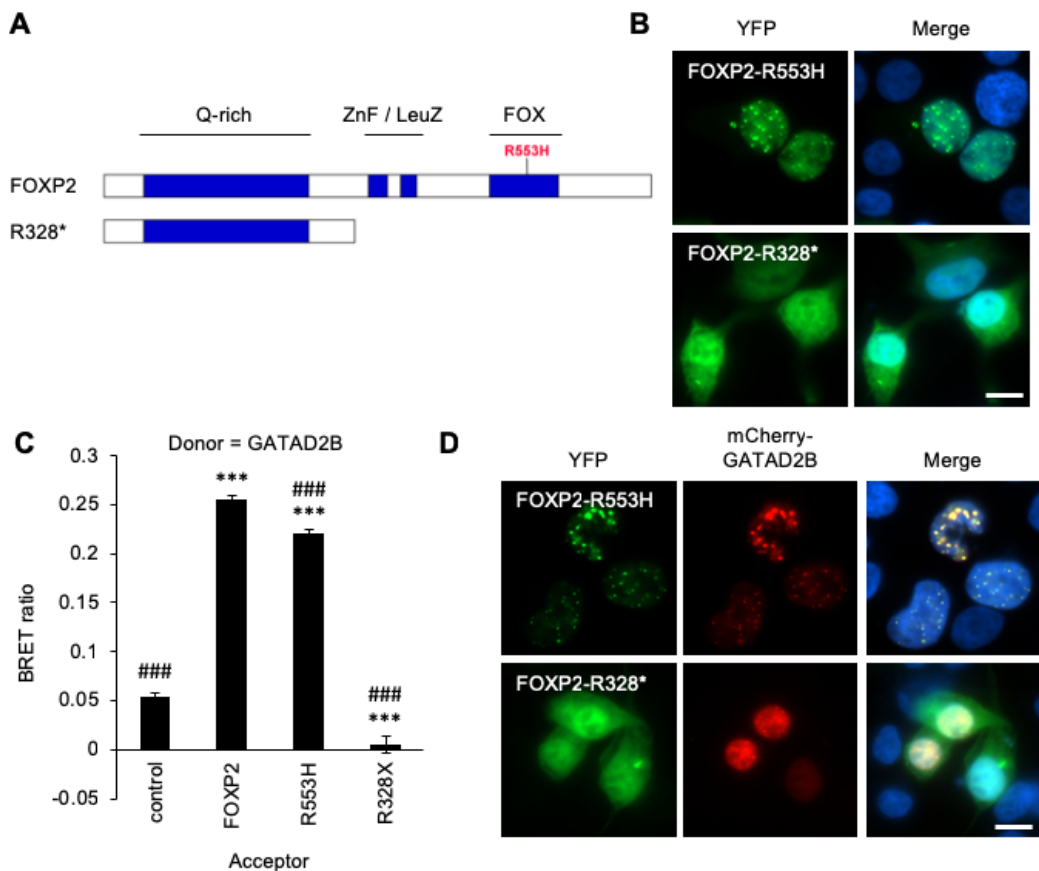


Figure 4.4. Effect of pathogenic FOXP2 variants on interaction with GATAD2B. (A) Schematic representing pathogenic FOXP2 variants found in people affected with a severe speech and language disorder. Domains are shown in blue (Q-rich = glutamine-rich region, ZnF = zinc finger, LeuZ = leucine zipper, FOX = forkhead box). The p.R553H missense variant is indicated on the wild-type protein. (B) Fluorescence imaging of cells transfected with YFP-tagged FOXP2 patient variants. Nuclei are stained with Hoechst 33342 (blue). Scale bar, 10 μ m. (C) BRET assays for interaction between GATAD2B and FOXP2 patient variants. Bars represent the corrected mean BRET ratio \pm S.D. of one experiment performed in triplicate (***) $P < 0.001$ compared to control, ### $P < 0.001$ compared to WT FOXP2, one-way ANOVA and *post-hoc* Tukey's test). (D) Fluorescence imaging of cells co-transfected with YFP-tagged FOXP2 variants and mCherry-tagged GATAD2B. Nuclei are stained with Hoechst 33342 (blue). Scale bar, 10 μ m.

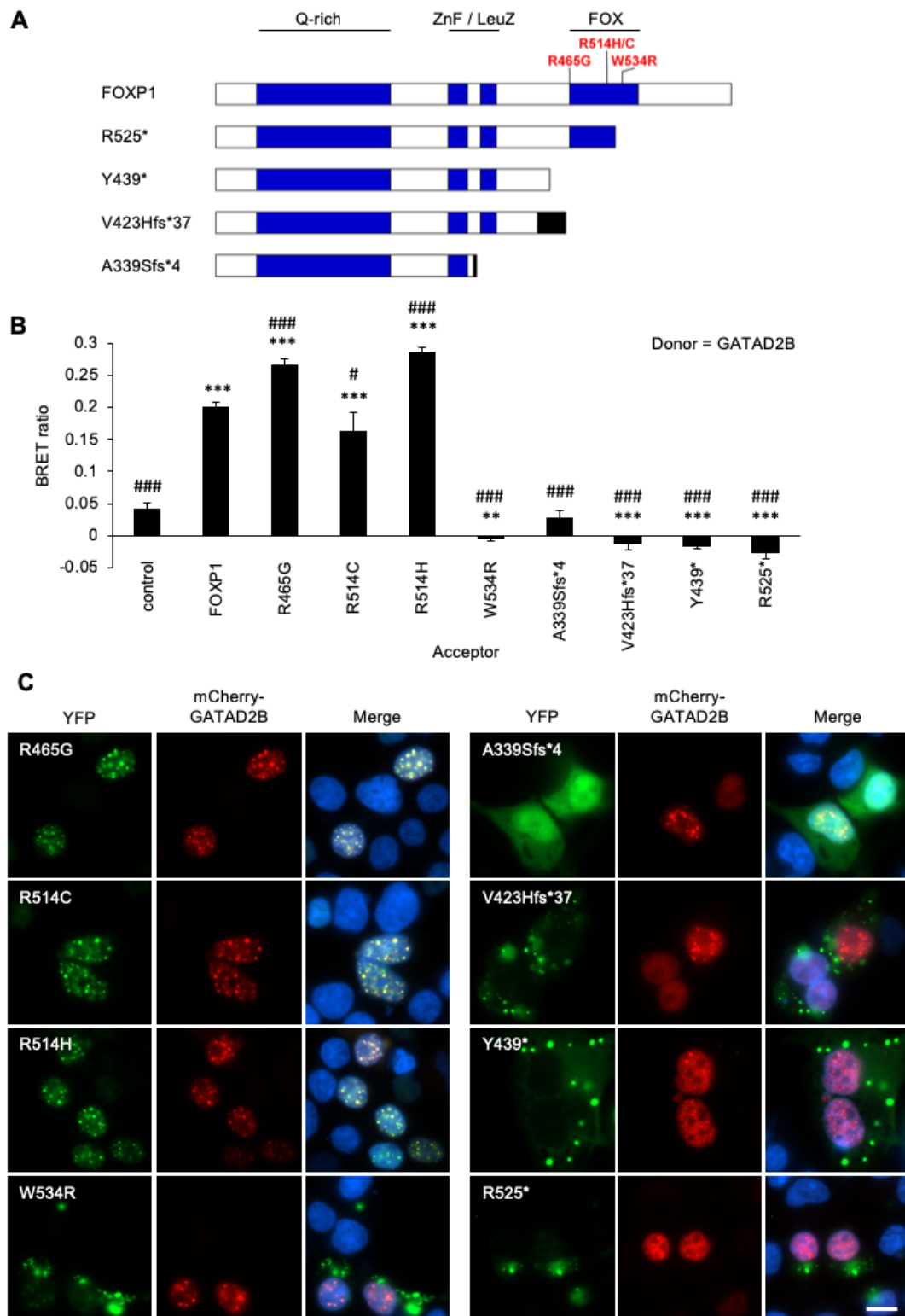


Figure 4.5 (previous page). Effect of pathogenic FOXP1 variants on interaction with GATAD2B. (A) Schematic representing pathogenic FOXP1 mutations found in patients with ID and language impairment. Domains are shown in blue (Q-rich = glutamine-rich region, ZnF = zinc finger, LeuZ = leucine zipper, FOX = forkhead box), and abnormal addition sequence caused by frameshift variants is shown in black. Missense variants are indicated on the wild-type protein. (B) BRET assays for interaction between GATAD2B and FOXP1 patient variants. Bars represent the corrected mean BRET ratio \pm S.D. of one experiment performed in triplicate (** $P < 0.01$ and *** $P < 0.001$ compared to control, # $P < 0.05$ and ### $P < 0.001$ compared to WT FOXP1, one-way ANOVA and *post-hoc* Tukey's test). (C) Fluorescence imaging of cells co-transfected with YFP-tagged FOXP1 missense (left panel; green) or truncating variants (right panel; green) and mCherry-tagged GATAD2B (red). Nuclei are stained with Hoechst 33342 (blue). Scale bar, 10 μ m.

4.3.4 EFFECTS OF PATHOGENIC GATAD2B VARIANTS ON INTERACTION WITH FOXP2/1/4

GATAD2B contains two conserved regions with well-studied functions (Fig 4.6A). The CR1 domain (residues 165-195) is required for incorporation of GATAD2B into the NuRD complex through protein-protein interactions with other subunits, including MDB2/3 (Brackertz et al., 2006). The CR2 domain (residues 340-480) is involved in the localisation of GATAD2B and bound proteins to specific nuclear loci and mediates direct interactions between GATAD2B and histone tails (Brackertz et al., 2006).

Chokas et al. (Chokas et al., 2010) performed co-immunoprecipitation assays to map the Foxp-binding region in mouse *Gatad2b*. It should be noted that these studies were performed using only the “repression domains” of mFoxp1 (residues 251-489) and mFoxp4 (residues 228-469), so it is uncertain whether the equivalent region of Foxp2 and/or full-length Foxp1/2/4 would exhibit the same binding properties. They found that the Foxp1/4 repression domains interacted with an N-terminal deletion beginning after the CR1 domain (equivalent to hGATAD2B residues 190-593) but not with the complementary C-terminal deletion (equivalent to hGATAD2B residues 1-190), suggesting that CR2, rather than CR1, is crucial for Foxp-interaction (Fig 4.6A). Interestingly, a fragment equivalent to hGATAD2B residues 1-415 also maintained the interaction, indicating that the ZnF motif within CR2 is not required for the interaction (Fig 4.6A).

For the current study, expression plasmids were generated for three pathogenic GATAD2B variants: two frameshifts (p.Q190Afs*34 and p.N195Kfs*30), which truncate the protein at the border of the CR1 domain, and a nonsense mutation (p.Q470*), which terminates within the CR2 domain (de Ligt et al., 2012; Willemsen et al., 2013) (Fig 4.6A). The patient mutations were expressed as fusions with mCherry and produced proteins at the expected molecular weights (Fig 4.6B). All three variants resulted in diffuse nuclear expression (Fig 4.6C), consistent with previous findings that the speckled expression pattern of GATAD2B requires the CR2 domain, which is absent in the two frameshift variants and shortened in Q470* (Brackertz et al., 2006). All three variants abolished the interaction with FOXP1/2/4, despite the clear co-localisation of variant GATAD2B proteins with each FOXP protein in the nucleus (Fig 4.7). These findings are consistent with a loss of function mechanism of disease for all three GATAD2B variants. The results also indicate that the CR2 domain is required for FOXP-interaction, consistent with previous experiments (Chokas et al., 2010). However, the lack of interaction by p.Q470* is unexpected, given that a shorter truncation of mouse *Gatad2b* (equivalent to hGATAD2B residues 1-415) maintained an interaction with the mouse Foxp1/4 repression domains. Possible reasons for this discrepancy are outlined in the Discussion.

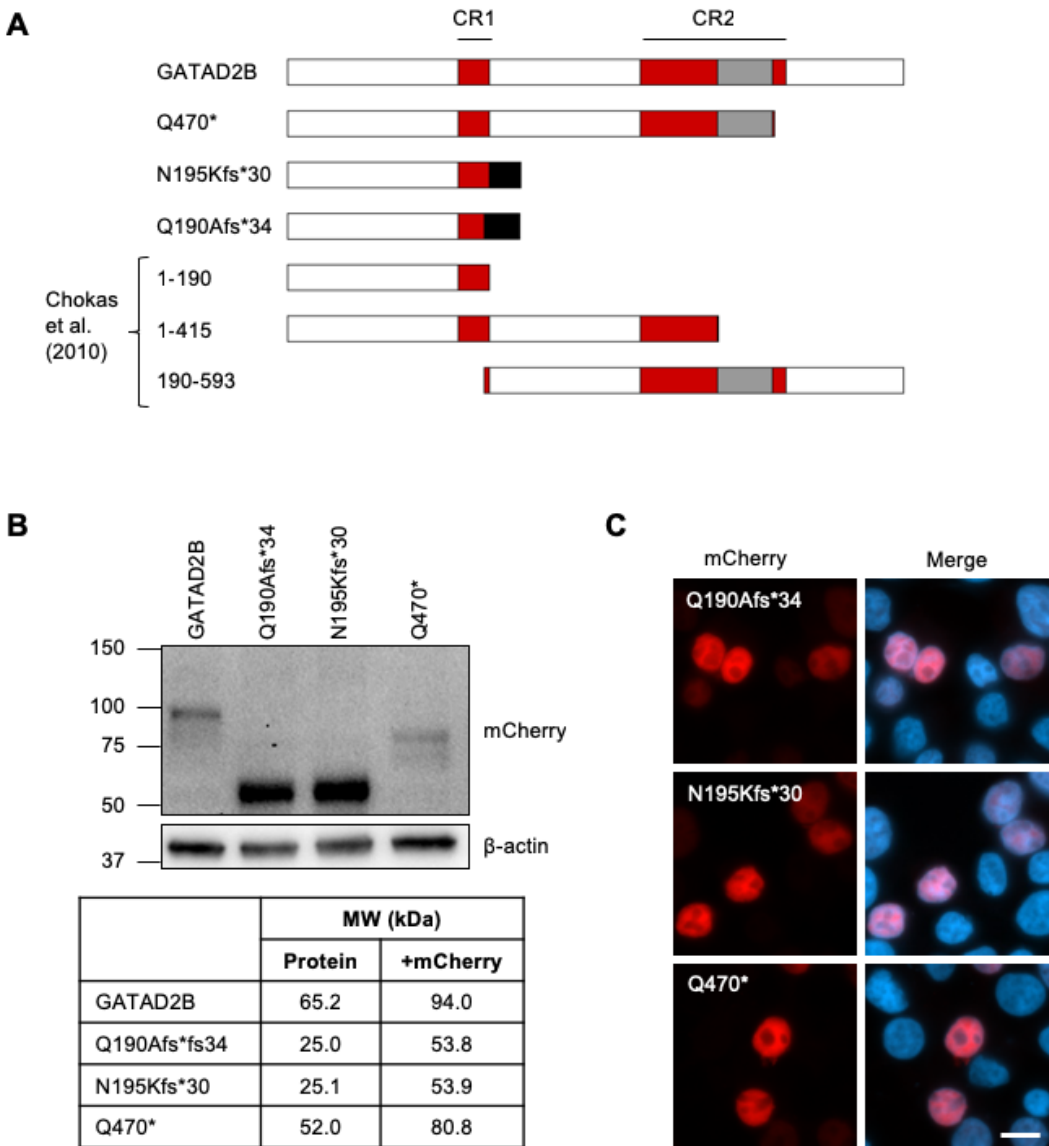
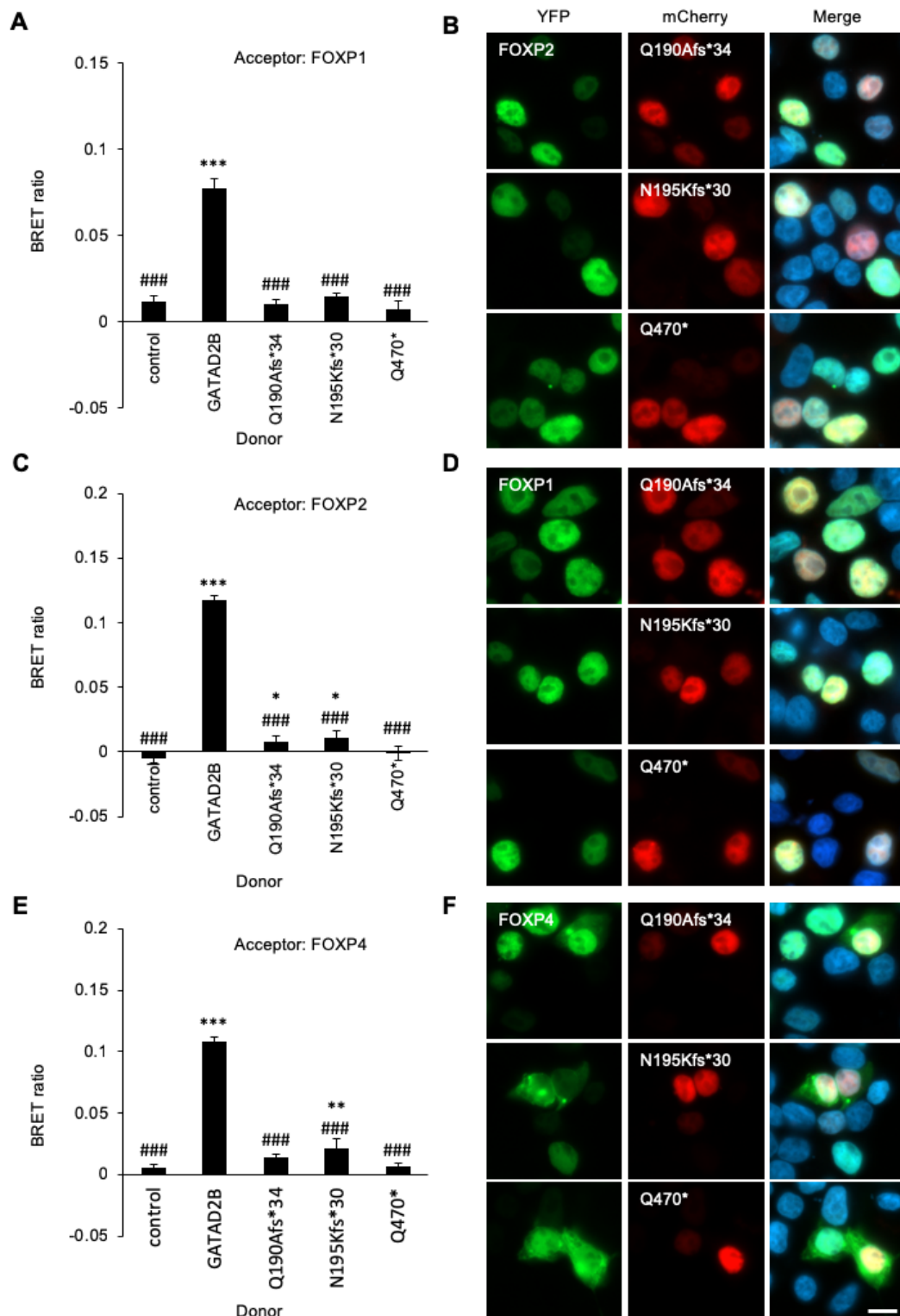


Figure 4.6. Pathogenic GATAD2B variants abolish nuclear speckles. (A) Schematic representing GATAD2B patient variants. The conserved regions CR1 and CR2 are shown in red. A GATA-type zinc finger within CR2 is shown in grey. Abnormal addition sequence caused by frameshift variants is shown in black. (B) Immunoblot of whole-cell lysates of cells expressing mCherry-tagged GATAD2B variants probed with anti-mCherry antibody. Blot was stripped and re-probed with anti-β-actin antibody to confirm equal loading. The predicted molecular weights of the mCherry-GATAD2B fusion proteins are indicated below. (C) Fluorescence imaging of cells transfected with mCherry-tagged GATAD2B variants (red). Nuclei are stained with Hoechst 33342 (blue). Scale bar, 10μm.



4.3.5 NO FURTHER EVIDENCE THAT FOXP PROTEINS INTERACTS WITH CHD3 OR KANSL1

Next, I investigated the putative interaction between FOXP proteins and CHD3, another NuRD complex member identified in the human foetal brain Y2H screening for FOXP2 interactors. The mCherry-tagged CHD3 protein was expressed at the expected molecular weight (Fig 4.8A) and exhibited a diffuse nuclear pattern of expression (Fig 4.8B), as previously described (Hu et al., 2015). BRET assays showed no interaction between CHD3 (as either donor or acceptor) and any FOXP protein (Fig 4.8C), despite at least partial co-localisation in the nucleus (Fig 4.8D).

Similarly, the NSL-complex member KANSL1 was expressed as an mCherry fusion protein at the expected molecular weight (Fig 4.9A) and localised to the nucleus (Fig 4.9B) in agreement with previous studies (Meunier et al., 2015). No evidence of interaction was observed in BRET assays between KANSL1 (as either donor or acceptor) and FOXP1/2/4 (Fig 4.9C), despite nuclear co-localisation (Fig 4.9D).

Figure 4.7 (previous page). Pathogenic GATAD2B variants prevent interaction with FOXP proteins. BRET assays for interaction between pathogenic GATAD2B variants and (A) FOXP2, (C) FOXP1 and (E) FOXP4. Bars represent the corrected mean BRET ratio \pm S.D. of one experiment performed in triplicate ($*P < 0.05$, $*P < 0.01$ and $***P < 0.001$ compared to control, $###P < 0.001$ compared to WT GATAD2B, one-way ANOVA and *post-hoc* Tukey's test). (B-D) Fluorescence imaging of cells co-transfected with mCherry-tagged GATAD2B variants (red) and YFP-tagged (B) FOXP2, (D) FOXP1 and (F) FOXP4. Nuclei are stained with Hoechst 33342 (blue). Scale bar, 10 μ m.

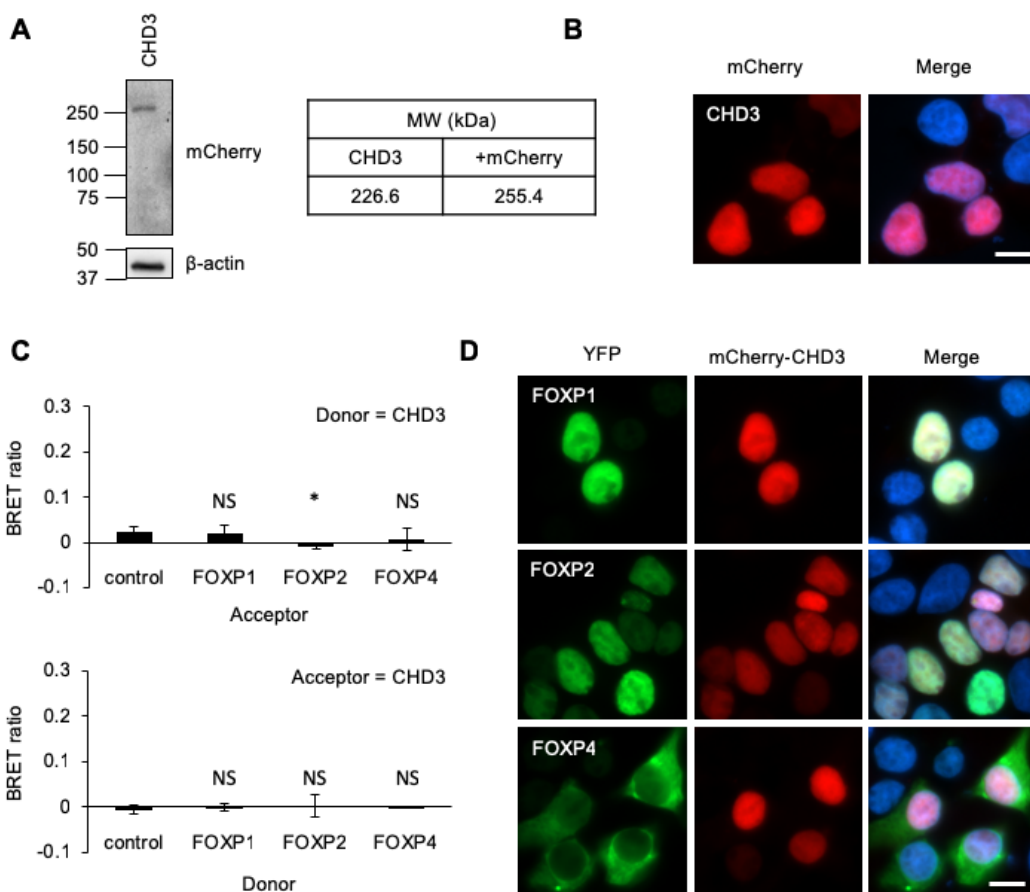


Figure 4.8. BRET does not detect an interaction between FOXP proteins and CHD3. (A) Immunoblot of whole-cell lysate of cells expressing mCherry-tagged CHD3 probed with anti-mCherry antibody. Blot was stripped and re-probed with anti- β -actin antibody. The predicted molecular weights of the mCherry-CHD3 fusion proteins is indicated. (B) Fluorescence imaging of cells transfected with mCherry-tagged CHD3 (red). Nuclei are stained with Hoechst 33342 (blue). Scale bar, 10 μ m. (C) BRET assay for interaction between CHD3 and FOXP1/2/4, with CHD3 as either donor (top) or acceptor (bottom). Bars represent the corrected mean BRET ratio \pm S.D. of one experiment performed in triplicate (* $P < 0.05$ and NS = not significant compared to control, Student's T-test). (D) Fluorescence imaging of cells co-transfected with YFP-tagged FOXP1/2/4 (green) and mCherry-tagged CHD3 (red). Nuclei are stained with Hoechst 33342 (blue). Scale bar, 10 μ m.

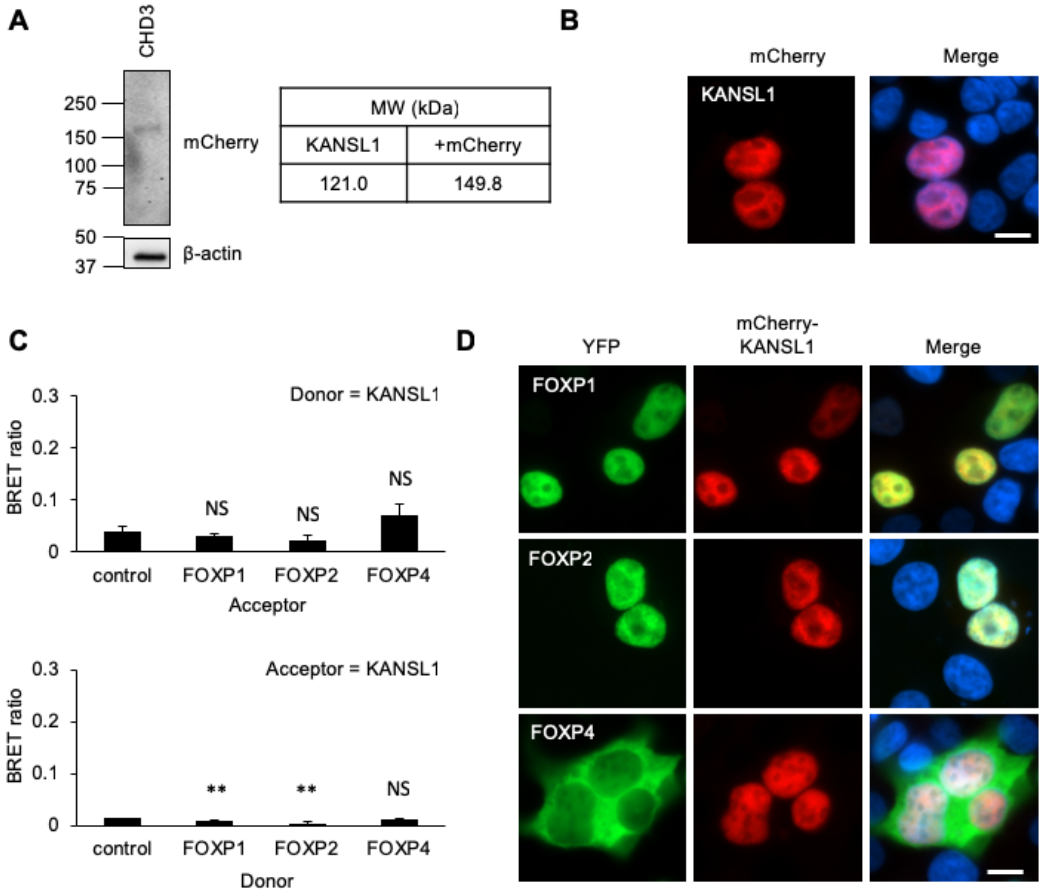


Figure 4.9. BRET does not detect an interaction between FOXP proteins and KANSL1. (A) Immunoblot of whole-cell lysate of cells expressing mCherry-tagged KANSL1 probed with anti-mCherry antibody. Blot was stripped and re-probed with anti- β -actin antibody to confirm equal loading. The predicted molecular weights of the mCherry-KANSL1 fusion proteins are indicated. (B) Fluorescence imaging of cells transfected with mCherry-tagged KANSL1 (red). Nuclei are stained with Hoechst 33342 (blue). Scale bar, 10 μ m. (C) BRET assay for interaction between KANSL1 and FOXP1/2/4, with KANSL1 as either donor (top) or acceptor (bottom). Bars represent the corrected mean BRET ratio \pm S.D. of one experiment performed in triplicate (** $P < 0.01$ and NS = not significant compared to control, Student's T-test). (D) Fluorescence imaging of cells co-transfected with YFP-tagged FOXP1/2/4 (green) and mCherry-tagged KANSL1 (red). Nuclei are stained with Hoechst 33342 (blue). Scale bar, 10 μ m.

4.4 DISCUSSION

The regulatory function of many transcription factors relies on interactions with other proteins, including chromatin remodelling factors. Here I have confirmed that the human FOXP1/2/4 transcription factors interact with GATAD2B, corroborating and extending previous discoveries in the orthologous mouse proteins (Chokas et al., 2010). The results here indicate that residues 260-422 are required for FOXP2-GATAD2B interaction, essentially replicating previous evidence for an interaction between mouse *Gatad2b* and the SD1 domain of mouse *Foxp1* (equivalent to hFOXP2 251-420) (Chokas et al., 2010). I also show that the leucine zipper motif is particularly important for the interaction. This motif has been previously shown to be essential for FOXP homo- and heterodimerisation (Li et al., 2004). Indeed, *FOXP* variants that disrupt dimerisation also abolish interaction with GATAD2B (Table 4.4). This may suggest that the same FOXP region is involved in both dimerisation and association with GATAD2B, or alternatively, that FOXP dimerisation is a necessary precondition for recruitment of GATAD2B and the NuRD complex.

Interestingly, several missense variants in *FOXP1/2* have been previously shown to form nuclear aggregates and to maintain the ability to bind wild-type FOXP proteins and translocate them into the nucleus – as a potential dominant negative effect (Sollis et al., 2016, 2017). These same variants also interact and co-localise with GATAD2B, although because GATAD2B exhibits a speckled expression pattern under normal circumstances, it is unclear how its localisation may be affected by the *FOXP* variants. Future work aimed at distinguishing abnormal aggregates from various normal nuclear structures that also produce a punctate localisation pattern (nuclear speckles, Cajal bodies, PML bodies etc.) might help to untangle these effects. For example, localisation in nuclear speckles could be confirmed by co-staining for pre-mRNA splicing factors indicative of these structures (e.g. SRSF2) (Spector and Lamond, 2011). Conversely, two *FOXP1* variants that cause cytoplasmic aggregation (p.V423Hfs*37, p.Y439*) were able to translocate WT FOXP proteins to the cytoplasm in previous studies (Lozano et al., 2015; Sollis et al., 2016), but did not have the same effect on GATAD2B here. GATAD2B therefore appears to be less susceptible to mislocalisation, perhaps because of strong interactions with other NuRD complex members that help to retain it within the nucleus.

I also investigated, for the first time, the effects of *GATAD2B* mutations on FOXP-interaction, including two frameshift variants (p.N195Kfs*30 and p.Q190Afs*35) that retain only the CR1 domain, and a nonsense variant (p.Q470*) that is truncated towards the end of the CR2 domain (Willemsen et al., 2013). Recently, two additional mutations have been reported, but are not characterised here: a p.K184Nfs*2 variant expected to behave similarly to the two frameshifts above, and a much shorter p.L28Mfs*18 variant that lacks all functional domains (Luo et al., 2017). All five variants have the potential to undergo nonsense-mediated decay (NMD) resulting in haploinsufficiency of *GATAD2B* in the affected patients. Indeed, immortalised lymphoblast cells from one patient exhibited significantly weaker expression of GATAD2B compared to healthy controls, providing evidence of degradation (Luo et al., 2017). Even if any truncated proteins were to be

expressed in patient cells, the experiments here demonstrate that none of the variants tested are capable of interacting with FOXP proteins and are therefore unlikely to exert abnormal effects on their functions.

Here, I find support for the hypothesis that the CR2 domain of GATAD2B is essential for FOXP-binding (Chokas et al., 2010). However, the results in this chapter do not agree with previous evidence that the C-terminal portion of CR2, including the zinc finger motif (residues 414-467), is not required for the interaction (Chokas et al., 2010). Indeed, the p.Q470* truncation, which terminates after the zinc finger, was unable to interact with FOXP1/2/4 in the BRET assays presented here. These contradictory results could arise from the fact that the BRET assays employed full-length FOXP proteins, whereas Chokas et al. (Chokas et al., 2010) used only the repression domains, which might exhibit dissimilar binding properties. Alternatively, it might reflect differences between the mouse and human proteins, or differences between the assays used in each study. For example, interactions may be influenced by the dynamic live cell environment of the BRET assays used here, in contrast to the cell-free pull-down assays in the earlier study (Chokas et al., 2010).

FOXP-GATAD2B interactions provide a potential molecular link between clinically distinct neurodevelopmental disorders that nonetheless share certain phenotypic features. Most notably, the *FOXP1*-, *FOXP2*- and *GATAD2B*-related disorders all involve severe speech and language impairment or delay (Table 4.5). The degree of speech delay varies considerably between patients but falls within a similar range for all three disorders. For example, first word acquisition is reported at 1.5-7 years for patients with a mutation in *FOXP2* (Morgan et al., 2017); 1.5-6 years for *FOXP1* mutations (Le Fevre et al., 2013), with at least one patient who was entirely non-verbal at 7 years 11 months (Sollis et al., 2017); and 2-8 years for *GATAD2B* mutations (Willemsen et al., 2013), with one patient who spoke no words at 12 years (Luo et al., 2017). One of the characteristics of *FOXP2*-related disorder is that expressive language is more severely affected than receptive language (Morgan et al., 2017), and similar observations have been made in patients with *FOXP1* variants (Le Fevre et al., 2013; Sollis et al., 2016). Detailed language assessment has not been performed in *GATAD2B*-related disorder, however anecdotal evidence is consistent with relatively spared receptive language in at least one patient, whose ‘comprehension of language was at a higher level’ (Willemsen et al., 2013), and another who ‘could understand simple instructions’ (Luo et al., 2017). Oromotor dysfunction (including CAS as well as non-linguistic oral motor dyspraxia) is a core symptom of *FOXP2*-related disorder (Morgan et al., 2017) and has also been reported in some people with *FOXP1* mutations (Le Fevre et al., 2013; Sollis et al., 2016). In both cases, oromotor dysfunction is sometimes associated with feeding problems during infancy (Horn et al., 2010; Le Fevre et al., 2013; Zimmerman and Maron, 2016). Feeding difficulties were also reported in a patient with a 1q21.3 microdeletion encompassing *GATAD2B* and 9 other genes (Willemsen et al., 2013), and another patient with a larger 1q21.3 microdeletion containing *GATAD2B* and 34 other genes (Tim-Aroon et al., 2017). Further studies are needed, however, to determine whether oromotor problems are a consequence of mutations affecting *GATAD2B* only.

Table 4.5. Shared phenotypic features across disorders.

	FOXP1	FOXP2	GATAD2B	KANSL1
Intellectual disability	+	< average verbal IQ	+	+
General motor delay	+	rare	+	mild
Oral motor difficulties	+	+	ND	+
Speech/language impairment/delay	+	+	+	+
Expressive language more severely affected	+	+	2 cases	+
Autistic features	some cases	rare	1 case	-
Behavioural problems	some cases	-	some cases	some cases
Hypotonia	+	1 mild case	+	+
Strabismus	some cases	1 case	+	some cases
Sleep problems	1 case	-	some cases	-
Sensory symptoms	sensory integration disorder (2 cases), sensitivity to temperature/textures (1 case)	-	high pain threshold (1 case)	-
Other features	urogenital anomalies (some cases)			epilepsy (some cases), urogenital anomalies (some cases), heart defects (some cases)

Based on the following references: FOXP1 (Le Fevre et al., 2013; Sollis et al., 2016, 2017), FOXP2 (Morgan et al., 2017), GATAD2B (Willemsen et al., 2013; Luo et al., 2017), KANSL1 (Koolen and de Vries, 2013). ND = no data.

Several aspects of *GATAD2B* disorder align more closely with the *FOXP1*- rather than *FOXP2*-related phenotype. Both *GATAD2B* and *FOXP1* variants lead to ID, as well as gross motor delays and often hypotonia during infancy (Le Fevre et al., 2013; Willemsen et al., 2013; Sollis et al., 2016, 2017). In contrast, people with *FOXP2* mutations are usually not intellectually disabled, although some individuals have lower than average non-verbal and especially verbal IQ (Vargha-Khadem et al., 1995; Morgan et al., 2017), and general motor deficits and hypotonia have only been reported in isolated cases (Reuter et al., 2017). Highly variable behavioural problems, including hyperactivity and obsessions, have occasionally been reported in patients with mutations in *GATAD2B* and *FOXP1*, but not *FOXP2* (Willemsen et al., 2013; Sollis et al., 2016, 2017; Luo et al., 2017; Morgan et al., 2017).

ASD (or autistic features) is a relatively frequent but not universal component of *FOXP1*-related disorder (Le Fevre et al., 2013; Sollis et al., 2016, 2017), but is reported in only a handful of *FOXP2* mutation cases (Reuter et al., 2017). ASD has not been formally

diagnosed in any patient with a *GATAD2B* mutation, although one patient was described as exhibiting an “autistic tendency” (Luo et al., 2017). It therefore remains to be seen whether autistic features form part of the core *GATAD2B* haploinsufficiency phenotype.

Interestingly, strabismus is consistently noted in patients with *GATAD2B* mutations (Willemsen et al., 2013), and has also been reported in several patients with *FOXP1* mutations (Sollis et al., 2016, 2017), as well as a single individual with a *FOXP2* mutation (Reuter et al., 2017). Other potential commonalities include sleeping problems, seen in two cases with *GATAD2B* mutations (Willemsen et al., 2013) and one with a *FOXP1* mutation (Sollis et al., 2017), as well as sensory symptoms that include an elevated pain threshold in one *GATAD2B* mutation case (Willemsen et al., 2013), sensory integration disorder in two patients with *FOXP1* mutations and heightened sensitivity to temperature and certain textures in a third (Sollis et al., 2016, 2017). These clinical features could potentially share a similar underlying defect.

FOXP4 mutations have not been definitively linked to disorder and were therefore not tested here. A homozygous recessively-inherited frameshift variant, expected to result in the complete absence of FOXP4, was recently reported in a patient with developmental delay, laryngeal hypoplasia, feeding difficulties and ventricular septal defect (Charng et al., 2016). However, the variant is of unknown significance as the patient also carried additional variants in other genes, and the *FOXP4* variant was absent from a mildly affected brother. If a *FOXP4*-related syndrome emerges in the future, it will be interesting to see how it compares to the *GATAD2B* phenotype.

These overlapping neurological symptoms support a potential cooperative role for *GATAD2B* and *FOXP1/2*, and possibly *FOXP4*, in the brain. However, the variation in the breadth and severity of these disorders suggests that these genes nonetheless have distinct functions in neurodevelopment. Expression studies in developing and adult rodent brain have shown that *FOXP1/2/4* have distinct but partially overlapping expression patterns, including in the striatum (*FOXP1/2/4*), hippocampus (*FOXP1/4*) and the Purkinje cells of the cerebellum (*FOXP2/4*) (Ferland et al., 2003; Takahashi et al., 2008; Tam et al., 2011). *FOXP4* is expressed throughout the cortex, whereas *FOXP1* is concentrated in layers 3-5 and *FOXP2* is restricted to layer 6 and in a small subset of layer 5 neurons (Ferland et al., 2003; Takahashi et al., 2008; Hisaoka et al., 2010). In adult mouse and human post-mortem tissue, *GATAD2B* is expressed throughout the brain, including at high levels in cortical, hippocampal and striatal neurons, and Purkinje cells (Uhlén et al., 2015; see URL¹; Lein et al., 2007; see URL²), suggesting that it co-occurs with different combinations of *FOXP*s in multiple brain regions. A previous study found that mouse *Gatad2b* co-operatively repressed *Foxp1* target genes in the lung (Chokas et al., 2010). It is therefore possible that similar co-

¹ Human Protein Atlas:

<https://www.proteinatlas.org/ENSG00000143614-GATAD2B/tissue>

² Allen Brain Atlas:

<http://mouse.brain-map.org/experiment/show/71021020>

repression occurs in neuronal populations expressing FOXP proteins together with GATAD2B.

Recent studies have begun to elucidate the role of the NuRD complex, and *GATAD2B* in particular, in neurodevelopment. NuRD-mediated repression plays a vital role in promoting synaptic connectivity (Yamada et al., 2014). In fact, neuronal knockdown of *simjang*, a *GATAD2A/B* orthologue in *Drosophila*, disrupted synapse development and impaired non-associative learning in affected flies (Willemsen et al., 2013). The NuRD complex has also been shown to regulate neural progenitor proliferation, radial migration, neuronal differentiation and cortical layer specification (Egan et al., 2013; Nitarska et al., 2016). Related functions have been identified among *FOXP2* transcriptional targets - differentiation, migration and synaptic plasticity (Spiteri et al., 2007; Vernes et al., 2007, 2011); and *FOXP1* targets - synapse, neuronal development, neuronal differentiation (Araujo et al., 2015). *FOXP1* is also involved in regulating cortical radial migration (Li et al., 2015b). Interactions between FOXP proteins and GATAD2B/NuRD may therefore be relevant for the regulation of multiple aspects of neurodevelopment.

The human brain Y2H screen also identified CHD3 and KANSL1 as potential FOXP-interaction partners, but these putative interactions could not be validated in confirmatory BRET assays. The lack of interaction seen in the BRET assays could be a false negative result. Even when two proteins are bound to one another, the BRET technique is sensitive to their conformations, which can place the fused RLuc- and YFP-tags too far away from one another to allow energy transfer (Deriziotis et al., 2014a). This could be particularly problematic in the case of the CHD3 protein, which is considerably longer (2000 amino acids) than FOXP1/2/4 (677, 715 and 680 amino acids, respectively). Another possibility is that the required binding sites are positioned at the N-terminus of the protein, where the RLuc- or YFP-tags could interfere with the interaction (Deriziotis et al., 2014a). Experiments that instead use C-terminal tags could help to resolve this question. Alternatively, the Y2H results may have been false positives - interactions that occur only in the artificial context of human proteins expressed in yeast cells. The interactions may not occur in the more physiological conditions of the BRET assay, which takes place in human cells against a background of other dynamic processes. These interactions might also require the presence of additional proteins that are not expressed at sufficient levels in the cell model used here. Chromatin remodelling complexes comprise multiple proteins with specialised functions, not all of which may interact directly with transcription factors. For example, FOXP proteins might only bind directly to GATAD2B, which then indirectly links them to the rest of the NuRD complex, including CHD3. On the other hand, indirect interactions are less likely to account for the detection of CHD3 and KANSL1 in Y2H assays, where the yeast cell environment lacks other human proteins with which the candidates may interact *in vivo* (Galletta and Rusan, 2015). Co-immunoprecipitation assays were attempted in order to clarify these issues, but optimising these assays proved to be difficult for FOXP proteins, as has been previously noted for transcription factor complexes in general (Klenova et al.,

2002). Future studies using alternative techniques may ascertain whether CHD3 and KANSL1 bind directly to FOXP proteins.

Although I could not demonstrate an interaction between FOXP proteins and either CHD3 or KANSL1 in my assays, both of these chromatin remodelling proteins remain interesting candidates for neurodevelopmental disorders involving speech and language impairment. In fact, *CHD3* variants have recently been identified in a number of patients exhibiting a broad spectrum of mild-severe ID and language deficits, including patients diagnosed with childhood apraxia of speech (Eising et al., 2018; Snijders Blok et al., 2018). Research is ongoing to assess the phenotype of these patients, with a focus on characterising their speech/language deficits. Koolen-de Vries syndrome, caused by mutations in *KANSL1*, also exhibits intriguing overlaps with *FOXP1/2*-related disorders (Table 4.5), including mild-moderate ID, severe speech delay, speech apraxia and childhood hypotonia (Koolen et al., 2012; Zollino et al., 2012; Morgan et al., 2018). Importantly, expressive language appears to be more severely affected than receptive language, as is also the case in *FOXP2*- and *FOXP1*-related disorders (Koolen et al., 2012). Interestingly, analyses of RNA-sequencing data from patient-derived cell lines found that differentially expressed genes were enriched in neuronal and synaptic processes, suggesting that KANSL1 may have functions in common with FOXP proteins and GATAD2B.

Knowledge of the broader FOXP interactome is still limited at this stage, although a recent mass spectrometry-based study has identified several novel FOXP-interactors (Estruch et al., 2018). In the meantime, valuable information has been obtained by validating interactions in cell models, and by investigating the effects of disease-related variants on these interactions.

5 MASS SPECTROMETRY-BASED SCREENING FOR TBR1-INTERACTING PROTEINS

ABSTRACT

TBR1 is a brain-specific transcription factor involved in the development of multiple brain regions and mutated in patients with autism spectrum disorder (ASD) and intellectual disability (ID), often accompanied by speech/language deficits. Among the few known interaction partners of TBR1 are FOXP1 and FOXP2, which are also implicated in speech/language-related disorders. In this study, affinity purification-mass spectrometry was employed to identify novel TBR1-interaction candidates in a stably-transfected HEK293 cell-line. After filtering for common contaminants, a set of 247 novel interaction candidates were detected, in addition to one known interactor (CASK). The putative interactome was enriched for processes such as transcriptional regulation, chromatin modification, RNA processing, DNA replication, cell cycle regulation and ubiquitination. In particular, epigenetic factors (including those with both positive and negative effects on transcription) were significantly overrepresented. There was also a significant overrepresentation of proteins encoded by ASD- and ID-related genes. As expected, given the established expression patterns of TBR1, the majority of detected proteins are expressed at high levels in the brain and localise predominantly to the cell nucleus. These results provide the first comprehensive assessment of the TBR1 interactome and highlight several promising avenues for investigating TBR1 function, as well as suggesting new candidates for language-related and other neurodevelopmental disorders.

5.1 INTRODUCTION

In previous chapters I investigated Mendelian speech/language-related disorders in order to identify genes and proteins that may be involved in the development of a language-capable brain, focussing on the FOXP1/2 transcription factors (TFs) in particular. Another way to uncover molecular networks involved in speech/language is to investigate the protein-protein interactions of known language-related proteins. Affinity purification-mass spectrometry (AP-MS) enables a comprehensive survey of the interaction network (interactome) of a particular protein. In the current chapter, AP-MS is employed to characterise the interactome of TBR1, a neuron-specific member of the T-box TF family with interesting phenotypic and molecular links to speech and language disorders.

Recurrent heterozygous *TBR1* disruptions (including whole gene deletions, missense variants, and truncating mutations) have been reported in sporadic cases of autism spectrum disorder (ASD) (Neale et al., 2012; O’Roak et al., 2012b, 2012a; De Rubeis et al., 2014; Iossifov et al., 2014). The associated phenotype often includes intellectual disability (ID) and developmental delay, and most patients exhibit some degree of language delay in addition to the general deficits in social communication associated with ASD (Deriziotis et al., 2014b). In addition, a 2q24.2 deletion encompassing only the *TBR1* gene was identified in a patient with ID and language delay, but without a diagnosis of autism (Palumbo et al., 2014). Interestingly, TBR1 interacts with the FOXP1/2 TFs (Deriziotis et al., 2014b). *FOXP2* is mutated in a severe speech and language disorder (Lai et al., 2001), and *FOXP1* is mutated in patients with ID and speech delay (Le Fevre et al., 2013; Sollis et al., 2016). The TBR1-FOXP2 interaction is abolished by pathogenic variants in either protein (Deriziotis et al., 2014b; den Hoed et al., 2018), and at least one of the known *FOXP1* pathogenic variants also disrupts TBR1-interaction (Sollis et al., 2017). TBR1 may therefore represent an important node in the network of proteins involved in speech and language disorders and related phenotypes. It is a promising focal point for further studies seeking to expand our knowledge of these networks.

TBR1 expression is restricted to the brain, where it is specific to post-mitotic neurons (Huang and Hsueh, 2015). Studies in mice reveal that *Tbr1* expression reaches its peak during embryogenesis, and gradually decreases postnatally (Bulfone et al., 1995). Around birth, it is highly expressed in the cerebral cortex, primarily in layer 6, but also in layer 2/3 and in a few neurons in layer 5 (Bulfone et al., 1995; McKenna et al., 2011). It is also expressed in embryonic amygdala (Remedios et al., 2007; Huang et al., 2014), hippocampus (Bulfone et al., 1995; Cipriani et al., 2016), olfactory bulb (Bulfone et al., 1998) and deep cerebellar nuclei (Fink et al., 2006). *Tbr1* appears to play an important regulatory role in the development of many of these brain regions. In the cortex, *Tbr1* controls both regional and laminar neuronal identity, driving differentiation towards frontal cortex and layer 6 cell fates, while suppressing caudal and layer 5 identity (Bedogni et al., 2010; McKenna et al., 2011). In the amygdala, *Tbr1* promotes cell migration, axonal outgrowth and the formation of inter- and intra-amygdalar connections (Remedios et al., 2007; Huang et al., 2014). Consistent with its role as a TF, TBR1 protein is typically localised to the nucleus of cells in which it is

expressed. There is also some evidence of cytoplasmic and synaptic localisation in certain neurons in postnatal and adult (but not embryonic) rat cortex, hippocampus and cerebellum, although the functional significance of this is unclear (Hong and Hsueh, 2007).

At the molecular level, TBR1 binds to target DNA loci via its T-box DNA-binding domain and recognises the T-box binding element AGGTGTGA (Jolma et al., 2013; Notwell et al., 2016). For some members of the T-box TF family, DNA-binding requires homodimerisation (Müller and Herrmann, 1997), and in vitro functional assays suggest this may also be true of TBR1, as disease-causing variants that prevent homodimerisation also disrupt transcriptional repression by TBR1 (Deriziotis et al., 2014b). TBR1 appears to function as either an activator or repressor of transcription. TBR1-binding sites identified by ChIPseq are enriched for both active (H2K27ac, H3K4me1) and repressive (H3K9me3, H3K27me3) chromatin marks (Notwell et al., 2016). Targets found to be upregulated by *Tbr1* in mouse neurons include *Auts2*, *Reln* and *Grin2b* (Bedogni et al., 2010; Chuang et al., 2014); while *Fezf2* and *Bcl11b* (*Ctip2*) are downregulated (McKenna et al., 2011). Several TBR1 target genes are also high-confidence candidates for ASD susceptibility, including *GRIN2B* (O’Roak et al., 2011) and *RELN* (De Rubeis et al., 2014). In a large-scale screen in mouse cortex, genes associated with ASD susceptibility were found to be enriched adjacent to *Tbr1* ChIP peaks and among genes that are differentially expressed in *Tbr1* knockout mice (Notwell et al., 2016).

The regulatory function of many TFs relies on interactions with other proteins, and it is likely that TBR1 is no exception. In addition to forming homodimers (Deriziotis et al., 2014b), TBR1 also interacts with CASK, a membrane-associated guanylate kinase primarily expressed at neuronal synapses (Hsueh et al., 2000). This interaction allows CASK to enter the nucleus, where it acts as a coactivator with TBR1 to promote expression of TBR1 target genes, such as *RELN* and *GRIN2B* (Hsueh et al., 2000; Wang et al., 2004b). As mentioned above, TBR1 also interacts with *FOXP2*, and its paralogues *FOXP1* and *FOXP4* (Deriziotis et al., 2014b), as well as *BCL11A* (a.k.a. *CTIP1*) (den Hoed et al., 2018). However, it is likely that the interaction network of TBR1 is much larger than the proteins identified so far. The AP-MS approach used here enables a more comprehensive characterisation of the TBR1 interactome.

5.2 METHODS

5.2.1 CELL CULTURE AND STABLE TRANSFECTION

HEK293 cells were cultured in DMEM supplemented with 10% foetal bovine serum. The coding sequence of TBR1 was amplified from a plasmid template with *Bgl*II and *Xho*I restriction sites and inserted with N-terminal double-FLAG and V5 tags into a puromycin-resistant pPyCAG vector (van den Berg et al., 2010). TBR1 forward primer (5' to 3', *Bgl*II site underlined): AGATCTCAGCTGGAGCACTGCCTTTC. Reverse primer (5' to 3', *Xho*I site underlined): CTCGAGCTAGCTGTGCGAGTAGAAGC.

To generate stable cell lines, the pPyCAG-2xFLAG/V5-TBR1 plasmid was linearised by *AdhI* digestion and transfected into HEK293 cells using GeneJuice (Merck-Millipore), according to the manufacturer's instructions. Single clones were isolated following selection with culture medium containing 10 μ M puromycin. The expression of tagged TBR1 protein in selected clones was confirmed using an anti-V5 antibody (Abcam, ab27671), by Western blotting (1:3,000) and by immunofluorescence (1:500). Selected stable cell lines were maintained in culture medium containing 5 μ M puromycin.

5.2.2 NUCLEAR EXTRACTION AND FLAG-TBR1 AFFINITY PURIFICATION

HEK293 cells stably expressing 2xFLAG/V5-TBR1, and untransfected control cells, were expanded to confluence in twenty 15 cm dishes, harvested by scraping in PBS and nuclear extracts were prepared following Dignam et al. (Dignam et al., 1983). Two separate nuclear extracts were prepared for each condition, and affinity purifications (APs) were performed in duplicate.

The AP procedure has been described previously (van den Berg et al., 2010). Briefly, nuclear extracts were dialysed into buffer C-100 (20 mM HEPES pH 7.6, 20% glycerol, 100 mM KCl, 1.5 mM MgCl₂, 0.2 mM EDTA) and 1.5 ml nuclear extract incubated with anti-FLAG M2 agarose beads (Sigma) for 3 hr at 4°C. Nuclear extract was supplemented with 225 units of benzonase (Novagen) to digest DNA and 50 μ g/ml ethidium bromide to inhibit DNA-protein associations (Lai and Herr, 1992). Beads were washed five times for 5 min with buffer C-100 containing 0.02% NP-40 (C-100*) and bound proteins eluted four times for 15 min at 4°C with buffer C-100* containing 0.2 mg/ml FLAG-tripeptide (Sigma). Elution of proteins was validated by Western blot, and the first two elutions were pooled for each condition. Proteins were TCA precipitated and separated by polyacrylamide gel electrophoresis.

The AP experiments in this chapter were performed with assistance from Martí Quevedo (MQ) at the Erasmus University Medical Centre, Rotterdam (Erasmus MC). Nuclear extractions were performed by ES; dialysis, anti-FLAG bead incubation and elution by MQ and ES; Western blotting by MQ (AP1) or ES (AP2); and TCA precipitation and protein separation by MQ.

5.2.3 MASS SPECTROMETRY

Mass spectrometry was performed by the Proteomics Centre at the Erasmus MC, as previously described (van den Berg et al., 2010). Briefly, 1D SDS-PAGE gel lanes were prepared by in-gel reduction with dithiothreitol, alkylation with iodoacetamide and digestion with trypsin. Nanoflow LC-MS/MS was performed on an 1100 series capillary LC system (Agilent Technologies) connected to an LTQ-Orbitrap mass spectrometer (Thermo). Mass spectra were acquired and searched against the UniProt human proteome database (UP000005640, accessed February 2016; The UniProt Consortium, 2017) using the Mascot search algorithm (version 2.5.2). Each protein identification was assigned a Mascot score,

equal to $-10 \cdot \log_{10}(P)$, where P is the probability that the observed match is a random event. Peptides with a Mascot score lower than 40 (i.e. $P > 10^{-4}$) were excluded. An emPAI score (exponentially modified protein abundance index) was also calculated for each protein hit, which incorporates the number of peptides identified per protein normalised by the theoretical number of peptides for that protein (Ishihama et al., 2005). This score corrects for the fact that, for the same number of molecules, larger proteins and proteins with many peptides in the preferred mass range for mass spectrometry will generate more observed peptides.

5.2.4 FILTERING

Preliminary data preparation was done using Microsoft Excel and R. Filtering was performed in Cytoscape (version 3.5.0). Contaminants, including human keratins, bovine serum proteins introduced during cell culture, and trypsin used for protein fragmentation, were removed from each list. For each experiment, non-specific hits were removed by retaining only those proteins detected in the TBR1-expressing cells and not in control cells. Protein hits were then filtered further by removing common background contaminants obtained from the Contaminant Repository for Affinity Purification (CRAPome) database (accessed April 2017) (Mellacheruvu et al., 2013). Employing a strict filtering approach, all proteins identified in >1 out of 30 control experiments were excluded. The control experiments were matched for similar experimental conditions to the present study (HEK293, FLAG-tag, agarose beads). Proteins that could not be mapped to the CRAPome database were also excluded. Only proteins that were replicated in two independent AP-MS experiments were selected for inclusion in the final list of confident interaction partners (Chapter 5 Appendix). The number of proteins remaining at each stage of filtering is shown in Table 5.1.

Table 5.1. Proteins remaining for each experiment after each filtering step.

	AP1	% of original	AP2	% of original
Raw data	866	100%	689	100%
After removal of control proteins	836	96.5%	540	78.4%
After removal of common contaminants	579	66.9%	363	52.7%
Replicated in two AP-MS experiments	248	28.6%	248	36.0%

AP = affinity purification, AP-MS = affinity purification-mass spectrometry.

5.2.5 GENE ONTOLOGY ANALYSIS

Gene ontology overrepresentation analysis was performed using the PANTHER Overrepresentation Test (release 20170413) (Mi et al., 2017) with annotations from the Gene Ontology database (released 2017-10-23). The consistent hits identified by mass spectrometry analysis were uploaded as the Analysed List ($n = 248$). The Reference List consisted of all genes expressed in HEK293 cells ($n = 12,095$) according to RNA-sequencing

data from the Human Protein Atlas (Uhlén et al., 2015). Terms with $p < 0.05$ were returned, after Bonferroni correction for multiple testing.

5.2.6 NETWORK ANALYSIS

Network analysis of the putative TBR1 interactome was performed in Cytoscape (version 3.5.0). Known interactions within the network were imported from the STRING database (version 10.5) (Szklarczyk et al., 2017), with a minimum required interaction score of 0.700 (high confidence) and allowing only interactions supported by experimental evidence or curated databases. The MCODE (Molecular Complex Detection) algorithm (version 1.4.1) (Bader and Hogue, 2003) was used to identify highly interconnected regions within the network.

5.2.7 OVERREPRESENTATION ANALYSES

Proteins were annotated as TFs according to a curated list of human sequence-specific DNA-binding TFs (Vaquerizas et al., 2009). The TF list used for the analyses in the current chapter consisted of all proteins defined by Vaquerizas et al. (2009) as probable TFs (classes 'a', 'b' or 'other') or as possible TFs that contain InterPro domains that are only ever found in TFs (class 'c'). Epigenetic factor status and complex membership were assigned according to the Epifactors database (accessed May 2017) (Medvedeva et al., 2015). ASD candidate genes ($n = 190$) were taken from the Simons Foundation Autism Research Initiative database for ASD (SFARI Gene 2.0, accessed May 2017; Abrahams et al., 2013). Genes with a SFARI score in category 1) High Confidence, 2) Strong Candidate or 3) Suggestive Evidence were included, while lower confidence categories were excluded. Genes related to syndromic forms of ASD were included. ID candidate genes ($n=748$) with a mutation identified in at least one patient were taken from the Radboud University Human Genetics Department diagnostic sequencing panel (version DG2.5; see URL¹). The list of schizophrenia genes ($n = 662$) was compiled from the literature by (Moen et al., 2017) and included all genes with a *de novo* loss-of-function or missense mutation in at least one patient. A two-tailed Fisher's exact test was employed to test for overrepresentation of TFs, epigenetic factors or neurodevelopmental disorder genes in the TBR1 interactome ($n = 248$), compared to all HEK293-expressed genes ($n = 12,095$; as for gene ontology above).

5.2.8 TISSUE EXPRESSION AND SUBCELLULAR LOCALISATION

Immunohistochemistry-based protein expression levels were obtained from the Human Protein Atlas for various brain regions and cell populations (Uhlén et al., 2015), excluding proteins with no tissue expression data, or where the reliability of the expression data was tagged as "Uncertain". Immunocytochemistry/immunofluorescence-based subcellular

¹ ID sequencing panel: <http://www.radboudumc.nl/en/patientenzorg/onderzoeken/exome-sequencing-diagnostics/exomepanelspreviousversions/intellectual-disability>

localisation data were also obtained from the Human Protein Atlas (Thul et al., 2017), excluding proteins for which the reliability of the subcellular localisation data was tagged as "Uncertain". Proteins were classified as nuclear if they were annotated with "Nucleus", "Nucleoplasm", "Nuclear bodies", "Nuclear membrane", "Nuclear speckles", "Nucleoli" or "Nucleoli fibrillary center".

5.3 RESULTS

5.3.1 IDENTIFICATION OF TBR1-INTERACTING PROTEINS

HEK293 cells were stably transfected under puromycin selection to express 2xFLAG/V5-tagged TBR1 protein. Seven clones were picked and assessed for TBR1 expression (Fig 5.1A). In all seven clones, Western blotting detected TBR1 at the expected molecular weight of ~77.5kDa. A second fainter non-specific band was detected in all clones at ~60kDa. As no alternative isoforms of TBR1 have been experimentally confirmed (The UniProt Consortium, 2017; see URL¹) this band is more likely due to degradation. Clone H7 was selected for all experiments in this study. Typical nuclear expression of TBR1 in clone H7 was confirmed by immunofluorescence (Fig 5.1B). Two independent affinity purifications (AP1 and AP2) were performed using an anti-FLAG antibody to isolate TBR1 and any bound proteins. In both experiments, TBR1 was successfully immunoprecipitated from clone H7 cells (Fig 5.1C). For each experiment, the first two elutions were pooled for proteomic analysis. After protein separation by polyacrylamide gel electrophoresis, Coomassie staining detected proteins at a broad range of masses, including a strong band corresponding to TBR1, at ~77.5kDa (Fig 5.1D). Protein detection was much weaker in the untransfected control cells and there was no band indicating the presence of TBR1.

After removing MS artefacts and common contaminants, and excluding TBR1 itself, a total of 579 proteins were identified in AP1 and 363 proteins in AP2. A total of 248 proteins were replicated in both experiments (42.8% of AP1 hits, 68.3% of AP2) and carried through for further analysis (Chapter 5 Appendix). Of the previously-reported TBR1 interactors, only CASK was identified in our AP-MS screen (average Mascot score = 192.5, average emPAI = 0.20). The FOXP1/2/4 TFs were not detected in our experiments, while BCL11A was detected in AP1 only, at relatively low levels (Mascot score = 71, emPAI = 0.07), and could not be replicated in the second experiment. The absence of these proteins is unlikely to be explained by a lack of expression, as all four are expressed in HEK293 cells: RNA sequencing studies in HEK293 have detected moderate expression of FOXP1 (11.7 transcripts per million [TPM]), FOXP2 (6.7 TPM), FOXP4 (19.1 TPM) and BCL11A (12.2 TPM), all of which were higher than the median expression level for all genes in HEK293 (4.1 TPM) (Uhlén et al., 2015). These expression values were also in a similar range as that observed for CASK (17.0 TPM), which was identified in the current AP-MS study.

¹ UniProt entry for TBR1: <https://www.uniprot.org/uniprot/Q16650>

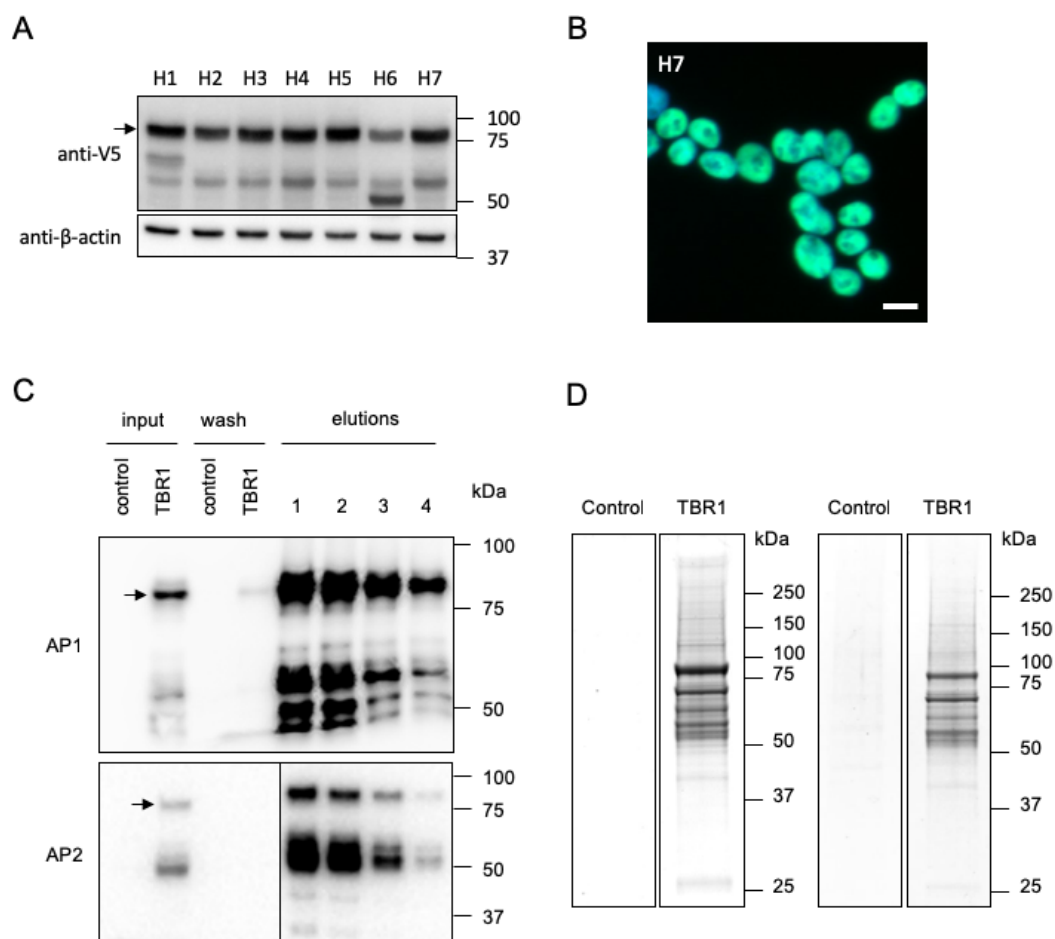


Figure 5.1. Cell preparation and affinity purification of TBR1. (A) Immunoblotting of whole-cell lysate from HEK293 cells stably transfected with 2xFLAG/V5-TBR1, using anti-V5 antibody (1:3,000). β -actin is also shown (1:10,000). Arrow shows ~77.5kDa band corresponding to TBR1. (B) Immunofluorescence staining of nuclear TBR1 expression (green) in the stably transfected cells using anti-V5 antibody (1:500) and Alexa 488 secondary antibody. Scale bar = 10 μ m (C) Affinity purification of TBR1-interacting proteins (AP1 and AP2). Western blots show total lysate (input) and washed proteins (wash) for empty HEK293 cells (control) and TBR1-containing stable cell lines (TBR1), and affinity-purified material (elutions; TBR1 stable cell lines only). Immunoblotting performed with anti-FLAG primary antibody (1:1000) and HRP-conjugated anti-mouse secondary antibody (1:2,000). Arrow shows ~77.5kDa band corresponding to TBR1. (D) Coomassie-stained SDS-polyacrylamide gel of affinity purifications of TBR1 and control cells.

5.3.2 THE TRANSCRIPTION FACTOR ROLE OF TBR1 IS SUPPORTED BY INTERACTOME FUNCTIONS

The PANTHER Overrepresentation Test was used to identify Gene Ontology terms that were significantly overrepresented in the TBR1 interactome. Genes expressed in HEK293 cells served as the reference list (n = 12,095). Overall, the enriched terms reflect the established role of TBR1 as a TF. Overrepresented *Molecular Functions* included TF activity, TF binding, DNA/chromatin binding, RNA binding, helicase activity and DNA-dependent ATPase activity (Table 5.2). Similarly, overrepresented *Biological Processes* included transcription, histone modification and chromatin organisation, mRNA splicing/processing and RNA export from the nucleus (Table 5.3). Proteins involved in negative regulation of transcription were overrepresented, whereas those involved in positive regulation of transcription were not. This suggests that, in the sample investigated in the current chapter, TBR1 interactors may play a larger role in transcriptional repression rather than activation, even though previous mouse brain expression studies and luciferase assays have demonstrated both up- and down-regulation of TBR1 target genes (Bedogni et al., 2010; McKenna et al., 2011; Chuang et al., 2014; Notwell et al., 2016). Other overrepresented biological processes related to DNA replication/repair, cell cycle regulation, and ubiquitination.

Table 5.2. Overrepresented terms in GO Molecular Function.

	HEK	Int	Exp	Enr	P value
transcription factor activity, core RNA polymerase binding	7	4	0.2	27.0	4.5E-02
› transcription factor activity, protein binding	514	30	10.9	2.8	1.7E-03
DNA-dependent ATPase activity	78	13	1.7	7.9	4.8E-05
› ATPase activity, coupled	276	19	5.8	3.3	2.3E-02
›› ATPase activity	339	22	7.2	3.1	1.1E-02
helicase activity	134	15	2.8	5.3	6.6E-04
chromatin binding	421	28	8.9	3.1	3.1E-04
transcription factor binding	475	26	10.1	2.6	3.0E-02
transcription factor activity, transcription factor binding	506	27	10.7	2.5	3.1E-02
double-stranded DNA binding	629	33	13.3	2.5	4.5E-03
› DNA binding	1912	87	40.5	2.2	1.7E-09
›› nucleic acid binding	3232	143	68.4	2.1	1.9E-19
RNA binding	1465	74	31.0	2.4	1.3E-09
sequence-specific DNA binding	756	35	16.0	2.2	3.2E-02

GO terms with ≥ 2 -fold enrichment and Bonferroni-adjusted p-value < 0.05 are shown. Terms are grouped in alternating white and grey bands, where terms in the same band form a hierarchical relationship in the ontology tree: the most specific child term is shown at the top and broader parent terms are listed below, with the › symbol indicating their relative level in the ontology. HEK = number of HEK293-expressed genes annotated with the GO term, Int = number of putative TBR1 interactors annotated with the GO term. Exp = expected number of genes with that term in the TBR1 interactome. Enr = fold enrichment of genes with that term in the TBR1 interactome (actual/expected).

Table 5.3. Overrepresented terms in GO Biological Process.

	HEK	Int	Exp	Enr	P value
tRNA transcription from RNA polymerase III promoter	6	5	0.1	39.4	1.7E-03
›› transcription, DNA-templated	2055	102	43.5	2.4	2.9E-14
››› nucleic acid-templated transcription	2056	102	43.5	2.3	3.0E-14
›››› RNA biosynthetic process	2065	102	43.7	2.3	4.1E-14
››››› RNA metabolic process	2804	128	59.3	2.2	6.2E-17
›››››› nucleic acid metabolic process	3253	149	68.8	2.2	4.3E-22
›››››› nucleobase-containing compound biosynthetic process	2337	111	49.5	2.2	9.7E-15
››››››› organic cyclic compound biosynthetic process	2472	112	52.3	2.1	2.2E-13
››››››› heterocycle biosynthetic process	2389	112	50.6	2.2	1.6E-14
››››››› aromatic compound biosynthetic process	2384	111	50.4	2.2	4.4E-14
››› gene expression	3071	141	65.0	2.2	3.5E-20
› tRNA transcription	7	5	0.2	33.8	3.6E-03
›› ncRNA transcription	90	14	1.9	7.4	8.7E-05
››› ncRNA metabolic process	517	28	10.9	2.6	4.5E-02
5S class rRNA transcription from RNA polymerase III type 1 promoter	6	5	0.1	39.4	1.7E-03
nuclear pore complex assembly	10	5	0.2	23.6	2.0E-02
› nuclear pore organisation	15	6	0.3	18.9	7.3E-03
protein K11-linked ubiquitination	27	9	0.6	15.8	6.7E-05
regulation of DNA-dependent DNA replication	44	8	0.9	8.6	4.0E-02
mRNA export from nucleus	99	15	2.1	7.2	3.7E-05
› RNA export from nucleus	122	15	2.6	5.8	5.5E-04
›› nuclear export	147	16	3.1	5.1	1.0E-03
››› nucleocytoplasmic transport	248	20	5.3	3.8	3.3E-03
›››› nuclear transport	253	20	5.4	3.7	4.5E-03
›››› RNA transport	170	19	3.6	5.3	4.5E-05
››››› establishment of RNA localisation	172	19	3.6	5.2	5.4E-05
››››› RNA localisation	191	21	4.0	5.2	9.3E-06
››››› nucleic acid transport	170	19	3.6	5.3	4.5E-05
›››››› nucleobase-containing compound transport	198	19	4.2	4.5	4.8E-04
› mRNA transport	137	19	2.9	6.6	1.4E-06
› mRNA-containing ribonucleoprotein complex export from nucleus	99	15	2.1	7.2	3.7E-05
›› ribonucleoprotein complex export from nucleus	117	15	2.5	6.1	3.2E-04
››› protein export from nucleus	137	15	2.9	5.2	2.4E-03
››› ribonucleoprotein complex localisation	119	15	2.5	6.0	4.0E-04
anaphase-promoting complex-dependent catabolic process	75	11	1.6	6.9	5.7E-03
regulation of cellular response to heat	75	11	1.6	6.9	5.7E-03

	HEK	Int	Exp	Enr	P value
regulation of gene silencing by miRNA	75	11	1.6	6.9	5.7E-03
› regulation of gene silencing by RNA	78	11	1.7	6.7	8.3E-03
regulation of chromosome segregation	86	12	1.8	6.6	3.0E-03
sister chromatid cohesion	111	15	2.4	6.4	1.6E-04
› cell cycle process	893	56	18.9	3.0	1.3E-09
› › cell cycle	1172	65	24.8	2.6	2.5E-09
› chromosome organisation	865	63	18.3	3.4	2.3E-14
› sister chromatid segregation	171	19	3.6	5.3	4.9E-05
› › nuclear chromosome segregation	212	19	4.5	4.2	1.4E-03
› › › chromosome segregation	253	20	5.4	3.7	4.5E-03
DNA-templated transcription, termination	107	13	2.3	5.7	4.7E-03
regulation of nuclear division	144	14	3.1	4.6	2.3E-02
› regulation of organelle organisation	975	44	20.6	2.1	1.1E-02
positive regulation of protein ubiquitination	161	15	3.4	4.4	1.7E-02
peptidyl-lysine modification	268	22	5.7	3.9	6.4E-04
regulation of chromosome organisation	274	22	5.8	3.8	9.4E-04
mRNA splicing, via spliceosome	272	21	5.8	3.7	3.4E-03
› mRNA processing	414	29	8.8	3.3	1.7E-04
› › RNA processing	815	50	17.2	2.9	7.5E-08
› › mRNA metabolic process	612	34	13.0	2.6	2.3E-03
› RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	272	21	5.8	3.7	3.4E-03
› › RNA splicing, via transesterification reactions	275	21	5.8	3.6	4.1E-03
› › › RNA splicing	368	33	7.8	4.2	3.1E-08
mitotic cell cycle process	608	42	12.9	3.3	1.2E-07
› mitotic cell cycle	657	45	13.9	3.2	2.7E-08
histone modification	311	21	6.6	3.2	2.8E-02
› covalent chromatin modification	412	29	8.7	3.3	1.5E-04
› › chromatin organisation	545	36	11.5	3.1	1.3E-05
cell division	446	29	9.4	3.1	7.9E-04
regulation of mitotic cell cycle phase transition	372	24	7.9	3.1	1.2E-02
› regulation of cell cycle phase transition	389	26	8.2	3.2	2.2E-03
› › regulation of cell cycle process	577	32	12.2	2.6	5.6E-03
› regulation of mitotic cell cycle	571	32	12.1	2.7	4.5E-03
DNA repair	450	29	9.5	3.1	9.5E-04
› DNA metabolic process	675	40	14.3	2.8	3.2E-05
› cellular response to DNA damage stimulus	669	33	14.2	2.3	4.5E-02
transcription from RNA polymerase II promoter	639	35	13.5	2.6	2.1E-03
negative regulation of transcription from RNA polymerase II promoter	655	34	13.9	2.5	1.1E-02
› negative regulation of transcription, DNA-templated	893	42	18.9	2.2	7.0E-03
› › negative regulation of cellular macromolecule biosynthetic process	1073	49	22.7	2.2	1.8E-03

	HEK	Int	Exp	Enr	P value
>>> negative regulation of cellular biosynthetic process	1154	49	24.4	2.0	1.4E-02
>>> negative regulation of macromolecule biosynthetic process	1115	49	23.6	2.1	5.4E-03
>> negative regulation of nucleic acid-templated transcription	923	42	19.5	2.2	1.6E-02
>>>> negative regulation of RNA metabolic process	993	44	21.0	2.1	1.8E-02
>>>> negative regulation of nucleobase-containing compound metabolic process	1106	51	23.4	2.2	7.0E-04
symbiosis, encompassing mutualism through parasitism	609	31	12.9	2.4	4.7E-02
positive regulation of nucleobase-containing compound metabolic process	1322	57	28.0	2.0	1.0E-03

See Table 5.2 legend.

Most overrepresented *Cellular Component* terms related to the nucleus and in particular the chromatin/chromosomes (Table 5.4). In contrast, cell membrane and endosome proteins were significantly underrepresented (Table 5.5). It should be noted that the inclusion of a nuclear extraction step in the current study is expected to select for nuclear proteins. Nonetheless, the overrepresented terms also reflect the typical nuclear localisation and DNA-binding TF activity of TBR1. Specific nuclear complexes were also overrepresented, including RNA polymerase complexes, DNA replication complexes, the nuclear pore, anaphase-promoting complex and multiple chromatin-remodelling complexes (SWI/SNF superfamily, histone methyltransferases).

Table 5.4. Overrepresented terms in GO Cellular Component.

	HEK	Int	Exp	Enr	P value
transcription factor TFIIC complex	6	5	0.1	39.4	2.9E-04
› RNA polymerase III transcription factor complex	7	5	0.2	33.8	6.1E-04
››› nuclear part	3684	188	78.0	2.4	2.0E-42
DNA replication factor C complex	6	4	0.1	31.5	1.2E-02
›› chromosomal part	718	55	15.2	3.6	1.0E-13
››› chromosome	820	60	17.4	3.5	2.4E-14
nuclear pore outer ring	10	6	0.2	28.4	1.2E-04
› nuclear pore	72	13	1.5	8.5	8.8E-06
anaphase-promoting complex	22	10	0.5	21.5	8.8E-08
››› transferase complex	688	44	14.6	3.0	7.5E-08
›››› catalytic complex	1214	63	25.7	2.5	1.9E-08
› nuclear ubiquitin ligase complex	39	13	0.8	15.8	5.6E-09
SWI/SNF superfamily-type complex	67	11	1.4	7.8	3.3E-04
› nuclear chromosome part	423	33	9.0	3.7	1.9E-07
›› nuclear chromosome	454	33	9.6	3.4	1.1E-06
››› nuclear lumen	3416	178	72.3	2.5	1.3E-39
›››› intracellular organelle lumen	4182	186	88.5	2.1	2.1E-32
› ATPase complex	83	11	1.8	6.3	2.6E-03
› nuclear chromatin	269	24	5.7	4.2	5.8E-06
›› chromatin	411	32	8.7	3.7	3.9E-07
host cell	61	9	1.3	7.0	9.4E-03
histone methyltransferase complex	64	9	1.4	6.7	1.4E-02
› nucleoplasm part	970	62	20.5	3.0	3.4E-12
›› nucleoplasm	2944	165	62.3	2.7	6.8E-39
nuclear matrix	97	13	2.1	6.3	2.7E-04
› nuclear periphery	123	19	2.6	7.3	3.9E-08
inclusion body	69	9	1.5	6.2	2.5E-02
nuclear chromosome, telomeric region	101	11	2.1	5.2	1.6E-02
›› chromosomal region	300	23	6.4	3.6	1.8E-04
DNA-directed RNA polymerase complex	127	12	2.7	4.5	2.6E-02
› RNA polymerase complex	129	12	2.7	4.4	3.0E-02
nuclear body	645	33	13.7	2.4	3.6E-03

See Table 5.2 legend.

Table 5.5. Underrepresented terms in GO Cellular Component.

	HEK	Int	Exp	Enr	P value
organelle subcompartment	1213	7	25.7	0.3	8.0E-03
integral component of membrane	2807	12	59.4	0.2	4.6E-13
› intrinsic component of membrane	2859	12	60.5	< 0.2	1.5E-13
›› membrane part	3686	28	78.0	0.4	1.1E-10
endosome	655	1	13.9	< 0.2	1.2E-02

All GO terms with Bonferroni-adjusted p-value < 0.05 are shown. For other details, see Table 5.2 legend.

5.3.3 CLUSTER ANALYSIS IDENTIFIES KNOWN TRANSCRIPTIONAL REGULATION COMPLEXES

The 248-protein TBR1 interactome was cross-referenced with the STRING database, to find known interactions within the network. There were 604 interactions amongst the 248 proteins, and each protein interacted with an average of 4.871 other proteins (Fig 5.2). The most connected hubs were the RNA polymerase proteins POL2RA (34 interactions) and POLR2B (31 interactions), and the nuclear cap-binding protein NCBP1 (31 interactions). Other network statistics are summarised in Table 5.6. The interactome was dominated by one large subnetwork of 154 proteins (Component 1), followed by two 3-protein clusters, five 2-protein clusters and 78 isolated proteins with established interactions.

Table 5.6. Network statistics

	Complete network	Component 1
Number of nodes (proteins)	248	154
Number of edges (interactions)	604	594
Average number of neighbours	4.871	7.714
Network density	0.020	0.050
Number of isolated nodes	78	0
Network centralisation	0.119	0.174
Network heterogeneity	1.322	0.872
Network clustering coefficient	0.336	0.521

Within Component 1, a cluster analysis was performed using the MCODE algorithm (Bader and Hogue, 2003) (Version 1.4.1) to identify highly interconnected regions, which tend to correspond to protein complexes or parts of pathways. Ten clusters were identified, which comprised several known protein complexes and accounted well for the overrepresented terms identified in the gene ontology analysis (Fig 5.2; Table 5.7). The largest group (Cluster 1a) comprised a set of proteins involved in (m)RNA transcription, processing and export, including multiples members of the RNA polymerase II complex, the spliceosome, various cleavage and polyadenylation factors and the nuclear pore complex. Other RNA-related complexes included the PELP1 rRNA processing complex (Cluster 3), the TF IIIC complex (Cluster 4) and the Integrator snRNA transcription complex (Cluster 6). Various transcriptional co-regulators were also identified in the cluster analysis, including several Polycomb group proteins (Cluster 1b), and members of the MLL histone methyltransferase complex (Cluster 3), the NuA4 histone acetyltransferase complex (Cluster 5) and others (Clusters 8, 9). Cluster 2 is dominated by DNA replication and repair complexes, including the replication factor C complex and the MRN complex (MRE11A-RAD50-NBN). The cluster analysis also highlighted several complexes involved in the cell cycle, including the anaphase-promoting complex (Cluster 1b), various chromosome segregation factors (Cluster 6) and cytoskeletal components involved in mitosis, such as the gamma-tubulin complex (Cluster 10). Interestingly, the anaphase-promoting complex also has E3-ubiquitin ligase activity (Peters, 2006), as do several other members of Cluster 1b.

This might suggest that TBR1 is targeted for ubiquitination, or that it plays a role in the ubiquitination of other proteins.

Combined with the gene ontology analysis above, these results suggest that TBR1 interacts with multiple proteins to regulate transcription of protein-coding genes as well as non-coding RNAs, and that this TF activity is mediated by interactions with co-regulators and chromatin modifying complexes, as might be expected. However, these results also suggest additional roles in DNA replication/repair, cell cycle regulation and ubiquitination.

Outside of Component 1 were several smaller subnetworks with diverse functions, including actin-associated signalling proteins (NCKAP1-CYFIP2-MYO9B), two CASK interactors (LIN7C and EPB41L2), ankyrins (ANK2-ANK3), mediator complex members (MED12-MED23), collagen-modifying proteins (PLOD1-COLGALT1) and mitochondrial import proteins (PMPCA-PMPCB).

Figure 5.2 (opposite page). Graphical network depiction of the putative TBR1 interactome.

Nodes represent proteins, connectors represent known interactions imported from the STRING database. The network comprises a large connected component of 154 proteins (Component 1; top), as well as 2 trios, 5 pairs and 78 isolated proteins (bottom). Highly interconnected clusters of proteins within Component 1, as identified by the MCODE algorithm, are grouped together and colour coded. Proteins encoded by ASD/ID-related genes are marked with a red border.

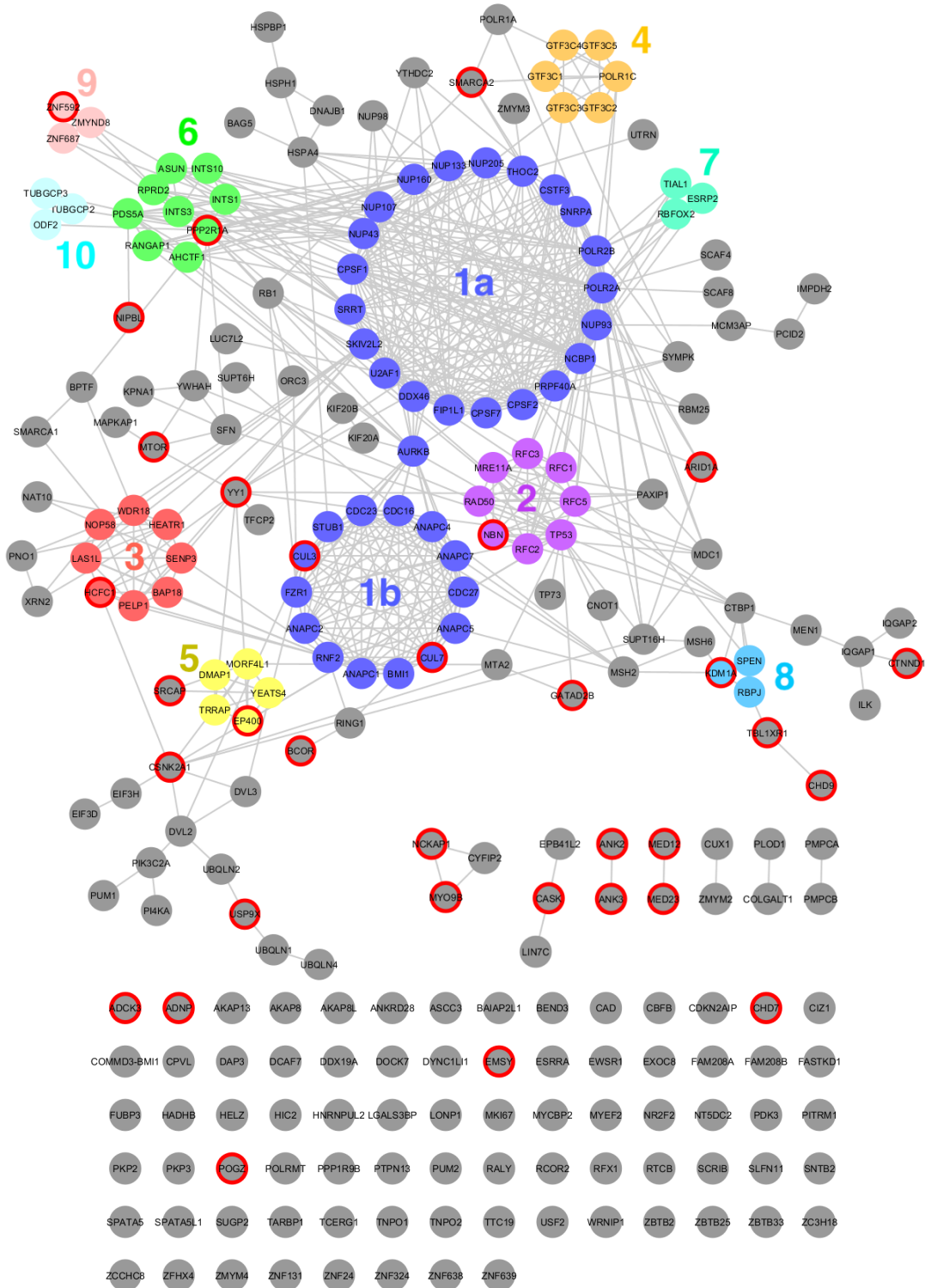


Table 5.7. Clusters of interconnected proteins in the TBR1 interactome.

Cluster	Major Functions	Proteins and Complexes
1a	mRNA transcription/processing/export	Nuclear pore complex: NUP107, NUP133, NUP160, NUP205, NUP43, NUP93 RNA polymerase II: POLR2A, POLR2B Cleavage and polyadenylation factor(s): CPSF1, CPSF2, CPSF7, CSTF3, FIP1L1 Spliceosome: NCBP1, PRPF40A, DDX46, SRRT, THOC2, SNRPA, SKIV2L2, U2AF1
1b	ubiquitination	Anaphase-promoting complex: ANAPC1, ANAPC2, ANAPC4, ANAPC5, ANAPC7, CDC16, CDC23, CDC27, FZR1 (cofactor of APC/C) Other ubiquitin E3-ligase complexes/cofactors: BMI1, RNF2, CUL3, CUL7, STUB1 Other: AURKB
2	DNA replication and repair	MRN complex: MRE11A, RAD50, NBN Replication factor C complex: RFC1, TFC2, RFC3, RFC5 Other: TP53
3	rRNA processing	5FMC complex: LAS1L, PELP1, SENP3, WDR18 MLL complex: HCFC1, BAP18, LAS1L, PELP1 Other: HEATR1 (nucleolar processing of pre-18S ribosomal RNA) NOP58 (nucleolar processing of 60S ribosomal subunit)
4	rRNA/tRNA transcription	Transcription factor IIIC complex: GTF3C1, GTF3C2, GTF3C3, GTF3C4, GTF3C5 Other: POLR1C
5	histone acetylation	Nu4a histone acetyltransferase complex: DMAP1, EP400, MORF4L1, TRRAP, YEATS4
6	snRNA transcription separation of sister chromatids	Integrator complex: INTS1, INTS3, INTS10, ASUN Other snRNA transcription cofactors: RPRD2 Separation of sister chromatids: AHCTF1, PPP2R1A, PDS5A, RANGAP1
7	RNA-binding	ESRP2, RBFOX2, TIAL1
8	regulation of transcription	KDM1A, RBPJ, SPEN
9	regulation of transcription	ZMYND8, ZNF592, ZNF687
10	cytoskeleton, mitosis	Gamma-tubulin complex: TUBGCP2, TUBGCP3 Other: ODF2

Clusters were identified within Component 1, using the MCODE (Molecular Complex Detection) algorithm (version 1.4.1). The major functions of each cluster are shown, alongside the proteins and known complexes that make up each cluster.

5.3.4 EPIGENETIC FACTORS ARE OVERREPRESENTED IN THE TBR1 INTERACTOME

Gene ontology and cluster analyses both pointed to the involvement of many of the putative TBR1 interactors in transcriptional regulation and epigenetic processes. This suggested two specific protein classes that might also be well-represented in the TBR1 interactome: a) other TFs and b) epigenetic factors, including chromatin remodelling factors and histone modifiers.

Epigenetic factors were found using the EpiFactors database (a manually curated database of epigenetic regulators, complexes and targets) (Medvedeva et al., 2015), identifying 52 proteins involved in histone modification and chromatin remodelling. These were significantly overrepresented in the TBR1 interactome, compared to a background list of HEK293-expressed genes (Table 5.8; Fisher's exact, $p = 7.7 \times 10^{-18}$). These proteins were involved in a range of processes with both positive and negative effects on transcription, including histone (de)methylation, (de)acetylation, ubiquitination, deSUMOylation and nucleosome restructuring (Table 5.9). The diverse interactions may help to explain the mixed activator/repressor functions that have been observed for TBR1 (Bedogni et al., 2010; McKenna et al., 2011; Chuang et al., 2014; Notwell et al., 2016), even though only negative regulation of transcription was significantly overrepresented in the gene ontology analysis. The epigenetic factors identified include multiple members of Nu4A, MLL, SWI/SNF and Polycomb complexes, as noted above, as well as the NuRD histone deacetylation/chromatin remodelling complex.

Table 5.8. Epigenetic factors are significantly overrepresented in the putative TBR1 interactome

		TBR1 interactome	
		+	-
Epigenetic factors	+	52	599
	-	195	11249

Fisher's exact test, two-tailed, $p = 7.7 \times 10^{-18}$

Table 5.9. Epigenetic factors identified in the TBR1 interactome

Gene	emPAI	Mascot	Function	Complex group	Specific complexes
ADNP	0.22	107	chromatin remod. cofactor	other	
ARID1A	0.075	112	chromatin remod. cofactor	SWI/SNF	BAF, nBAF, npBAF, PBAF, SWI/SNF_Brg1, SWI/SNF_Brm, WINAC, bBAF
AURKB	0.7	56.5	histone phosphorylation	other	
*C17orf49	0.75	117.5	histone methylation read	MLL	CHD8, MLL2/3, MLL4/WBP7
BCOR	0.225	427.5	Polycomb group protein	PcG/PcG-like	BCOR
BMI1	0.185	81.5	Polycomb group protein	PcG/PcG-like	PRC1
BPTF	0.28	1109.5	chromatin remodelling	ISWI	NuRF
CHD7	0.195	662.5	chromatin remodelling	other	
CHD9	0.06	199	chromatin remodelling	other	
CSNK2A1	0.275	134	histone modification	PcG/PcG-like	RING2-FBRS
CTBP1	0.495	147.5	chromatin remodelling	CoREST	LSD-CoREST
CUL3	0.13	145.5	histone ubiquitination	other	
DMAP1	0.315	164	chromatin remodelling	HAT; SRCAP	NuA4, NuA4-related, SRCAP
*C11orf30	0.225	335.5	histone methyl. cofactor	other	
EP400	0.07	295.5	chromatin remodelling, histone acetylation	HAT; SWR	SWR, NuA4, NuA4-related
GATAD2B	0.395	249.5	histone modification read	HDAC	NuRD
GTF3C4	0.365	292	histone acetylation	other	
HCFC1	0.11	265	chromatin remodelling	HAT; MLL; HMT	NSL, COMPASS, MLL-HCF, CHD8, MLL2/3, COMPASS-like MLL1,2, MLL4/WBP7
KDM1A	0.54	447.5	histone demethylation	CoREST; HDAC	NuRD, BHC, SCL
LAS1L	0.265	212	histone methyl. cofactor histone acetyl. cofactor	MLL	CHD8, MLL2/3, MLL4/WBP7

Gene	emPAI	Mascot	Function	Complex group	Specific complexes
MDC1	0.05	144	histone modification read	other	
MEN1	0.11	84	histone methyl. cofactor	MLL; HMT	Menin-associated_HMT, MLL-HCF, CHD8, MLL2/3, COMPASS-like MLL1,2, MLL4/WBP7
MORF4L1	0.165	57.5	histone modification read	HAT	NuA4
MSH6	0.15	205	histone modification read	other	
MTA2	0.41	305.5	histone deacetyl. cofactor	HDAC	NuRD
NAT10	0.03	54	histone acetylation	other	
NBN	0.08	86	chromatin remodelling	other	
NIPBL	0.045	193.5	histone deacetyl. cofactor	other	
PAXIP1	0.08	114	histone methyl. cofactor	MLL	CHD8, MLL2/3, MLL4/WBP7, COMPASS-like MLL3,4
PELP1	0.14	163	histone methylation, histone acetyl. cofactor	MLL	CHD8, MLL2/3, MLL4/WBP7
POGZ	0.16	245	histone methylation read	other	
RBI	0.045	57.5	chromatin remodelling, histone ubiquitination	CREST-BRG1; L3MBTL1	CREST-BRG1, L3MBTL1
RING1	0.515	177	histone ubiquitination, Polycomb group protein	PcG/PcG-like	PRC1, BCOR, RING2-L3MBTL2, RING2-FBRS
RNF2	0.46	151	histone ubiquitination	PcG/PcG-like; MLL	PRC1, BCOR, RING2-L3MBTL2, RING2-FBRS, CHD8, MLL2/3, MLL4/WBP7
SEN3	0.28	203	histone deSUMOylation, histone acetyl. cofactor	MLL	CHD8, MLL2/3, MLL4/WBP7
SMARCA1	0.4	423.5	chromatin remodelling, histone deacetylation	SWI/SNF; ISWI	NuRF, CERF

Gene	emPAI	Mascot	Function	Complex group	Specific complexes
SMARCA2	0.205	392.5	histone modification read	SWI/SNF	BAF, nBAF, npBAF, WINAC, bBAF, SWI/SNF BRM-BRG1
SPEN	0.095	416.5	histone deacetyl. cofactor	other	
SRCAP	0.065	262	chromatin remodelling, histone deacetylation	HAT; SRCAP	NuA4-related, SRCAP
SUPT16H	0.265	332	histone modification read	SWI/SNF; FACT	WINAC, FACT
SUPT6H	0.11	258.5	histone demethyl. cofactor	other	
TBL1XR1	0.27	73		other	
TP53	1.38	434.5	histone acetyl. cofactor	other	
TRRAP	0.32	1269.5	histone acetyl. cofactor	HAT; SWR	SWR, PCAF, TFIIIC-HAT, NuA4, SAGA, NuA4-related, STAGA
YEATS4	0.31	65.5	histone acetyl. cofactor	HAT; SRCAP	NuA4, NuA4-related, SRCAP
YY1	0.38	78.5	chromatin remod. cofactor	Ino80	Ino80
ZBTB33	0.165	140	histone acetyl. cofactor, histone dimethyl. cofactor	other	
ZMYM2	0.77	999.5	histone deacetyl. cofactor	CoREST; HDAC	BHC, LSD-CoREST
ZMYM3	0.1	164	histone deacetyl. cofactor	HDAC	BHC
ZMYND8	0.155	253	histone deacetyl. cofactor	other	
ZNF592	0.065	113.5	histone deacetyl. cofactor	other	
ZNF687	0.145	219.5	histone deacetyl. cofactor	other	

*Alternative gene symbols: C17orf49 a.k.a. BAP18, C11orf30 a.k.a. EMSY. The emPAI (exponentially modified protein abundance index) and Mascot score are shown as the mean of two AP experiments. Epigenetic factor function and complex membership is from the EpiFactors Database (Medvedeva et al., 2015).

The TBR1 interactome was also cross-referenced with a published list of human TFs (Vaquerizas et al., 2009). There were 26 confirmed or probable TFs amongst the TBR1-interacting proteins (excluding TBR1 itself, Table 5.10). However, this did not constitute a significant overrepresentation of TFs (Table 5.11; Fisher's exact test, $p = 0.31$). Note that six proteins (ADNP, TP53, YY1, ZBTB33, ZNF592, ZNF687) appeared in both the epigenetic factors and TF lists.

Table 5.10. Transcription factors identified in the TBR1 interactome.

Gene	emPAI	Mascot
ADNP	0.22	107
AHCTF1	0.2	525
CBFB	0.36	41.5
CIZ1	0.07	79.5
CUX1	0.1	214
ESRRA	0.285	95
HIC2	0.155	124
NR2F2	0.15	73
RBPJ	0.145	53.5
RCOR2	0.185	150.5
RFX1	0.34	376
TFCP2	0.34	195
TP53	1.38	434.5
TP73	0.16	96.5
USF2	0.375	126
YY1	0.38	78.5
ZBTB2	0.25	161.5
ZBTB25	0.14	83
ZBTB33	0.165	140
ZFHX4	0.145	619
ZNF131	0.365	261
ZNF24	0.215	77
ZNF324	0.09	52
ZNF592	0.065	113.5
ZNF639	0.235	79.5
ZNF687	0.145	219.5

The emPAI (exponentially modified protein abundance index) and Mascot score are shown as the mean of two AP experiments.

Table 5.11. Transcription factors are not significantly overrepresented in the putative TBR1 interactome

	TBR1 interactome		
	+	-	
Transcription factors	+	26	1039
	-	221	10809

Fisher's exact test, two-tailed, $p = 0.31$

5.3.5 TBR1 INTERACTOME IS ENRICHED FOR ASD/ID-RELATED GENES

TBR1 variants have been identified in patients with ASD and/or ID (Neale et al., 2012; O’Roak et al., 2012b, 2012a; De Rubeis et al., 2014; Iossifov et al., 2014; Palumbo et al., 2014) and *TBR1* might therefore be expected to interact with other proteins that are disrupted in these disorders. The *TBR1* interactome was tested for overrepresentation of ASD/ID-related proteins, using HEK293-expressed genes as a reference list. There were 11 proteins in the *TBR1* interactome that are encoded by ASD candidate genes. This was a significant overrepresentation, compared to a background list of HEK293-expressed genes (Table 5.12; Fisher’s exact test, $p = 6.1 \times 10^{-5}$). There were also 23 ID-related genes in the *TBR1* interactome, which also constituted a significant overrepresentation (Table 5.13; Fisher’s exact test, $p = 0.0094$). There was some overlap between the two lists: *ADNP*, *MTOR*, *POGZ* and *TBLIXR1* have been implicated in ID syndromes in addition to ASD. A complete list of ID/ASD-related proteins in the putative *TBR1* interactome ($n = 32$) can be found in Table 5.14. Notably, many are implicated in disorders that also include speech/language delay or impairment, or vocal abnormalities such as high-pitched/hypernasal speech or a distinctive cry during infancy (Table 5.14).

Table 5.12. Proteins encoded by ASD candidate genes are significantly overrepresented in the putative *TBR1* interactome

		TBR1 interactome	
		+	-
Autism spectrum disorders	+	11	117
	-	236	11731

Fisher’s exact test, two-tailed, $p = 6.1 \times 10^{-5}$

Table 5.13. Proteins encoded by ID-related genes are significantly overrepresented in the putative *TBR1* interactome

		TBR1 interactome	
		+	-
Intellectual disability	+	23	625
	-	224	11223

Fisher’s exact test, two-tailed, $p = 0.0094$

Table 5.14. ID and ASD-related proteins in the TBR1 interactome.

Gene	emPAI	Mascot	SFARI	ID disorder (OMIM #)	Speech/language
ADCK3	0.095	85		Coenzyme Q10 deficiency, primary 4 (612016)	dysarthric speech ¹
ADNP	0.22	107	1S	Mental retardation, autosomal dominant 28 (615873)	language impairment
ANK2	0.05	74	1		-
ANK3	0.04	260		Mental retardation, autosomal recessive 37 (615493)	speech delay ²
ARID1A	0.075	112		Coffin-Siris syndrome 2 (614607)	speech delay
BCOR	0.225	427.5		Microphthalmia, syndromic 2 (300166)	-
CASK	0.2	192.5		Mental retardation and microcephaly with pontine and cerebellar hypoplasia (300749) FG syndrome 4 / Mental retardation, with or without nystagmus (300422)	speech absent
CHD7	0.195	662.5		CHARGE syndrome (214800)	speech delay ³
CSNK2A1	0.275	134		Okur-Chung neurodevelopmental syndrome (617062)	speech delay/absent
CTNND1	0.26	298		Autism with intellectual disability (O'Roak et al., 2012b; no OMIM number)	-
CUL3	0.13	145.5	2		-
CUL7	0.05	107	3		-
C11orf30 (EMSY)	0.225	335.5	3		-
EP400	0.07	295.5	3		-
GATAD2B	0.395	249.5		Mental retardation, autosomal dominant 18 (615074)	poor speech
HCFC1	0.11	265		Mental retardation, X-linked 3 (309541)	speech delay ⁴
KDM1A	0.54	447.5		Cleft palate, psychomotor retardation and distinctive facial features (616728)	speech delay
MED12	0.065	163.5		Lujan-Fryns syndrome (309520) Ohdo syndrome, X-linked (300895) Opitz-Kaveggia syndrome (305450)	hypernasal speech - high-pitched voice
MED23	0.06	102.5		Mental retardation, autosomal recessive 18 (614249)	-

Gene	emPAI	Mascot	SFARI	ID disorder (OMIM #)	Speech/language
MTOR	0.155	458.5	3	Smith-Kingsmore syndrome (616638)	speech delay ⁵
MYO9B	0.13	324.5	3		-
NBN	0.08	86		Nijmegen breakage syndrome (251260)	-
NCKAP1	0.385	496	3		-
NIPBL	0.045	193.5		Cornelia de Lange syndrome 1 (122470)	language delay, low-pitched, growling cry in infancy
POGZ	0.16	245	1S	White-Sutton syndrome (616364)	speech/ language delay ⁶
PPP2R1A	0.17	123		Mental retardation, autosomal dominant 36 (616362)	speech absent/ impaired
SMARCA2	0.205	392.5		Nicolaidis-Baraitser syndrome (601358)	speech absent/ impaired
SRCAP	0.065	262		Floating-Harbor syndrome (136140)	expressive language delay
TBL1XR1	0.27	73	3	Mental retardation, autosomal dominant 41 (616944) Pierpont syndrome (602342)	speech/ language delay
USP9X	0.065	188.5		Mental retardation, X-linked 99 (300919)	-
YY1	0.38	78.5		Gabriele-de Vries syndrome (617557)	speech delay

The emPAI (exponentially modified protein abundance index) and Mascot score are shown as the mean of two AP experiments. Involvement in autism spectrum disorder (ASD) is based on the SFARI Gene database and categorised according to candidate confidence: 1 = High Confidence, 2 = Strong Candidate, 3 = Suggestive Evidence, S = syndromic form. Intellectual disability (ID) disorders are taken from the Radboud University Human Genetics Department sequencing panel (DG2.5x), supported by OMIM numbers where available. Clinical features related to speech/language are taken from OMIM clinical synopses, or from the numbered references: ¹ Blumkin et al. (2014), Liu et al. (2014); ² Iqbal et al. (2013); ³ Zentner et al. (2010); ⁴ Jolly et al. (2015); ⁵ Baynam et al. (2015); ⁶ Stessman et al. (2016).

The overrepresentation of ASD/ID-related proteins in the TBR1 interactome could have been a secondary consequence of a general enrichment of neurodevelopmental proteins. To investigate the specificity of the ASD/ID results, the overrepresentation test was repeated with a set of candidate genes for a third neurodevelopmental disorder: schizophrenia. From a set of 662 schizophrenia candidates, 13 were identified in the putative TBR1 interactome, but this did not represent a significant overrepresentation compared to the background list of HEK293-expressed proteins (Table 5.15; Fisher's exact test, $p = 0.16$). This suggests that the TBR1 interactome is specifically enriched for ASD/ID-related proteins, rather than neurodevelopmental proteins in general, although it should be noted that the inclusion criteria for these lists of ASD, ID and schizophrenia genes were not identical (see Methods), which could introduce bias when comparing overrepresentation for the three disorders.

The overrepresentation of ASD/ID proteins could be partly driven by the preponderance of chromatin modifying proteins, which are enriched among genes implicate in ASD (De Rubeis et al., 2014) and ID (Iwase et al., 2017). Indeed, over half of the ID/ASD-related proteins (18/31) were also in the epigenetic factors list. Two of these (ADNP, YY1) were also classified as TFs, but none of the remaining TFs were associated with an ID/ASD syndrome. In general, the ID/ASD-related proteins are spread across the TBR1 interactome network, and do not appear to be particularly concentrated in any of the identified clusters (Fig. 5.2). However, some interactions do occur amongst the ID/ASD-related proteins that may hint at shared pathways, e.g. the SWI/SNF chromatin remodelling complex members ARID1A and SMARCA2 (ADNP can also act as a SWI/SNF cofactor) (Mandel and Gozes, 2007) and the ubiquitin ligase-associated cullin proteins (CUL3, CUL7).

Table 5.15. Proteins encoded by schizophrenia candidate genes are not significantly overrepresented in the putative TBR1 interactome

		TBR1 interactome	
		+	-
Schizophrenia	+	13	409
	-	234	11439

Fisher's exact test, two-tailed, $p = 0.16$

5.3.6 BRAIN EXPRESSION OF TBR1-INTERACTING PROTEINS

TBR1 expression is brain-specific and largely restricted to neuronal cells (Bulfone et al., 1995). As the AP-MS experiment was performed in HEK293 cells, it was important to investigate whether the proteins identified in the screen are also expressed at high levels in the brain, and particularly in neurons. Immunohistochemistry-based normal human tissue expression data were obtained from the Human Protein Atlas, including semi-quantitative assessments ("High, Medium, Low, Not Detected") of protein expression in multiple brain regions and cell populations: caudate (neuronal and glial cells), cerebellum (Purkinje cells,

and cells in the granular and molecular layers), cerebral cortex (neuronal, glial, endothelial and neuropil cells) and hippocampus (neuronal and glial cells) (Uhlén et al., 2015).

Reliable protein expression data were available for 184 out of 248 putative TBR1 interactors. Of these, 175 (95%) exhibited some ("Low", "Medium" or "High") expression in at least one cell population in the brain, suggesting that brain-expressed proteins were well detected in the AP-MS experiment, despite the use of HEK293 cells. By comparison, 7,414 of 12,095 HEK-expressed proteins had reliable brain expression data in the Human Protein Atlas, of which 6,720 (91%) exhibited some expression in at least one cell population. More than half (55%) of the reliably assayed TBR1-interaction candidates were expressed at a "High" level in at least one cell population, compared to 39% of background HEK-expressed proteins.

A greater percentage of putative TBR1-interactors were expressed at high levels in neurons than in non-neuronal cell populations (Fig 5.3). However, this trend was also true of HEK293-expressed proteins in general, suggesting that the higher level of neuronal detection may be due to higher overall expression levels in neurons or to differences in the method of quantification used for different cell types in the Human Protein Atlas, rather than a specific property of TBR1-interacting proteins.

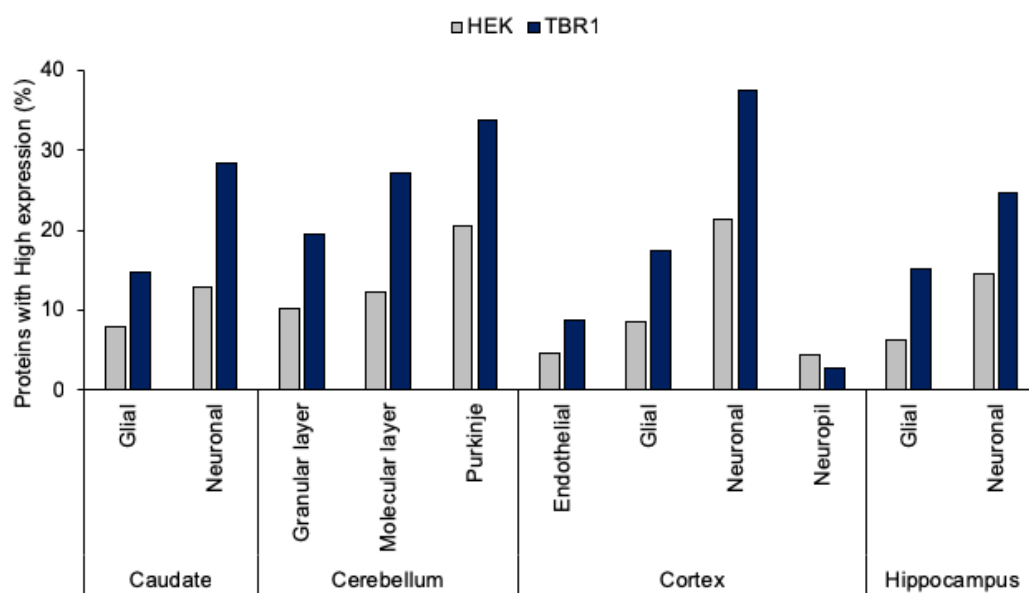


Figure 5.3. Proteins expression in different cell populations in the brain. Grey bars indicate the percentage of all HEK-expressed proteins detected at high levels of expression in each cell type. Blue bars indicate the percentage of putative TBR1-interactors expressed at high levels in the same cell types. Expression data from Human Protein Atlas, based on immunohistochemistry in normal human brain tissue (Uhlén et al., 2015).

5.3.7 SUBCELLULAR LOCALISATION OF TBR1-INTERACTING PROTEINS

The 2xFLAG/V5-TBR1 construct was localised to the nucleus (Fig 5.1B), in agreement with the typical TBR1 expression pattern reported in the literature (Deriziotis et al., 2014b). For this reason, a nuclear extraction step was included in the AP-MS protocol, to maximise the concentration and therefore optimise the detection of TBR1 and its interaction partners, which were also hypothesised to be predominantly nuclear.

To confirm this, the subcellular localisation of the putative TBR1 interactors was investigated using data from the Human Protein Atlas (Thul et al., 2017). Reliable subcellular localisation information was available for 222 of the 248 proteins in the putative TBR1 interactome. Of these, 107 proteins (48%) were expressed only within the nuclear compartment, while a further 74 proteins (33%) were expressed in the nuclear compartment in addition to another part of the cell. Therefore, as expected, the vast majority of TBR1 interaction candidates ($181/222 = 82\%$) were expressed at least partly in the nucleus.

On the other hand, 41 proteins (18%) were expressed only in extra-nuclear compartments, such as the cytosol, plasma membrane, vesicles or mitochondria (Table 5.16). These also include several cell cycle-related proteins, as identified above, which are likely to be expressed during mitotic phases when the nucleus does not form a distinct compartment. Given that the AP-MS procedure was performed using nuclear extracts, this number of non-nuclear proteins is perhaps surprising. One potential explanation is that these proteins may in fact exhibit some nuclear expression that has not been identified in the Human Protein Atlas data. They could also be translocated to the nucleus when co-expressed with TBR1, as is the case for CASK. Alternatively, these proteins might represent cytoplasmic material that was not completely excluded during the nuclear extraction. In the latter case, cytoplasmic proteins might only have the potential to interact with nuclear TBR1 when the normal structure of the cell is broken down.

Table 5.16. Non-nuclear proteins in the TBR1 interactome.

	Localisation	Function/process
ADCK3	mitochondria	kinase; ubiquinone biosynthesis
AKAP13	cytosol	guanine-nucleotide releasing factor
ANAPC1	vesicles	cell cycle, cell division, mitosis, ubiquitin conjugation pathway
ANAPC4	intermediate filaments	cell cycle, cell division, mitosis, ubiquitin conjugation pathway
ANK3	plasma membrane	-
ASCC3	Golgi apparatus, cytosol	helicase, hydrolase; DNA damage, DNA repair, transcription, transcription regulation
BAG5	vesicles	chaperone
BAIAP2L1	plasma membrane, cytosol	actin-binding
CHD9	plasma membrane, cytosol	chromatin regulator, DNA-binding, helicase, hydrolase; transcription, transcription regulation
CNOT1	cytosol	developmental protein, repressor; RNA-mediated gene silencing, transcription, transcription regulation, translation regulation
COLGALT1	vesicles	glycosyltransferase, transferase
CPVL	endoplasmic reticulum	carboxypeptidase, hydrolase, protease
CTNND1	plasma membrane	cell adhesion, transcription, transcription regulation, Wnt signalling pathway
CYFIP2	endoplasmic reticulum, plasma membrane, cytosol	apoptosis, cell adhesion
DVL3	intermediate filaments, midbody ring, centrosome	developmental protein; Wnt signalling protein
DYNC1LI1	centrosome, cytosol	motor protein; cell cycle, cell division, host-virus interaction, mitosis, transport
EIF3D	cytosol	initiation factor, RNA-binding; protein biosynthesis
EIF3H	cytosol	initiation factor; protein biosynthesis
EXOC8	cytosol	exocytosis, protein transport, transport
HADHB	mitochondria	acetyltransferase, transferase; fatty acid metabolism, lipid metabolism
HSPBP1	vesicles, cytosol	-
ILK	cell junctions, focal adhesion sites	kinase, serine/threonine-protein kinase, transferase
IMPDH2	cytosol, rods & rings	DNA-binding, oxidoreductase, RNA-binding; GMP biosynthesis, purine biosynthesis
IQGAP1	plasma membrane, cell junctions	calmodulin-binding
IQGAP2	vesicles, plasma membrane	calmodulin-binding
MTOR	vesicles, cytosol	kinase, serine/threonine-protein kinase, transferase
MYO9B	cytosol	actin-binding, calmodulin-binding, GTPase activation, motor protein, myosin
NCKAP1	cytosol	-

	Localisation	Function/process
ODF2	centrosome	developmental protein; differentiation, spermatogenesis
PITRM1	mitochondria	hydrolase, metalloprotease, protease
PMPCA	mitochondria	hydrolase, metalloprotease, protease
PMPCB	mitochondria	hydrolase, metalloprotease, protease
POLRMT	mitochondria	nucleotidyltransferase, transferase; transcription
PPP2R1A	cytosol	activator, DNA-binding, repressor; apoptosis, biological rhythms, cell cycle, host-virus interaction, necrosis, transcription, transcription regulation
PTPN13	plasma membrane, cytosol	hydrolase, protein phosphatase
PUM2	cytosol	RNA-binding; translation regulation
SFN	cytosol	-
SPATA5	cytosol	developmental protein; differentiation, spermatogenesis
TTC19	mitochondria	cell cycle, cell division, electron transport, respiratory chain, transport
UBQLN2	plasma membrane, cytosol	autophagy
UTRN	nucleoplasm, plasma membrane	actin-binding

Subcellular localisation annotations from Human Protein Atlas (Thul et al., 2017). Function/process keywords from UniProt (The UniProt Consortium, 2017). Note that several proteins not detected in the nucleus in the Human Protein Atlas are tagged with DNA-related UniProt keywords, suggesting that they may in fact occur in the nucleus under certain conditions.

5.4 DISCUSSION

In this study, mass spectrometry-based proteomics was used to characterise the TBR1 interactome. The AP-MS screen identified CASK, a known interactor of TBR1 in neurons, demonstrating that the HEK293 AP-MS model is capable of detecting physiologically relevant TBR1-interactors. Furthermore, 247 novel interaction candidates were identified. As expected for a TF, the putative TBR1 interactome was enriched for proteins involved in transcriptional regulation, chromatin modification and RNA processing, but also included DNA replication, cell cycle and ubiquitination-related proteins.

In keeping with the established role of TBR1 in ASD, autism-related gene products were significantly overrepresented in the TBR1 interactome. These included three high confidence candidate genes (*ADNP*, *ANK2*, *POGZ*) and eight other ASD candidates. Similarly, the TBR1 interactome was enriched for ID-related genes, reflecting previous findings of ID in patients with TBR1 mutations (Deriziotis et al., 2014b; Hamdan et al., 2014). Several proteins were associated with both ASD and ID. Together, these results place TBR1 within a complex network of interacting proteins with roles in two distinct but frequently co-morbid neurodevelopmental disorders.

The AP-MS screen replicated the previously published interaction between TBR1 and CASK, a protein kinase with roles in ion channel and synaptic adhesion molecule organisation, dendritic spine formation and axonal outgrowth (Hsueh, 2006). CASK is localised at the plasma membrane and synapses but is translocated to the nucleus through interaction with TBR1, where it functions as a coactivator for TBR1 target genes such as *RELN* and *GRIN2B* (Hsueh et al., 2000; Wang et al., 2004b). The current study identified an additional putative interaction between TBR1 and LIN7C, a known CASK-interacting protein. LIN7C is one of three alternative subunits (with LIN7A and LIN7B) that forms a tripartite complex with CASK and APBA1; this complex has been hypothesised to control synaptic adhesion and synaptic vesicle exocytosis (Butz et al., 1998). LIN7C has not been observed in the nucleus (Shelly et al., 2003), so it is not clear whether it too can be translocated by TBR1. Instead, affinity purification of LIN7C interaction might reflect a temporary association as TBR1 makes contact with a CASK-LIN7C complex before transporting CASK to the nucleus.

The other known TBR1 interactors could not be detected in this study: FOXP1/2/4 (Deriziotis et al., 2014b) and BCL11A (den Hoed et al., 2018). One possible explanation could be that these interactions are relatively weak or transient and may not survive the cell lysis and multiple washing steps in the affinity purification. AP-MS screens for interactors of FOXP1/2/4 (Li et al., 2015a; Estruch et al., 2018) and BCL11A (Xu et al., 2013) have also failed to detect TBR1, however these were performed in cell populations in which TBR1 is not expressed (HEK293 and erythroid cells, respectively).

On the other hand, connections between TBR1 and FOXP/BCL11A are supported by the identification of a number of shared interactors in the current study. For example, 92/248 (37%) of the putative TBR1 interactors were also identified in a recent AP-MS study of the

FOXP1/2/4 interactomes (Estruch et al., 2018). Three of these were subsequently confirmed to interact with multiple FOXP proteins using BRET assays (NR2F2-FOXP1/2, YY1-FOXP1/2/4 and ZMYM2-FOXP1/2/4) (Estruch et al., 2018). Two other putative TBR1 interactors - CTBP1 and GATAD2B - have been previously shown to interact with FOXP1/2/4 (Li et al., 2004; Chokas et al., 2010; Deriziotis et al., 2014a; Estruch et al., 2016a) (Chapter 4). Similarly, AP-MS screens for BCL11A interactors in erythroid cells have identified multiple interactors potentially shared by TBR1 (MTA2, GATAD2B, KDM1A, TBL1XR1, ZBTB33, BCOR, SMARCA2, MSH2), of which three (MTA2, KDM1A, ZBTB33) were confirmed by co-immunoprecipitation (Sankaran et al., 2008; Xu et al., 2013). BCL11A has also been shown to interact with NR2F2 (Chan et al., 2013). These overlaps suggest that TBR1 interacts with other TFs involved in neurodevelopmental disorders (including speech/language disorders) and that they may recruit shared mechanisms for transcriptional regulation.

Epigenetic factors were overrepresented in the TBR1 interactome and constituted more than 20% of all proteins detected (52/248). These included multiple members of several well-described chromatin remodelling complexes, including NuRD, CoREST, Nu4A and SWI/SNF. Notably, the AP-MS screen identified epigenetic factors with predominantly positive effects on transcription (e.g. histone acetylation), as well as those with predominantly negative effects on transcription (e.g. histone deacetylation). This supports previous findings in ChIP-seq experiments, where TBR1-binding sites were enriched for both active (H3K27ac, H3K4me1) and repressive (H3K9me3, H3K27me3) chromatin marks (Notwell et al., 2016), as well as evidence of both up- and down- regulation of Tbr1 target genes in mouse neurons (Bedogni et al., 2010; McKenna et al., 2011; Chuang et al., 2014).

In addition to transcriptional regulation and chromatin modification, RNA-processing proteins were also overrepresented in the TBR1 interactome and included spliceosome subunits and various cleavage and polyadenylation factors. Interestingly, TBR1 is not the first T-box TF family member to be implicated in post-transcriptional regulation via RNA-binding (Kumar P et al., 2014). A proteomic screen for TBX3-interacting proteins identified a number of mRNA splicing factors and other proteins involved in RNA metabolism (Kumar P et al., 2014). The authors then demonstrated direct binding of TBX3 to certain RNAs containing T-box binding elements, and implicated TBX3 in the regulation of alternative splicing. Prior evidence that T-box TFs may be involved in post-transcriptional regulation therefore raises the possibility of a similar role for TBR1.

Another somewhat unexpected finding is that cell-cycle and ubiquitin pathway proteins, including members of the anaphase promoting complex, are overrepresented in the TBR1 interactome. The anaphase promoting complex (APC) is a cell cycle-regulated E3 ubiquitin ligase complex that targets proteins for degradation by the proteasome, and is required for sister-chromatid separation, exit of cells from mitosis and subsequent DNA replication in S phase (Peters, 2006). Some TFs have been identified as targets of the APC: for example, the FOXM1 TF is degraded in late M or early G1 phase, and its depletion is required for regulated entry into S phase (Park et al., 2008). The T-box family TF TBX2 is also cell cycle

regulated and exhibits very low expression throughout mitosis, although the mechanisms regulating this depletion are not well understood (Bilican and Goding, 2006). TBR1 itself is specifically expressed in post-mitotic projection neurons, suggesting that its expression is also suppressed during mitosis. The large number of putative interactions between TBR1 and the APC (as well as other ubiquitin ligase complexes and cofactors) might indicate that TBR1 is ubiquitinated for degradation in mitotic cells. The combination of TBR1 overexpression and continuing cell division in the HEK293 cell model might stimulate APC-mediated ubiquitination and degradation of TBR1 in an effort to deplete TBR1 and allow mitosis to progress.

Another unexpected process identified in the TBR1 interactome is DNA replication and repair. This may arise due to links between DNA repair and transcription. The replication factor C complex plays an essential role in transcription-coupled nucleotide excision repair, where transcription-blocking DNA errors are rapidly removed from the transcribed strand when encountered by RNA polymerase II (Overmeer et al., 2010). The replication factor C complex was therefore probably immunoprecipitated along with RNA polymerase II proteins.

On the other hand, a more direct involvement of TBR1 in DNA replication/repair cannot be ruled out. It has been proposed that some TFs may have a secondary role in the formation of non-transcriptional complexes on the chromosome, including the replication initiation complex (Kohzaki and Murakami, 2005). The T-box TF Tbx2 has been detected at replication foci during S phase (Bilican and Goding, 2006). T-box TFs might therefore play a role in regulating DNA replication. The current study also identified the TP53 protein, which has roles in both transcriptional regulation and the DNA damage response (Meek, 2004).

The design of the current study specifically targets nuclear TBR1, in keeping with the exclusively nuclear TBR1 expression seen under most conditions (Deriziotis et al., 2014b). As expected therefore, the majority of TBR1 interaction candidates are expressed in the nucleus. But despite using a nuclear extraction step to selection for nuclear proteins, the current study also detected a number of proteins that do not appear to be expressed in the nucleus. It is not clear whether these cytoplasmic proteins were carried over as an artefact in the nuclear extract and subsequent cell lysis, and therefore may not interact with TBR1 under normal physiological conditions; or whether they might translocate to the nucleus in the presence of TBR1, where they may fulfil other functions, as is the case with CASK. A final possibility is that these proteins might interact with a potential small amount of cytoplasmic TBR1, which was also carried over despite the nuclear extraction. There is some evidence of cytoplasmic and synaptic expression of Tbr1 in certain neurons in postnatal and adult (but not embryonic) rat cortex, hippocampus and cerebellum (Hong and Hsueh, 2007), although this finding does not seem to have been replicated in other studies. Any residual cytoplasmic TBR1 detected in the current study might interact with cytoplasmic proteins *in situ*. This explanation seems unlikely, however, given the strong nuclear localisation observed here for the 2xFLAG/V5-TBR1 construct in the HEK293 stable transfection model (Fig 5.1B).

One limitation of the present study is that it was performed in HEK293 cells, which may not reflect the physiological conditions of neurons, where TBR1 is almost exclusively expressed. It may be that a more accurate description of the TBR1 interactome could be obtained in the future using more physiologically relevant (i.e. neuronal) tissue, such as dissected brain tissue or primary neuron cultures from mice, or neurons derived from human induced pluripotent stem cells expressing tagged TBR1. A TBR1-specific antibody could alternatively be used to precipitate endogenous TBR1 in these cells or tissues. Nonetheless, the majority of the proteins detected in the current study are co-expressed in several of the same brain regions as TBR1 and many are highly expressed in neurons. Therefore, the HEK293 model employed here is a valuable first step in understanding the TBR1 interactome, particularly as a hypothesis-building exercise. Further experiments using alternative methods, such as co-immunoprecipitation or BRET assays, will be important for validating these putative interactors and characterising the nature of the interactions, yielding novel insights into TBR1-related mechanisms in brain development.

5.5 APPENDIX

Protein	UniProt ID	Mascot score			emPAI		
		Mean	AP1	AP2	Mean	AP1	AP2
ADCK3	Q8NI60	85	52	118	0.095	0.05	0.14
ADNP	Q9H2P0	107	134	80	0.22	0.08	0.36
AHCTF1	Q8WYP5	525	587	463	0.2	0.23	0.17
AKAP13	Q12802	248.5	348	149	0.095	0.11	0.08
AKAP8	O43823	103.5	58	149	0.11	0.09	0.13
AKAP8L	Q9ULX6	197.5	186	209	0.255	0.2	0.31
ANAPC1	Q9H1A4	411.5	335	488	0.25	0.21	0.29
ANAPC2	Q9UJX6	175.5	108	243	0.21	0.11	0.31
ANAPC4	Q9UJX5	261	206	316	0.21	0.19	0.23
ANAPC5	Q9UJX4	284	266	302	0.385	0.31	0.46
ANAPC7	Q9UJX3	364.5	401	328	0.46	0.46	0.46
ANK2	Q01484	74	83	65	0.05	0.06	0.04
ANK3*	Q12955	260	351	169	0.04	0.05	0.03
ANKRD28	O15084	177.5	301	54	0.17	0.25	0.09
ARID1A	O14497	112	79	145	0.075	0.05	0.1
ASCC3	Q8N3C0	245.5	243	248	0.09	0.09	0.09
ASUN	Q9NVM9	90	112	68	0.125	0.17	0.08
AURKB	Q96GD4	56.5	54	59	0.7	0.27	1.13
BAG5	Q9UL15	156	124	188	0.245	0.13	0.36
BAIAP2L1	Q9UHR4	130	109	151	0.18	0.18	0.18
BAP18	Q8IXM2	117.5	112	123	0.75	0.75	0.75
BCOR	Q6W2J9	427.5	793	62	0.225	0.42	0.03
BEND3	Q5T5X7	119	178	60	0.125	0.18	0.07
BMI1	P35226	81.5	66	97	0.185	0.18	0.19
BPTF	Q12830	1109.5	1318	901	0.28	0.32	0.24
CAD	P27708	936.5	883	990	0.43	0.4	0.46
CASK	O14936	192.5	181	204	0.2	0.16	0.24
CBFB	Q13951	41.5	43	40	0.36	0.36	0.36
CDC16	Q13042	170.5	103	238	0.22	0.19	0.25
CDC23	Q9UJX2	233	284	182	0.35	0.38	0.32
CDC27	P30260	254	205	303	0.25	0.23	0.27
CDKN2AIP	Q9NXV6	243.5	372	115	0.305	0.44	0.17
CHD7	Q9P2D1	662.5	749	576	0.195	0.19	0.2
CHD9	Q3L8U1	199	153	245	0.06	0.04	0.08
CIZ1	Q9ULV3	79.5	68	91	0.07	0.07	0.07
CNOT1	A5YKK6	458.5	287	630	0.165	0.1	0.23
COLGALT1	Q8NBJ5	75.5	82	69	0.09	0.09	0.09
COMMD3-BMI1	-	81.5	66	97	0.185	0.18	0.19
CPSF1	Q10570	314.5	403	226	0.16	0.22	0.1
CPSF2	Q9P2I0	103	71	135	0.115	0.07	0.16
CPSF7*	Q8N684	98.5	80	117	0.215	0.17	0.26
CPVL	Q9H3G5	164.5	203	126	0.485	0.42	0.55
CSNK2A1	P68400	134	170	98	0.275	0.32	0.23
CSTF3	Q12996	158.5	114	203	0.165	0.12	0.21
CTBP1	Q13363	147.5	182	113	0.495	0.83	0.16

Protein	UniProt ID	Mascot score			emPAI		
		Mean	AP1	AP2	Mean	AP1	AP2
CTNND1	O60716	298	380	216	0.26	0.32	0.2
CUL3	Q13618	145.5	149	142	0.13	0.15	0.11
CUL7	Q14999	107	154	60	0.05	0.07	0.03
CUX1	P39880	214	284	144	0.1	0.12	0.08
CYFIP2	Q96F07	463	421	505	0.355	0.25	0.46
DAP3	P51398	67.5	56	79	0.105	0.1	0.11
DCAF7	P61962	82.5	88	77	0.22	0.27	0.17
DDX19A	Q9NUU7	145.5	50	241	0.25	0.13	0.37
DDX46*	Q7L014	87.5	58	117	0.075	0.06	0.09
DMAP1	Q9NPF5	164	71	257	0.315	0.2	0.43
DNAJB1	P25685	119	49	189	0.59	0.39	0.79
DOCK7	Q96N67	254	372	136	0.105	0.16	0.05
DVL2	O14641	162.5	183	142	0.18	0.18	0.18
DVL3	Q92997	94	120	68	0.11	0.13	0.09
DYNC1LI1	Q9Y6G9	89	106	72	0.17	0.18	0.16
EIF3D	O15371	148.5	117	180	0.22	0.22	0.22
EIF3H	O15372	139	73	205	0.46	0.17	0.75
EMSY	Q7Z589	335.5	564	107	0.225	0.38	0.07
EP400	Q96L91	295.5	509	82	0.07	0.12	0.02
EPB41L2*	O43491	163.5	238	89	0.155	0.22	0.09
ESRP2	Q9H6T0	153.5	187	120	0.18	0.18	0.18
ESRRA	P11474	95	110	80	0.285	0.23	0.34
EWSR1	Q01844	126.5	127	126	0.22	0.16	0.28
EXOC8	Q8IYI6	59.5	61	58	0.04	0.04	0.04
FAM208A	Q9UK61	615.5	790	441	0.315	0.38	0.25
FAM208B	Q5VWN6	259	438	80	0.09	0.14	0.04
FASTKD1	Q53R41	172	155	189	0.18	0.14	0.22
FIP1L1*	Q6UN15	73	93	53	0.075	0.1	0.05
FUBP3	Q96I24	413.5	433	394	0.68	0.77	0.59
FZR1	Q9UM11	123	148	98	0.23	0.19	0.27
GATAD2B	Q8WXI9	249.5	376	123	0.395	0.63	0.16
GTF3C1	Q12789	612	513	711	0.25	0.19	0.31
GTF3C2	Q8WUA4	294	410	178	0.315	0.46	0.17
GTF3C3	Q9Y5Q9	305.5	329	282	0.25	0.29	0.21
GTF3C4	Q9UKN8	292	248	336	0.365	0.27	0.46
GTF3C5	Q9Y5Q8	191.5	161	222	0.27	0.27	0.27
HADHB	P55084	339	272	406	0.76	0.54	0.98
HCFC1	P51610	265	342	188	0.11	0.14	0.08
HEATR1	Q9H583	72.5	82	63	0.03	0.03	0.03
HELZ	P42694	154.5	239	70	0.06	0.09	0.03
HIC2	Q96JB3	124	155	93	0.155	0.21	0.1
HNRNPUL2	Q1KMD3	153.5	216	91	0.165	0.21	0.12
HSPA4*	P34932	118	86	150	0.125	0.11	0.14
HSPBP1	Q9NZL4	77.5	90	65	0.22	0.27	0.17
HSPH1*	Q92598	112.5	92	133	0.1	0.1	0.1
ILK	Q13418	132	81	183	0.19	0.12	0.26

Protein	UniProt ID	Mascot score			emPAI		
		Mean	AP1	AP2	Mean	AP1	AP2
IMPDH2	P12268	76.5	70	83	0.13	0.13	0.13
INTS1	Q8N201	499	630	368	0.195	0.23	0.16
INTS10	Q9NVR2	107	133	81	0.145	0.17	0.12
INTS3	Q68E01	191	269	113	0.145	0.21	0.08
IQGAP1	P46940	275	92	458	0.135	0.08	0.19
IQGAP2	Q13576	825	771	879	0.455	0.4	0.51
KDM1A	O60341	447.5	478	417	0.54	0.57	0.51
KIF20A	O95235	156	147	165	0.12	0.14	0.1
KIF20B	Q96Q89	114.5	164	65	0.045	0.06	0.03
KPNA1	P52294	103.5	87	120	0.2	0.11	0.29
LAS1L	Q9Y4W2	212	139	285	0.265	0.17	0.36
LGALS3BP	Q08380	255.5	218	293	0.37	0.34	0.4
LIN7C*	Q9NUP9	84.5	58	111	0.43	0.24	0.62
LONP1	P36776	213.5	137	290	0.215	0.11	0.32
LUC7L2	Q9Y383	110	68	152	0.19	0.15	0.23
MAPKAP1	Q9BPZ7	139.5	195	84	0.175	0.24	0.11
MCM3AP	O60318	133	175	91	0.06	0.08	0.04
MDC1	Q14676	144	242	46	0.05	0.09	0.01
MED12	Q93074	163.5	262	65	0.065	0.1	0.03
MED23	Q9ULK4	102.5	112	93	0.06	0.06	0.06
MEN1	O00255	84	87	81	0.11	0.11	0.11
MKI67	P46013	120	188	52	0.03	0.05	0.01
MORF4L1	Q9UBU8	57.5	64	51	0.165	0.22	0.11
MRE11A*	P49959	390	332	448	0.585	0.43	0.74
MSH2	P43246	276.5	61	492	0.32	0.06	0.58
MSH6	P52701	205	105	305	0.15	0.09	0.21
MTA2	O94776	305.5	271	340	0.41	0.35	0.47
MTOR	P42345	458.5	502	415	0.155	0.18	0.13
MYCBP2	O75592	172	260	84	0.025	0.04	0.01
MYEF2	Q9P2K5	254.5	272	237	0.41	0.46	0.36
MYO9B*	Q13459	324.5	404	245	0.13	0.19	0.07
NAT10	Q9H0A0	54	42	66	0.03	0.03	0.03
NBN	O60934	86	126	46	0.08	0.12	0.04
NCBP1	Q09161	77	83	71	0.07	0.07	0.07
NCKAP1	Q9Y2A7	496	523	469	0.385	0.42	0.35
NIPBL	Q6KC79	193.5	242	145	0.045	0.05	0.04
NOP58	Q9Y2X3	281.5	244	319	0.455	0.38	0.53
NR2F2	P24468	73	70	76	0.15	0.15	0.15
NT5DC2	Q9H857	78.5	63	94	0.11	0.11	0.11
NUP107	P57740	326.5	301	352	0.195	0.16	0.23
NUP133	Q8WUM0	202	164	240	0.13	0.1	0.16
NUP160	Q12769	199	108	290	0.115	0.07	0.16
NUP205	Q92621	410	382	438	0.15	0.12	0.18
NUP43	Q8NFB3	104	68	140	0.205	0.16	0.25
NUP93*	Q8N1F7	242	200	284	0.285	0.23	0.34
NUP98	P52948	339	301	377	0.235	0.12	0.35

Protein	UniProt ID	Mascot score			emPAI		
		Mean	AP1	AP2	Mean	AP1	AP2
ODF2	Q5BJF6	112.5	104	121	0.125	0.14	0.11
ORC3	Q9UBD5	105.5	87	124	0.135	0.12	0.15
PAXIP1	Q6ZW49	114	124	104	0.08	0.08	0.08
PCID2	Q5JVF3	54	67	41	0.11	0.15	0.07
PDK3	Q15120	82.5	101	64	0.15	0.15	0.15
PDS5A	Q29RF7	202.5	85	320	0.125	0.04	0.21
PELP1	Q8IZL8	163	189	137	0.14	0.14	0.14
PI4KA	P42356	162	239	85	0.06	0.08	0.04
PIK3C2A	O00443	302	423	181	0.135	0.2	0.07
PITRM1	Q5JRX3	210	191	229	0.18	0.18	0.18
PKP2	Q99959	223	329	117	0.185	0.26	0.11
PKP3	Q9Y446	195	316	74	0.21	0.34	0.08
PLOD1	Q02809	130.5	99	162	0.145	0.12	0.17
PMPCA	Q10713	141	110	172	0.245	0.18	0.31
PMPCB	O75439	202.5	224	181	0.38	0.42	0.34
PNO1	Q9NRX1	73	43	103	0.25	0.1	0.4
POGZ	Q7Z3K3	245	353	137	0.16	0.23	0.09
POLR1A	O95602	167	63	271	0.08	0.03	0.13
POLR1C	O15160	72	43	101	0.265	0.21	0.32
POLR2A	P24928	319	341	297	0.14	0.14	0.14
POLR2B	P30876	348	323	373	0.24	0.21	0.27
POLRMT	O00411	128.5	177	80	0.075	0.1	0.05
PPP1R9B	Q96SB3	188.5	102	275	0.2	0.07	0.33
PPP2R1A	P30153	123	191	55	0.17	0.27	0.07
PRPF40A	O75400	146.5	76	217	0.145	0.06	0.23
PTPN13	Q12923	314	459	169	0.095	0.14	0.05
PUM1	Q14671	159	189	129	0.14	0.14	0.14
PUM2	Q8TB72	120	122	118	0.12	0.1	0.14
RAD50*	Q92878	1151.5	834	1469	1.135	0.76	1.51
RALY	Q9UKM9	61	61	61	0.29	0.29	0.29
RANGAP1*	P46060	267.5	174	361	0.435	0.22	0.65
RB1	P06400	57.5	73	42	0.045	0.06	0.03
RBFOX2	O43251	92	116	68	0.24	0.26	0.22
RBM25	P49756	124.5	133	116	0.12	0.14	0.1
RBPJ	Q06330	53.5	62	45	0.145	0.07	0.22
RCOR2	Q8IZ40	150.5	214	87	0.185	0.25	0.12
RFC1	P35251	243.5	229	258	0.175	0.13	0.22
RFC2	P35250	222.5	192	253	0.62	0.49	0.75
RFC3	P40938	252	248	256	0.65	0.71	0.59
RFC5	P40937	222	135	309	0.66	0.39	0.93
RFX1	P22670	376	423	329	0.34	0.36	0.32
RING1	Q06587	177	209	145	0.515	0.68	0.35
RNF2*	Q99496	151	175	127	0.46	0.52	0.4
RPRD2	Q5VT52	117	130	104	0.065	0.09	0.04
RTCB	Q9Y310	115	143	87	0.19	0.19	0.19
SCAF4	O95104	83.5	54	113	0.065	0.02	0.11

Protein	UniProt ID	Mascot score			emPAI		
		Mean	AP1	AP2	Mean	AP1	AP2
SCAF8	Q9UPN6	55.5	54	57	0.065	0.02	0.11
SCRIB	Q14160	241	358	124	0.13	0.2	0.06
SENP3	Q9H4L4	203	242	164	0.28	0.34	0.22
SFN*	P31947	89.5	66	113	0.405	0.25	0.56
SKIV2L2	P42285	469	533	405	0.35	0.42	0.28
SLFN11	Q7Z7L1	72.5	64	81	0.06	0.06	0.06
SMARCA1	P28370	423.5	436	411	0.4	0.4	0.4
SMARCA2	P51531	392.5	280	505	0.205	0.13	0.28
SNRPA	P09012	101.5	74	129	0.69	0.52	0.86
SNTB2	Q13425	183.5	138	229	0.32	0.25	0.39
SPATA5	Q8NB90	194	233	155	0.18	0.22	0.14
SPATA5L1	Q9BVQ7	160.5	174	147	0.22	0.27	0.17
SPEN	Q96T58	416.5	765	68	0.095	0.17	0.02
SRCAP	Q6ZRS2	262	305	219	0.065	0.07	0.06
SRRT*	Q9BXP5	188	60	316	0.195	0.06	0.33
STUB1	Q9UNE7	282.5	168	397	1.255	0.57	1.94
SUGP2	Q8IX01	312.5	443	182	0.22	0.3	0.14
SUPT16H*	Q9Y5B9	332	488	176	0.265	0.42	0.11
SUPT6H	Q7KZ85	258.5	321	196	0.11	0.14	0.08
SYMPK	Q92797	394	339	449	0.285	0.23	0.34
TARBP1	Q13395	58.5	40	77	0.03	0.02	0.04
TBL1XR1	Q9BZK7	73	94	52	0.27	0.28	0.26
TCERG1	O14776	203	217	189	0.155	0.17	0.14
TFCP2	Q12800	195	230	160	0.34	0.43	0.25
THOC2	Q8NI27	148	185	111	0.08	0.1	0.06
TIAL1	Q01085	144.5	111	178	0.405	0.26	0.55
TNPO1	Q92973	50.5	54	47	0.04	0.04	0.04
TNPO2	O14787	50.5	54	47	0.04	0.04	0.04
TP53*	P04637	434.5	451	418	1.38	1.55	1.21
TP73	O15350	96.5	106	87	0.16	0.1	0.22
TRRAP	Q9Y4A5	1269.5	1324	1215	0.32	0.35	0.29
TTC19	Q6DKK2	121.5	78	165	0.255	0.16	0.35
TUBGCP2	Q9BSJ2	188.5	165	212	0.17	0.13	0.21
TUBGCP3	Q96CW5	166	126	206	0.17	0.15	0.19
U2AF1	Q01081	126.5	146	107	0.405	0.56	0.25
UBQLN1	Q9UMX0	122.5	175	70	0.195	0.29	0.1
UBQLN2	Q9UHD9	173.5	277	70	0.19	0.28	0.1
UBQLN4	Q9NRR5	158.5	207	110	0.255	0.35	0.16
USF2	Q15853	126	185	67	0.375	0.56	0.19
USP9X	Q93008	188.5	177	200	0.065	0.07	0.06
UTRN	P46939	213.5	271	156	0.05	0.07	0.03
WDR18	Q9BV38	143	148	138	0.31	0.31	0.31
WRNIP1	Q96S55	201	207	195	0.245	0.3	0.19
XRN2	Q9H0D6	180	184	176	0.175	0.19	0.16
YEATS4	O95619	65.5	44	87	0.31	0.26	0.36
YTHDC2	Q9H6S0	197.5	232	163	0.13	0.15	0.11

Protein	UniProt ID	Mascot score			emPAI		
		Mean	AP1	AP2	Mean	AP1	AP2
YWHAH	Q04917	89.5	66	113	0.405	0.25	0.56
YY1	P25490	78.5	88	69	0.38	0.56	0.2
ZBTB2	Q8N680	161.5	159	164	0.25	0.25	0.25
ZBTB25	P24278	83	69	97	0.14	0.14	0.14
ZBTB33	Q86T24	140	221	59	0.165	0.24	0.09
ZC3H18	Q86VM9	83.5	103	64	0.095	0.13	0.06
ZCCHC8	Q6NZY4	120.5	201	40	0.185	0.23	0.14
ZFHX4	Q86UP3	619	647	591	0.145	0.16	0.13
ZMYM2	Q9UBW7	999.5	1060	939	0.77	0.73	0.81
ZMYM3	Q14202	164	156	172	0.1	0.09	0.11
ZMYM4	Q5VZL5	1348.5	1151	1546	1.055	0.83	1.28
ZMYND8	Q9ULU4	253	383	123	0.155	0.23	0.08
ZNF131	P52739	261	174	348	0.365	0.3	0.43
ZNF24	P17028	77	107	47	0.215	0.35	0.08
ZNF324	O75467	52	60	44	0.09	0.11	0.07
ZNF592	Q92610	113.5	180	47	0.065	0.1	0.03
ZNF638	Q14966	321.5	420	223	0.125	0.17	0.08
ZNF639	Q9UID6	79.5	69	90	0.235	0.29	0.18
ZNF687	Q8N1G0	219.5	250	189	0.145	0.16	0.13

*Proteins identified in 1 of the 30 control experiments in the Contaminant Repository for Affinity Purification database. All other proteins were undetected in all 30 control experiments.

6 TBR1 INTERACTS WITH TRANSCRIPTIONAL REGULATORS DISRUPTED IN ASD/ID-RELATED DISORDERS

ABSTRACT

The TBR1 transcription factor is recurrently mutated in individuals with autism spectrum disorder (ASD), intellectual disability (ID) and language deficits. Deciphering the role of TBR1 in these disorders will involve an understanding of its molecular interactions. In addition to forming homodimers, TBR1 has been shown to interact with several other proteins, including CASK, FOXP1/2/4 and BCL11A. In the previous chapter, 247 new interaction candidates were identified using affinity purification-mass spectrometry. In this chapter, validation was attempted for a subset of candidates previously implicated in ASD and ID, as well as their paralogues. Five novel TBR1-interaction partners were confirmed (GATAD2B, BCOR, ADNP, NR2F1 and NR2F2), while a sixth candidate (CYFIP2) showed evidence of interaction only with a cytoplasmic variant of TBR1. The effects of *TBR1* variants on the new interactions were assessed and compared to previous findings for known interactors. Variants had contrasting effects on different interactions: in particular, variants within the T-box domain of TBR1 disrupted interactions with FOXP1/2/4 and ADNP (and to a lesser extent NR2F1/2) but left other interactions unaffected. Further mapping experiments pointed to two discrete protein-binding regions within the TBR1 protein – the T-box domain and a C-terminal region (residues 394-568) – which were of varying importance for different interactions. *GATAD2B* and *ADNP* mutations also displayed the potential to disrupt TBR1-interaction. Two previously-characterised regions of GATAD2B (CR1 and CR2) were required for TBR1-interaction and a putative TBR1-binding site on ADNP (residues 644-730) was also identified. Taken together, these results establish new links in a growing network of interacting proteins involved in neurodevelopmental disorders and demonstrate that mutations affecting one protein may have deleterious effects on the function of other proteins in the network.

6.1 INTRODUCTION

The neuron-specific TBR1 transcription factor (OMIM 604616) is recurrently mutated in cases of autism spectrum disorder (ASD), often involving intellectual disability (ID), developmental delay and language impairment (Neale et al., 2012; O’Roak et al., 2012b, 2012a; Deriziotis et al., 2014b; De Rubeis et al., 2014; Iossifov et al., 2014). Previous studies have shown that TBR1 can form homodimers (Deriziotis et al., 2014b) and interact with several other proteins, including CASK (Hsueh et al., 2000), FOXP1/2/4 (Deriziotis et al., 2014b) and BCL11A (den Hoed et al., 2018). In Chapter 5 of this thesis, I described an affinity purification-mass spectrometry (AP-MS) study aimed at identifying new TBR1-interaction partners. In addition to confirming the known interaction with CASK, I identified 247 novel interaction candidates, with a significant overrepresentation of proteins that are disrupted in ID and ASD. Because AP-MS experiments can be vulnerable to false-positive results, in the current chapter I select a subset of these candidates for validation and follow-up.

I also assess how the confirmed interactions are affected by *TBR1* variants identified in patients with ASD. Eleven *TBR1* variants have been thoroughly characterised in previous studies (Table 6.1, Table 6.2; Fig 6.1) (Deriziotis et al., 2014b; den Hoed et al., 2018). These include a *de novo* nonsense variant (p.S351*) and a *de novo* frameshift variant (p.A136Pfs*80), which are both likely to undergo nonsense-mediated decay (NMD). In previous studies, these truncating variants exhibited abnormal cytoplasmic localisation, abolished transcriptional repression and disrupted interactions with WT (wild-type) TBR1, CASK, FOXP2 and BCL11A (Deriziotis et al., 2014b; den Hoed et al., 2018). Thus, haploinsufficiency is the likely mechanism of disorder for these variants, even if a small proportion of truncated protein escapes NMD. Four *de novo* missense variants of TBR1 (p.K228E, p.W271C, p.N374H, p.K389E) were found to preserve transcriptional repression, while forming abnormal aggregates within the nucleus (Deriziotis et al., 2014b; den Hoed et al., 2018). The variants retained the ability to homodimerise and to interact with CASK and BCL11A, but translocated WT TBR1, CASK and BCL11A into nuclear aggregates, suggesting a possible dominant-negative effect (Deriziotis et al., 2014b; den Hoed et al., 2018). The same variants abolished interaction between TBR1 and FOXP2 (Deriziotis et al., 2014b; den Hoed et al., 2018). In contrast, a fifth *de novo* missense variant of TBR1, p.W271R, displayed normal localisation and TF activity and retained interactions with WT TBR1, CASK, BCL11A and FOXP2, suggesting that it may not be causative, or may have subtler effects on protein function (den Hoed et al., 2018).

Functional experiments have also been conducted for four private TBR1 variants identified in patients with ASD but inherited from an unaffected parent (p.Q178E, p.V356M, p.Q418R, p.P542R). Three of the variants (p.Q178E, p.V356M, p.P542R) had no effect on any aspect of TBR1 function (Deriziotis et al., 2014b; den Hoed, 2016; den Hoed et al., 2018), indicating that they may not be causative for the disorder. The fourth, p.Q418R, had mixed effects, abolishing interaction with FOXP2 and BCL11A but not WT TBR1 or CASK, while maintaining normal nuclear localisation and transcriptional repression (Deriziotis et

al., 2014b; den Hoed et al., 2018). These findings suggest that the p.Q418R variant might contribute to the ASD phenotype by disrupting a subset of protein-protein interactions.

I investigate how novel TBR1 interactions might be affected by these eleven variants and compare the effects to the existing data for TBR1 homodimerisation and interaction with CASK, FOXP2 and BCL11A. In some cases, variants in the new interactors themselves have been identified in patients with neurodevelopmental disorders, so I also investigate the effects of those variants on interaction with TBR1.

Two regions of TBR1 have been identified as important for protein-protein interactions (Fig 6.1; Table 6.1). The T-box DNA-binding domain (residues 213-393) mediated the FOXP2-interaction and was sufficient for some degree of TBR1 homodimerisation and BCL11A-interaction (Deriziotis et al., 2014b; den Hoed et al., 2018). However, for full homodimerisation and BCL11A-interaction, a C-terminal region (residues 394-567) was also required (Deriziotis et al., 2014b; den Hoed et al., 2018). An overlapping binding region has been identified for CASK (residues 342-682) (Hsueh et al., 2000). In the current chapter, I investigate whether the same or different regions of TBR1 are important for binding the novel interactors.

Table 6.1. Previous functional characterisation of TBR1 variants.

cDNA	Protein	PolyPhen2	Ref	Loc	Repr	Interactions			
						CASK	TBR1	FOXP2	BCL11A
<i>de novo</i> missense									
c.682A>G	p.K228E	Prob. 0.986	1	N*	+				
c.811T>C	p.W271R	Prob. 0.999	2	N	+				
c.813G>T	p.W271C	Prob. 0.999	3	N*	+				
c.1120A>C	p.N374H	Prob. 0.990	4	N*	+				
c.1165A>G	p.K389E	Prob. 0.973	3	N*	+				
rare inherited missense									
c.532C>G	p.Q178E	Benign 0.178	1	N	+				
c.1066G>A	p.V356M	Prob. 0.961	5	N	+				
c.1253A>G	p.Q418R	Benign 0.043	1	N	+				
c.1625C>G	p.P542R	Prob. 0.992	1	N	+				
<i>de novo</i> truncations									
c.405delC	p.A136Pfs*80	NA	1	N+C	-				
c.1049dupC	p.S351*	NA	1	N+C	-				
synthetic truncations									
NA	p.N394*	NA	1	N+C	-				
NA	p.S568*	NA	1	N+C	-				

Variants are annotated based on the following reference sequences: transcript NM_006593.3, protein NP_006584.1. PolyPhen2 (HumVar) predictions: prob. = probably damaging, NA = not applicable.. References: ¹Deriziotis et al. (2014b), ²Hamdan et al. (2014), ³O'Roak et al. (2014), ⁴Neale et al. (2012), ⁵Bacchelli et al. (2003). Localisation: N = nuclear, C = cytoplasmic, N+C = mixed localisation, * = forms aggregates. Repr = repressive function, intact (+) or impaired (-). Interactions: green = interaction; red = no interaction; orange = decreased interaction signal; grey = no data.

Table 6.2. Phenotypic description of patients with *TBR1* variants.

Variant	Ref	Patient no.	Inh.	Sex	Age	ASD	ID	IQ		Speech/language			
								V	NV	Level	Delay	Regr.	Motor delay
p.K228E	1	13814.p1	-	M	7y	+	+	75	78	fluent	+	-	ND
p.W271R	2	121.83	-	F	10y	+	+	ND	ND	non-verbal	+	-	+
p.W271C	3	214-17068-1	-	M	5y	+	-	94	88	phrase	+	ND	ND
p.N374H	4	09C86232A	-	F	ND	+	ND	69	74	ND	ND	ND	ND
p.K389E	3	220-9833-201**	-	F	8y	+	+	ND	42	single word	+	ND	ND
p.Q178E	1	12994.p1	pat.	M	11y	+	-	96	99	fluent	+	-	ND
p.Q178E	1	14332.p1	pat.	M	4y	+	+	28	39	non-verbal	+	-	ND
p.V356M	5	-(sib-pair)	mat.	M	ND	+	ND	ND	ND	ND	ND	ND	ND
p.V356M	5	-(sib-pair)	mat.	F	ND	+	ND	ND	ND	ND	ND	ND	ND
p.Q418R	1	13702.p1	mat.	M	5y	+	+/-	58	86	fluent	+	-	ND
p.P542R	1	13060.p1	pat.	M	6y	+	-	98	81	fluent	+	+/-	ND
p.A136P fs*80	1	11480.p1	-	M	7y	+	+	24	41	non-verbal	+	+	ND
p.S351*	1	13796.p1	-	F	8y	+	+	69	63	phrase	+	-	ND

Variants are annotated based on protein reference NP_006584.1. References: ¹ Deriziotis et al. (2014b), ² Hamdan et al. (2014), ³ O’Roak et al. (2014), ⁴ Neale et al. (2012), ⁵ Bacchelli et al. (2003). Patient numbers from original publications. **This patient had an *FGF22* missense variant c.205G>A, p.V84M in addition to the *TBR1* variant. Inh. = inheritance (paternal or maternal). V/NV = verbal/nonverbal IQ. Regr. = regression. ND = no data, +/- indicates a possible or borderline case.

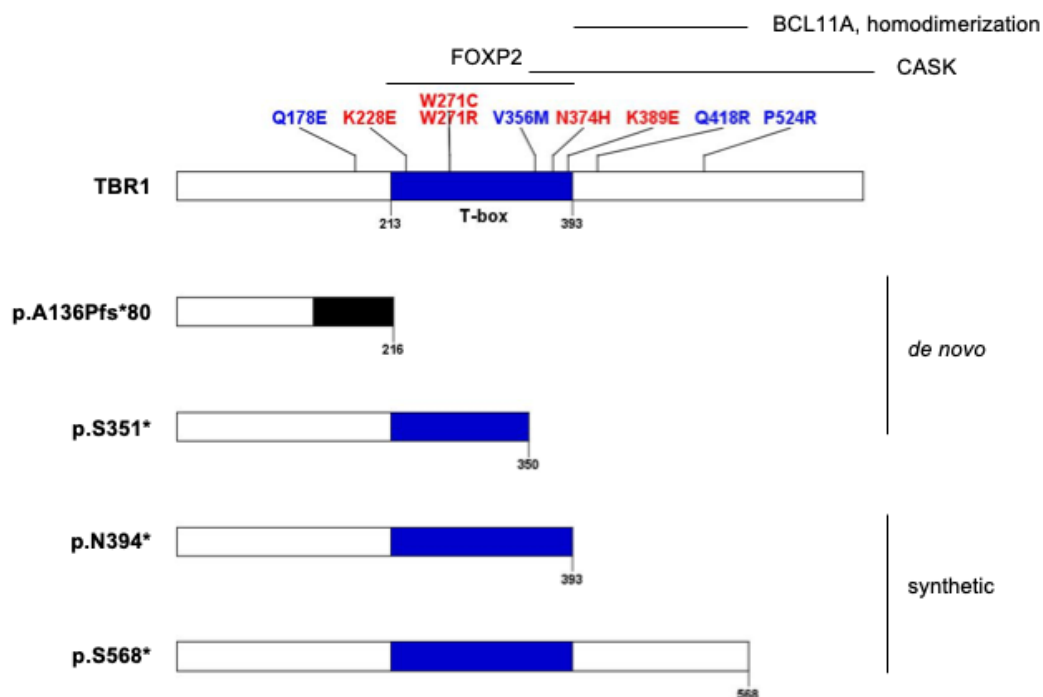


Figure 6.1. Patient-derived and synthetic *TBR1* variants. Schematic representations of TBR1 missense and truncating variants used in this study. Truncations are either *de novo* patient mutations or synthetic constructs. Missense variants are *de novo* (red) or rare inherited (blue). The T-box DNA-binding domain is also shown. Horizontal bars indicate the proposed binding regions for TBR1-homodimerisation and CASK- FOXP2- and BCL11A-interaction (Hsueh et al., 2000; Deriziotis et al., 2014b; den Hoed et al., 2018).

6.2 METHODS

6.2.1 DNA CONSTRUCTS

TBR1 was synthesised by GenScript USA, as previously described (Deriziotis et al., 2014b). *CTBP1*, *CTBP2*, *YY1*, *NR2F1*, *NR2F2*, *ZMYM2* were cloned as previously described (Deriziotis et al., 2014a; Estruch et al., 2016a, 2018). *KDM1A*, *GATAD2B*, *NCKAP1*, *CSNK2A1*, *TBL1XR1*, *CTNND1*, *BCOR*, *ADNP*, *SMARCA2*, *CYFIP1* and *CYFIP2* were amplified by PCR from human foetal cDNA (see Table 6.3 for primer sequences). Open reading frames of these genes were subcloned into pLuc, pYFP and a modified pmCherry-C1 expression vector (Clontech). Generation of constructs with *GATAD2B* and *TBR1* variants has been described previously (Chapter 4; Deriziotis et al., 2014b; den Hoed et al., 2018). *ADNP* variants and the *TBR1* NLS mutant were generated using the QuikChange II Site-Directed Mutagenesis Kit (Agilent; see Table 6.4 for SDM primer sequences). All constructs were verified by Sanger sequencing. The putative nuclear localisation signal (NLS) within *TBR1* was predicted using two programs: cNLSmapper (Kosugi et al., 2009) and NLStradamus (Nguyen Ba et al., 2009).

Table 6.3. Cloning primer sequences

	Forward primer (5'-3')	Site	Reverse primer (5'-3')	Site
KDM1A	<u>AGATCTCGTTATCTGGG</u> AAGAAGGCGGC	BglII	<u>TCTAGATCACATGCTTGG</u> GGACTGCT	XbaI
GATAD2B	<u>GGATCCTGGATAGAAT</u> GACAGAAGATGC	BamHI	<u>TCTAGATTATTTCTGTCC</u> ACTGATGG	XbaI
NCKAP1	<u>GTCGACGGTCGCGCTCA</u> GTGCTGCAGCC	Sall	<u>TCTAGATTTATGCAGAAG</u> ATGTAACAC	XbaI
YY1	<u>GGATCCTGGCCTCGGGC</u> GACACC	BamHI	<u>TCTAGATCACTGGTTGTT</u> TTTGGCCTTAGCATG	XbaI
CSNK2A1	<u>GGATCCACTCGGGACCC</u> GTGCCAAGCAG	BamHI	<u>TCTAGATTACTGCTGAGC</u> GCCAGCGG	XbaI
TBL1XR1	<u>GAATTCAGTATAAGCA</u> GTGATGAGGT	EcoRI	<u>CCCGGGCTATTTCCGAAG</u> GTCTAATAC	XmaI
CTNND1	<u>AGATCTCGGACGACTCA</u> GAGGTGGAGTCGACCG CCAGCATCTT	BglII	<u>TCTAGATTAAATCTTCTG</u> CATCAAGGGTGTTGTTC CTTCAATGGCTCC	XbaI
BCOR	<u>CTCGAGTCTCTCAGCAA</u> CCCCCTGTA	XhoI	<u>GTCGACTTACCAGTAGTT</u> GTCTGAGG	Sall
ADNP	<u>AGATCTCCTTCCAACCTT</u> CCTGTCAACAA	BglII	<u>TCTAGATTAGGCCTGTTG</u> GCTGCTCA	XbaI
SMARCA2	<u>AAGCTTCTCCACGCCCA</u> CAGACCCTGGTGCGAT	HindIII	<u>TCTAGATTACTCATCATC</u> CGTCCCCTTCTTCTGA CTGTTC	XbaI
CTBP1	<u>GAGATCTCGGGCAGCTC</u> GCACTTGCTCAAC	BglII	<u>GCTAGACTACAACCTGGT</u> CACTGGCGTGG	XbaI
NR2F1	<u>GGATCCTGGCAATGGTA</u> GTTAGCAGCTG	BamHI	<u>TCTAGACTAGGAGCACTG</u> GATGGACATG	XbaI

	Forward primer (5'-3')	Site	Reverse primer (5'-3')	Site
NR2F2	<u>AGATCTTGGCAATGGTA</u> GTCAGCACG	BglII	<u>TCTAGATTATTGAATTGC</u> CATATACGGCCAGTT	XbaI
ZMYM2	<u>GGATCCTGGACACAAG</u> TTCAGTGGGAGG	BamHI	<u>GCTAGCTTAGTCTGTGTC</u> TTCATCCAGTTC	NheI
CYFIP1	<u>GTCGACGGGCGGCCCA</u> GGTGA CTCTGGA	Sall	<u>TCTAGATTCAGCTGCTGG</u> CGAGGGACT	XbaI
CYFIP2	<u>GTCGACGGACCACGCA</u> CGTCACCCTGGA	Sall	<u>TCTAGATTTAGCAAGTGG</u> TGGCCAAGG	XbaI

Table 6.4. SDM primer sequences

	Forward primer (5'-3')	Reverse primer (5'-3')
<i>ADNP</i> variants		
T443A	CATATTTTCCACTTTTGAGCTG AGGAAGTGTTACCTGGG	CCCAGGTAACACTTCCTCAGCT CAAAAGTGGAAAATATG
G1094V	GCTGCTCAGTTTAACTACGGC TAAGCTGCCATG	CATGGCAGCTTAGCCGTAGTTA AACTGAGCAGC
S404*	GACTTTAACTGGCCCTATGAG AGAGAAGAGGCATTAG	CTAATGCCTCTTCTCTCATA GGGCCAGTTAAAGTC
R644*	ACTTGGTGCCTCTCTCATAAG TGATGTGCAAGTG	CACTTGCACATCACTTATGAGA GAGGCACCAAGT
R730*	ATCATCATCTAACTTTTCATTTT TTCAGTAAGGGAAATTCCATT TGCTC	GAGCAAATGGAATTTCCCTTAC TGAAAAAATGAAAGTTAGATG ATGAT
L1057*	CAATCTGGGGGTTAGATTAGC GCTCATCATTCTCA	TGAGAATGATGAGCGCTAATCT AACCCCCAGATTG
<i>TBRI</i> variant		
<i>TBRI</i>-NLS-mut	ATTTACGAGCAGGCCGGGGG GGGGGGGATCTCGCCGGCCG	CGGCCGCGAGATCCCCCCCCC CCCGCCTGCTCGTAAAT

6.2.2 FLUORESCENCE MICROSCOPY

Cells were seeded onto coverslips coated with poly-L-lysine (Sigma-Aldrich) and were fixed 24 h post-transfection using 4% paraformaldehyde (Electron Microscopy Sciences) for 10 min at room temperature. YFP and mCherry fusion proteins were visualised by direct fluorescence. Nuclei were visualised with Hoechst 33342 (Invitrogen). Fluorescence images were obtained using an Axio Imager M2 upright microscope (Zeiss).

6.2.3 BIOLUMINESCENCE RESONANCE ENERGY TRANSFER (BRET)

BRET assays were performed as previously described (Deriziotis et al., 2014a, 2014b). In summary, cells were transfected with pairs of *Renilla* luciferase and YFP-fusion proteins in 96-well plates. *Renilla* luciferase and YFP were used as control proteins (with a C-terminal NLS for nuclear interactions, without an NLS for cytoplasmic interactions). EnduRen luciferase substrate (Promega) was added to cells 48 h after transfection at a final concentration of 60µM and incubated for 4 h. Emission measurements were taken with a

TECAN F200PRO microplate reader using the Blue1 and Green1 filters and corrected BRET ratios were calculated as follows: $[\text{Green1}_{(\text{experimental condition})}/\text{Blue1}_{(\text{experimental condition})} - \text{Green1}_{(\text{control condition})}/\text{Blue1}_{(\text{control condition})}]$. YFP fluorescence was measured separately, with excitation at 485 nm and emission at 535 nm, to quantify expression of the YFP-fusion proteins.

6.2.4 IMMUNOFLUORESCENCE

Immunofluorescence was performed on embryonic (E16.5) and adult (P101) mouse brain sections. Brains were dissected, fixed in 4% paraformaldehyde, cryopreserved in a sucrose gradient (15-30%) and embedded in OCT. Sagittal sections were prepared at a thickness of 4 μm using a Leica CM1950 cryostat. Tissue was permeabilised in acetone and blocked using 10% donkey serum in PBS for 1 hr at RT. Antibodies were diluted in 2% donkey serum in PBS. Primary antibodies were applied overnight at 4°C: goat anti-p66 β (ab111248, Abcam; 1:50 dilution), goat anti-ADNP (PA5-47792, Thermo Fisher; 1:40) and rabbit anti-TBR1 (ab31940, Abcam; 1:1000). Secondary antibodies were incubated for 1 hour at RT: donkey anti-goat Alexa 594 and donkey anti-rabbit Alexa 488 (both Invitrogen; 1:500). After staining, slides were washed and mounted with VectaShield Antifade Mounting Medium with DAPI (Vector Laboratories). Images were obtained using an Axio Imager M2 upright fluorescence microscope (Zeiss).

6.3 RESULTS

6.3.1 PART 1: VALIDATION OF SIX NOVEL TBR1 INTERACTORS

6.3.1.1 *Selection of TBR1 interaction candidates*

248 putative TBR1-interacting proteins were identified by AP-MS (see Chapter 5). Of these, 32 were encoded by genes with reported variants in syndromes involving ASD/ID and were therefore considered most likely to share a role with TBR1 in the aetiology of neurodevelopmental disorders. These proteins were ranked by exponentially modified protein abundance index (emPAI), an approximate measure of relative protein abundance that takes into account the size of the protein (Ishihama et al., 2005), averaged across the two AP-MS experiments (Chapter 5). Ten of the highest-ranked candidates were selected for validation and further functional characterisation: KDM1A, GATAD2B, NCKAP1, YY1, CSNK2A1, TBL1XR1, CTNND1, BCOR, ADNP and SMARCA2 (Table 6.5). Due to a change in the HGNC approved gene symbol of the 9th-ranked protein (EMSY or C11orf30) during the course of this project, there was a mismatch between the name used in the TBR1 interactome list and that used in the SFARI database (Abrahams et al., 2013). As a result, this particular protein was mistakenly excluded from the validation and follow-up experiments, despite it being of interest for ASD. However, this omission does not have any impact on the overall interpretation and conclusions of this chapter.

Table 6.5. Top-ranked ID/ASD-related proteins identified in the putative TBR1 interactome, by mean emPAI score.

Protein	emPAI	Mascot	Disorder (OMIM #)	SFARI	
KDM1A	0.54	447.5	ID	Cleft palate, psychomotor retardation and distinctive facial features (616728)	-
GATAD2B	0.395	249.5	ID	Mental retardation, autosomal dominant 18 (615074)	-
NCKAP1	0.385	496	ASD		2
YY1	0.38	78.5	ID	Gabriele-de Vries syndrome (617557)	-
CSNK2A1	0.275	134	ID	Okur-Chung neurodevelopmental syndrome (617062)	-
TBL1XR1	0.27	73	ID/ ASD	Mental retardation, autosomal dominant 41 (602342) Pierpont syndrome (616944)	3
CTNND1	0.26	298	ASD	Autism with intellectual disability (O'Roak et al., 2012b; no OMIM number)	-
BCOR	0.225	427.5	ID	Microphthalmia, syndromic 2 (300166)	-
C11orf30 (EMSY)	0.225	335.5	ASD	-	3
ADNP	0.22	107	ID/ ASD	Helsmoortel-van der Aa syndrome (615873)	1S
SMARCA2	0.205	392.5	ID	Nicolaides-Baraitser syndrome (601358)	-

The emPAI (exponentially modified protein abundance index) and Mascot score are shown as the mean of two AP experiments. Involvement in autism spectrum disorder (ASD) is based on the SFARI Gene database and categorised according to candidate confidence: 1 = High Confidence, 2 = Strong Candidate, 3 = Suggestive Evidence, S = syndromic form. Intellectual disability (ID) disorders are taken from the Radboud University Human Genetics Department sequencing panel (DG2.5x), supported by OMIM numbers where available.

TBR1 has previously been shown to interact with three TFs that are mutated in neurodevelopmental disorders affecting speech and language: FOXP2, FOXP1 and BCL11A (Deriziotis et al., 2014b; den Hoed et al., 2018). Thus, I hypothesised that these TFs may have additional interactors in common with TBR1. The putative TBR1 interactome contained 5 previously reported FOXP1/2-interactors: CTBP1, GATAD2B, NR2F2, YY1, ZMYM2 (Table 6.6) (Li et al., 2004; Chokas et al., 2010; Deriziotis et al., 2014a; Estruch et al., 2016a, 2018; Chapter 4). NR2F1 also interacts with BCL11A (Chan et al., 2013). Therefore, in addition to those proteins listed as high-ranked candidates above, three further

candidates were selected for further investigation: CTBP1, NR2F2 and ZMYM2 (Table 6.6). Follow-up experiments were also performed for CTBP2, which is 77% identical to CTBP1 at the amino-acid level and shares its interactions with FOXP1/2/4 (Estruch et al., 2016a); as well as NR2F1, which is 85% identical to NR2F2 and shares its interactions with FOXP1/2/4 (Estruch et al., 2016a) and BCL11A (Chan et al., 2013) (Table 6.6). Although these two candidates were each detected in only one of the two AP-MS experiments (Chapter 5), they were considered promising candidates due to their similarity to CTBP1 and NR2F2 and their conserved interactions with other TFs.

Table 6.6. Putative TBR1 interactors shared with FOXP1/2 and BCL11A.

Protein	emPAI	Mascot	FOXP1/2	BCL11A
GATAD2B	0.395	249.5	Chokas et al. (2010) Chapter 4	-
YY1	0.38	78.5	Estruch et al. (2018)	-
ZMYM2	0.77	999.5	Estruch et al. (2018)	-
CTBP1	0.495	147.5	Li et al. (2004) Deriziotis et al. (2014a) Estruch et al. (2016a)	-
CTBP2	-	-	Estruch et al. (2016a)	-
NR2F1	-	-	Estruch et al. (2018)	Chan et al. (2013)
NR2F2	0.15	73	Estruch et al. (2018)	Chan et al. (2013)

The emPAI (exponentially modified protein abundance index) and Mascot score are shown as the mean of two AP experiments.

6.3.1.2 Expression of TBR1 interaction candidates

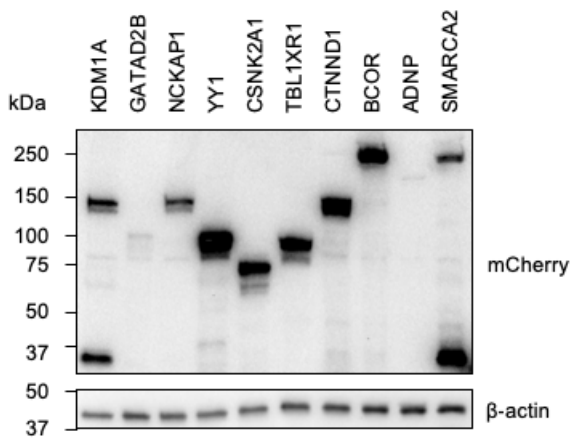
The 10 high-ranked TBR1 interaction candidates were expressed with N-terminal mCherry tags in HEK293 cells and produced proteins close to the expected molecular weights (Fig 6.2A). However, it should be noted that additional bands at lower molecular weights were present for all proteins of interest, perhaps due to degradation. Three proteins exhibited a diffuse nuclear pattern of expression (Fig 6.2B), in keeping with previous reports: KDM1A (Humphrey et al., 2001), ADNP (Mandel and Gozes, 2007), and SMARCA2 (Koga et al., 2009). GATAD2B and BCOR were also nuclear, but with a distinctive speckled appearance consistent with prior reports (Fig 6.2B) (Brackertz et al., 2002; Buchberger et al., 2013). CSNK2A1 was mostly nuclear, with some weaker expression in the cytoplasm; a similar localisation pattern has been seen in at least one previous study (Bae et al., 2016). YY1 was expressed in the nucleus, as in earlier studies (Lee et al., 2013), but also in the cytoplasm. YY1 has been previously detected in cytoplasmic vesicles, giving a pattern of expression similar to that observed here (Thul et al., 2017; see URL¹). NCKAP1 was approximately evenly distributed between the nucleus and cytoplasm (Fig 6.2B).

¹ Human Protein Atlas entry for YY1: <https://www.proteinatlas.org/ENSG00000100811-YY1/cell>

In the current study, TBL1XR1 localisation was predominantly cytoplasmic (Fig 6.2B). Nuclear localisation of TBL1XR1 has been reported in most cells and tissues, however NIH-3T3 cells exhibited predominantly cytoplasmic localisation (Zhang et al., 2006), suggesting that intracellular localisation patterns may be cell line-dependent. The TBL1XR1 localisation pattern in HEK293 cells has not been previously reported. It is also possible that the addition of an N-terminal tag (in this case mCherry) might interfere with the nuclear localisation of TBL1XR1 by affecting the LisH domain near the start of the protein, which has been implicated in nuclear import (Gerlitz et al., 2005).

CTNND1 has 32 different isoforms produced by alternative splicing and alternative initiation sites, which differ in their subcellular localisation (Aho et al., 2002; Krakstad et al., 2004). The experiments in this chapter were performed using CTNND1 isoform 1A, which contains an NLS and has been previously reported to localise equally to the nucleus and cytoplasm (Aho et al., 2002; Krakstad et al., 2004). Nonetheless, in the current study, CTNND1 isoform 1A was exclusively cytoplasmic (Fig 6.2B). This may reflect differences in the specific set of CTNND1-interacting proteins present in the cells used here, as prior reports demonstrate that CTNND1 localisation is influenced by interactions with other proteins (e.g. NANOS1 and CHD1) (Krakstad et al., 2004; Strumane et al., 2006).

CTBP1, CTBP2, NR2F1, NR2F2, ZMYM2 (the interaction candidates that are shared with FOXP1/2 and BCL11A) also produced proteins at the expected molecular weights (Fig 6.3A), although lower MW bands were seen for CTBP1/2, which may result from degradation. Consistent with previous reports, CTBP1 was localised mostly to the cytoplasm, while CTBP2 was restricted to the nucleus (Bergman et al., 2006; Verger et al., 2006), as were NR2F1 and NR2F2 (Fig 6.3B) (Yamazaki et al., 2009). ZMYM2 exhibited a punctate pattern of expression in the nucleus, as in prior studies (Fig 6.3B) (Kunapuli et al., 2006).

A

	MW (kDa)	
	Protein	+mCherry
KDM1A	92.9	121.7
GATAD2B	65.3	94.1
NCKAP1	128.8	157.6
YY1	44.7	73.5
CSNK2A1	45.1	73.9
TBL1XR1	55.6	84.4
CTNND1	105.0	133.8
BCOR	192.2	221.0
ADNP	122.8	151.6
SMARCA2	181.3	210.1

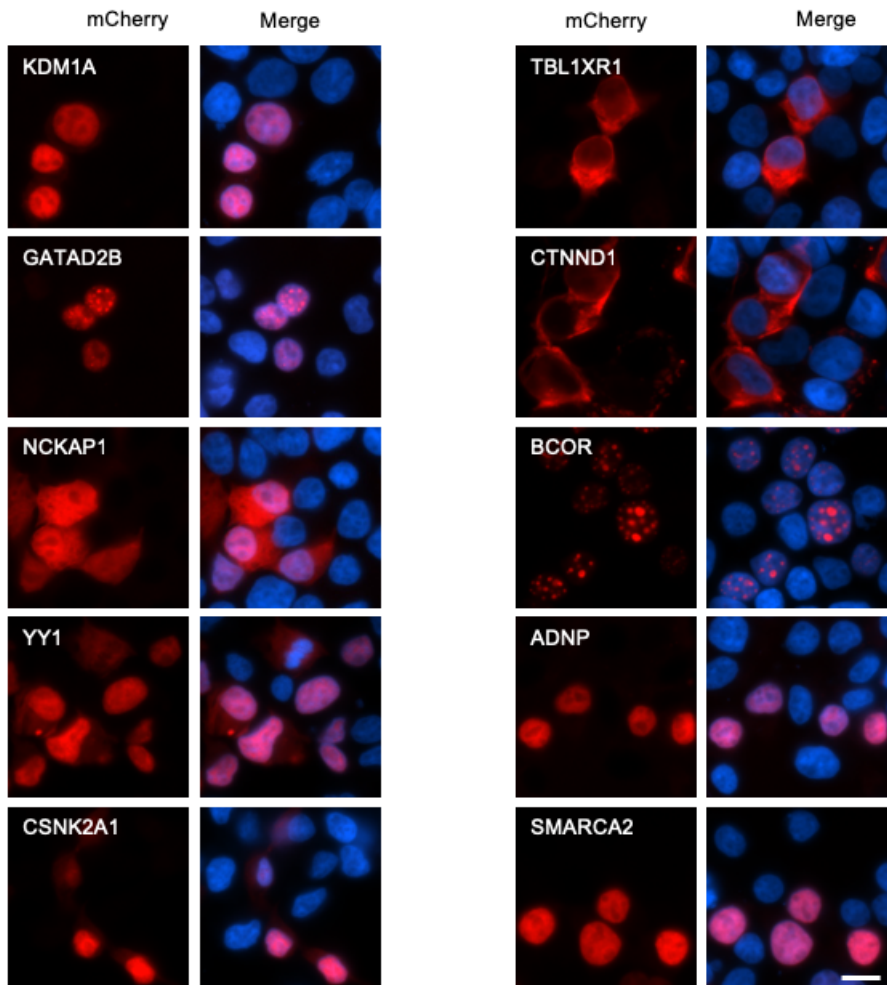
B

Figure 6.2 (opposite page). Expression of TBR1 interaction candidates. (A) Left panel: Immunoblotting of whole-cell lysates from HEK293 cells transfected with TBR1 interaction candidates fused to mCherry. β -actin served as a loading control. Right panel: expected molecular weights of candidates alone and as mCherry fusion proteins. (B) Fluorescence microscopy images of HEK293 cells transfected with TBR1 interaction candidates (fused to mCherry, red). Nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μ m.

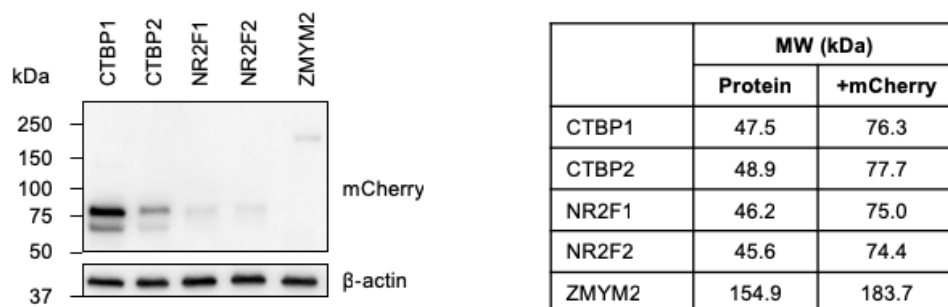
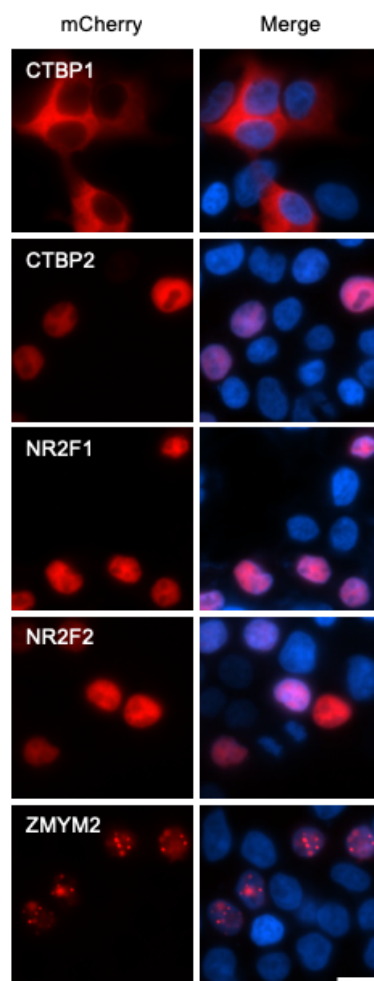
A**B**

Figure 6.3 (opposite page). Expression of TBR1 interaction candidates previously found to interact with FOXP1/2 or BCL11A. (A) Left panel: Immunoblotting of whole-cell lysates from HEK293 cells transfected with TBR1 interaction candidates fused to mCherry. β -actin served as a loading control. Right panel: expected molecular weights of candidates alone and as mCherry fusion proteins. (B) Fluorescence microscopy images of HEK293 cells transfected with TBR1 interaction candidates (fused to mCherry, red). Nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μ m.

6.3.1.3 Validation of five TBR1-interacting proteins

Renilla luciferase (Rluc)-fusion proteins were generated for the fifteen TBR1 interaction candidates and tested for interactions with YFP-TBR1 using BRET assays in HEK293 cells. Positive interactions were detected between TBR1 and three of the highly-ranked candidates: GATAD2B, BCOR and ADNP (Fig 6.4A). In addition, out of the five candidates selected on the basis of known interactions with FOXP1/2 and BCL11A, two showed positive interactions with TBR1 using this method: NR2F1 and NR2F2 (Fig 6.5A). Significant BRET signals were also observed for CTBP2 and ZMYM2, which may indicate weak interactions with TBR1, however the magnitude of these signals was considered too low to warrant further investigation here. Each of the confirmed interactors was expressed in the nucleus and co-localised with TBR1 in co-transfection experiments (Fig 6.4B, 6.5B). There was no evidence of an interaction with the other 8 candidates. TBL1XLR1, CTNND1 and CTBP1 remained predominantly cytoplasmic when co-transfected with TBR1, which could account for the observed lack of interaction (Fig 6.4B, 6.5B). On the other hand, the remaining 5 candidates (KDM1A, NCKAP1, YY1, CSNK2A1, SMARCA2) showed no evidence of interaction despite at least partial co-localisation within the nucleus (Fig 6.4, 6.5).

Immunofluorescence experiments were attempted in mouse brain tissue, to determine whether the interactors were co-expressed with Tbr1 in the same brain regions. During the time period available for this study, suitable antibodies and staining protocols could be optimised for Tbr1 and Adnp but not for the other candidates. Immunofluorescence staining was performed in embryonic (E16.5) and adult (P101) mouse brain sections, where Adnp was observed throughout the brain (Fig 6.6). In E16.5 cortex, Tbr1 was observed in a medial layer that likely corresponds to the cortical plate, as previously shown in mice at a similar developmental stage (Fig 6.6A) (Hevner et al., 2001). Considerable co-localisation with Adnp was seen in this layer. In adult cortex, Tbr1 was detected in multiple layers, but with a concentration of TBR1+ cells in layer 6 (Fig 6.6B). The separation of layers was not as distinct as in some studies that, for example, have shown TBR1 expression largely restricted to layer 6 at postnatal day P3 (Cánovas et al., 2015). However, several studies in older mice closer to the time point investigated here, such as P20-25 (Brunjes and Osterberg, 2015) and adult (Uhlén et al., 2015; see URL¹), do show more widely-spread TBR1 expression similar to Fig 6.6B. Co-localisation with Adnp could be seen in multiple cells. Overlapping expression was also detected in the Purkinje cells of the cerebellum (Fig 6.6C). It is therefore probable that Tbr1 and Adnp are co-expressed, with the potential to interact, in multiple neuronal populations in both developing and adult brain.

¹ Mouse Brain Atlas Atlas entry for TBR1 <https://www.proteinatlas.org/ENSG00000136535-TBR1/tissue/mouse+brain>

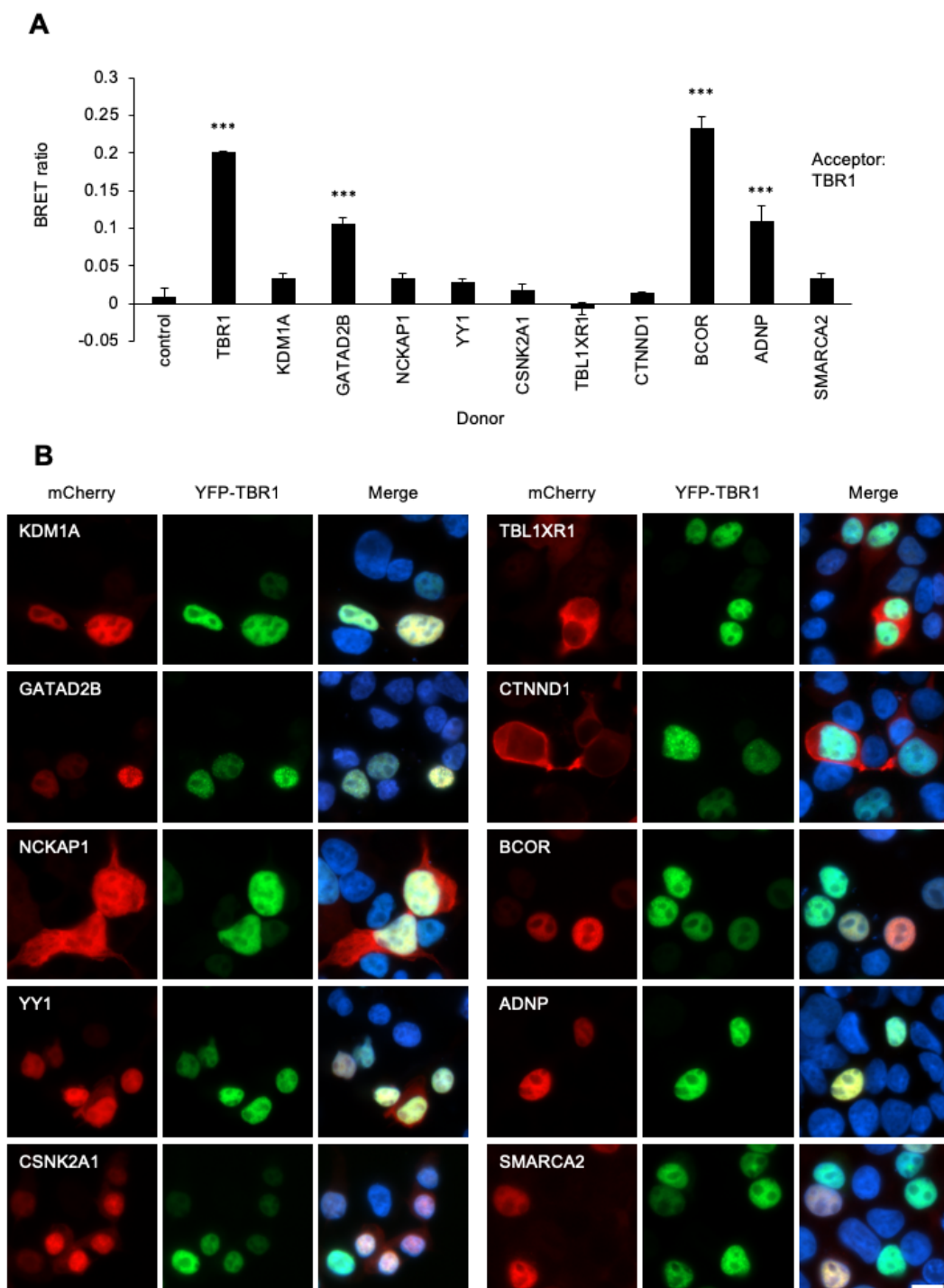
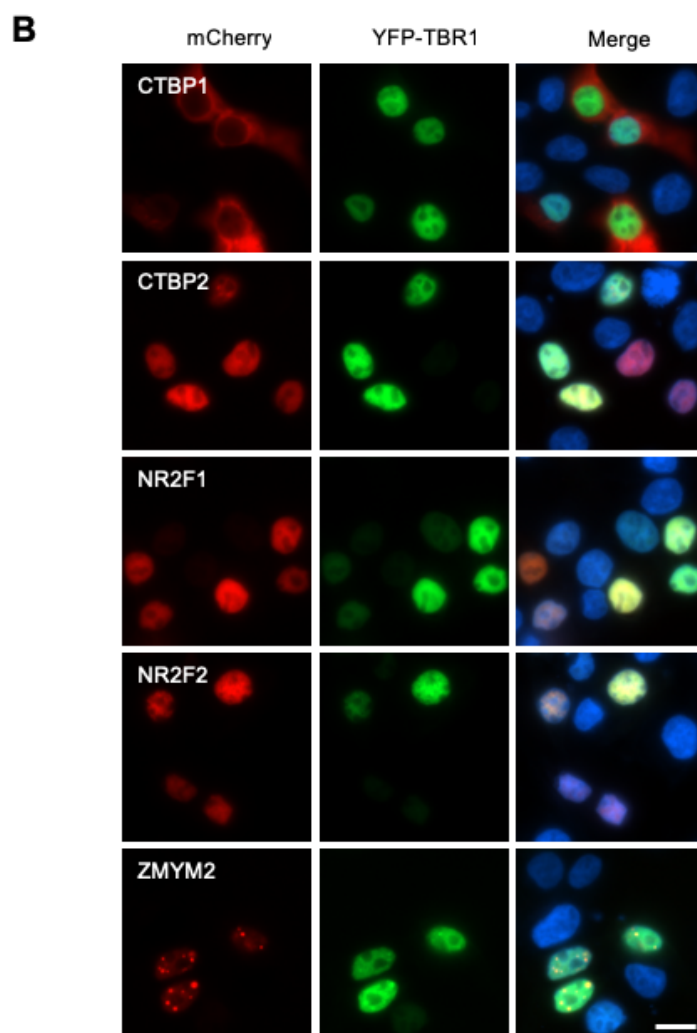
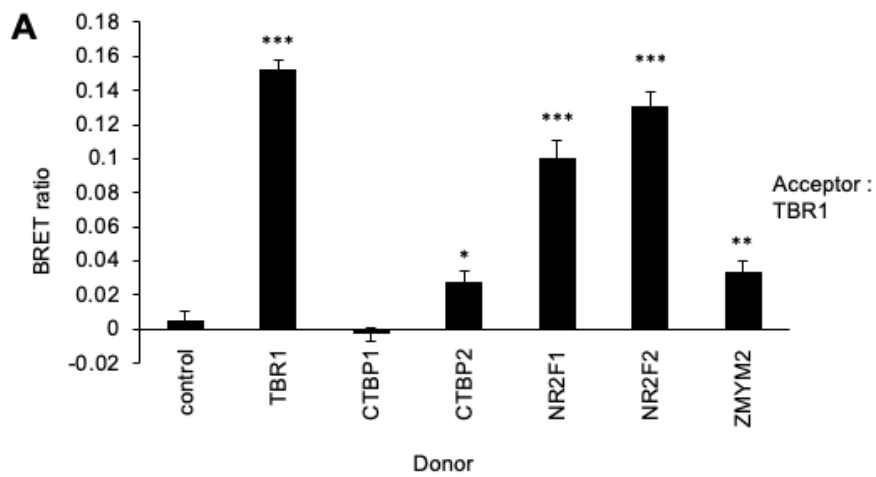


Figure 6.4 (previous page). TBR1 interacts with GATAD2B, BCOR and ADNP. (A) BRET assay for interaction between TBR1 and ten interaction candidates. Bars represent the corrected mean BRET ratios \pm SD of one experiment performed in triplicate ($***P < 0.001$ compared to control, Student's T-test). (B) Fluorescence microscopy images of HEK293 cell transfected with TBR1 (fused to YFP, green) and interaction candidates (fused to mCherry, red). Nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μ m.

Figure 6.5 (opposite page). TBR1 interacts with NR2F1/2. (A) BRET assay for interaction between TBR1 and five interaction candidates. Bars represent the corrected mean BRET ratios \pm SD of one experiment performed in triplicate ($*P < 0.05$, $**P < 0.01$ and $***P < 0.001$ compared to control, Student's T-test). (B) Fluorescence microscopy images of HEK293 cell transfected with TBR1 (fused to YFP, green) and interaction candidates (fused to mCherry, red). Nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μ m.



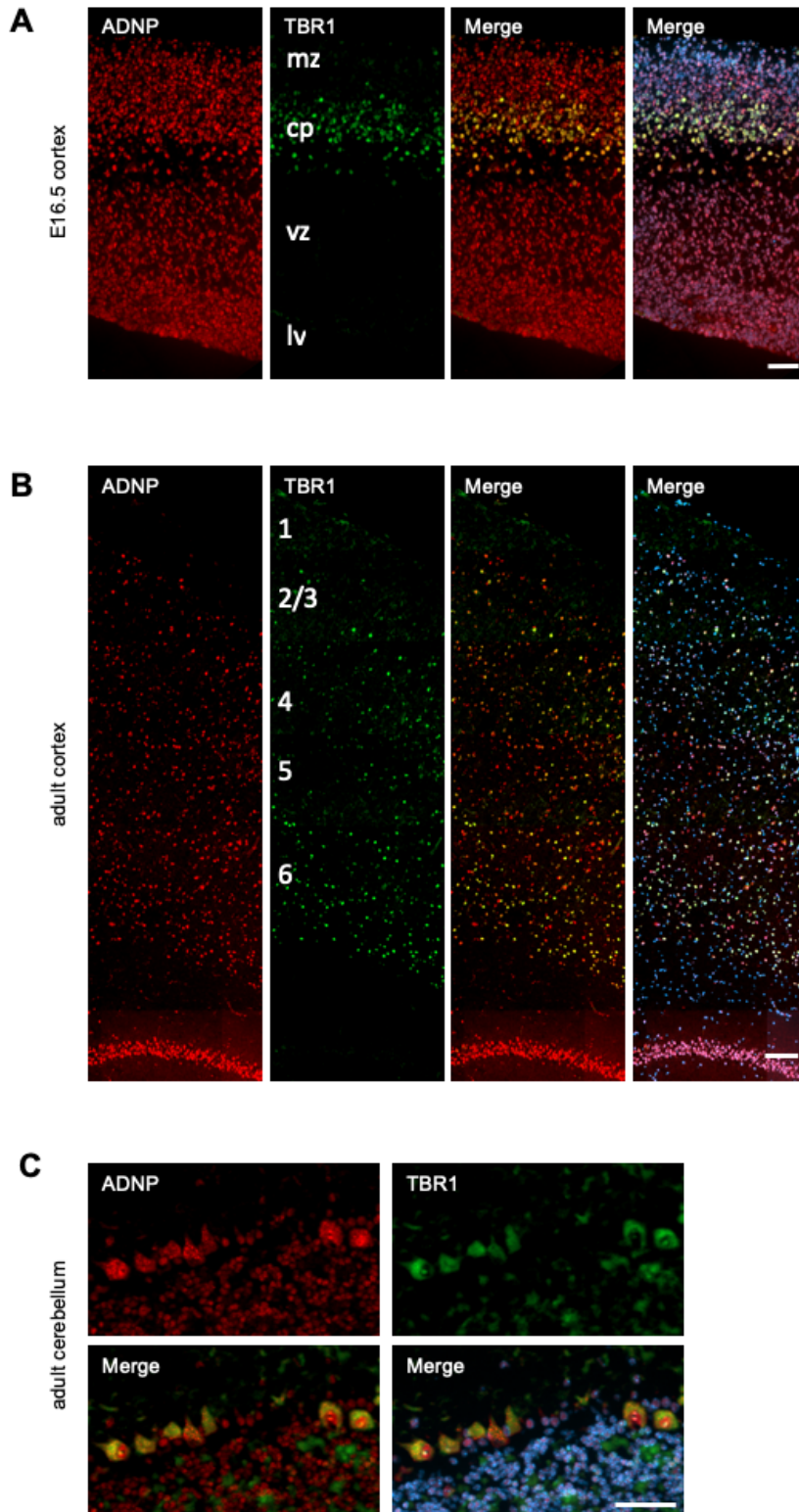
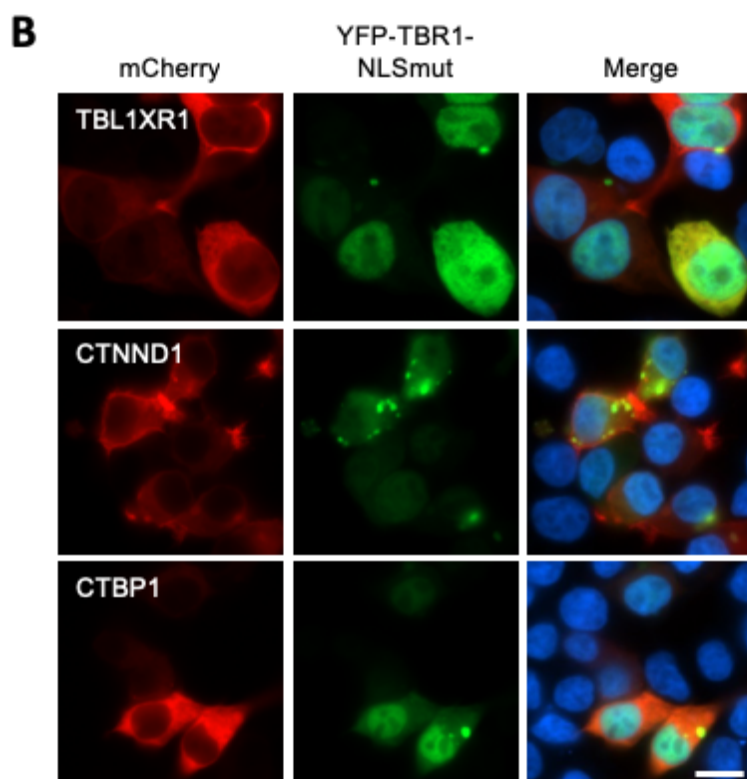
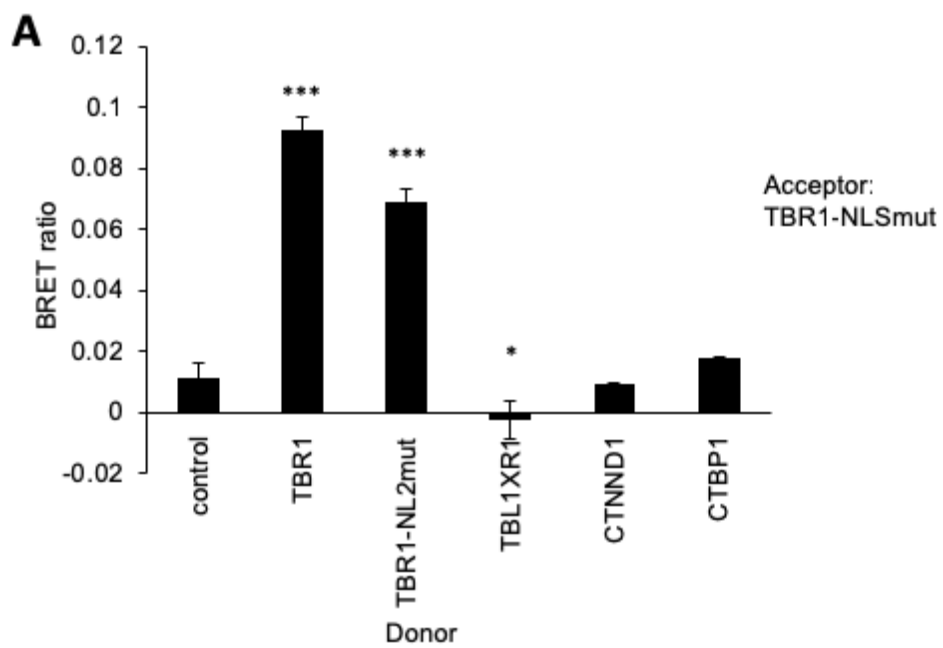


Figure 6.6 (opposite page). Tbr1 is co-expressed with Adnp in developing and adult mouse brain. Immunofluorescence images of mouse brain tissue stained for expression of ADNP (red) and TBR1 (green). (A) Embryonic (E16.5) cortex, scale bar = 50 μm . (B) Adult (P101) cortex, scale bar = 100 μm . (C) Adult (P101) cerebellum, scale bar = 50 μm .

6.3.1.4 *Cytoplasmic localisation of TBR1 does not enable interaction with TBL1XR1, CTNND1 or CTBP1*

Three TBR1 interaction candidates – TBL1XR1, CTNND1 and CTBP1 – were expressed predominantly in the cytoplasm in the current study and displayed no interaction with nuclear-expressed TBR1. However, there is at least one previous report of cytoplasmic TBR1 localisation (Hong and Hsueh, 2007), and it is possible that any potential TBR1-interactions occurring in the cytoplasm might be missed in the preceding BRET assays, where the YFP-TBR1 fusion protein was exclusively nuclear. To test for potential cytoplasmic interactions, a synthetic mutation was introduced in the TBR1 expression plasmid to abolish a predicted NLS (residues 636-639, KR₃RR > GGGG), producing a variant protein with partly cytoplasmic expression (TBR1-NLSmut; Fig 6.7B). The TBR1-NLSmut protein dimerised with itself and with WT TBR1 (Fig 6.7A), but gave no evidence of interaction with TBL1XR1, CTNND1 or CTBP1 (Fig 6.7A) despite some co-localisation (Fig 6.7B). Thus, there is no further evidence that these candidates interact with either nuclear or cytoplasmic TBR1.

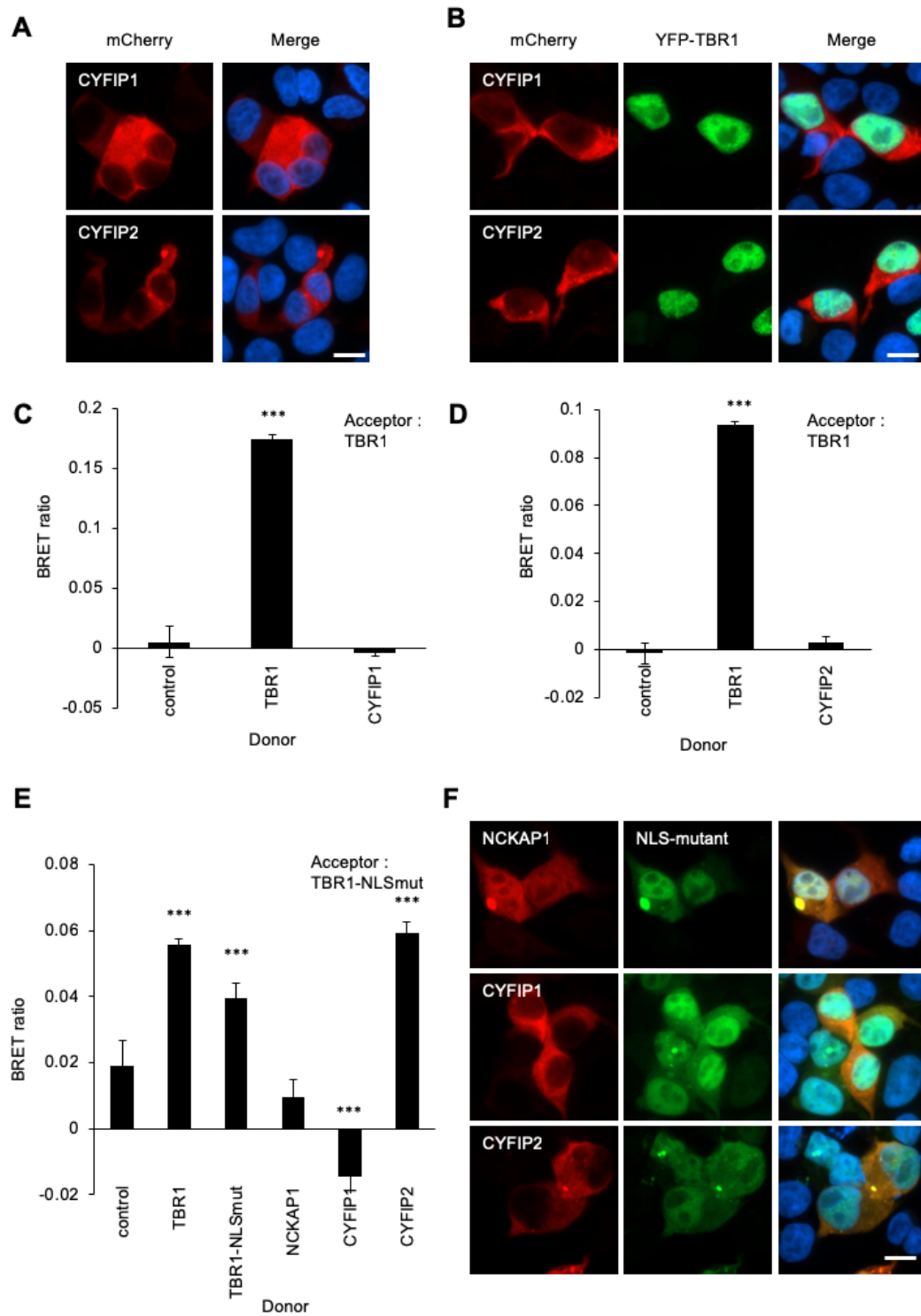
Figure 6.7 (opposite page). Cytoplasmic candidates do not interact with a cytoplasmic mutant form of TBR1. (A) BRET assay for interaction between TBR1-NLS mutant and cytoplasm-expressed interaction candidates. Bars represent the corrected mean BRET ratios \pm SD of one experiment performed in triplicate ($*P < 0.05$ and $***P < 0.001$ compared to control, Student's T-test). (B) Fluorescence microscopy images of HEK293 cells co-transfected with interaction candidates (fused to mCherry, red) and TBR1-NLS mutant (fused to YFP, green). Nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μ m.



6.3.1.5 Cytoplasmic TBR1 variant interacts with CYFIP2

Although an interaction between TBR1 and NCKAP1 could not be verified, it was notable that CYFIP2, which forms a complex with NCKAP1 (Innocenti et al., 2004), was also identified by AP-MS as a potential TBR1 interactor (Chapter 5). The close homologue CYFIP1 was also detected in one of the two AP-MS experiments. I therefore hypothesised that TBR1 may interact with CYFIP1 and/or CYFIP2. Both proteins localised predominantly to the cytoplasm (Fig 6.8A) and did not interact or co-localise with WT TBR1 (Fig 6.8B, 6.8C, 6.8D). In contrast, CYFIP2 did exhibit an interaction with the cytoplasmic TBR1-NLSmut variant, although NCKAP1 and CYFIP1 did not (Fig 6.8E). Both CYFIP2 and NCKAP1 co-localised with TBR1-NLSmut, including in cytoplasmic aggregates (Fig 6.8F). These results indicate that TBR1 could interact with CYFIP2 in the cytoplasm, and that this might form an indirect connection with NCKAP1.

Figure 6.8 (opposite page). Cytoplasmic TBR1 mutant binds to CYFIP2, a known NCKAP1-interaction partner. (A) Fluorescence microscopy images of HEK293 cells transfected with CYFIP1/2 (fused to mCherry, red). Nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μ m. (B) Fluorescence microscopy images of HEK293 cells co-transfected with TBR1 (fused to YFP, green) and CYFIP1/2 (fused to mCherry, red). Nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μ m. (C, D) BRET assays for interaction between TBR1 and CYFIP1/2. Bars represent the corrected mean BRET ratios \pm SD of one experiment performed in triplicate ($***P < 0.001$ compared to control, Student's T-test). (E) BRET assays for interaction between TBR1-NLS mutant and NCKAP1 or CYFIP1/2. Bars represent the corrected mean BRET ratios \pm SD of one experiment performed in triplicate ($***P < 0.001$ compared to control, Student's T-test). (F) Fluorescence microscopy images of HEK293 cells co-transfected with TBR1-NLS mutant (fused to YFP, green) and NCKAP1 or CYFIP1/2 (fused to mCherry, red). Nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μ m.



6.3.2 PART 2: EFFECTS OF TBR1 VARIANTS ON CONFIRMED INTERACTIONS

Further BRET assays were performed to examine whether *TBR1* variants identified in individuals with ASD (Table 6.1) would affect interactions with GATAD2B, ADNP and NR2F1/2. I tested the 6 *de novo* variants (p.A136Pfs*80, p.S351*, p.K228E, p.W271C, p.N374H, p.K389E) and 1 inherited variant (p.Q418R) that have previously been shown to disrupt interaction with CASK, BCL11A and/or FOXP2, hypothesising that these variants might also affect the novel TBR1-interactions. Also included were the 1 *de novo* variant (p.W271R) and 3 inherited variants (p.Q178E, p.V356M, p.P542R) that did not abolish interactions in prior tests, based on the possibility that these variants may affect previously untested interactions. The patient variants as well as two synthetic truncations (p.N394*, p.S568*) were employed to map the GATAD2B-, ADNP- and NR2F1/2-binding sites. Further investigation of the TBR1-BCOR interaction was performed by Joery den Hoed. This work has been described in detail elsewhere (den Hoed, 2016), but is also summarised in Table 6.7.

6.3.2.1 *TBR1* variants affect GATAD2B-interaction

All 5 *de novo* TBR1 missense variants (p.K228E, p.W271R, p.W271C, p.N374H, p.K389E) retained the interaction with GATAD2B, and the p.K228E and p.N374H variants exhibited a slightly increased BRET ratio compared to WT TBR1 (Fig 6.9A; Table 6.7). Both TBR1 and GATAD2B occurred in nuclear speckles/aggregates when these were present (Fig 6.9B). It is unclear whether this pattern represents the normal speckled expression of WT GATAD2B, or whether GATAD2B may be mislocalised by interaction with abnormally aggregated TBR1 variants (p.K228E, p.W271C, p.N374H, p.K389E). If the latter interpretation is correct, then these *de novo* TBR1 missense variants may interfere with the function of GATAD2B, echoing similar dominant negative effects on WT TBR1 and BCL11A (Deriziotis et al., 2014b; den Hoed et al., 2018). Three of the rare inherited TBR1 variants (p.Q178E, p.V265M, p.P524R) also interacted and co-localised with GATAD2B (Fig 6.9A, 6.9B; Table 6.7). However, the p.Q418R variant abolished the GATAD2B-interaction (Fig 6.9A; Table 6.7), as has been previously demonstrated for FOXP2- (Deriziotis et al., 2014b) and BCL11A-interaction (den Hoed et al., 2018). This lends further support for a pathogenic role for p.Q418R in neurodevelopmental disorder, through the disruption of multiple protein-protein interactions.

The TBR1-GATAD2B interaction was abolished by the *de novo* truncating TBR1 variants (p.A136Pfs*80, p.S351*) (Fig 6.10A; Table 6.7). While both truncations were present in the nucleus, they did not appear to co-localise with GATAD2B in nuclear speckles (Fig 6.10B). Synthetic TBR1 truncations (p.N394* and p.S568*) were used to map the GATAD2B-binding site. The p.N394* truncation, which terminates immediately after the T-box domain, did not interact with GATAD2B, while the longer p.S568* truncated protein exhibited a strong interaction (Fig 6.10A). In agreement with this finding, only p.S568* exhibited punctate co-localisation with GATAD2B in the nucleus (Fig 6.10B). These results indicate that residues 394-568 of TBR1 are required for interaction with GATAD2B (Fig

6.17). Notably, identical or overlapping regions have been identified as important for CASK-binding (342-682) and BCL11A-binding (394-568), as well as TBR1 homodimerisation (394-568) (Hsueh et al., 2000; Deriziotis et al., 2014b; den Hoed et al., 2018). The p.Q418R variant also lies within this region and may disrupt the GATAD2B-binding site. Because the TBR1-GATAD2B interaction tolerates multiple missense variants within the T-box domain, it is likely that the T-box is not important for GATAD2B-binding.

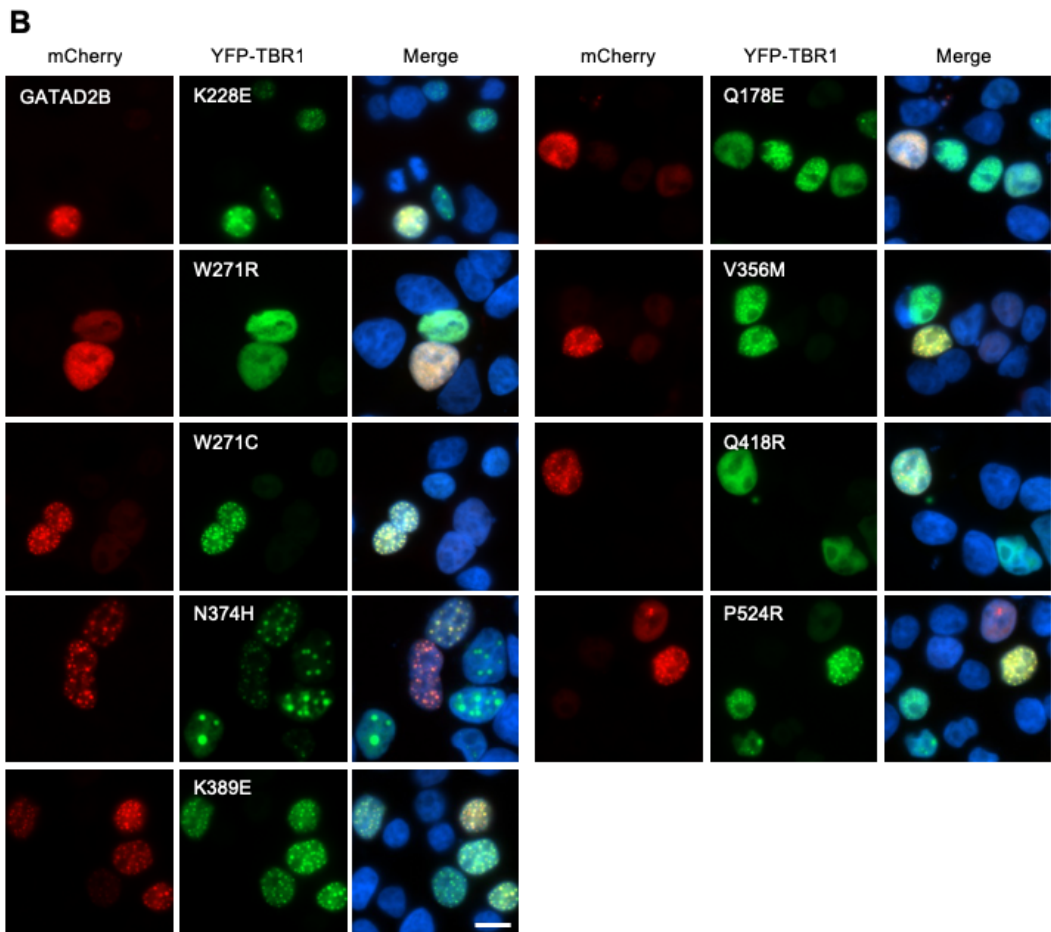
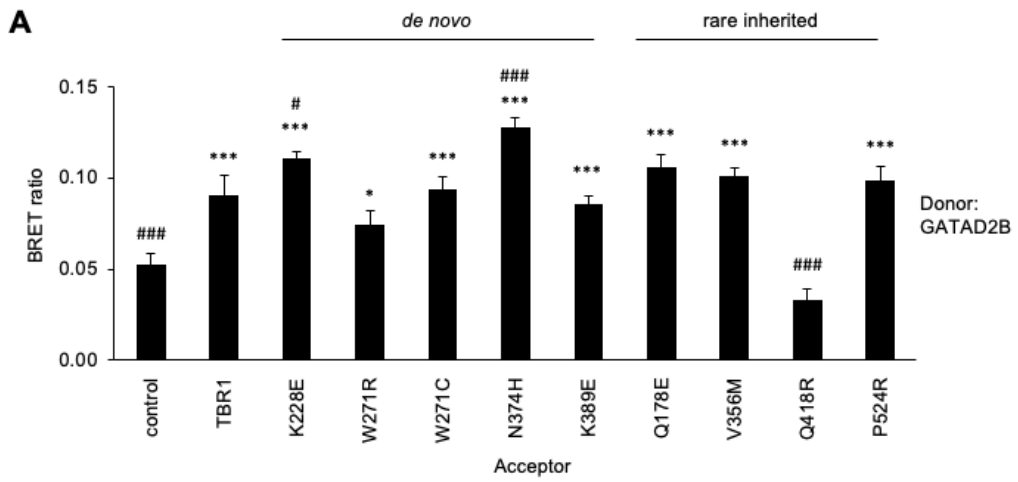


Figure 6.9 (opposite page). The majority of *TBR1* missense variants do not disrupt interaction with GATAD2B. (A) BRET assay for interaction between *de novo* and rare inherited *TBR1* missense variants and GATAD2B. Bars represent the corrected mean BRET ratios \pm SD of one experiment performed in triplicate (* $P < 0.05$ and *** $P < 0.001$ compared to control, # $P < 0.05$ and ### $P < 0.001$ compared to WT *TBR1*, one-way ANOVA and *post-hoc* Tukey's test). (B) Fluorescence microscopy images of HEK293 cells co-transfected with GATAD2B (fused to mCherry, red) and *TBR1* variants (fused to YFP, green). Nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μ m.

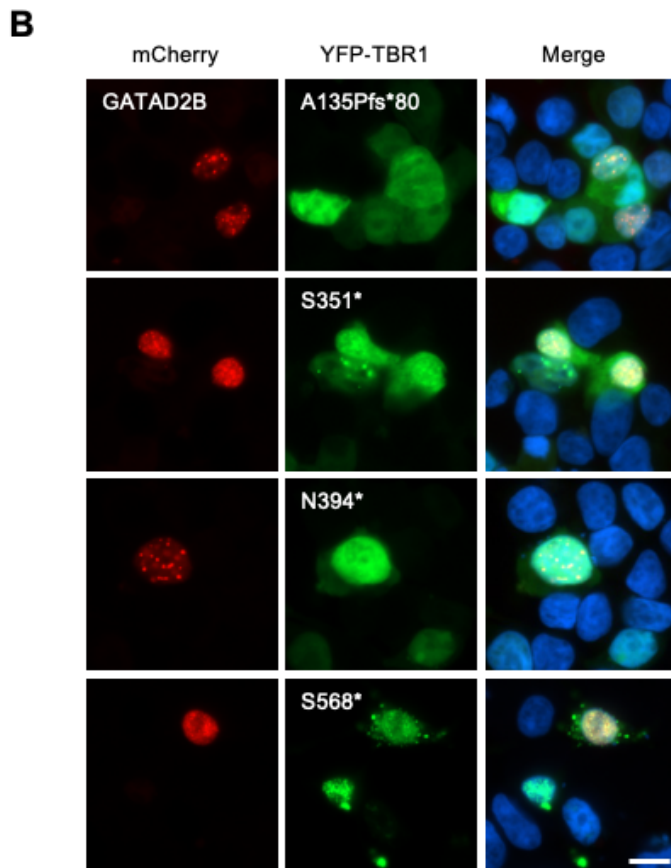
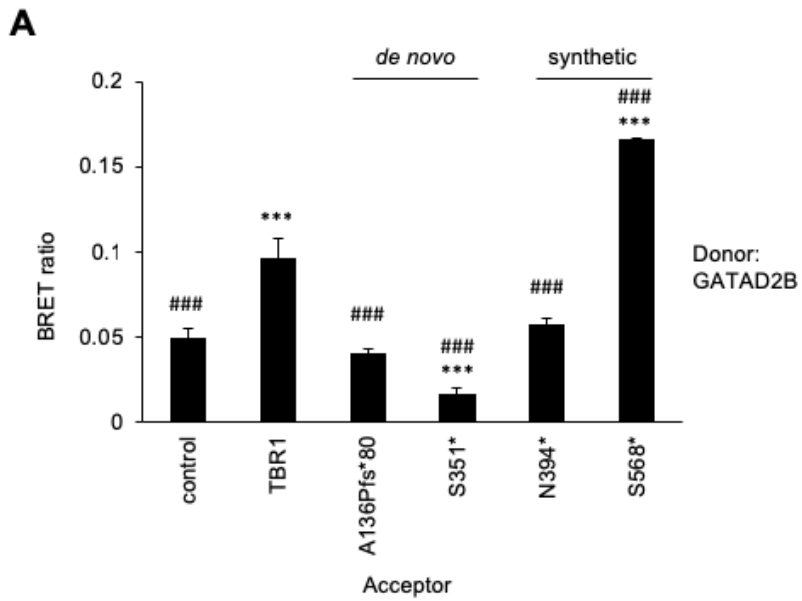


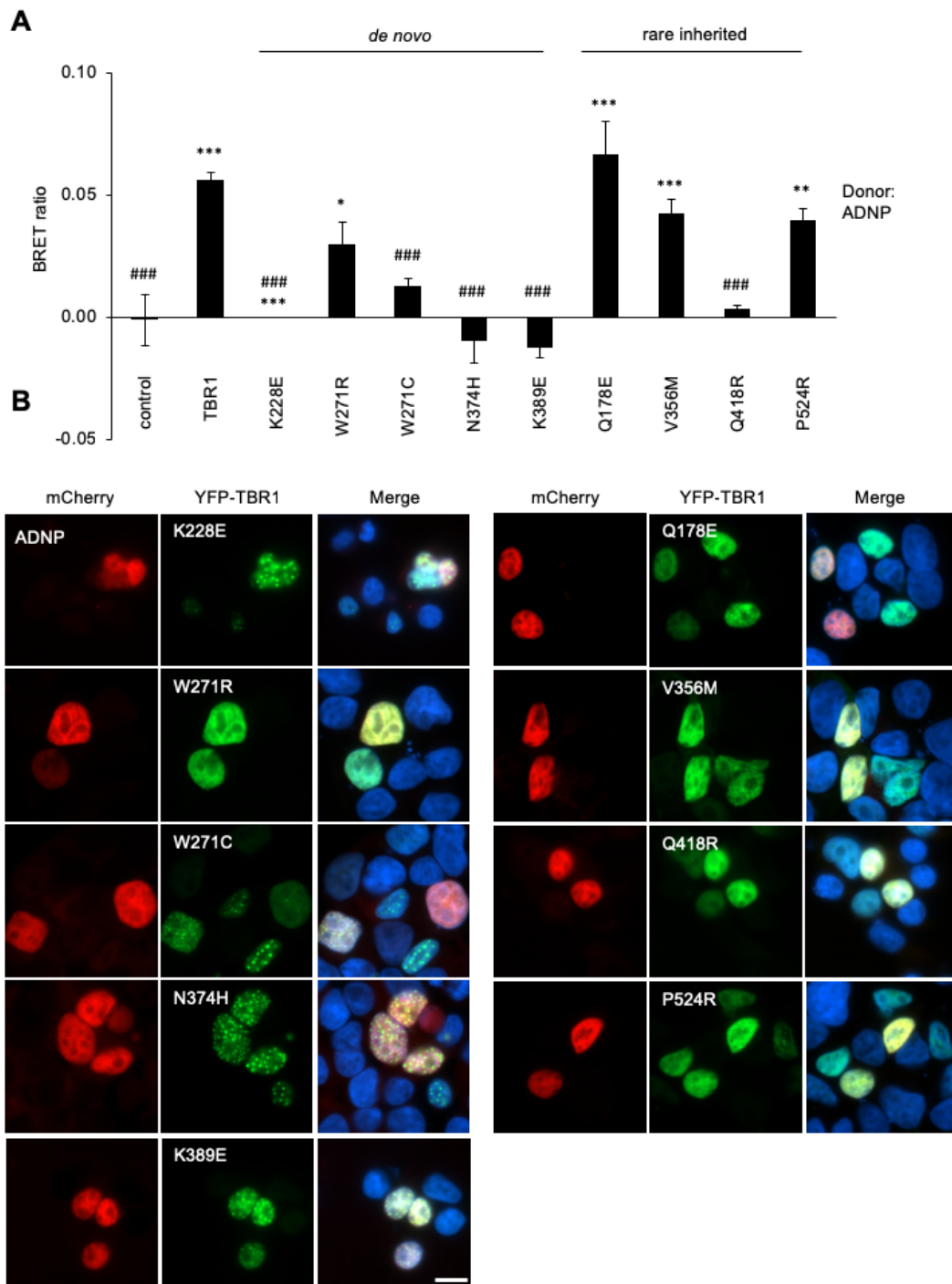
Figure 6.10 (opposite page). TBR1 truncations disrupt interaction with GATAD2B. (A) BRET assay for interaction between *de novo* patient-derived and synthetic TBR1 truncations and GATAD2B. Bars represent the corrected mean BRET ratios \pm SD of one experiment performed in triplicate ($***P < 0.001$ compared to control, $####P < 0.001$ compared to WT TBR1, one-way ANOVA and *post-hoc* Tukey's test). (B) Fluorescence microscopy images of HEK293 cells co-transfected with GATAD2B (fused to mCherry, red) and TBR1 variants (fused to YFP, green). Nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μ m.

6.3.2.2 *TBR1* variants affect ADNP-interaction

Four *de novo* TBR1 missense variants (p.K228E, p.W271C, p.N374H, p.K389E) abolished interaction with ADNP (Fig 6.11A; Table 6.7). These variants also formed nuclear aggregates, while ADNP retained a diffuse nuclear expression pattern (Fig 6.11B). The loss of ADNP interaction could disrupt potential shared functions of TBR1 and ADNP and might contribute to the ASD phenotype in these patients. Notably, the same variants had a very similar effect on FOXP1/2-interaction (Deriziotis et al., 2014b; den Hoed et al., 2018). Only the p.W271R variant retained the interaction, setting it apart from the p.W271C variant at the same residue (Fig 6.11A). The same distinction was previously seen in FOXP1/2-binding experiments, where p.W271C but not p.W271R abolished the interaction (den Hoed et al., 2018). Three rare inherited variants (p.Q178E, p.V356M, p.P524R) interacted and co-localised normally with ADNP (Fig 6.11A, 6.11B; Table 6.7), providing further evidence against a pathogenic role for these variants. However, the interaction was disturbed by p.Q418R (Fig 6.11A; Table 6.7), as has also been seen for FOXP2 (Deriziotis et al., 2014b), BCL11A (den Hoed et al., 2018) and now GATAD2B (Fig 6.9A).

The TBR1-ADNP interaction was also abolished by the *de novo* truncating TBR1 variants (p.A136Pfs*80, p.S351*; Fig 6.12A; Table 6.7), which lack all or part of the T-box domain. On the other hand, the two synthetic TBR1 truncations (p.N394*, p.S568*) did interact with ADNP (Fig 6.12A; Table 6.7). Nonetheless, ADNP remained consistently nuclear, even when co-expressed with the partly cytoplasmic TBR1 truncations (Fig 6.12B). Overall, these findings suggest that the T-box domain, rather than the C-terminal region, is primary involved in ADNP-binding (Fig 6.17). The fact that the p.Q418R variant also abolished the interaction may indicate a second binding site in that region. Alternatively, the p.Q418R variant may lead to conformational changes that affect the structure of the protein as a whole, rather than blocking a specific binding surface.

Figure 6.11 (opposite page). *De novo* missense variants in TBR1 disrupt interaction with ADNP. (A) BRET assay for interaction between TBR1 variants and ADNP. Bars represent the corrected mean BRET ratios \pm SD of one experiment performed in triplicate (* $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$ compared to control, #### $P < 0.001$ compared to WT TBR1, one-way ANOVA and *post-hoc* Tukey's test). (B) Fluorescence microscopy images of HEK293 cells co-transfected with ADNP (fused to mCherry, red) and TBR1 variants (fused to YFP, green). Nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μ m.



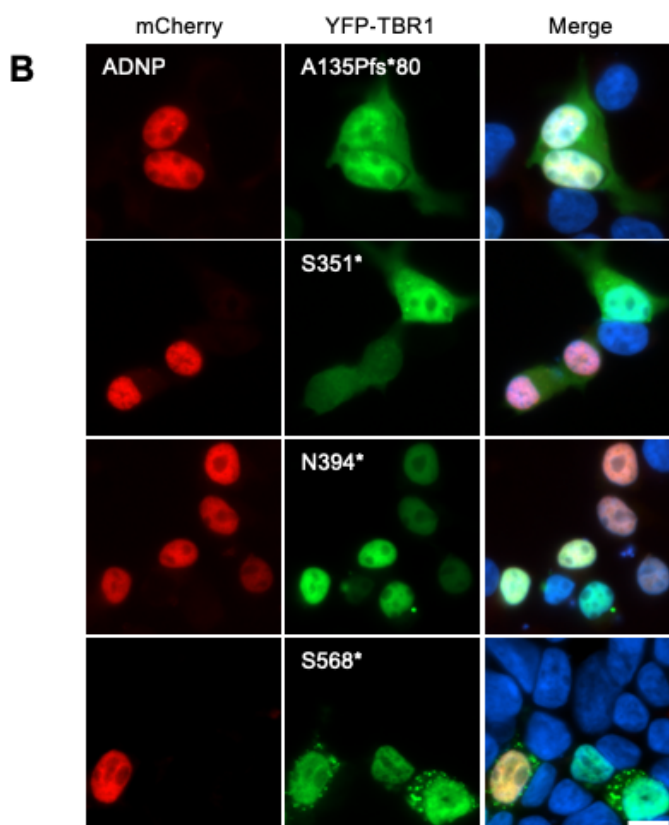
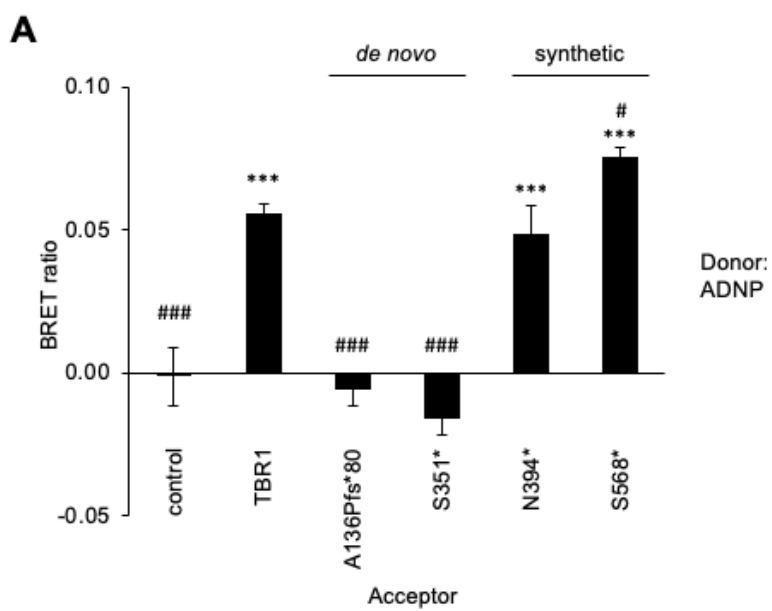


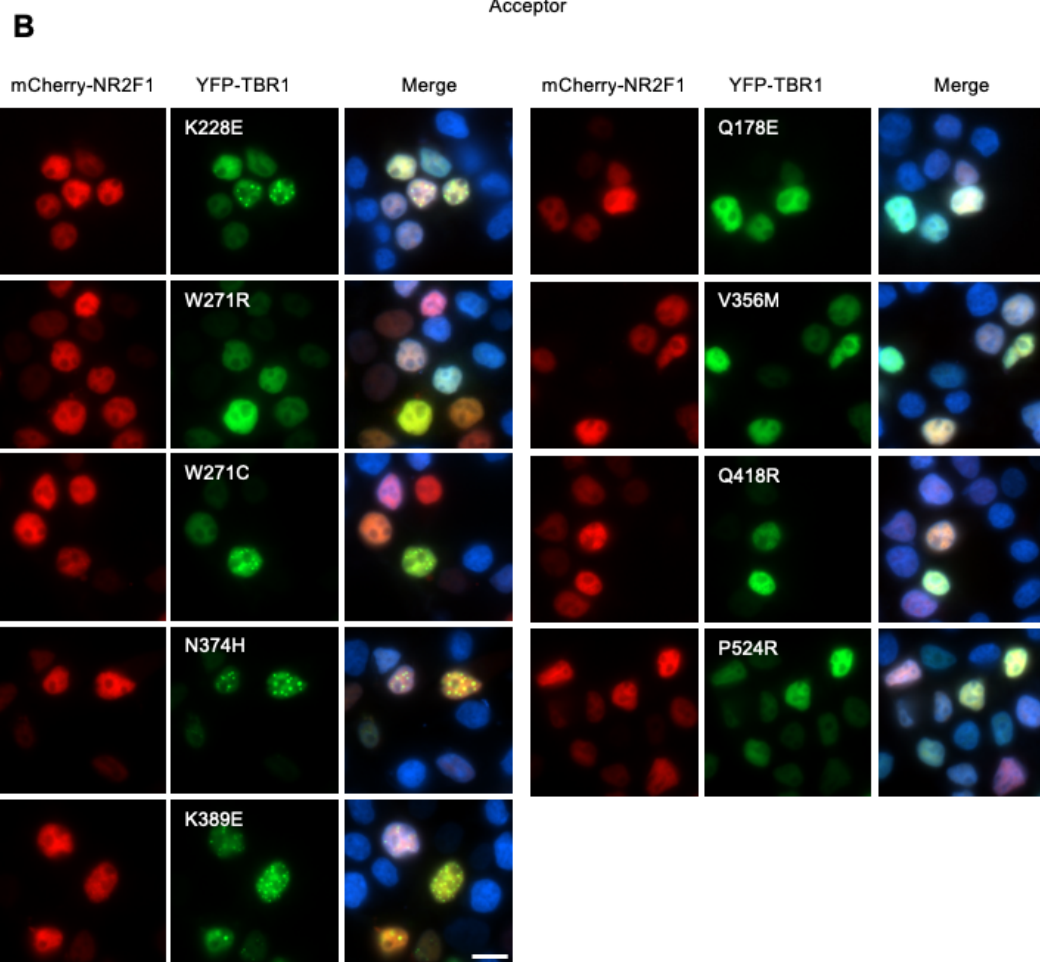
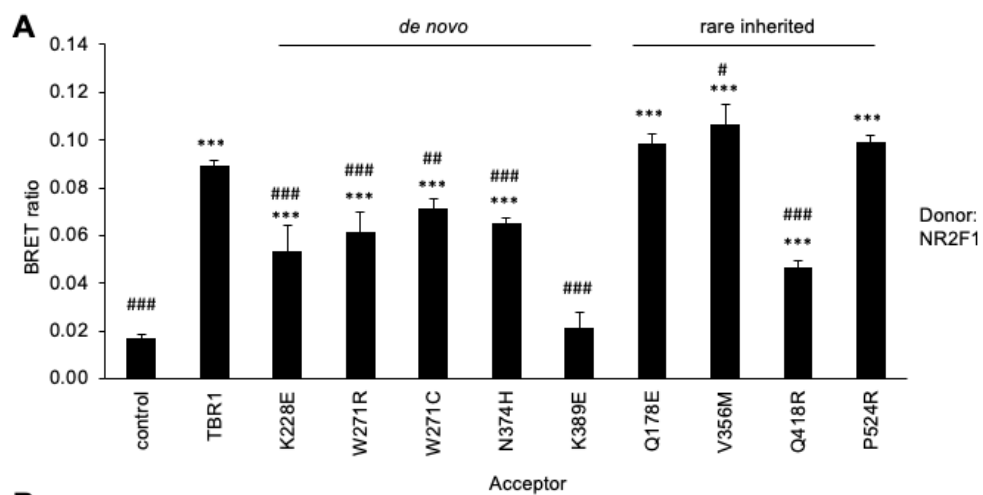
Figure 6.12 (opposite page). The T-box domain of TBR1 is required for interaction with ADNP. (A) BRET assay for interaction between *de novo* patient-derived and synthetic TBR1 truncations and ADNP. Bars represent the corrected mean BRET ratios \pm SD of one experiment performed in triplicate (** $P < 0.001$ compared to control, # $P < 0.01$ and #### $P < 0.001$ compared to WT TBR1, one-way ANOVA and *post-hoc* Tukey's test). (B) Fluorescence microscopy images of HEK293 cells co-transfected with ADNP (fused to mCherry, red) and TBR1 variants (fused to YFP, green). Nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μ m.

6.3.2.3 *TBR1* variants affect NR2F1/2-interaction

The p.K389E variant abolished interactions with NR2F1 and NR2F2 while the other *de novo* TBR1 missense variants exhibited a slightly reduced interaction signal: p.K228E (significant for NR2F1 only), p.W271R, p.W271C, p.N374H (Fig 6.13A, 6.14A; Table 6.7). Notably, NR2F1/2 consistently exhibited diffuse nuclear localisation, and did not appear to be mislocalised in the presence of aggregate-forming *de novo* TBR1 variants (p.K228E, p.W271C, p.N374H, p.K389E) (Fig 6.13B, 6.14B). The rare inherited variant p.Q418R decreased, but did not completely abolish, the interactions with NR2F1/2, while the other rare inherited variants retained these interactions (Fig 6.13A, 6.15A; Table 6.7).

The *de novo* truncating TBR1 variants (p.A136Pfs*80, p.S351*) both abolished the TBR1-NR2F1/2 interactions (Fig 6.15A, 6.16A; Table 6.7). The shorter synthetic truncated protein (p. N394*), which terminates just after the T-box, exhibited reduced interaction, whereas the longer synthetic truncated protein (p.S568*) exhibited strong interaction (Fig 6.15A, 6.16A; Table 6.7). NR2F1 and NR2F2 were exclusively nuclear in all co-transfection conditions (Fig. 6.15B, 6.16B). These data suggest that the TBR1-NR2F1/2 interactions may involve both the T-box domain and the C-terminal region (residues 394-568) of TBR1 (Fig 6.17). Within the T-box, however, different residues (e.g. p.K389) may be of greater importance for interaction with NR2F1/2, compared to those involved in FOXP2- and ADNP-binding.

Figure 6.13 (opposite page). The p.K389E TBR1 variant disrupts interaction with NR2F1. (A) BRET assay for interaction between *de novo* and rare inherited TBR1 missense variants and NR2F1. Bars represent the corrected mean BRET ratios \pm SD of one experiment performed in triplicate (*** P < 0.001 compared to control, # P < 0.05 and #### P < 0.001 compared to WT TBR1, one-way ANOVA and *post-hoc* Tukey's test). **(B)** Fluorescence microscopy images of HEK293 cells co-transfected with NR2F1 (fused to mCherry, red) and TBR1 variants (fused to YFP, green). Nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μ m.



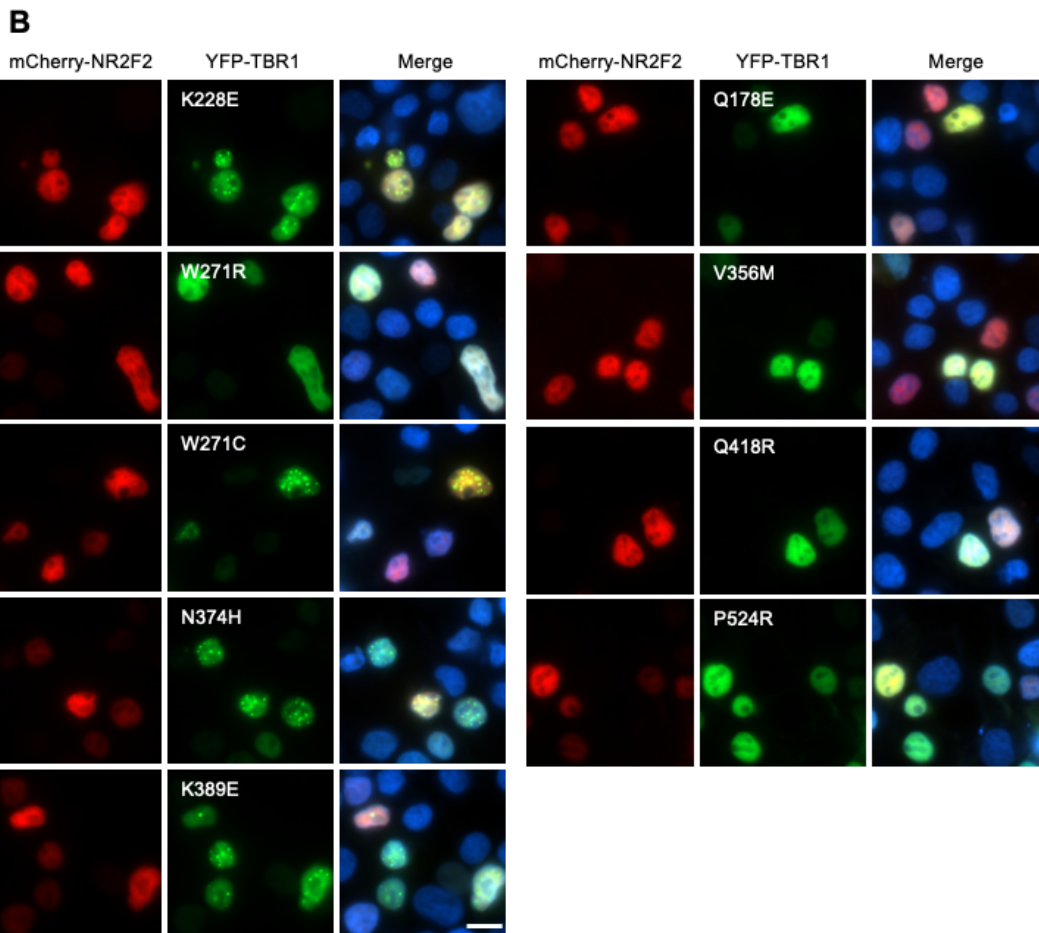
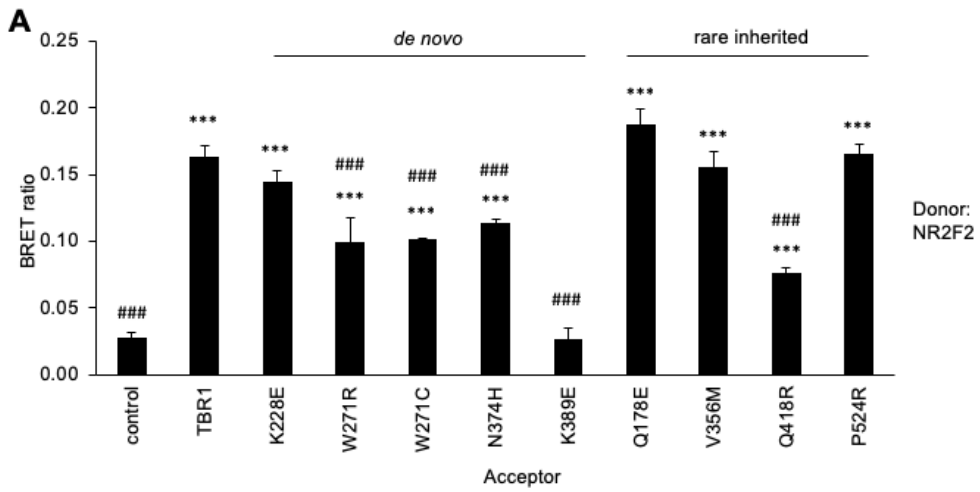


Figure 6.14 (opposite page). The p.K389E TBR1 variant disrupts interaction with NR2F2. (A) BRET assay for interaction between *de novo* and rare inherited TBR1 missense variants and NR2F2. Bars represent the corrected mean BRET ratios \pm SD of one experiment performed in triplicate (*** $P < 0.001$ compared to control, #### $P < 0.001$ compared to WT TBR1, one-way ANOVA and *post-hoc* Tukey's test). **(B)** Fluorescence microscopy images of HEK293 cells co-transfected with NR2F2 (fused to mCherry, red) and TBR1 variants (fused to YFP, green). Nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μ m.

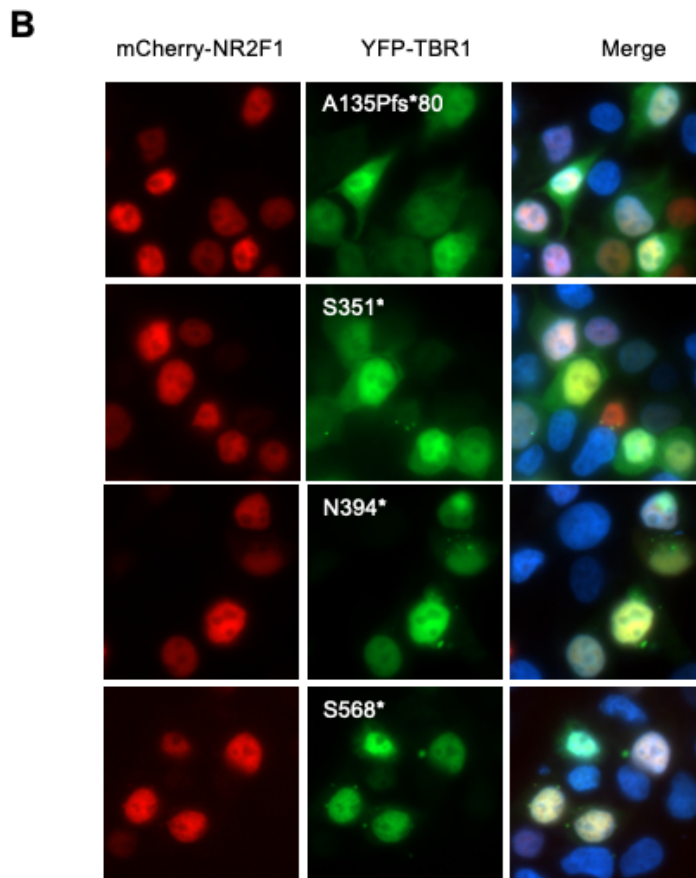
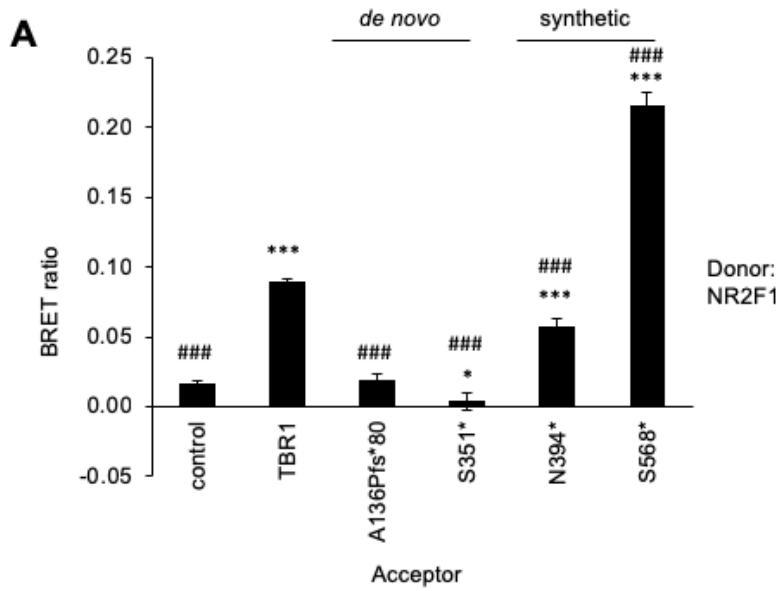


Figure 6.15 (opposite page). The T-box domain of TBR1 is required for interaction with NR2F1. (A) BRET assay for interaction between *de novo* patient-derived and synthetic TBR1 truncations and NR2F1. Bars represent the corrected mean BRET ratios \pm SD of one experiment performed in triplicate ($*P < 0.05$ and $***P < 0.001$ compared to control, $####P < 0.001$ compared to WT TBR1, one-way ANOVA and *post-hoc* Tukey's test). **(B)** Fluorescence microscopy images of HEK293 cells co-transfected with NR2F1 (fused to mCherry, red) and TBR1 variants (fused to YFP, green). Nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μ m.

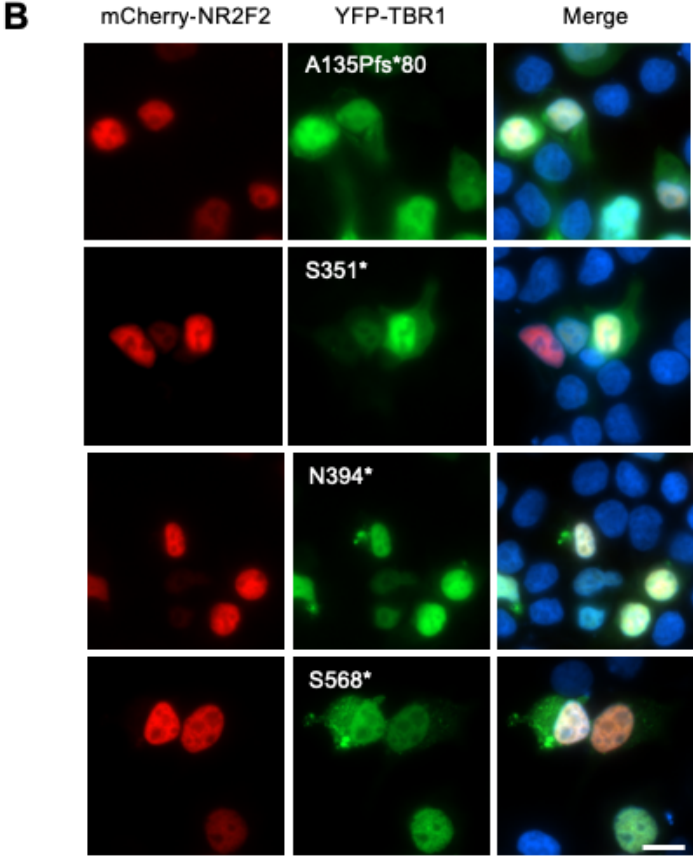
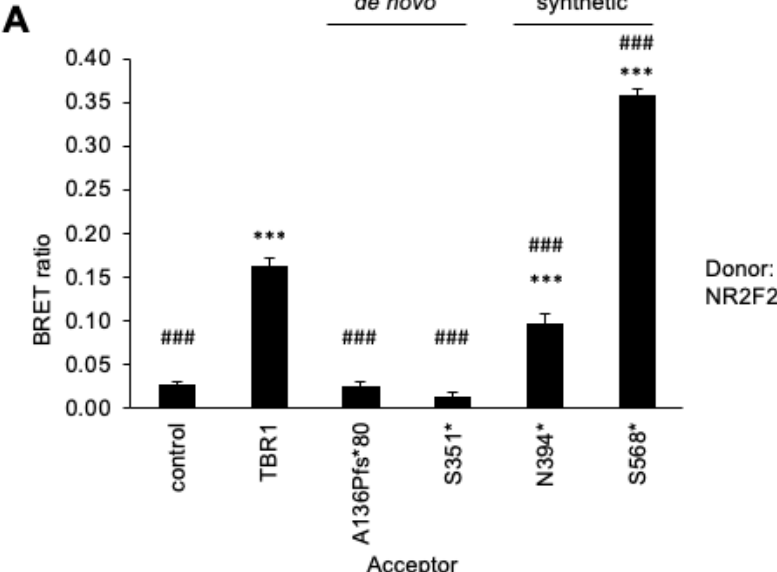


Figure 6.16 (opposite page). The T-box domain of TBR1 is required for interaction with NR2F2. (A) BRET assay for interaction between *de novo* patient-derived and synthetic TBR1 truncations and NR2F2. Bars represent the corrected mean BRET ratios \pm SD of one experiment performed in triplicate ($***P < 0.001$ compared to control, $###P < 0.001$ compared to WT TBR1, one-way ANOVA and *post-hoc* Tukey's test). (B) Fluorescence microscopy images of HEK293 cells co-transfected with NR2F2 (fused to mCherry, red) and TBR1 variants (fused to YFP, green). Nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μ m.

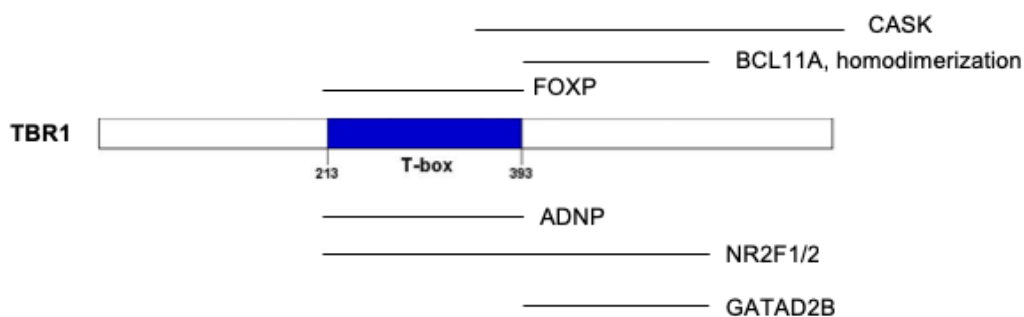


Figure 6.17 – Proposed binding sites for novel TBR1-interactors. Schematic representation of TBR1 protein. Horizontal bars indicate the proposed binding regions for the novel interaction partners ADNP, GATAD2B and NR2F1/2. Binding regions for previously established interactions are shown for comparison (Hsueh et al., 2000; Deriziotis et al., 2014b; den Hoed et al., 2018).

Table 6.7. Effects of *TBR1* variants on novel and previously established interactions.

Protein	Pred	Ref	Loc	Repr	Interactions								
					CASK	TBR1	FOXP2	BCL11A	BCOR	GATAD2B	ADNP	NR2F1	NR2F2
<i>de novo</i> missense													
p.K228E	Prob. 0.986	1	N*	+	Green	Green	Red	Green	Green	Green	Red	Yellow	Green
p.W271R	Prob. 0.999	2	N	+	Green	Yellow	Green	Yellow	Green	Green	Green	Yellow	Yellow
p.W271C	Prob. 0.999	3	N*	+	Green	Green	Red	Green	Green	Green	Red	Yellow	Yellow
p.N374H	Prob. 0.990	4	N*	+	Green	Green	Red	Green	Green	Green	Red	Yellow	Yellow
p.K389E	Prob. 0.973	3	N*	+	Green	Yellow	Red	Green	Green	Green	Red	Yellow	Yellow
rare inherited missense													
p.Q178E	Benign 0.178	1	N	+	Green	Green	Green	Green	Green	Green	Green	Green	Green
p.V356M	Prob. 0.961	5	N	+	Green	Green	Green	Green	Green	Green	Green	Green	Green
p.Q418R	Benign 0.043	1	N	+	Green	Green	Red	Green	Yellow	Red	Green	Yellow	Green
p.P542R	Prob. 0.992	1	N	+	Green	Green	Green	Green	Green	Green	Green	Green	Green
<i>de novo</i> truncations													
p.A136Pfs*80	NA	1	N+C	-	Red	Red	Red	Red	Red	Red	Red	Red	Red
p.S351*	NA	1	N+C	-	Red	Red	Red	Red	Red	Red	Red	Red	Red
synthetic truncations													
p.N394*	NA	1	N+C	-	Grey	Yellow	Green	Green	Yellow	Red	Green	Yellow	Green
p.S568*	NA	1	N+C	-	Grey	Green	Green	Green	Yellow	Red	Green	Yellow	Green

Variables are annotated based on protein reference NP_006584.1. PolyPhen2 (HumVar) predictions: prob. = probably damaging, NA = not applicable. References: ¹Deriziotis et al. (2014b), ²Hamdan et al. (2014), ³O'Roak et al. (2014), ⁴Neale et al. (2012), ⁵Bacchelli et al. (2003). Localisation: N = nuclear, C = cytoplasmic, N+C = mixed localisation, * = forms aggregates. Repr = repressive function, intact (+) or impaired (-) Interactions: green = interaction; red = no interaction; orange = decreased interaction signal; grey = no data.

6.3.3 PART 3: EFFECTS OF GATAD2B AND ADNP VARIANTS ON TBR1-INTERACTION

6.3.3.1 Pathogenic GATAD2B variants disrupt interaction with TBR1

As detailed in Chapter 4 of this thesis, GATAD2B is a transcriptional repressor and member of the NuRD chromatin-remodelling complex. It contains two main functional domains: CR1 (residues 165-195), required for incorporation of GATAD2B into the NuRD complex through protein-protein interactions with MBD2/3 (Brackertz et al., 2002); and CR2 (residues 340-480), which contains a GATA-type zinc finger (residues 414-467), and is involved in localising the NuRD complex to specific loci (Feng et al., 2002). GATAD2B variants cause a syndrome involving ID, motor delay, poor speech, neonatal hypotonia and characteristic facial features (OMIM 615074). All reported variants to date are heterozygous *de novo* truncations that disrupt one or both of the functional domains (de Ligt et al., 2012; Willemsen et al., 2013; Luo et al., 2017). These include two frameshifts (p.Q190Afs*34 and p.N195Kfs*30) that truncate the protein at the border of the CR1 domain, and a nonsense mutation (p.Q470*) that terminates within the CR2 domain, but after the zinc finger (Fig 6.18A) (de Ligt et al., 2012; Willemsen et al., 2013). Here, the effect of these pathogenic variants on interaction with TBR1 was assessed.

All three GATAD2B variants produced truncated proteins at the expected sizes and inhibited the normal speckled nuclear localisation of the WT protein (Fig 6.18B), as previously shown in Chapter 4 of this thesis (Fig 4.6). The short frameshift variants (p.Q190Afs*34, p.N195Kfs*30) abolished interaction with TBR1, while the p.Q470* variant exhibited a reduced interaction (Fig 6.18C). Each of the variants co-localised with TBR1 in the nucleus, but without the normal speckled pattern seen with WT GATAD2B (Fig 6.18D). While all of the GATAD2B variants tested are likely to undergo NMD, the data presented here indicate that any surviving variant GATAD2B protein would have impaired TBR1-binding function. It is tempting to speculate that this might impair the recruitment of the NuRD complex to TBR1 target loci and potentially disrupt transcriptional repression.

Previous work identified the CR2 domain (residues 340-480; see Fig 4.6) of GATAD2B as essential for interaction with FOXP TFs (Chokas et al., 2010) (Chapter 4). The current study suggests that CR2 is also important for GATAD2B-TBR1 interaction. However, unlike the GATAD2B-FOXP interaction (Fig 4.7), loss of the final portion of the domain (470-480) did not entirely abolish the interaction with TBR1. Because both of the GATAD2B frameshift variants (p.Q190Afs*34, p.N195Kfs*30) also disrupt the final portion of CR1, it is possible that an intact CR1 domain might be sufficient for the observed partial interaction. I therefore tested a synthetic p.T200* GATAD2B truncation that terminates several amino acids after the end of the CR1 domain. This truncation also completely abolished the interaction, indicating that CR1 is insufficient for TBR1-binding (Fig. 6.18E). Thus, both FOXP- and TBR1-interaction are likely to involve the CR2 domain of GATAD2B, consistent with its established role in directing the NuRD complex to target loci (Brackertz et al., 2006).

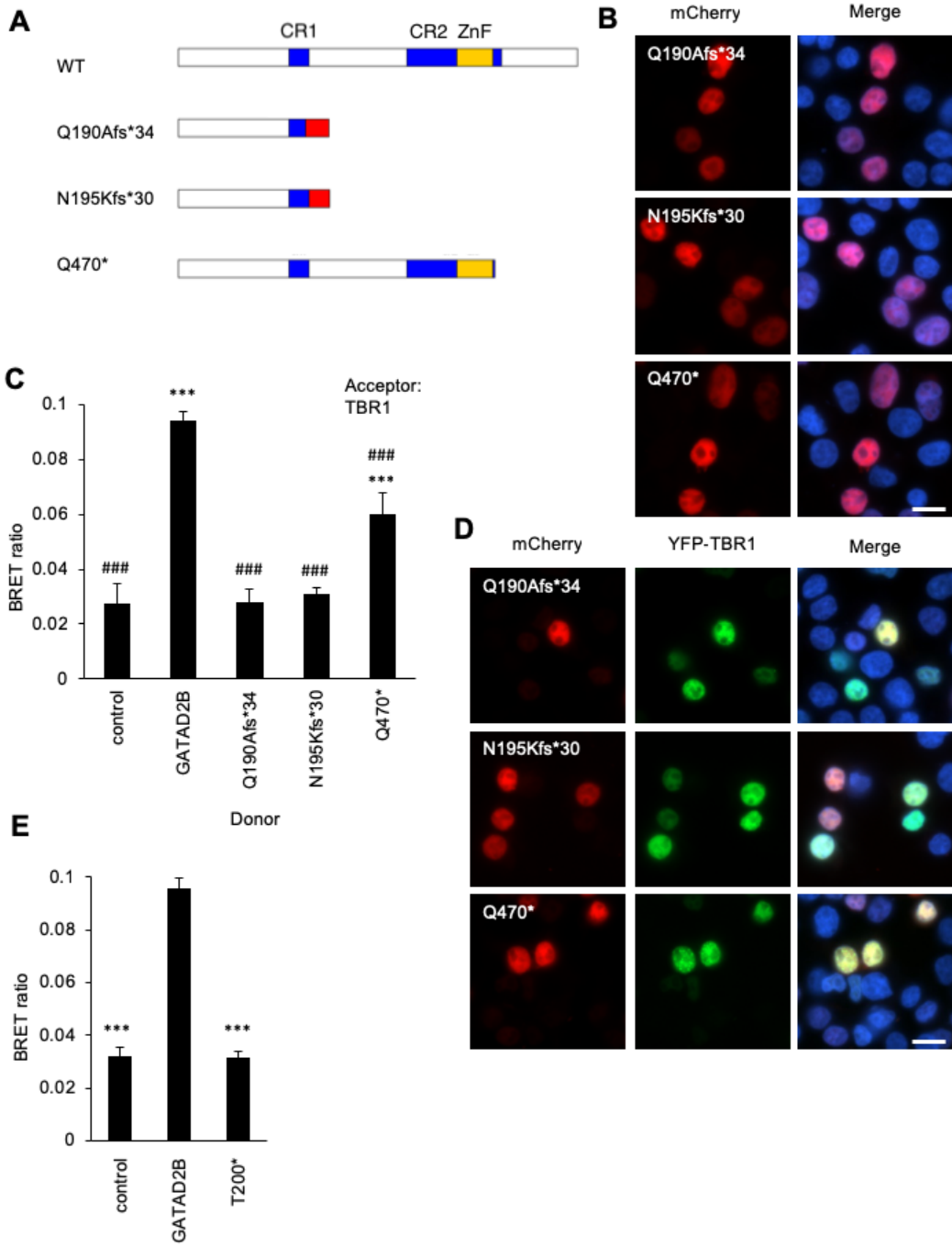


Figure 6.18 (opposite page). Pathogenic truncating variants in GATAD2B disrupt interaction with TBR1. (A) Schematic representations of GATAD2B variants used in this assay, displaying functional domains: conserved regions 1 and 2 (CR1/2) and GATA-type zinc finger (ZnF). The frameshift variants p.Q190Afs*34 and p.N195Kfs*30 disrupt the last few amino acids of CR1, while p.Q470* terminates after the ZnF, within CR2. Red segments indicate alternative amino acid sequences generated by the frameshifts. (B) Fluorescence microscopy images of HEK293 cells transfected with GATAD2B variants (fused to mCherry, red). Nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μ m. (C) BRET assay for interaction between TBR1 and GATAD2B variants. Bars represent the corrected mean BRET ratios \pm SD of one experiment performed in triplicate (** $P < 0.001$ compared to control, #### $P < 0.001$ compared to WT GATAD2B, one-way ANOVA and *post-hoc* Tukey's test). (D) Fluorescence microscopy images of HEK293 cells co-transfected with GATAD2B variants (fused to mCherry, red) and WT TBR1 (fused to YFP, green). Nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μ m. (E) BRET assay for interaction between TBR1 and GATAD2B p.T200* synthetic truncation. Bars represent the corrected mean BRET ratios \pm SD of one experiment performed in triplicate (** $P < 0.001$ compared to WT GATAD2B, Student's T-test).

6.3.3.2 *ADNP* zinc finger domains required for *TBR1*-interaction

Next, I investigated variants in *ADNP* and their effects on *TBR1*-interaction. *ADNP* is a TF and interaction partner of the BAF chromatin-remodelling complex. It contains several functional domains, including an 8-amino acid neuroprotective peptide; a DNA-binding homeobox domain; and 9 C2H2-type zinc fingers, which have been hypothesised to assist with DNA-binding (Vandeweyer et al., 2014) (Fig 6.19A). The C-terminal portion of the protein is important for interaction with the BAF chromatin-remodelling complex (Mandel and Gozes, 2007).

ADNP is a high-confidence ASD candidate and variants have been estimated to account for 0.17% of ASD cases (Helsmoortel et al., 2014). Specifically, *ADNP* mutations cause Helsmoortel-van der Aa syndrome (HVDAS; OMIM 615873), a disorder characterised by ASD, mild-to-severe ID, speech impairment and developmental delay, often accompanied by hypotonia, feeding difficulties, visual problems and recurrent infections (Helsmoortel et al., 2014). All of the *ADNP* variants reported in HVDAS are heterozygous truncations: including the four nonsense mutations characterised here (p.S404*, p.R644*, p.R730*, p.L1057*), which lead to truncated forms of *ADNP* of various lengths (Table 6.8; Fig 6.19A). Interestingly, almost the entire length of the *ADNP* coding sequence (encoding residues 68-1102) lies on the final exon. All four of the nonsense mutations studied in this chapter occur within this final exon, and could therefore escape NMD (Vandeweyer et al., 2014). Also included were two rare missense variants of unknown inheritance status (p.T443A and p.G1094V), identified in ASD patients who do not fit the typical HVDAS phenotype (Table 6.8).

The effect of these 6 *ADNP* variants has not previously been assessed at the molecular level. Here, all variants produced proteins at the expected molecular weight (Fig 6.19B). The two missense *ADNP* variants retained a nuclear pattern of expression (Fig 6.19C), co-localised with *TBR1*, and showed normal interactions (Fig 6.20). The two longer truncated *ADNP* variants (p.R730* and p.L1057*) were also nuclear (Fig 6.19C) and interacted with *TBR1* (Fig 6.20). In contrast, the shorter truncated variants (p.S404* and p.R644*) exhibited both nuclear and cytoplasmic expression (Fig 6.19C). Paradoxically, the p.R644* truncation clearly abolished interaction with *TBR1*, while the shorter p.S404* truncation did not significantly affect the interaction (Fig 6.20). One possible explanation could be that the p.R644* truncation exposes a region that alters the conformation of *ADNP* or otherwise blocks interaction with *TBR1*, while the p.S404* truncation removes the inhibiting region and allows some degree of interaction to take place. Alternatively, given that the apparent S404* interaction exhibits a lower BRET signal and is less significant than the other interacting *ADNP* variants, the validity of this interaction is less confident and could be a false positive result. It may be noted that expression of *ADNP* was consistently weak (and evident in a smaller proportion of cells) compared to other proteins tested in this chapter (Fig 6.2A), which may contribute to a greater degree of background noise and therefore an increased potential for false results. Further experiments, perhaps using an alternative technique, may help to clarify this contradiction in the future.

Putting the uncertain p.S404* result to one side, the results here seem to identify ADNP residues 644-730 as an important TBR1-binding region. This region encompasses the 9th and part of the 8th ZnF-domains. The exact functions of the ZnF motifs in ADNP are not well understood, although such motifs are known to be involved in both DNA and protein binding (Cassandri et al., 2017). It is therefore not clear whether these motifs interact directly with TBR1, or if ZnF-mediated DNA-binding is a prerequisite for interaction with TBR1. The cytoplasmic mislocalisation of truncations lacking these ZnFs may indicate disturbed DNA-binding. Further studies using N-terminal truncations could confirm whether the earlier ZnFs are also necessary for nuclear localisation and/or TBR1-interaction - it may be that the full set is required. Interestingly, the homeobox domain, missing in p.R730*, does not appear to be necessary for nuclear localisation or interaction with TBR1.

Table 6.8. Summary of ADNP variants tested in this chapter.

Variant	Ref	Inheritance	Phenotype
T443A	1	unknown (parental DNA unavailable)	high-functioning ASD
G1094V	2	unknown (parental DNA unavailable)	language/developmental delay, ADD, mild hypotonia
S404*	3	<i>de novo</i>	HVDAS
R644*	3	<i>de novo</i>	HVDAS
R730*	4	<i>de novo</i>	HVDAS
L1057*	5	<i>de novo</i>	HVDAS

Variants are annotated based on protein reference NP_001269460.1. References: ¹Alvarez-Mora et al. (2016); ²D’Gama et al. (2015); ³Helsmoortel et al. (2014); ⁴Krajewska-Walasek et al. (2016); ⁵O’Roak et al. (2014). HVDAS = Helsmoortel-van der Aa syndrome.

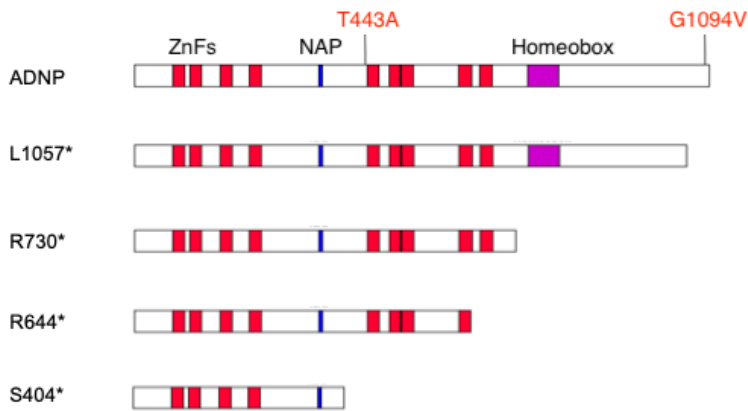
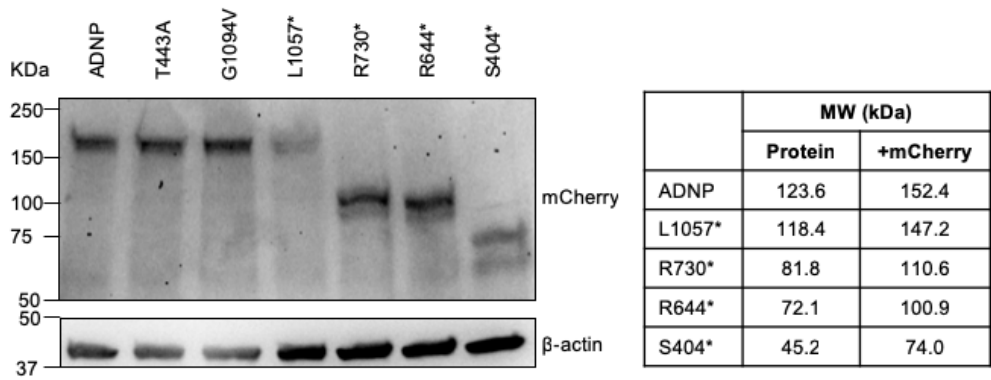
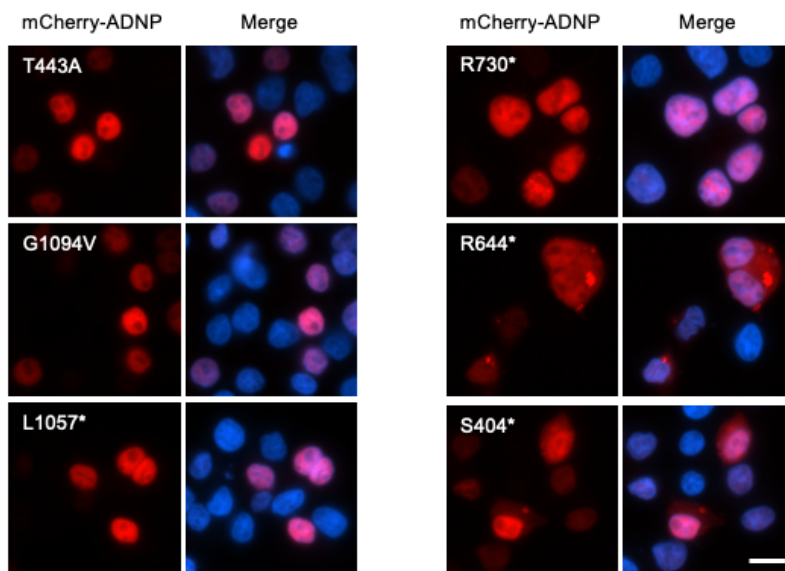
A**B****C**

Figure 6.19 (opposite page). Characterisation of ADNP variants. (A) Schematic representations of ADNP missense variants (red text) and truncations used in this assay, displaying functional domains: C2H2-type zinc fingers 1-9 (ZnF, red segments), neuroprotective peptide (NAP, blue segment) and DNA-binding homeobox domain (purple segment). (B) Left panel: Immunoblotting of whole-cell lysates from HEK293 cells transfected with ADNP variants fused to mCherry. β -actin served as a loading control. Right panel: expected molecular weights of candidates alone and as mCherry fusion proteins. (C) Fluorescence microscopy images of HEK293 cells transfected with ADNP variants fused to mCherry. Nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μ m.

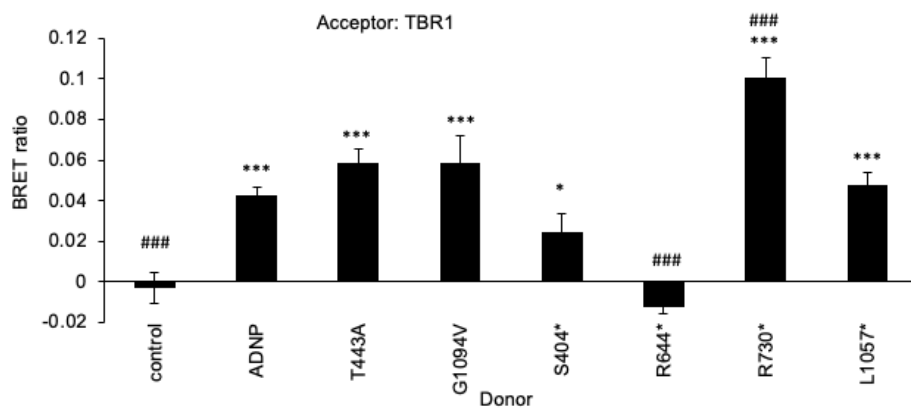
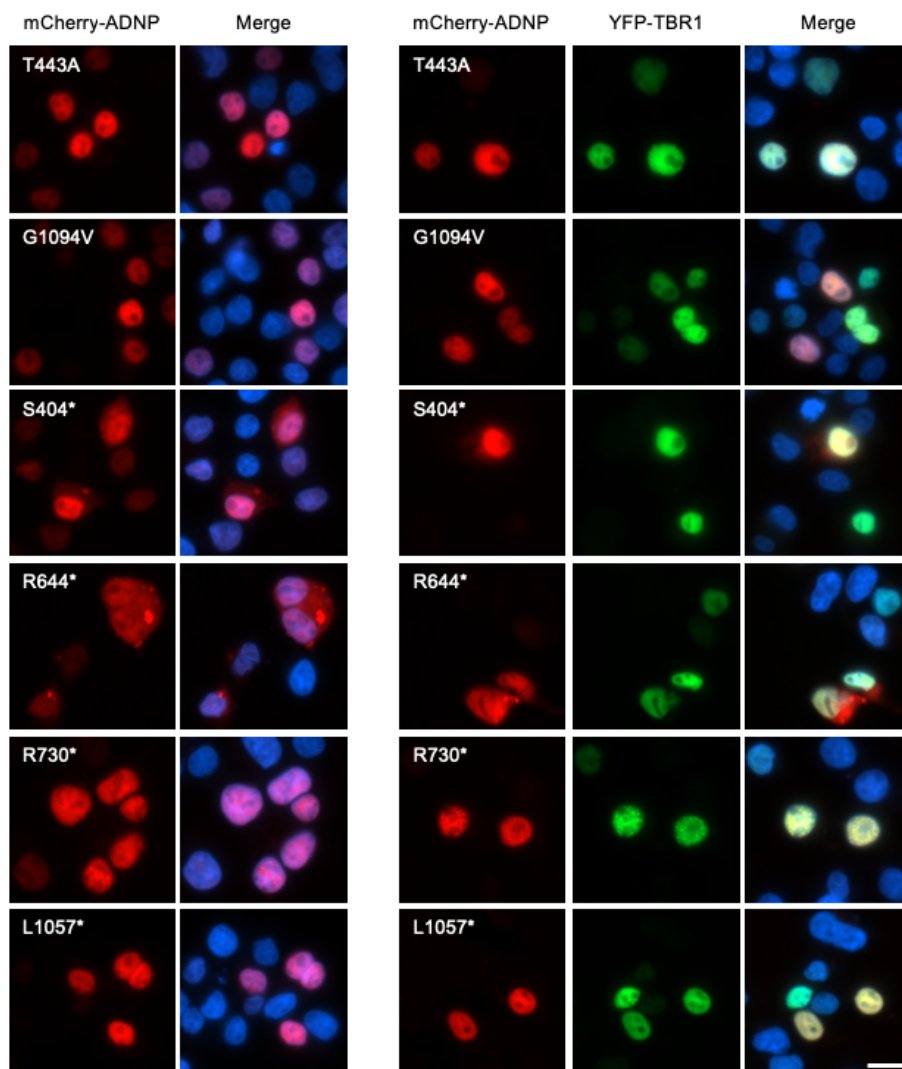
A**B**

Figure 6.20 (opposite page). Cytoplasmic ADNP variants are unable to interact with TBR1. (A) BRET assay for interaction between TBR1 and ADNP variants. Bars represent the corrected mean BRET ratios \pm SD of one experiment performed in triplicate (* $P < 0.05$ and *** $P < 0.001$ compared to control, ### $P < 0.001$ compared to WT ADNP, one-way ANOVA and *post-hoc* Tukey's test). (B) Fluorescence microscopy images of HEK293 cells co-transfected with ADNP variants (fused to mCherry, red) and TBR1 (fused to YFP, green). Nuclei were stained with Hoechst 33342 (blue). Scale bar = 10 μ m.

6.4 DISCUSSION

TBR1 mutations lead to a neurodevelopmental disorder (Deriziotis et al., 2014b), as do mutations in its previously established interactors: FOXP1 (Sollis et al., 2016; Meerschaut et al., 2017), FOXP2 (Lai et al., 2001; Morgan et al., 2017), BCL11A (Dias et al., 2016) and CASK (Moog et al., 2011). The work presented in this chapter confirms novel interactions between TBR1 and a further four proteins implicated in ID/ASD-related syndromes – GATAD2B (de Ligt et al., 2012), BCOR (Ng et al., 2004), ADNP (Helsmoortel et al., 2014) and NR2F1 (Bosch et al., 2014). Interactions were also found with NR2F2, a close homologue of NR2F1, and with the synaptic protein CYFIP2. Fig 6.21A shows the network of TBR1-interactors confirmed to date, including previously established interactions amongst these proteins. Several TBR1-interactors are linked to ASD/autistic features, ID or language impairment, with several examples of co-morbidity of two or more of these phenotypes (Fig 6.21B), though it should be noted that interaction candidates involved in neurodevelopmental disorders have been deliberately selected for validation, particular in the current chapter. In addition, there are several phenotypic features that are not associated with *TBR1* but are characteristic of mutations in several of its interacting proteins, including visual problems (*CASK*, *FOXP1*, *BCL11A*, *BCOR*, *GATAD2B*, *ADNP*, *NR2F1*); hypotonia (*CASK*, *FOXP1*, *GATAD2B*); cardiovascular defects (*BCOR*, *ADNP*, *NR2F2*) and urogenital abnormalities (*FOXP1*, *BCOR*). The narrower phenotype of individuals with *TBR1* variants is likely due to the fact that *TBR1* is exclusively expressed in the brain, while other interactors are also expressed in several other tissues (Uhlén et al., 2015).

TBR1 variants identified in sporadic ASD cases had a deleterious impact on multiple interactions. Nonsense/frameshift variants, which truncate both the T-box and C-terminal region, are the most severe, abolishing all interactions tested so far (Table 6.7). However, missense variants had variable effects. Four *de novo* variants (p.K228E, p.W271C, p.N374H, p.K389E) have now been found to abolish FOXP2- and ADNP-binding, while preserving interactions with WT TBR1, CASK, BCL11A, BCOR and GATAD2B. These variants typically reduce the interactions with NR2F1 and NR2F2, but do not abolish them, with the notable exception of p.K389E. The *de novo* p.W271R variant and the majority of rare inherited variants do not disrupt any interactions and exhibit normal localisation and TF activity. These may be benign mutations incidental to the reported phenotype. For several interactions, however, p.W271R gave a reduced BRET signal compared to wildtype TBR1, perhaps explained by the decreased expression level previously demonstrated for this variant (den Hoed et al., 2018). The rare inherited p.Q418R variant, which lies outside the T-box, consistently disrupts or reduces TBR1 interactions, leaving only TBR1 homodimerisation and CASK-interaction unaffected (Table 6.7). In all other investigations thus far, this variant appears normal, preserving transcriptional repression and diffuse nuclear localisation (Deriziotis et al., 2014b), and it is predicted to be benign by *in silico* methods (Table 6.7). It may be that this residue interacts directly with some but not all TBR1-interactors. Alternatively, the substitution might alter protein conformation in such a way that it blocks other important residues from interacting.

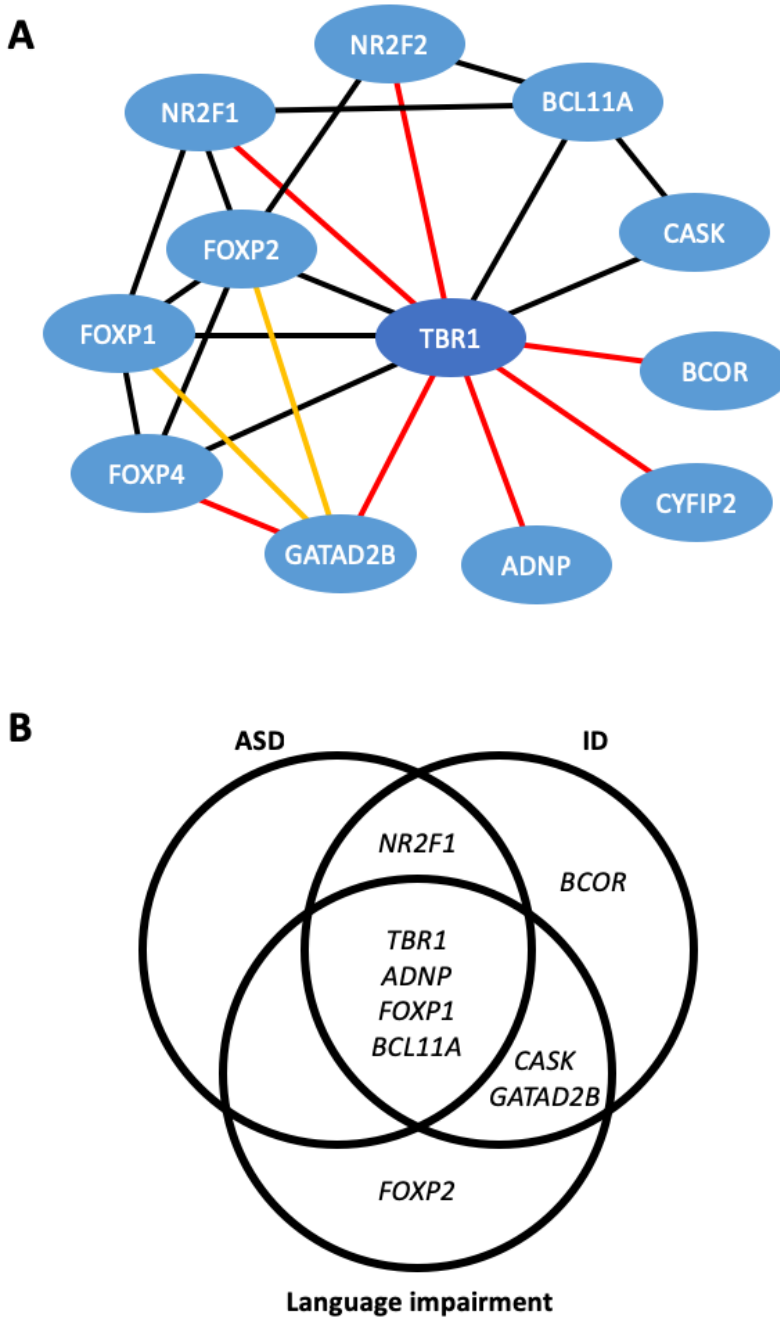


Figure 6.21. An expanded network of TBR1-interactors. (A) Confirmed TBR1-interactors identified in this and previous studies. Connectors: black = previously reported interactions, red = interactions identified in this thesis, yellow = interactions previously identified in other species but confirmed for human proteins in this thesis. Note that the TBR1-BCOR interaction was identified in the AP-MS screen presented here but was validated and further characterised by den Hoed (2016). (B) Neurodevelopmental phenotypes caused by mutations in TBR1-interactors.

The experiments presented here, together with a review of previous studies, show that different proteins interact with different regions of TBR1 (Fig 6.17). Interaction with ADNP, like FOXP2 (Deriziotis et al., 2014b), primarily involves the T-box domain (residues 213-393) and is disrupted by *de novo* missense variants within that region. The C-terminal region of TBR1 (residues 394-567) appears to be the major binding site for BCOR and GATAD2B. Notably, the T-box was not sufficient even for a partial interaction with BCOR and GATAD2B, unlike WT TBR1 and BCL11A. On the other hand, interactions with NR2F1 and NR2F2 were affected both by missense variants in the T-box and by the p.N394* truncation, indicating the involvement of both the T-box and C-terminal regions. It is interesting to note that the interaction partners that are most disrupted by T-box mutations - FOXP2, ADNP, and to a lesser extent, NR2F1/2 - involve DNA-binding TFs, while those that are least affected - BCL11A, BCOR, CASK and GATAD2B - are cofactors that do not have direct DNA-binding domains. Perhaps interactions between TBR1 and other TFs require both proteins to be bound to DNA, while interactions between TBR1 and non-TF proteins is DNA-independent.

Three *GATAD2B* variants, identified in patients with ID (de Ligt et al., 2012; Willemsen et al., 2013), were assessed for their effects on TBR1-interaction. The two shorter frameshift variants of *GATAD2B* (p.Q190Afs*34 and p.N195Kfs*30) disrupted the interaction with TBR1, but p.Q470* retained some residual interaction. These results indicate that while both the CR1 and CR2 domains of *GATAD2B* are required for TBR1-interaction, the final portion of CR2 after the ZnF (which is truncated in p.Q470*) may not be essential. This finding is in contrast to the interaction of *GATAD2B* with FOXP proteins, where p.Q470* was also unable to interact (Chapter 4). While helpful in mapping the TBR1-binding site of *GATAD2B*, the differences between the *GATAD2B* variants are unlikely to have much *in vivo* significance. As all three variants occur before the final exon, they are likely to undergo NMD and therefore produce little or no truncated protein in the affected patients. Instead, these mutations probably lead to disorder through haploinsufficiency. Indeed, the patients with these three mutations, and a fourth patient with a whole gene deletion of *GATAD2B*, displayed a remarkably similar phenotype with comparable developmental delay, ID, speech delay and characteristic facial features (Willemsen et al., 2013).

GATAD2B is ubiquitously expressed, and exhibits strong neuronal expression in the cerebral cortex, hippocampus, caudate and cerebellum (including Purkinje cells) (Uhlén et al., 2015; see URL¹), overlapping considerably with areas of TBR1 expression. *GATAD2B* is a member of the repressive NuRD chromatin remodelling complex. Another NuRD subunit, MTA2, was also present in the putative TBR1 interactome (Chapter 5). Other T-box TFs (e.g. TBX5) have also been shown to interact with NuRD complex members (Waldron et al., 2016). Recruitment of the NuRD complex by TBR1 is a plausible

¹ Human Protein Atlas entry for *GATAD2B*: <https://www.proteinatlas.org/ENSG00000143614-GATAD2B/tissue>

mechanism for TBR1-mediated repression, which might be facilitated by TBR1-GATAD2B interaction.

ADNP truncation mutations varied in their effects on TBR1-interaction. Longer nuclear-expressed truncations (p.L1057*, p.R730*) interacted with TBR1, while the shorter partially cytoplasmic p.R644* truncation showed no interaction. Conversely, the shortest truncation (p.S404*) did appear to interact with TBR1 – an unexpected result that may require further investigation in future. However, despite these differences at the molecular level, a review of the original clinical reports revealed no obvious phenotypic differences between patient variants that interacted with TBR1 and those that did not. For example, patients carrying the non-interacting p.R644* variant (Helsmoortel et al., 2014) exhibited developmental delay, hypotonia, behavioural problems and dysmorphic facial features comparable to those observed with the longer p.R730* variant (Krajewska-Walasek et al., 2016), which retained TBR1-interaction. The ADNP missense variants p.T443A and p.G1094V did not have any effects on localisation or interaction with TBR1 in the assays used here. While there is a possibility that these variants disrupt other aspects of ADNP function, there is also reason to believe that these variants may be benign, or responsible for a milder phenotype. The ADNP p.T443A variant was identified in a cohort of patients with high-functioning ASD without severe comorbid conditions (Alvarez-Mora et al., 2016), which would likely exclude patients with a typical HVDAS presentation. The ADNP p.G1094V variant was identified in post-mortem tissue from a patient with a history of language and developmental delay, attention deficit disorder and mild hypotonia (D’Gama et al., 2015). There was no medical record of non-neurological symptoms or facial features characteristic of HVDAS. Furthermore, the patient also carried a mutation in the *PQBPI* gene, which is implicated in Renpenning syndrome, an X-linked disorder involving ID and autistic features in addition to multiple non-neurological features (Stevenson et al., 2005). Indeed, the specific *PQBPI* mutation in this case (p.P244L) was previously reported as a potential causative mutation in twin brothers with ID (Redin et al., 2014). It is therefore unclear whether the *ADNP* missense mutations are causative in these two cases. The functional experiments presented in this chapter suggest that they may not be.

ADNP is a ubiquitously expressed TF, with strong neuronal expression detected in cerebral cortex, hippocampus and cerebellum (Fig 6.6; Uhlén et al., 2015; see URL¹). There is therefore good reason to believe that TBR1 and ADNP may interact *in vivo*. Studies in developing mice have shown that ADNP primarily represses endoderm-related genes, while activating genes involved in neurogenesis, including *Neurod1*, *Neurog1* and *Ngfr* (Mandel et al., 2007). It is plausible that ADNP and TBR1 may interact to co-regulate genes involved in neurodevelopment. The C-terminal region of ADNP directly interacts with members of the BAF chromatin remodelling complex (Mandel and Gozes, 2007) – including ARID1A, SMARCA4 and SMARCC2 - which play a key role in neurodevelopment. Mutations in

¹ Human Protein Atlas entry for ADNP: <https://www.proteinatlas.org/ENSG00000101126-ADNP/tissue>

multiple BAF subunits give rise to neurodevelopmental disorders, including Coffin Siris syndrome (*ARID1A*, *ARID1B*, *SMARCA4*, *SMARCE1* and *SMARCB1*) and Nicolaides-Baraitser syndrome (*SMARCA2*) (Santen et al., 2012). Two BAF complex members (*ARID1A*, *SMARCA2*) were also present in the putative TBR1 interactome (Chapter 5), although only *SMARCA2* was tested by BRET, and it did not validate. Future studies may determine whether TBR1 interacts with additional BAF complex subunits.

TBR1 variants had identical effects on ADNP- and FOXP2-interactions (Table 6.7), suggesting that these two TFs may bind to TBR1 in a similar manner. The current study suggests that one or more C2H2-type ZnF domains in ADNP may be required for interaction with TBR1. However, while the FOXP2s contain a similar C2H2-type ZnF, implicated in FOXP2-dimerisation, previous studies have shown that this domain is not necessary for TBR1-interaction (Deriziotis et al., 2014b). An alternative hypothesis is that functional DNA-binding of ADNP or FOXP2s may be the major prerequisite for TBR1-interaction. Indeed, FOXP2 variants that disrupt DNA-binding also prevent interaction with TBR1 (Deriziotis et al., 2014b; Sollis et al., 2017), while the non-interacting p.R644* ADNP variant identified here was also partially mislocalised to the cytoplasm, perhaps due to defective DNA-binding.

NR2F1 and NR2F2 are paralogous TFs with both activating and repressive roles in transcriptional regulation (Lin et al., 2011; Alfano et al., 2014). They are expressed throughout the body, but NR2F1 predominates in neural tissues, while NR2F2 predominates in the internal organs, including the heart (Lin et al., 2011). This is consistent with the patterns of symptoms observed in human disorders resulting from mutations of the two genes. Mutations in *NR2F2* cause congenital cardiac abnormalities (congenital heart defects, multiple types, 4; OMIM 615779) (Al Turki et al., 2014), while mutations in *NR2F1* cause Bosch-Boonstra-Schaaf optic atrophy syndrome (OMIM 615722), a disorder involving delayed development, moderate ID, optic atrophy and cerebral visual impairment (Bosch et al., 2014). Nonetheless, both proteins have been implicated in multiple aspects of brain development, including cortical patterning (Zhou et al., 2001; Armentano et al., 2007; Tomassy et al., 2010) and amygdala morphogenesis (Tang et al., 2012), suggesting links to similar functions of TBR1.

An interaction was identified between the synthetic TBR1 NLS-mutant, designed to mimic the cytoplasmic expression of TBR1 previously observed in mature rat neurons (Hong and Hsueh, 2007), and the cytoplasmic/synaptic CYFIP2 protein. CYFIP2 forms part of the WAVE1 complex with ABI2, BRK1, NCKAP1 and WASF1 (WAVE1), which regulates actin cytoskeleton reorganisation; and also acts as an adapter between FMRP and EIF4E in a translation inhibition complex (De Rubeis et al., 2013). Members of both complexes have been implicated in neurodevelopmental disorders, including NCKAP1 (ASD) (Iossifov et al., 2012) and FMRP (Fragile X syndrome, OMIM 300624). Indeed, mutations in synapse-expressed genes have emerged as a recurrent cause of ASD (De Rubeis et al., 2014). The relevance of a TBR1-CYFIP2 interaction is not clear. TBR1 may have additional functions in the cytoplasm/synapse, beyond its typical role as a TF. Another possibility is that TBR1

might encourage nuclear localisation of CYFIP2. Indeed, nuclear expression of CYFIP2 has been observed in response to treatment with the nuclear export inhibitor leptomycin-B (Jackson et al., 2007). However, there was no evidence of interaction between nuclear TBR1 and CYFIP2, or of translocation of CYFIP2 to the nucleus in the manner previously observed for CASK (Hsueh et al., 2000). The findings in this chapter highlight a potential functional difference between the closely related CYFIP1 and CYFIP2 proteins. Perhaps surprisingly there was evidence of an interaction between the TBR1-NLS mutant and CYFIP2, but not CYFIP1, despite 86% identity between the two proteins. Interestingly, different binding properties have been reported before for CYFIP1 vs CYFIP2: the fragile-X-related proteins FMRP, FXR1 and FXR2 all interact with CYFIP2, but only FMRP interacts with CYFIP1 (Schenck et al., 2001). Further work will be required to elucidate the significance of cytoplasmic TBR1 expression, and the observed interaction with CYFIP2.

A number of interaction candidates originally identified by AP-MS in Chapter 5 of this thesis could not be validated using the BRET assay: KDM1A, NCKAP1, YY1, CSNK2A1, TBL1XR1, CTNND1, SMARCA2, CTBP1, CTBP2 and ZMYM2. This may reflect the differing sensitivity and specificity of the two methods. These unconfirmed interaction candidates could be false positives detected in the AP-MS screen. Despite filtering against both internal control experiments and the Contaminant Repository for Affinity Purifications (CRAPome), non-specific proteins may still be present in the final TBR1 interactome list. False-positive interactions might also be detected between DNA-bound TBR1 and relatively distant DNA-binding proteins that are physically connected by an intervening stretch of DNA. Despite blocking protein-DNA interactions in the AP-MS assay using ethidium bromide and digesting DNA bridges with benzonase before precipitation, the carry-over of non-specific DNA-binding proteins cannot be excluded. Similarly, the unvalidated candidates could have been immunoprecipitated via interaction with an intermediary protein that in turn binds to TBR1. In this case, indirect interactors may not themselves come into close enough proximity with TBR1 to generate a significant BRET signal. Finally, it is possible that some of the unvalidated candidates here do actually interact with TBR1 but gave false negative results in the BRET assay. False negatives can occur in BRET if the conformation of either or of both putative interacting proteins prevents the N-terminal luciferase- and YFP- tags from making close contact for energy transfer to occur. Alternatively, the tags themselves might block important binding surfaces on either protein, which would also lead to false negative results. Future investigations using other methods (e.g. co-immunoprecipitation) will be necessary to untangle these alternative explanations.

In conclusion, the work presented here considerably expands the known TBR1 interactome with the confirmation of 6 novel interactors, many of which are implicated in ID, ASD and language impairment. These include TFs and chromatin modifiers involved in both positive and negative regulation of transcription, supporting dual roles for TBR1 in regulating gene expression. Several interactions may be important in brain development. There is also some evidence of an additional role for TBR1 in the cytoplasm, which could be explored further in future research.

7 GENERAL DISCUSSION

7.1 BACKGROUND

Language is a complex phenomenon found only in humans, but which probably emerges through a unique combination of neurobiological pathways with much deeper evolutionary histories. As such, a large number of molecular processes are likely to be involved in the development of the human language capacity. Research into the genetics of speech and language has sought to uncover these processes by pinpointing the genes and proteins responsible for disorders such as specific language impairment/developmental language disorder (SLI/DLD), childhood apraxia of speech (CAS), stuttering and dyslexia, which primarily affect language-related abilities and are often defined by the absence of other deficits. However, relatively few genes have been directly and reproducibly implicated in such disorders. The most firmly established is *FOXP2*, whereas most other promising candidates, identified through linkage, targeted and genome-wide association studies, whole exome/genome sequencing and cytogenetic methods, are yet to be confirmed in further studies. Given the complexity of language phenotypes, it is likely that a large number of key molecules remain to be discovered.

In the Introduction (Chapter 1), I outlined several reasons why it may not be possible for some language-related genes to be identified by only searching for mutations in disorders that exclusively affect speech/language and that alternative approaches can play an important role in elucidating the relevant molecular pathways. Therefore, this thesis considered disorders that influence speech and language to a significant degree while also having broader effects on neurological, anatomical and/or physiological development and sought to further expand language-related networks by investigating protein-protein interactions.

7.2 INTERACTIONS AMONG LANGUAGE-RELATED PROTEINS

The **first major aim of this thesis** was to identify and characterise novel protein-protein interactions involving known language-related proteins. *FOXP2*, the most firmly-established language-related protein (Morgan et al., 2017), served as a starting point for these investigations. Previously-established interaction partners of *FOXP2* include *FOXP1* (the focus of Chapters 2-3) and a third close homologue, *FOXP4* (Li et al., 2004; Deriziotis et al., 2014a); the *TBR1* transcription factor (the focus of Chapters 5-6) (Deriziotis et al., 2014b) and the transcriptional co-repressors *CTBP1* and *CTBP2* (Li et al., 2004; Estruch et al., 2016a).

In **Chapter 4**, I sought to validate putative interactions between *FOXP2* and three chromatin remodelling proteins – *GATAD2B*, *CHD3* and *KANSL1* – which were identified in a yeast-two hybrid (Y2H) assay using human *FOXP2* as bait (Estruch et al., 2016b). The mouse orthologue of *GATAD2B* had also been identified in a second published Y2H assay using a fragment of mouse *Foxp2* as bait and was further shown to interact with mouse

Foxp1 (Chokas et al., 2010). The link to language was strengthened by reports of severe speech delay in addition to intellectual disability (ID) and neonatal hypotonia in patients with heterozygous *GATAD2B* mutations (de Ligt et al., 2012; Willemsen et al., 2013; Luo et al., 2017). I confirmed that interactions also occur between human FOXP1/2/4 and *GATAD2B* and demonstrated the importance of the FOXP leucine zipper motif, as well as the CR1 and CR2 domains of *GATAD2B*, for this interaction. Recruitment of the NuRD complex via interaction with *GATAD2B* might be one mechanism by which FOXP transcription factors repress transcription of target genes in order to regulate neurodevelopment.

On the other hand, my experiments did not find evidence for an interaction between FOXP2 and either CHD3 or KANSL1. In any case, there is growing support from clinical genetics studies for the importance of both of these genes for speech and language development. Eising et al. (2018) identified a *de novo* *CHD3* missense variant in a patient with childhood apraxia of speech. Additional *de novo* variants were then detected in 34 other cases, leading to a distinct neurodevelopmental disorder with speech delay in 100% of patients, in addition to ID of varying severity, facial dysmorphisms and other features (Snijders Blok et al., 2018). Interestingly, expressive language was more affected than receptive language and speech problems also included dysarthria, apraxia and oromotor difficulties, showing some parallels with major characteristics of *FOXP2*-related disorder (Morgan et al., 2017). Furthermore, non-linguistic findings in several of the individuals with *CHD3* mutations overlap with phenotypes observed in *FOXP1*-related disorder, including hypotonia, autistic features, visual problems and genital abnormalities in males (Meerschaut et al., 2017; Snijders Blok et al., 2018) (Chapters 2-3). Similarly, the *KANSL1*-related Koolen-De Vries syndrome includes speech impairment alongside ID, autistic features, hypotonia and broader developmental abnormalities (Koolen et al., 2012; Zollino et al., 2012). A deeper analysis of communicative aspects in four individuals with *KANSL1* variants, as well as 25 with larger 17q21.31 microdeletions encompassing *KANSL1*, identified apraxia in 100% of patients, dysarthria in 93% and stuttering in 17% (Morgan et al., 2018). Therefore, even without independent validation of previously suggested interactions with the FOXP2 protein, CHD3 and KANSL1 remain relevant for understanding the molecular mechanisms underlying speech and language.

In **Chapter 5**, I investigated the TBR1 transcription factor, a promising candidate for the study of speech/language genetics because it interacts with FOXP1/2 and is mutated in a neurodevelopmental disorder that includes autism spectrum disorder (ASD, itself associated with pragmatic deficits in communication) and varying degrees of ID, with several patients also exhibiting significant speech delay (Deriziotis et al., 2014b). Aside from the FOXP2s and a handful of other interaction partners – CASK (Hsueh et al., 2000) and BCL11A (den Hoed et al., 2018) – the TBR1 interactome was not well understood. I therefore performed an affinity purification-mass spectrometry (AP-MS) screen for novel TBR1 interactors. This experiment identified 247 novel TBR1-interaction candidates and confirmed the interaction with CASK. The putative interactome was enriched for proteins involved in transcriptional

regulation (including a significant overrepresentation of epigenetic factors) as well as other processes such as RNA processing, cell-cycle regulation and ubiquitination. There was a significant overrepresentation of proteins mutated in both ASD and ID syndromes, including several that also involve speech/language delay.

In **Chapter 6**, I prioritised these candidates based on known involvement in neurodevelopmental disorders and/or established interactions with speech/language-related proteins, to select 15 candidates for validation and follow-up experiments *in vitro*. Five novel TBR1-interacting proteins were confirmed: GATAD2B, BCOR, ADNP, NR2F1 and NR2F2, transcription factors and chromatin remodelling proteins with both positive and negative effects on transcription and important roles in brain development. The first four of these are also implicated in ID, ASD and/or language impairment (Ng et al., 2004; Willemsen et al., 2013; Bosch et al., 2014; Helsmoortel et al., 2014). Mapping experiments showed that different regions of TBR1 were important for different protein-protein interactions. ADNP-binding primarily involved the T-box domain, which had also been previously shown to mediate FOXP2-interaction (Deriziotis et al., 2014b). Meanwhile, the C-terminal region of TBR1 was the major site mediating interaction with GATAD2B, as it was for BCOR (den Hoed, 2016), BCL11A (den Hoed et al., 2018) and CASK, as well as for TBR1 homodimerisation (Deriziotis et al., 2014b). Interaction with NR2F1 and NR2F2, on the other hand, appeared to require both the T-box and the C-terminal region to be intact. The identification of GATAD2B as a TBR1-interactor is especially interesting, given that both proteins interact with FOXP1/2/4 (Chapter 4). The CR1 and CR2 domains of GATAD2B were required for TBR1-binding, although the C-terminal portion of the CR2 domain was less essential than for interaction with the FOXP2s. Mapping experiments for the TBR1-ADNP interaction identified a region of ADNP containing two zinc finger domains as an important binding site, but also yielded some paradoxical results that require further investigation. A sixth candidate, the synaptic protein CYFIP2, interacted with a synthetic variant of TBR1 localised to the cytoplasm, which might hint at additional roles for TBR1 in the cytoplasm.

7.3 AN EMERGING NETWORK OF TRANSCRIPTIONAL REGULATORS IN NEURODEVELOPMENT

Taken together, the work in this thesis has expanded the network of interacting proteins involved in speech- and language-related disorders (Fig 7.1). The new interactions identified here and in other studies (Deriziotis et al., 2014b; den Hoed et al., 2018; Estruch et al., 2018) indicate a high degree of interconnectedness amongst these proteins. Notably, these interaction studies do not point to a highly specialised network of proteins involved exclusively in language. Rather, proteins mutated in disorders that involve primary speech/language deficits (e.g. FOXP2) interact with proteins that are mutated in broader neurodevelopmental disorders (e.g. FOXP1, TBR1, GATAD2B). The distinction between speech/language disorders and, for example, ID or ASD is not always sharp, and there is a high rate of co-morbidity between ID/ASD and speech/language delay (Tomblin, 2011). In

some cases, speech/language impairment might represent the less severe end of a spectrum of disorders, involving many of the same genes and pathways as more severe or generalised neurodevelopmental disorders. Interestingly, sequencing of individuals with CAS has identified probable pathogenic mutations in several genes previously associated with broader phenotypes, including *KAT6A* (developmental delay, language delay, hypotonia, distinctive facial features), *SETBP1* (developmental delay, language delay), *ZFHX4* (developmental delay, distinctive facial features) and *TNRC6B* (ASD) (Eising et al., 2018). Similarly, some patients with *CHD3* mutations have only borderline ID and may receive a formal diagnosis of CAS, while others have much more profound cognitive impairment (Snijders Blok et al., 2018). The accumulating evidence suggests that the development of speech/language abilities is controlled not by a distinct and self-contained network of highly specialised genes, but rather by a much larger network of neurodevelopmental genes that each impact speech/language development to varying degrees and at varying levels of specificity. The process of teasing apart the relative contributions of each of these genes will be an ongoing task for the field.

The language-related proteins characterised in this thesis predominantly consist of transcription factors (e.g. *FOXP1/2/4*, *TBR1*) and chromatin remodelling factors (e.g. *GATAD2B*, *KANSL1*, *CHD3*). The putative *TBR1* interactome was also enriched for proteins involved in transcriptional regulation in general and chromatin-modifying factors in particular (Chapter 5), while the validated interactions included transcription factors and regulatory cofactors (*ADNP*, *BCOR*, *NR2F1*, *NR2F2*) and chromatin modifiers (*GATAD2B*) (Chapter 6). Evidence in the literature suggests a prominent role for transcription factors and chromatin remodelling proteins in multiple neurodevelopmental disorders, including ASD (De Rubeis et al., 2014) and ID/DD (Iwase et al., 2017). Regarding language-related disorders, proteins robustly implicated in CAS include the transcription factors *FOXP2* (Lai et al., 2001) and *BCL11A* (Peter et al., 2014), while Eising et al. (2018) identified variants of pathogenic significance both in transcription factors (*ZFHX4*, *MKL2*, *SETBP1*) and in chromatin modifying proteins (*CHD3*, *SETD1A*, *WDR5*, *KAT6A*) in a CAS cohort. On the other hand, exome sequencing of probands with SLI did not seem to support an important role for transcription factors or chromatin-modifying proteins (Chen et al., 2017), with the top candidate mutations identified in *ERC1* (regulator of neurotransmitter release), *GRIN2A* (NMDA receptor subunit) and *SRPX2* (surface receptor involved in synapse formation). While it is difficult to make a direct comparison between these two differently-designed studies, it may be that the most important molecular processes differ between specific types of speech/language disorder.

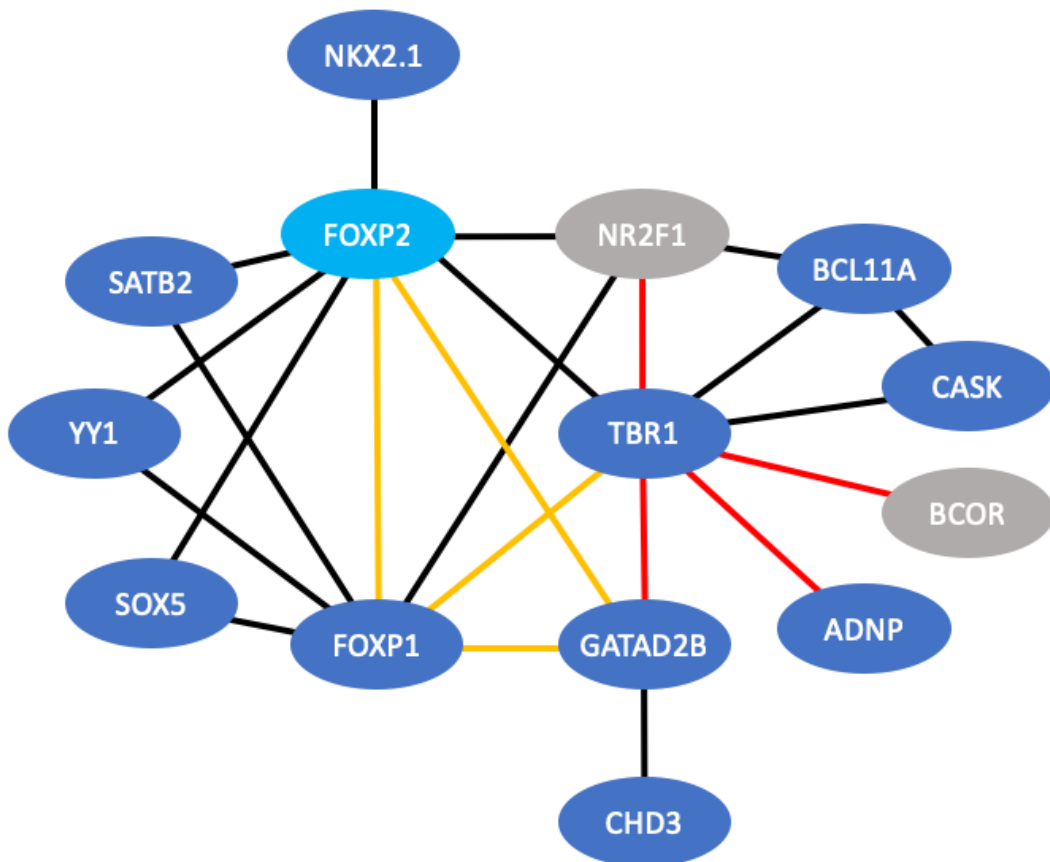


Figure 7.1. Network of interacting proteins in speech- and language-related disorders. Proteins involved in neurodevelopmental disorders (NDDs) are coloured: light blue = language-specific disorder, dark blue = broader NDD with speech/language deficits, grey = NDD without evidence of speech/language deficits. Connectors: black = previously reported interactions, red = interactions identified in this thesis, yellow = previously-identified interactions that were further characterised in this thesis. Note that the TBR1-BCOR interaction was identified in the AP-MS screen presented here but was validated and further characterised by den Hoed (2016).

7.4 CHARACTERISING MUTATIONS IN LANGUAGE-RELATED DISORDERS

The **second major aim of this thesis** was to investigate the consequences of mutations affecting speech/language-related proteins, at phenotypic and molecular levels, including the mutations' effects on interactions with other proteins in the network.

In **Chapters 2 and 3**, I investigated 7 *de novo* missense and truncating mutations in *FOXP1*, identified in patients with ID and speech impairment, with or without autistic features (OMIM 613670). The two chapters included detailed clinical reports of six patients and a review of published phenotypic information for a further four (reported in Hamdan et al., 2010; O'Roak et al., 2011; Srivastava et al., 2014; Lozano et al., 2015). The clinical features matched the characteristic phenotype described in the literature (Le Fevre et al., 2013), including motor delays and ID of varying severity in all patients, and autistic features reported in 6/8 patients for whom there was data. The new case reports also highlighted several novel or rarely-reported features, including ophthalmological and urogenital abnormalities (Pariani et al., 2009; Bekheirnia et al., 2017) and sensory integration disorder. Language impairment or delay was seen in all patients (Table 2.2, Table 3.3), with expressive language more severely affected than receptive language in at least three (Table 3.3). Poor articulation was also noted in several patients. Indeed, the centrality of language impairment to the *FOXP1* phenotype has been confirmed by a recent large phenotyping study published after the work in Chapters 2 and 3 (Meerschaut et al., 2017). The authors compiled phenotypic data on 25 novel and 23 previously-reported patients with *FOXP1* haploinsufficiency, including the three patients reported in Chapter 2 (Sollis et al., 2016) and three others discussed in detail in this thesis (originally reported by Hamdan et al., 2010; O'Roak et al., 2011; Lozano et al., 2015). They identified speech/language delay in 100% of patients, making it a more consistent feature than ID (98%; 65% mild-moderate, 33% severe), motor delay (96%) or autistic features (75%). Expressive language delay was specifically noted (97% of patients) as well as articulation problems (90%), poor grammar (82%), feeding difficulties (69%) and oromotor dysfunction (52%) – features that are also common in patients with *FOXP2* mutations (Morgan et al., 2017). The study by Meerschaut et al. also confirmed a high prevalence of ophthalmological (77%) and urogenital (55%) anomalies, though not sensory integration disorder, in patients with *FOXP1* mutations. Thus, the reports in this thesis as well as other recent phenotyping efforts support a prominent and distinctive form of speech impairment in *FOXP1*-related disorder, even as the range of associated clinical features expands to include other aspects of development.

At the molecular level, *FOXP1* mutations caused aberrant subcellular localisation, abolished transcriptional repression and disrupted multiple protein-protein interactions. *FOXP1* variants fell into two classes: the first group (including a frameshift, a stop-gain and a missense variant) abolished interaction with wild-type FOXP1 and FOXP2, consistent with a haploinsufficiency mechanism, while the second group (including 3 missense variants and a stop-gain variant) maintained these interactions but translocated the wild-type proteins into nuclear and cytoplasmic aggregates, suggesting a possible dominant negative effect. Despite the contrasting molecular properties observed in my assays, the two groups did not

correspond to obvious phenotypic subtypes, with similar clinical features and a comparable degree of variability in ID and speech delay severity within each group. Nor did the severity of the disorder noticeably differ between missense and truncating variants. In fact, there appeared to be little genotype-phenotype correlation overall: even for the identical p.R514H variant, one patient (Patient 2, Chapter 3) showed mild ID and could speak in 2-to-3-word sentences at 3.5 years, while another (Patient 3, Chapter 3) had moderate-to-severe ID and no speech at ~8 years. These observations were borne out in the phenotyping study by Meerschaut et al. (2017), who found no significant phenotypic differences between null mutations (deletions, truncating mutations, splice site mutations) and missense variants, with the one exception that prominent digit pads were more common in patients with missense variants. As for the p.R514H variant above, they also found wide phenotypic variation between four patients with an identical p.R525* variant. Thus, it seems that phenotypic severity is not predictable from the type of mutation, suggesting that other genetic and/or environment factors play roles in modifying *FOXP1*-related disorder.

In **Chapter 4**, I investigated the consequences of known mutations for the interactions between *FOXP1/2* and *GATAD2B*. I tested the *FOXP1* mutations characterised in Chapters 2 and 3, in addition to several previously-reported *FOXP2* mutations. The same two classes of *FOXP1* variants identified in Chapters 2-3 emerged here. Those that eliminated interaction with wild-type *FOXP1/2* also abolished *GATAD2B*-binding, suggesting loss-of-function through haploinsufficiency. However, the variants that maintained *FOXP1/2*-interactions were divided. Those that formed cytoplasmic aggregates were unable to interact with *GATAD2B*, which remained in the nucleus, while those that formed nuclear aggregates maintained interaction and co-localisation with *GATAD2B*. It could not be determined whether the punctate co-localisation was driven by the normal nuclear speckling of *GATAD2B* or by the abnormal aggregation of the *FOXP* variants, so the consequences for *GATAD2B* function remain uncertain without further experiments. Nevertheless, it is clear that at least some *FOXP* mutations are able to disrupt interaction with *GATAD2B*.

All of the *GATAD2B* mutations discussed in this thesis (and the only ones reported thus far to my knowledge) are deletions or truncating variants that are likely to undergo nonsense-mediated mRNA decay (NMD), pointing to haploinsufficiency as the most likely mechanism of disorder (Willemsen et al., 2013; Luo et al., 2017; Tim-Aroon et al., 2017). The fact that no *GATAD2B* missense variants have been reported to date might suggest that they are benign or lead to a milder phenotype. I tested three truncating variants and found that even if there could be escape from NMD, leading to expression of truncated proteins, all three would have lost the capacity to interact with the *FOXP* transcription factors. At the phenotypic level, I compared *GATAD2B* haploinsufficiency cases from the literature and noted several features in common with *FOXP1*- and *FOXP2*-related disorders, including severe speech and language impairment or delay, feeding difficulties that may be due to oromotor dysfunction and anecdotal evidence that expressive language is more severely affected than receptive language (Willemsen et al., 2013; Luo et al., 2017; Tim-Aroon et al., 2017). Other features were shared only with the *FOXP1*-related phenotype, including ID,

motor delay, hypotonia, ophthalmological abnormalities and possible autistic behaviours (Willemsen et al., 2013; Luo et al., 2017). These shared symptoms might be explained by a cooperative role for the FOXP transcription factors and GATAD2B in normal brain development, which might be impaired with overlapping consequences by mutations in either protein.

In **Chapter 6**, I tested the effects of TBR1 variants on 4 of the novel interactions that were validated in that chapter (GATAD2B, ADNP, NR2F1 and NR2F2). I also compared these to unpublished results for BCOR, a fifth candidate from the AP-MS screen (den Hoed, 2016), as well as results previously reported for TBR1 homodimerisation and interactions with FOXP2, CASK (Deriziotis et al., 2014b) and BCL11A (den Hoed et al., 2018). Several TBR1 mutations disrupted interactions with one or more proteins, but individual mutations had distinct effects on different interactions. For example, the p.K389E variant located in the T-box domain abolished interactions with FOXP2, ADNP, NR2F1 and NR2F2, while retaining interactions with GATAD2B, BCOR and BCL11A (Table 6.7). However, there were no clear correlations between these different molecular properties and phenotypic variation. I then investigated the effects of GATAD2B variants on interaction with TBR1 and found that two out of three abolished the interaction, contrasting with the fact that all three variants abolished interaction with FOXP1/2/4 (Chapter 4). However, this difference is probably not biologically meaningful, given that the GATAD2B variants are likely to undergo NMD. I also selected six *ADNP* variants from the literature for characterisation, including four stop-gain variants associated with Helsmoortel-van der Aa syndrome (which involves ASD, mild-to-severe ID, speech impairment and developmental delay) (Helsmoortel et al., 2014; O’Roak et al., 2014; Krajewska-Walasek et al., 2016) and two missense variants identified in patients with less precisely-defined neurodevelopmental disorders (D’Gama et al., 2015; Alvarez-Mora et al., 2016). Notably, since the stop-gain variants are in the last exon of *ADNP*, they are predicted to escape NMD, yielding expression of truncated proteins in the cell. The two shortest truncated *ADNP* proteins exhibited abnormal cytoplasmic localisation, and one of these was unable to interact with TBR1, while the longer truncated proteins and those with missense variants maintained nuclear localisation and interaction with TBR1. Further studies will be required to understand the mechanisms underlying *ADNP*-related disorders, but it is notable that at least one mutation disrupted interaction with TBR1.

Overall, I characterised mutations in several of the speech/language-related proteins shown in Fig 7.1. In many cases the mutations disrupted interactions with other proteins in the network. These findings may help to elucidate the pathogenic mechanisms that lead to these disorders and could potentially account for some of the commonalities between the different phenotypes.

7.5 THE IMPORTANCE OF FUNCTIONAL CHARACTERISATION

In investigating protein-coding variants in this thesis, one of the major themes that emerged was the challenge of proving a causative relationship between a variant and a

disorder. Most of the variants discussed here were identified through genetic investigations in individuals with speech/language-related disorders, either by linkage and targeted re-sequencing, or more recently, by whole exome/genome sequencing. However, simply identifying a non-synonymous mutation in a person with a disorder is not sufficient to establish a causal relationship, especially for variants identified by whole exome/genome sequencing, where many variants of unknown function may be detected in a single individual, not all of which will be relevant for the disorder in question. On the other side of the equation, the presence of a variant in controls suggests that it is benign, although control datasets (e.g. ExAC) may contain individuals with low-penetrance disease variants and even undiagnosed cases (Acuna-Hidalgo et al., 2016).

Establishing *de novo* status by sequencing of unaffected parents makes a variant more likely to be pathogenic. The established rate of point mutations in humans makes it unlikely for an irrelevant *de novo* variant to be identified in a gene with a known causal link to the observed phenotype (Sunyaev, 2012). For variants in novel genes, however, *de novo* status is not usually considered sufficient to assert causality but provides strong support to candidates that can be validated through follow-up experiments (Sunyaev, 2012). The existence of prior known mutations in the same gene can provide some support for pathogenicity, but this may not be definitive. Stronger evidence is provided by recurrent observation of the same specific mutation in multiple unrelated cases (Sunyaev, 2012). A good example can be found in Chapter 3 of this thesis, in which an identical p.R514H *FOXP1* variant was identified in 3 patients, with a different variant affecting the same residue (p.R514C) in a fourth patient (Chapter 2).

Further characterisation requires interpretation of the actual consequences of the mutation on e.g. protein sequence and structure. Truncating variants that are susceptible to NMD are highly likely to lead to a loss of function through haploinsufficiency, while missense variants require further investigation. Prediction algorithms (e.g. PolyPhen-2, SIFT, MutationTaster) are used to calculate the probability that a missense variant is pathogenic, but this kind of probabilistic approach cannot provide definitive evidence (Sunyaev, 2012). Another method of assessing the functional impact is to map a mutation onto the 3D structure of a protein in order to predict deleterious effects on protein stability (e.g. in Chapter 2). Knowledge of pathogenic variants in a paralogue can also be used to support pathogenicity – for example in Chapters 2 and 3, causative roles for several *FOXP1* variants were supported by known pathogenic mutations at homologous sites in other FOX family transcription factors.

However, the most direct method for assessing pathogenicity is with functional experiments, including protein characterisation assays. In Chapters 2, 3, 4 and 6, fluorescence imaging was employed to identify changes in protein localisation caused by *FOXP1*, *FOXP2*, *GATAD2B*, *TBRI* and *ADNP* variants. The BRET assay was used in the same chapters to investigate the effects of known or potential pathogenic variants on protein-protein interactions in live cells. Other aspects of protein function require specific functional assays that may be applicable only to a certain protein class. As many of the proteins studied in this thesis are transcription factors, luciferase reporter assays provided a well-developed

functional test for transcriptional regulation, demonstrating impaired transcriptional repression by *FOXP1* variants in Chapters 2 and 3. As chromatin-modifiers such as CHD3 have emerged as a new protein class of interest for speech/language disorders, new techniques are being employed to investigate the effects of specific mutations (Snijders Blok et al., 2018).

In some cases, functional experiments confirmed the predicted pathogenicity of a variant and helped to explain the mechanism. For example, several *FOXP1* variants abolished interaction with wild-type proteins, while others preserved the interaction but led to mislocalisation and abnormal aggregation. In other cases, such as *FOXP1* p.I107T, functional assays showed no effect on localisation, transcriptional repression or protein-protein interactions, contradicting *in silico* predictions and showing that the variant was likely to be benign (Chapter 2) – subsequently backed up by the discovery of the same variant in controls (Eising et al., 2018). A possible complementary example is the p.Q418R variant in *TBR1*, which was predicted to be benign by *in silico* methods (Deriziotis et al., 2014b), but which disrupted multiple interactions (Chapter 6). Another major advantage of functional experiments is the ability to dissect the functions of different protein regions/domains and to distinguish subtypes of mutations in the same gene and therefore understand more about the underlying mechanisms of disorders, as with the two classes of *FOXP1* mutations identified in Chapter 2 and 3 leading to either haploinsufficiency or dominant negative effects.

Detailed phenotyping is also extremely important. If a variant in a known gene is found in a patient with a very different phenotype to that which has been observed before, this might be an indication that it is not a true causative mutation. For example, the p.I107T *FOXP1* variant, which was later shown to be probably benign (Chapter 2), was initially identified in a patient with CAS and apparently without ID (Worthey et al., 2013), unlike the vast majority of *FOXP1* variants in the literature (Le Fevre et al., 2013; Meerschaut et al., 2017). Thorough phenotypic characterisation may also improve our ability to identify genotype-phenotype correlations and draw connections between different functional classes of mutations and different clinical presentations.

Ultimately, functional experiments are still the best methods of assessing pathogenicity. For language disorders, given the relatively small number of candidate genes and variants identified to date, this approach will probably continue to be feasible. However, as the number of cases under investigation increases and new variants are identified, more high-throughput methods may become necessary. For example, induced pluripotent stem cells and CRISPR-Cas9 genome editing technology have both been explored as ways of increasing the speed and versatility of functional validation experiments (Acuna-Hidalgo et al., 2016).

7.6 CONCLUSIONS AND FUTURE DIRECTIONS

The work in this thesis has contributed to our understanding of a network of interacting proteins that are all implicated in neurodevelopmental disorders affecting the development

of a language-capable brain. Crucially, variants in several of these proteins not only disrupt the functions of the mutant proteins, but also their interactions with other proteins in the network. These disrupted interactions may help to explain those phenotypic features that are shared across several distinct yet overlapping disorders.

Larger-scale association and sequencing projects will identify additional common and rare variants that affect speech and language and explain a larger percentage of the variation in these abilities. Functional characterisation of variants will continue to play a crucial role in assessing pathogenicity and in explaining the mechanisms by which they lead to disorders or influence normal variation in speech/language ability. Because speech/language probably did not arise in humans as a completely new and distinct function in the brain, but rather draws on many other abilities and cognitive processes, complementary lines of evidence such as those explored in this thesis will likely continue to be important. It remains to be seen how truly distinct the aetiologies of speech/language disorders are from broader cognitive disorders such as ID and ASD. More detailed phenotyping of speech and language features in a variety of neurodevelopmental and cognitive disorders will be one essential step in understanding the commonalities and differences, alongside other approaches, such as investigations of animal communication systems and vocal learning, to provide a better understanding of the evolutionary basis of language-related processes.

REFERENCES

- Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA, Menashe I, Wadkins T, Banerjee-Basu S, Packer A. 2013. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism* 4:36.
- Acuna-Hidalgo R, Veltman JA, Hoischen A. 2016. New insights into the generation and role of de novo mutations in health and disease. *Genome Biol* 17:241–241.
- Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. *Nat Methods* 7:248–249.
- Aho S, Levänsuo L, Montonen O, Kari C, Rodeck U, Uitto J. 2002. Specific sequences in p120ctn determine subcellular distribution of its multiple isoforms involved in cellular adhesion of normal and malignant epithelial cells. *J Cell Sci* 115:1391–1402.
- Al Turki S, Manickaraj AK, Mercer CL, Gerety SS, Hitz M-P, Lindsay S, D'Alessandro LCA, Swaminathan GJ, Bentham J, Arndt A-K, Louw J, Low J, et al. 2014. Rare variants in NR2F2 cause congenital heart defects in humans. *Am J Hum Genet* 94:574–585.
- Alfano C, Magrinelli E, Harb K, Studer M. 2014. The nuclear receptors COUP-TF: a long-lasting experience in forebrain assembly. *Cell Mol Life Sci* 71:43–62.
- Alvarez-Mora MI, Calvo RE, Puig ON, Madrigal I, Quintela I, Amigo J, Martinez-Elurbe D, Linder-Lucht M, Aznar GL, Carracedo A, Mila M, Rodriguez-Revenga L. 2016. Comprehensive molecular testing in patients with high functioning autism spectrum disorder. *Mutat Res* 784–785:46–52.
- Anthoni H, Zucchelli M, Matsson H, Müller-Myhsok B, Fransson I, Schumacher J, Massinen S, Onkamo P, Warnke A, Griesemann H, Hoffmann P, Nopola-Hemmi J, et al. 2007. A locus on 2p12 containing the co-regulated MRPL19 and C2ORF3 genes is associated to dyslexia. *Hum Mol Genet* 16:667–677.
- Araujo DJ, Anderson AG, Berto S, Runnels W, Harper M, Ammanuel S, Rieger MA, Huang H-C, Rajkovich K, Loerwald KW, Dekker JD, Tucker HO, et al. 2015. FoxP1 orchestration of ASD-relevant signaling pathways in the striatum. *Genes Dev* 29:2081–2096.
- Armentano M, Chou S-J, Tomassy GS, Leingärtner A, O'Leary DDM, Studer M. 2007. COUP-TFI regulates the balance of cortical patterning between frontal/motor and sensory areas. *Nat Neurosci* 10:1277–1286.
- Bacchelli E, Blasi F, Biondolillo M, Lamb JA, Bonora E, Barnby G, Parr J, Beyer KS, Klauck SM, Poustka A, Bailey AJ, Monaco AP, et al. 2003. Screening of nine candidate genes for autism on chromosome 2q reveals rare nonsynonymous variants in the cAMP-GEFII gene. *Mol Psychiatry* 8:916–924.
- Bacon C, Rappold GA. 2012. The distinct and overlapping phenotypic spectra of FOXP1 and FOXP2 in cognitive disorders. *Hum Genet* 131:1687–1698.
- Bacon C, Schneider M, Le Magueresse C, Froehlich H, Sticht C, Gluch C, Monyer H, Rappold GA. 2015. Brain-specific Foxp1 deletion impairs neuronal development and causes autistic-like behaviour. *Mol Psychiatry* 20:632–639.
- Bader GD, Hogue CWV. 2003. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4:2.
- Bae JS, Park S-H, Jamiyandorj U, Kim KM, Noh SJ, Kim JR, Park HJ, Kwon KS, Jung SH, Park HS, Park B-H, Lee H, et al. 2016. CK2 α /CSNK2A1 Phosphorylates SIRT6 and Is Involved in the Progression of Breast Carcinoma and Predicts Shorter Survival of Diagnosed Patients. *Am J Pathol* 186:3297–3315.
- Baris I, Arisoy AE, Smith A, Agostini M, Mitchell CS, Park SM, Halefoglu AM, Zengin E, Chatterjee VK, Battaloglu E. 2006. A novel missense mutation in human TTF-2 (FKHL15)

gene associated with congenital hypothyroidism but not athyreosis. *J Clin Endocrinol Metab* 91:4183–4187.

- Basta J, Rauchman M. 2015. The nucleosome remodeling and deacetylase complex in development and disease. *Transl Res* 165:36–47.
- Baynam G, Overkov A, Davis M, Mina K, Schofield L, Allcock R, Laing N, Cook M, Dawkins H, Goldblatt J. 2015. A germline MTOR mutation in Aboriginal Australian siblings with intellectual disability, dysmorphism, macrocephaly, and small thoraces. *Am J Med Genet A* 167:1659–1667.
- Bedogni F, Hodge RD, Elsen GE, Nelson BR, Daza RAM, Beyer RP, Bammler TK, Rubenstein JLR, Hevner RF. 2010. *Tbr1* regulates regional and laminar identity of postmitotic neurons in developing neocortex. *Proc Natl Acad Sci USA* 107:13129–13134.
- Bekheirnia MR, Bekheirnia N, Bainbridge MN, Gu S, Coban ZA, Gambin T, Janzen NK, Jhangiani SN, Muzny DM, Michael M, Brewer ED, Elenberg E, et al. 2017. Whole-exome sequencing in the molecular diagnosis of individuals with congenital anomalies of the kidney and urinary tract and identification of a new causative gene. *Genet Med* 19:412–420.
- Berg DLC van den, Snoek T, Mullin NP, Yates A, Bezstarosti K, Demmers J, Chambers I, Poot RA. 2010. An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell* 6:369–381.
- Bergen SE, Petryshen TL. 2012. Genome-wide association studies of schizophrenia: does bigger lead to better results? *Curr Opin Psychiatry* 25:76–82.
- Bergman LM, Morris L, Darley M, Mirnezami AH, Gunatilake SC, Blaydes JP. 2006. Role of the unique N-terminal domain of CtBP2 in determining the subcellular localisation of CtBP family proteins. *BMC Cell Biol* 7:35.
- Bernier R, Golzio C, Xiong B, Stessman HA, Coe BP, Penn O, Witherspoon K, Gerds J, Baker C, Vulto-van Silfhout AT, Schuurs-Hoeijmakers JH, Fichera M, et al. 2014. Disruptive CHD8 mutations define a subtype of autism early in development. *Cell* 158:263–276.
- Berry FB, Tamimi Y, Carle MV, Lehmann OJ, Walter MA. 2005. The establishment of a predictive mutational model of the forkhead domain through the analyses of FOXC2 missense mutations identified in patients with hereditary lymphedema with distichiasis. *Hum Mol Genet* 14:2619–2627.
- Beysen D, De Jaegere S, Amor D, Bouchard P, Christin-Maitre S, Fellous M, Touraine P, Grix AW, Hennekam R, Meire F, Oyen N, Wilson LC, et al. 2008. Identification of 34 novel and 56 known FOXL2 mutations in patients with Blepharophimosis syndrome. *Hum Mutat* 29:E205-219.
- Bilican B, Goding CR. 2006. Cell cycle regulation of the T-box transcription factor *tbx2*. *Exp Cell Res* 312:2358–2366.
- Blumkin L, Leshinsky-Silver E, Zerem A, Yosovich K, Lerman-Sagie T, Lev D. 2014. Heterozygous Mutations in the ADCK3 Gene in Siblings with Cerebellar Atrophy and Extreme Phenotypic Variability. *JIMD Rep* 12:103–107.
- Bosch DG, Boonstra FN, Gonzaga-Jauregui C, Xu M, De JL, Jhangiani S, Wiszniewski W, Muzny DM, Yntema HG, Pfundt R, Vissers LE, Spruijt L, et al. 2014. NR2F1 mutations cause optic atrophy with intellectual disability. *Am J Hum Genet* 94:303–309.
- Brackertz M, Boeke J, Zhang R, Renkawitz R. 2002. Two highly related p66 proteins comprise a new family of potent transcriptional repressors interacting with MBD2 and MBD3. *J Biol Chem* 277:40958–40966.
- Brackertz M, Gong Z, Leers J, Renkawitz R. 2006. p66alpha and p66beta of the Mi-2/NuRD complex mediate MBD2 and histone interaction. *Nucleic Acids Res* 34:397–406.

- Brice G, Mansour S, Bell R, Collin JRO, Child AH, Brady AF, Sarfarazi M, Burnand KG, Jeffery S, Mortimer P, Murday VA. 2002. Analysis of the phenotypic abnormalities in lymphoedema-distichiasis syndrome in 74 patients with FOXC2 mutations or linkage to 16q24. *J Med Genet* 39:478–483.
- Brückner A, Polge C, Lentze N, Auerbach D, Schlattner U. 2009. Yeast two-hybrid, a powerful tool for systems biology. *Int J Mol Sci* 10:2763–2788.
- Brunjes PC, Osterberg SK. 2015. Developmental Markers Expressed in Neocortical Layers Are Differentially Exhibited in Olfactory Cortex. *PLoS ONE* 10:e0138541.
- Buchberger E, El Harchi M, Payrhuber D, Zommer A, Schauer D, Simonitsch-Klupp I, Bilban M, Brostjan C. 2013. Overexpression of the transcriptional repressor complex BCL-6/BCoR leads to nuclear aggregates distinct from classical aggregates. *PLoS ONE* 8:e76845.
- Bulfone A, Smiga SM, Shimamura K, Peterson A, Puellas L, Rubenstein JL. 1995. T-brain-1: a homolog of Brachyury whose expression defines molecularly distinct domains within the cerebral cortex. *Neuron* 15:63–78.
- Bulfone A, Wang F, Hevner R, Anderson S, Cutforth T, Chen S, Meneses J, Pedersen R, Axel R, Rubenstein JL. 1998. An olfactory sensory map develops in the absence of normal projection neurons or GABAergic interneurons. *Neuron* 21:1273–1282.
- Butz S, Okamoto M, Südhof TC. 1998. A tripartite protein complex with the potential to couple synaptic vesicle exocytosis to cell adhesion in brain. *Cell* 94:773–782.
- Cánovas J, Berndt FA, Sepúlveda H, Aguilar R, Veloso FA, Montecino M, Oliva C, Maass JC, Sierralta J, Kukuljan M. 2015. The Specification of Cortical Subcerebral Projection Neurons Depends on the Direct Repression of TBR1 by CTIP1/BCL11a. *J Neurosci* 35:7552–7564.
- Carr CW, Moreno-De-Luca D, Parker C, Zimmerman HH, Ledbetter N, Martin CL, Dobyns WB, Abdul-Rahman OA. 2010. Chiari I malformation, delayed gross motor skills, severe speech delay, and epileptiform discharges in a child with FOXP1 haploinsufficiency. *Eur J Hum Genet* 18:1216–1220.
- Carvill GL, Heavin SB, Yendle SC, McMahon JM, O’Roak BJ, Cook J, Khan A, Dorschner MO, Weaver M, Calvert S, Malone S, Wallace G, et al. 2013. Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. *Nat Genet* 45:825–830.
- Cassandri M, Smirnov A, Novelli F, Pitolli C, Agostini M, Malewicz M, Melino G, Raschellà G. 2017. Zinc-finger proteins in health and disease. *Cell Death Discov* 3:17071.
- Chabout J, Sarkar A, Patel SR, Radden T, Dunson DB, Fisher SE, Jarvis ED. 2016. A Foxp2 Mutation Implicated in Human Speech Deficits Alters Sequencing of Ultrasonic Vocalisations in Adult Male Mice. *Front Behav Neurosci* 10:197.
- Chan CM, Fulton J, Montiel-Duarte C, Collins HM, Bharti N, Wadelin FR, Moran PM, Mongan NP, Heery DM. 2013. A signature motif mediating selective interactions of BCL11A with the NR2E/F subfamily of orphan nuclear receptors. *Nucleic Acids Res* 41:9663–9679.
- Charng W-L, Karaca E, Coban Akdemir Z, Gambin T, Atik MM, Gu S, Posey JE, Jhangiani SN, Muzny DM, Doddapaneni H, Hu J, Boerwinkle E, et al. 2016. Exome sequencing in mostly consanguineous Arab families with neurologic disease provides a high potential molecular diagnosis rate. *BMC Med Genomics* 9:42.
- Chen XS, Reader RH, Hoischen A, Veltman JA, Simpson NH, Francks C, Newbury DF, Fisher SE. 2017. Next-generation DNA sequencing identifies novel gene variants and pathways involved in specific language impairment. *Sci Rep* 7:46105.

- Chokas AL, Trivedi CM, Lu MM, Tucker PW, Li S, Epstein JA, Morrisey EE. 2010. Foxp1/2/4-NuRD interactions regulate gene expression and epithelial injury response in the lung via regulation of interleukin-6. *J Biol Chem* 285:13304–13313.
- Chuang H-C, Huang T-N, Hsueh Y-P. 2014. Neuronal excitation upregulates Tbr1, a high-confidence risk gene of autism, mediating Grin2b expression in the adult brain. *Front Cell Neurosci* 8:280.
- Cipriani S, Nardelli J, Verney C, Delezoide A-L, Guimiot F, Gressens P, Adle-Biassette H. 2016. Dynamic Expression Patterns of Progenitor and Pyramidal Neuron Layer Markers in the Developing Human Hippocampus. *Cereb Cortex* 26:1255–1271.
- De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, Kou Y, Liu L, Fromer M, Walker S, Singh T, Klei L, et al. 2014. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 515:209–215.
- De Rubeis S, Pasciuto E, Li KW, Fernández E, Di Marino D, Buzzi A, Ostroff LE, Klann E, Zwartkruis FJT, Komiyama NH, Grant SGN, Poujol C, et al. 2013. CYFIP1 coordinates mRNA translation and cytoskeleton remodeling to ensure proper dendritic spine formation. *Neuron* 79:1169–1182.
- Deciphering Developmental Disorders Study. 2015. Large-scale discovery of novel genetic causes of developmental disorders. *Nature* 519:223–228.
- Deriziotis P, Fisher SE. 2017. Speech and Language: Translating the Genome. *Trends Genet* 33:642–656.
- Deriziotis P, Graham SA, Estruch SB, Fisher SE. 2014a. Investigating protein-protein interactions in live cells using bioluminescence resonance energy transfer. *J Vis Exp*.
- Deriziotis P, O’Roak BJ, Graham SA, Estruch SB, Dimitropoulou D, Bernier RA, Gerdts J, Shendure J, Eichler EE, Fisher SE. 2014b. De novo TBR1 mutations in sporadic autism disrupt protein functions. *Nat Commun* 5:4954.
- Devanna P, Middelbeek J, Vernes SC. 2014. FOXP2 drives neuronal differentiation by interacting with retinoic acid signaling pathways. *Front Cell Neurosci* 8:305.
- D’Gama AM, Pochareddy S, Li M, Jamuar SS, Reiff RE, Lam AN, Sestan N, Walsh CA. 2015. Targeted DNA Sequencing from Autism Spectrum Disorder Brains Implicates Multiple Genetic Mechanisms. *Neuron* 88:910–917.
- Dias C, Estruch SB, Graham SA, McRae J, Sawiak SJ, Hurst JA, Joss SK, Holder SE, Morton JEV, Turner C, Thevenon J, Mellul K, et al. 2016. BCL11A Haploinsufficiency Causes an Intellectual Disability Syndrome and Dysregulates Transcription. *Am J Hum Genet* 99:253–274.
- Dignam JD, Lebovitz RM, Roeder RG. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* 11:1475–1489.
- Dosman CF, Andrews D, Goulden KJ. 2012. Evidence-based milestone ages as a framework for developmental surveillance. *Paediatr Child Health* 17:561–568.
- Egan CM, Nyman U, Skotte J, Streubel G, Turner S, O’Connell DJ, Rraklli V, Dolan MJ, Chadderton N, Hansen K, Farrar GJ, Helin K, et al. 2013. CHD5 is required for neurogenesis and has a dual role in facilitating gene expression and polycomb gene repression. *Dev Cell* 26:223–236.
- Eicher JD, Powers NR, Miller LL, Akshoomoff N, Amaral DG, Bloss CS, Libiger O, Schork NJ, Darst BF, Casey BJ, Chang L, Ernst T, et al. 2013. Genome-wide association study of shared components of reading disability and language impairment. *Genes Brain Behav* 12:792–801.
- Eising E, Carrion-Castillo A, Vino A, Strand EA, Jakielski KJ, Scerri TS, Hildebrand MS, Webster R, Ma A, Mazoyer B, Francks C, Bahlo M, et al. 2018. A set of regulatory genes

co-expressed in embryonic human brain is implicated in disrupted speech development. *Mol Psychiatry* (in press).

- Ercan-Sencicek AG, Davis Wright NR, Sanders SJ, Oakman N, Valdes L, Bakkaloglu B, Doyle N, Yrigollen CM, Morgan TM, Grigorenko EL. 2012. A balanced t(10;15) translocation in a male patient with developmental language disorder. *Eur J Med Genet* 55:128–131.
- Estruch SB, Graham SA, Chinnappa SM, Deriziotis P, Fisher SE. 2016a. Functional characterisation of rare FOXP2 variants in neurodevelopmental disorder. *J Neurodev Disord* 8:44.
- Estruch SB, Graham SA, Deriziotis P, Fisher SE. 2016b. The language-related transcription factor FOXP2 is post-translationally modified with small ubiquitin-like modifiers. *Sci Rep* 6:20911.
- Estruch SB, Graham SA, Quevedo M, Vino A, Dekkers DHW, Deriziotis P, Sollis E, Demmers J, Poot RA, Fisher SE. 2018. Proteomic analysis of FOXP proteins reveals interactions between cortical transcription factors associated with neurodevelopmental disorders. *Hum Mol Genet* 27:1212–1227.
- Feller C, Prestel M, Hartmann H, Straub T, Söding J, Becker PB. 2012. The MOF-containing NSL complex associates globally with housekeeping genes, but activates only a defined subset. *Nucleic Acids Res* 40:1509–1522.
- Feng Q, Cao R, Xia L, Erdjument-Bromage H, Tempst P, Zhang Y. 2002. Identification and functional characterisation of the p66/p68 components of the MeCP1 complex. *Mol Cell Biol* 22:536–546.
- Ferland RJ, Cherry TJ, Preware PO, Morrissey EE, Walsh CA. 2003. Characterisation of *Foxp2* and *Foxp1* mRNA and protein in the developing and mature brain. *J Comp Neurol* 460:266–279.
- Feuk L, Kalervo A, Lipsanen-Nyman M, Skaug J, Nakabayashi K, Finucane B, Hartung D, Innes M, Kerem B, Nowaczyk MJ, Rivlin J, Roberts W, et al. 2006. Absence of a paternally inherited FOXP2 gene in developmental verbal dyspraxia. *Am J Hum Genet* 79:965–972.
- Fink AJ, Englund C, Daza RAM, Pham D, Lau C, Nivison M, Kowalczyk T, Hevner RF. 2006. Development of the deep cerebellar nuclei: transcription factors and cell migration from the rhombic lip. *J Neurosci* 26:3066–3076.
- Fisher SE, Vargha-Khadem F, Watkins KE, Monaco AP, Pembrey ME. 1998. Localisation of a gene implicated in a severe speech and language disorder. *Nat Genet* 18:168–170.
- Francks C, Paracchini S, Smith SD, Richardson AJ, Scerri TS, Cardon LR, Marlow AJ, MacPhie IL, Walter J, Pennington BF, Fisher SE, Olson RK, et al. 2004. A 77-kilobase region of chromosome 6p22.2 is associated with dyslexia in families from the United Kingdom and from the United States. *Am J Hum Genet* 75:1046–1058.
- French CA, Fisher SE. 2014. What can mice tell us about *Foxp2* function? *Curr Opin Neurobiol* 28:72–79.
- French CA, Jin X, Campbell TG, Gerfen E, Groszer M, Fisher SE, Costa RM. 2012. An aetiological *Foxp2* mutation causes aberrant striatal activity and alters plasticity during skill learning. *Mol Psychiatry* 17:1077–1085.
- Fukai R, Hiraki Y, Yofune H, Tsurusaki Y, Nakashima M, Saitsu H, Tanaka F, Miyake N, Matsumoto N. 2015. A case of autism spectrum disorder arising from a de novo missense mutation in POGZ. *J Hum Genet* 60:277–279.
- Gabut M, Samavarchi-Tehrani P, Wang X, Slobodeniuc V, O’Hanlon D, Sung H-K, Alvarez M, Talukder S, Pan Q, Mazzoni EO, Nedelec S, Wichterle H, et al. 2011. An alternative splicing switch regulates embryonic stem cell pluripotency and reprogramming. *Cell* 147:132–146.

- Galletta BJ, Rusan NM. 2015. A yeast two-hybrid approach for probing protein-protein interactions at the centrosome. *Methods Cell Biol* 129:251–277.
- Gerlitz G, Darhin E, Giorgio G, Franco B, Reiner O. 2005. Novel functional features of the Lis-H domain: role in protein dimerisation, half-life and cellular localisation. *Cell Cycle* 4:1632–1640.
- Gialluisi A, Newbury DF, Wilcutt EG, Olson RK, DeFries JC, Brandler WM, Pennington BF, Smith SD, Scerri TS, Simpson NH, SLI Consortium, Luciano M, et al. 2014. Genome-wide screening for DNA variants associated with reading and language traits. *Genes Brain Behav* 13:686–701.
- Gialluisi A, Visconti A, Willcutt EG, Smith SD, Pennington BF, Falchi M, DeFries JC, Olson RK, Francks C, Fisher SE. 2016. Investigating the effects of copy number variants on reading and language performance. *J Neurodev Disord* 8:17.
- Gingras A-C, Gstaiger M, Raught B, Aebersold R. 2007. Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol* 8:645–654.
- Girirajan S, Brkanac Z, Coe BP, Baker C, Vives L, Vu TH, Shafer N, Bernier R, Ferrero GB, Silengo M, Warren ST, Moreno CS, et al. (2011) Relative burden of large CNVs on a range of neurodevelopmental phenotypes. *PLOS Genet* 7:e 1002334.
- Graham SA, Deriziotis P, Fisher SE. 2015. Insights into the genetic foundations of human communication. *Neuropsychol Rev* 25:3–26.
- Graham SA, Fisher SE. 2015. Understanding Language from a Genomic Perspective. *Annu Rev Genet* 49:131–160.
- Groszer M, Keays DA, Deacon RMJ, Bono JP de, Prasad-Mulcare S, Gaub S, Baum MG, French CA, Nicod J, Coventry JA, Enard W, Fray M, et al. 2008. Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits. *Curr Biol* 18:354–362.
- Hamdan FF, Daoud H, Rochefort D, Piton A, Gauthier J, Langlois M, Foomani G, Dobrzaniecka S, Krebs MO, Joobar R, Lafrenière RG, Lacaille JC, et al. 2010. De novo mutations in FOXP1 in cases with intellectual disability, autism, and language impairment. *Am J Hum Genet* 87:671–678.
- Hamdan FF, Srour M, Capo-Chichi J-M, Daoud H, Nassif C, Patry L, Massicotte C, Ambalavanan A, Spiegelman D, Diallo O, Henrion E, Dionne-Laporte A, et al. 2014. De novo mutations in moderate or severe intellectual disability. *PLoS Genet* 10:e1004772.
- Hannula-Jouppi K, Kaminen-Ahola N, Taipale M, Eklund R, Nopola-Hemmi J, Kääriäinen H, Kere J. 2005. The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. *PLoS Genet* 1:e50.
- Helsmoortel C, Vulto-van AS, Coe BP, Vandeweyer G, Rooms L, Van J den E, Schuurs-Hoeijmakers JH, Marcelis CL, Willemsen MH, Vissers LE, Yntema HG, Bakshi M, et al. 2014. A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. *Nat Genet* 46:380–384.
- Hevner RF, Shi L, Justice N, Hsueh Y, Sheng M, Smiga S, Bulfone A, Goffinet AM, Campagnoni AT, Rubenstein JL. 2001. Tbr1 regulates differentiation of the preplate and layer 6. *Neuron* 29:353–366.
- Hisaoka T, Nakamura Y, Senba E, Morikawa Y. 2010. The forkhead transcription factors, Foxp1 and Foxp2, identify different subpopulations of projection neurons in the mouse cerebral cortex. *Neuroscience* 166:551–563.
- Hoed J den. 2016. Investigating protein networks of language-related transcription factors. (Master's thesis).
- Hoed J den, Sollis E, Venselaar H, Estruch SB, Deriziotis P, Fisher SE. 2018. Functional characterisation of TBR1 variants in neurodevelopmental disorder. *Sci Rep* 8:14279–14279.

- Hong C-J, Hsueh Y-P. 2007. Cytoplasmic distribution of T-box transcription factor Tbr-1 in adult rodent brain. *J Chem Neuroanat* 33:124–130.
- Horn D, Kapeller J, Rivera-Brugués N, Moog U, Lorenz-Depiereux B, Eck S, Hempel M, Wagenstaller J, Gawthroppe A, Monaco AP, Bonin M, Riess O, et al. 2010. Identification of FOXP1 deletions in three unrelated patients with mental retardation and significant speech and language deficits. *Hum Mutat* 31:E1851-1860.
- Hsueh Y-P. 2006. The role of the MAGUK protein CASK in neural development and synaptic function. *Curr Med Chem* 13:1915–1927.
- Hsueh YP, Wang TF, Yang FC, Sheng M. 2000. Nuclear translocation and transcription regulation by the membrane-associated guanylate kinase CASK/LIN-2. *Nature* 404:298–302.
- Hu Y, Liu X, Zhang A, Zhou H, Liu Z, Chen H, Jin M. 2015. CHD3 facilitates vRNP nuclear export by interacting with NES1 of influenza A virus NS2. *Cell Mol Life Sci* 72:971–982.
- Huang T-N, Chuang H-C, Chou W-H, Chen C-Y, Wang H-F, Chou S-J, Hsueh Y-P. 2014. Tbr1 haploinsufficiency impairs amygdalar axonal projections and results in cognitive abnormality. *Nat Neurosci* 17:240–247.
- Huang T-N, Hsueh Y-P. 2015. Brain-specific transcriptional regulator T-brain-1 controls brain wiring and neuronal activity in autism spectrum disorders. *Front Neurosci* 9:406.
- Humphrey GW, Wang Y, Russanova VR, Hirai T, Qin J, Nakatani Y, Howard BH. 2001. Stable histone deacetylase complexes distinguished by the presence of SANT domain proteins CoREST/kiaa0071 and Mta-L1. *J Biol Chem* 276:6817–6824.
- Innocenti M, Zucconi A, Disanza A, Frittoli E, Areces LB, Steffen A, Stradal TEB, Di Fiore PP, Carlier M-F, Scita G. 2004. Abi1 is essential for the formation and activation of a WAVE2 signalling complex. *Nat Cell Biol* 6:319–327.
- Iossifov I, O’Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, Stessman HA, Witherspoon KT, Vives L, Patterson KE, Smith JD, Paeppe B, et al. 2014. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 515:216–221.
- Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, Yamrom B, Lee Y-H, Narzisi G, Leotta A, Kendall J, Grabowska E, et al. 2012. De novo gene disruptions in children on the autistic spectrum. *Neuron* 74:285–299.
- Iqbal Z, Vandeweyer G, Voet M van der, Waryah AM, Zahoor MY, Besseling JA, Roca LT, Vulto-van Silfhout AT, Nijhof B, Kramer JM, Van der Aa N, Ansar M, et al. 2013. Homozygous and heterozygous disruptions of ANK3: at the crossroads of neurodevelopmental and psychiatric disorders. *Hum Mol Genet* 22:1960–1970.
- Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M. 2005. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* 4:1265–1272.
- Ito YA, Footz TK, Berry FB, Mirzayans F, Yu M, Khan AO, Walter MA. 2009. Severe molecular defects of a novel FOXC1 W152G mutation result in aniridia. *Invest Ophthalmol Vis Sci* 50:3573–3579.
- Iwase S, Bérubé NG, Zhou Z, Kasri NN, Battaglioli E, Scandaglia M, Barco A. 2017. Epigenetic Etiology of Intellectual Disability. *J Neurosci* 37:10773–10782.
- Jackson RS, Cho Y-J, Stein S, Liang P. 2007. CYFIP2, a direct p53 target, is leptomycin-B sensitive. *Cell Cycle* 6:95–103.
- Jolly LA, Nguyen LS, Domingo D, Sun Y, Barry S, Hancarova M, Plevova P, Vlckova M, Havlovicova M, Kalscheuer VM, Graziano C, Pippucci T, et al. 2015. HCFC1 loss-of-function mutations disrupt neuronal and neural progenitor cells of the developing brain. *Hum Mol Genet* 24:3335–3347.

- Jolma A, Yan J, Whittington T, Toivonen J, Nitta KR, Rastas P, Morgunova E, Enge M, Taipale M, Wei G, Palin K, Vaquerizas JM, et al. 2013. DNA-binding specificities of human transcription factors. *Cell* 152:327–339.
- Kang C, Riazuddin S, Mundorff J, Krasnewich D, Friedman P, Mullikin JC, Drayna D. 2010. Mutations in the lysosomal enzyme-targeting pathway and persistent stuttering. *N Engl J Med* 362:677–685.
- Kawase C, Kawase K, Taniguchi T, Sugiyama K, Yamamoto T, Kitazawa Y, Alward WL, Stone EM, Nishimura DY, Sheffield VC. 2001. Screening for mutations of Axenfeld-Rieger syndrome caused by FOXC1 gene in Japanese patients. *J Glaucoma* 10:477–482.
- Klenova E, Chernukhin I, Inoue T, Shamsuddin S, Norton J. 2002. Immunoprecipitation techniques for the analysis of transcription factor complexes. *Methods* 26:254–259.
- Koga M, Ishiguro H, Yazaki S, Horiuchi Y, Arai M, Niizato K, Iritani S, Itokawa M, Inada T, Iwata N, Ozaki N, Ujike H, et al. 2009. Involvement of SMARCA2/BRM in the SWI/SNF chromatin-remodeling complex in schizophrenia. *Hum Mol Genet* 18:2483–2494.
- Kohzaki H, Murakami Y. 2005. Transcription factors and DNA replication origin selection. *Bioessays* 27:1107–1116.
- Koolen DA, Kramer JM, Neveling K, Nillesen WM, Moore-Barton HL, Elmslie FV, Toutain A, Amiel J, Malan V, Tsai AC, Cheung SW, Gilissen C, et al. 2012. Mutations in the chromatin modifier gene KANSL1 cause the 17q21.31 microdeletion syndrome. *Nat Genet* 44:639–641.
- Koolen DA, Vries BB de. 2013. KANSL1-Related Intellectual Disability Syndrome. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJ, Stephens K, Amemiya A, editors. *GeneReviews®*, Seattle (WA): University of Washington, Seattle,.
- Kornilov SA, Rakhlin N, Kuposov R, Lee M, Yrigollen C, Caglayan AO, Magnuson JS, Mane S, Chang JT, Grigorenko EL. 2016. Genome-Wide Association and Exome Sequencing Study of Language Disorder in an Isolated Population. *Pediatrics* 137:.
- Kosugi S, Hasebe M, Tomita M, Yanagawa H. 2009. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc Natl Acad Sci USA* 106:10171–10176.
- Krajewska-Walasek M, Jurkiewicz D, Piekutowska-Abramczuk D, Kucharczyk M, Chrzanowska KH, Jezela-Stanek A, Ciara E. 2016. Additional data on the clinical phenotype of Helsmoortel-Van der Aa syndrome associated with a novel truncating mutation in ADNP gene. *Am J Med Genet A* 170:1647–1650.
- Krakstad BF, Ardawatia VV, Aragay AM. 2004. A role for Galpha12/Galpha13 in p120ctn regulation. *Proc Natl Acad Sci USA* 101:10314–10319.
- Kumar P P, Franklin S, Emechebe U, Hu H, Moore B, Lehman C, Yandell M, Moon AM. 2014. TBX3 regulates splicing in vivo: a novel molecular mechanism for Ulnar-mammary syndrome. *PLoS Genet* 10:e1004247.
- Kunapuli P, Kasyapa CS, Chin S-F, Caldas C, Cowell JK. 2006. ZNF198, a zinc finger protein rearranged in myeloproliferative disease, localizes to the PML nuclear bodies and interacts with SUMO-1 and PML. *Exp Cell Res* 312:3739–3751.
- Laffin JJS, Raca G, Jackson CA, Strand EA, Jakielski KJ, Shriberg LD. 2012. Novel candidate genes and regions for childhood apraxia of speech identified by array comparative genomic hybridisation. *Genet Med* 14:928–936.
- Lai AY, Wade PA. 2011. Cancer biology and NuRD: a multifaceted chromatin remodelling complex. *Nat Rev Cancer* 11:588–596.
- Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. 2001. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413:519–523.

- Lai JS, Herr W. 1992. Ethidium bromide provides a simple tool for identifying genuine DNA-independent protein associations. *Proc Natl Acad Sci USA* 89:6958–6962.
- Le Fevre AK, Taylor S, Malek NH, Horn D, Carr CW, Abdul-Rahman OA, O'Donnell S, Burgess T, Shaw M, Gecz J, Bain N, Fagan K, et al. 2013. FOXP1 mutations cause intellectual disability and a recognizable phenotype. *Am J Med Genet A* 161A:3166–3175.
- Lee K, Na W, Maeng J-H, Wu H, Ju B-G. 2013. Regulation of DU145 prostate cancer cell growth by Scm-like with four mbt domains 2. *J Biosci* 38:105–112.
- Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF, Boguski MS, Brockway KS, Byrnes EJ, Chen L, Chen L, et al. 2007. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445:168–76.
- Lemke JR, Lal D, Reinthaler EM, Steiner I, Nothnagel M, Alber M, Geider K, Laube B, Schwake M, Finsterwalder K, Franke A, Schilhabel M, et al. 2013. Mutations in GRIN2A cause idiopathic focal epilepsy with rolandic spikes. *Nat Genet* 45:1067–1072.
- Lennon PA, Cooper ML, Peiffer DA, Gunderson KL, Patel A, Peters S, Cheung SW, Bacino CA. 2007. Deletion of 7q31.1 supports involvement of FOXP2 in language impairment: clinical report and review. *Am J Med Genet A* 143A:791–798.
- Lesca G, Rudolf G, Bruneau N, Lozovaya N, Labalme A, Boutry-Kryza N, Salmi M, Tsintsadze T, Addis L, Motte J, Wright S, Tsintsadze V, et al. 2013. GRIN2A mutations in acquired epileptic aphasia and related childhood focal epilepsies and encephalopathies with speech and language dysfunction. *Nat Genet* 45:1061–1066.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
- Li S, Weidenfeld J, Morrissey EE. 2004. Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions. *Mol Cell Biol* 24:809–822.
- Li X, Wang W, Wang J, Malovannaya A, Xi Y, Li W, Guerra R, Hawke DH, Qin J, Chen J. 2015a. Proteomic analyses reveal distinct chromatin-associated and soluble transcription factor complexes. *Mol Syst Biol* 11:775.
- Li X, Xiao J, Fröhlich H, Tu X, Li L, Xu Y, Cao H, Qu J, Rappold GA, Chen J-G. 2015b. Foxp1 regulates cortical radial migration and neuronal morphogenesis in developing cerebral cortex. *PLoS ONE* 10:e0127671.
- Ligt J de, Willemsen MH, Bon BWM van, Kleefstra T, Yntema HG, Kroes T, Vulto-van Silfhout AT, Koolen DA, Vries P de, Gilissen C, Rosario M del, Hoischen A, et al. 2012. Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med* 367:1921–1929.
- Lin F-J, Qin J, Tang K, Tsai SY, Tsai M-J. 2011. Coup d'Etat: an orphan takes control. *Endocr Rev* 32:404–421.
- Liu Y-T, Hershenson J, Plagnol V, Fawcett K, Duberley KEC, Preza E, Hargreaves IP, Chalasani A, Laurá M, Wood NW, Reilly MM, Houlden H. 2014. Autosomal-recessive cerebellar ataxia caused by a novel ADCK3 mutation that elongates the protein: clinical, genetic and biochemical characterisation. *J Neurol Neurosurg Psychiatry* 85:493–498.
- Lozano R, Viano A, Lozano C, Fisher SE, Deriziotis P. 2015. A de novo FOXP1 variant in a patient with autism, intellectual disability and severe speech and language impairment. *Eur J Hum Genet* 23:1702–1707.
- Luo X, Zou Y, Tan B, Zhang Y, Guo J, Zeng L, Zhang R, Tan H, Wei X, Hu Y, Zheng Y, Liang D, et al. 2017. Novel *GATAD2B* loss-of-function mutations cause intellectual disability in two unrelated cases. *Journal of Human Genetics* 62:513–516.
- MacDermot KD, Bonora E, Sykes N, Coupe A-M, Lai CSL, Vernes SC, Vargha-Khadem F, McKenzie F, Smith RL, Monaco AP, Fisher SE. 2005. Identification of FOXP2 truncation

as a novel cause of developmental speech and language deficits. *Am J Hum Genet* 76:1074–1080.

- Mandel S, Gozes I. 2007. Activity-dependent neuroprotective protein constitutes a novel element in the SWI/SNF chromatin remodeling complex. *J Biol Chem* 282:34448–34456.
- Mandel S, Rechavi G, Gozes I. 2007. Activity-dependent neuroprotective protein (ADNP) differentially interacts with chromatin to regulate genes essential for embryogenesis. *Dev Biol* 303:814–824.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303.
- McKenna WL, Betancourt J, Larkin KA, Abrams B, Guo C, Rubenstein JLR, Chen B. 2011. *Tbr1* and *Fezf2* regulate alternate corticofugal neuronal identities during neocortical development. *J Neurosci* 31:549–564.
- Medvedeva YA, Lennartsson A, Ehsani R, Kulakovskiy IV, Vorontsov IE, Panahandeh P, Khimulya G, Kasukawa T, FANTOM Consortium, Drabløs F. 2015. EpiFactors: a comprehensive database of human epigenetic factors and complexes. *Database (Oxford)* 2015:bav067.
- Meek DW. 2004. The p53 response to DNA damage. *DNA Repair (Amst)* 3:1049–1056.
- Meerschaut I, Rochefort D, Revençu N, Pêtre J, Corsello C, Rouleau GA, Hamdan FF, Michaud JL, Morton J, Radley J, Rague N, García-Miñaur S, et al. 2017. FOXP1-related intellectual disability syndrome: a recognisable entity. *J Med Genet* 54:613–623.
- Mellacheruvu D, Wright Z, Couzens AL, Lambert J-P, St-Denis NA, Li T, Miteva YV, Hauri S, Sardi ME, Low TY, Halim VA, Bagshaw RD, et al. 2013. The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Methods* 10:730–736.
- Meng H, Smith SD, Hager K, Held M, Liu J, Olson RK, Pennington BF, DeFries JC, Gelernter J, O'Reilly-Pol T, Somlo S, Skudlarski P, et al. 2005. DCDC2 is associated with reading disability and modulates neuronal development in the brain. *Proc Natl Acad Sci USA* 102:17053–17058.
- Meunier S, Shvedunova M, Van Nguyen N, Avila L, Vernos I, Akhtar A. 2015. An epigenetic regulator emerges as microtubule minus-end binding and stabilizing factor in mitosis. *Nat Commun* 6:7889.
- Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD. 2017. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res* 45:D183–D189.
- Moen MJ, Adams HHH, Brandsma JH, Dekkers DHW, Akinci U, Karkampouna S, Quevedo M, Kockx CEM, Ozgür Z, IJcken WFJ van, Demmers J, Poot RA. 2017. An interaction network of mental disorder proteins in neural stem cells. *Transl Psychiatry* 7:e1082.
- Moog U, Kutsche K, Kortüm F, Chilian B, Bierhals T, Apeshiotis N, Balg S, Chassaing N, Coubes C, Das S, Engels H, Van Esch H, et al. 2011. Phenotypic spectrum associated with CASK loss-of-function mutations. *J Med Genet* 48:741–751.
- Morgan A, Fisher SE, Scheffer I, Hildebrand M. 2017. FOXP2-Related Speech and Language Disorders. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJ, Stephens K, Amemiya A, editors. *GeneReviews®*, Seattle (WA): University of Washington, Seattle,.
- Morgan AT, Haaften LV, Van KH, Edley C, Mei C, Tan TY, Amor D, Fisher SE, Koolen DA. 2018. Early speech development in Koolen de Vries syndrome limited by oral praxis and hypotonia. *Eur J Hum Genet* 26:75–84.

- Müller CW, Herrmann BG. 1997. Crystallographic structure of the T domain-DNA complex of the Brachyury transcription factor. *Nature* 389:884–888.
- Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, Lin C-F, Stevens C, Wang L-S, Makarov V, Polak P, Yoon S, et al. 2012. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 485:242–245.
- Neveling K, Feenstra I, Gilissen C, Hoefsloot LH, Kamsteeg E-J, Mensenkamp AR, Rodenburg RJT, Yntema HG, Spruijt L, Vermeer S, Rinne T, Gassen KL van, et al. 2013. A post-hoc comparison of the utility of sanger sequencing and exome sequencing for the diagnosis of heterogeneous diseases. *Hum Mutat* 34:1721–1726.
- Newbury DF, Fisher SE, Monaco AP. 2010. Recent advances in the genetics of language impairment. *Genome Med* 2:6.
- Newbury DF, Winchester L, Addis L, Paracchini S, Buckingham L-L, Clark A, Cohen W, Cowie H, Dworzynski K, Everitt A, Goodyer IM, Hennessy E, et al. 2009. CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *Am J Hum Genet* 85:264–272.
- Ng D, Thakker N, Corcoran CM, Donnai D, Perveen R, Schneider A, Hadley DW, Tiftt C, Zhang L, Wilkie AO, Van J der S, Gorlin RJ, et al. 2004. Oculofaciocardiodental and Lenz microphthalmia syndromes result from distinct classes of mutations in BCOR. *Nat Genet* 36:411–416.
- Nguyen Ba AN, Pogoutse A, Provart N, Moses AM. 2009. NLStradamus: a simple Hidden Markov Model for nuclear localisation signal prediction. *BMC Bioinformatics* 10:202.
- Nitarska J, Smith JG, Sherlock WT, Hillege MMG, Nott A, Barshop WD, Vashisht AA, Wohlschlegel JA, Mitter R, Riccio A. 2016. A Functional Switch of NuRD Chromatin Remodeling Complex Subunits Regulates Mouse Cortical Development. *Cell Rep* 17:1683–1698.
- Notwell JH, Heavner WE, Darbandi SF, Katzman S, McKenna WL, Ortiz-Londono CF, Tastad D, Eckler MJ, Rubenstein JLR, McConnell SK, Chen B, Bejerano G. 2016. TBR1 regulates autism risk genes in the developing neocortex. *Genome Res* 26:1013–1022.
- O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, Karakoc E, Mackenzie AP, Ng SB, Baker C, Rieder MJ, Nickerson DA, et al. 2011. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet* 43:585–589.
- O'Roak BJ, Stessman HA, Boyle EA, Witherspoon KT, Martin B, Lee C, Vives L, Baker C, Hiatt JB, Nickerson DA, Bernier R, Shendure J, et al. 2014. Recurrent de novo mutations implicate novel genes underlying simplex autism risk. *Nat Commun* 5:5595–5595.
- O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG, Carvill G, Kumar A, Lee C, Ankenman K, Munson J, Hiatt JB, et al. 2012a. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 338:1619–1622.
- O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, Levy R, Ko A, Lee C, Smith JD, Turner EH, Stanaway IB, et al. 2012b. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 485:246–250.
- Overmeer RM, Gourdin AM, Giglia-Mari A, Kool H, Houtsmuller AB, Siegal G, Fouteri MI, Mullenders LHF, Vermeulen W. 2010. Replication factor C recruits DNA polymerase delta to sites of nucleotide excision repair but is not required for PCNA recruitment. *Mol Cell Biol* 30:4828–4839.
- Pagnamenta AT, Bacchelli E, Jonge MV de, Mirza G, Scerri TS, Minopoli F, Chiocchetti A, Ludwig KU, Hoffmann P, Paracchini S, Lowy E, Harold DH, et al. 2010. Characterisation of a family with rare deletions in CNTNAP5 and DOCK4 suggests novel risk loci for autism and dyslexia. *Biol Psychiatry* 68:320–328.

- Palka C, Alfonsi M, Mohn A, Cerbo R, Guanciali Franchi P, Fantasia D, Morizio E, Stuppia L, Calabrese G, Zori R, Chiarelli F, Palka G. 2012. Mosaic 7q31 deletion involving FOXP2 gene associated with language impairment. *Pediatrics* 129:e183-188.
- Palumbo O, Fichera M, Palumbo P, Rizzo R, Mazzolla E, Cocuzza DM, Carella M, Mattina T. 2014. TBR1 is the candidate gene for intellectual disability in patients with a 2q24.2 interstitial deletion. *Am J Med Genet A* 164A:828-833.
- Paracchini S, Thomas A, Castro S, Lai C, Paramasivam M, Wang Y, Keating BJ, Taylor JM, Hacking DF, Scerri T, Francks C, Richardson AJ, et al. 2006. The chromosome 6p22 haplotype associated with dyslexia reduces the expression of KIAA0319, a novel gene involved in neuronal migration. *Hum Mol Genet* 15:1659-1666.
- Pariani MJ, Spencer A, Graham JJ, Rimoin DL. 2009. A 785kb deletion of 3p14.1p13, including the FOXP1 gene, associated with speech delay, contractures, hypertonia and blepharophimosis. *Eur J Med Genet* 52:123-127.
- Park HJ, Costa RH, Lau LF, Tyner AL, Raychaudhuri P. 2008. Anaphase-promoting complex/cyclosome-CDH1-mediated proteolysis of the forkhead box M1 transcription factor is critical for regulated entry into S phase. *Mol Cell Biol* 28:5162-5171.
- Peter B, Matsushita M, Oda K, Raskind W. 2014. De novo microdeletion of BCL11A is associated with severe speech sound disorder. *Am J Med Genet A* 164A:2091-2096.
- Peters J-M. 2006. The anaphase promoting complex/cyclosome: a machine designed to destroy. *Nat Rev Mol Cell Biol* 7:644-656.
- Pfeifer GP. 2006. Mutagenesis at methylated CpG sequences. *Curr Top Microbiol Immunol* 301:259-281.
- Pflieger KDG, Eidne KA. 2006. Illuminating insights into protein-protein interactions using bioluminescence resonance energy transfer (BRET). *Nat Methods* 3:165-174.
- Raza MH, Mattera R, Morell R, Sainz E, Rahn R, Gutierrez J, Paris E, Root J, Solomon B, Brewer C, Basra MAR, Khan S, et al. 2015. Association between Rare Variants in AP4E1, a Component of Intracellular Trafficking, and Persistent Stuttering. *Am J Hum Genet* 97:715-725.
- Redin C, Gérard B, Lauer J, Herenger Y, Muller J, Quartier A, Masurel-Paulet A, Willems M, Lesca G, El-Chehadeh S, Le Gras S, Vicaire S, et al. 2014. Efficient strategy for the molecular diagnosis of intellectual disability using targeted high-throughput sequencing. *J Med Genet* 51:724-736.
- Remedios R, Huilgol D, Saha B, Hari P, Bhatnagar L, Kowalczyk T, Hevner RF, Suda Y, Aizawa S, Ohshima T, Stoykova A, Tole S. 2007. A stream of cells migrating from the caudal telencephalon reveals a link between the amygdala and neocortex. *Nat Neurosci* 10:1141-1150.
- Reuter MS, Riess A, Moog U, Briggs TA, Chandler KE, Rauch A, Stampfer M, Steindl K, Gläser D, Joset P, DDD Study, Krumbiegel M, et al. 2017. FOXP2 variants in 14 individuals with developmental speech and language disorders broaden the mutational and clinical spectrum. *J Med Genet* 54:64-72.
- Rice GM, Raca G, Jakielski KJ, Laffin JJ, Iyama-Kurtycz CM, Hartley SL, Sprague RE, Heintzelman AT, Shriberg LD. 2012. Phenotype of FOXP2 haploinsufficiency in a mother and son. *Am J Med Genet A* 158A:174-181.
- Roll P, Rudolf G, Pereira S, Royer B, Scheffer IE, Massacrier A, Valenti M-P, Roedel-Trevisiol N, Jamali S, Beclin C, Seegmuller C, Metz-Lutz M-N, et al. 2006. SRPX2 mutations in disorders of language cortex and cognition. *Hum Mol Genet* 15:1195-1207.
- Roll P, Vernes SC, Bruneau N, Cillario J, Ponsolle-Lenfant M, Massacrier A, Rudolf G, Khalife M, Hirsch E, Fisher SE, Szeppetowski P. 2010. Molecular networks implicated in

speech-related disorders: FOXP2 regulates the SRPX2/uPAR complex. *Hum Mol Genet* 19:4848–4860.

- Ronan JL, Wu W, Crabtree GR. 2013. From neural development to cognition: unexpected roles for chromatin. *Nat Rev Genet* 14:347–359.
- Saleem RA, Banerjee-Basu S, Berry FB, Baxeavanis AD, Walter MA. 2003. Structural and functional analyses of disease-causing missense mutations in the forkhead domain of FOXC1. *Hum Mol Genet* 12:2993–3005.
- Sankaran VG, Menne TF, Xu J, Akie TE, Lettre G, Van Handel B, Mikkola HKA, Hirschhorn JN, Cantor AB, Orkin SH. 2008. Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science* 322:1839–1842.
- Santen GWE, Kriek M, Attikum H van. 2012. SWI/SNF complex in disorder: SWITching from malignancies to intellectual disability. *Epigenetics* 7:1219–1224.
- Schenck A, Bardoni B, Moro A, Bagni C, Mandel JL. 2001. A highly conserved protein family interacting with the fragile X mental retardation protein (FMRP) and displaying selective interactions with FMRP-related proteins FXR1P and FXR2P. *Proc Natl Acad Sci USA* 98:8844–8849.
- Sen P, Yang Y, Navarro C, Silva I, Szafranski P, Kolodziejska KE, Dharmadhikari AV, Mostafa H, Kozakewich H, Kearney D, Cahill JB, Whitt M, et al. 2013. Novel FOXF1 mutations in sporadic and familial cases of alveolar capillary dysplasia with misaligned pulmonary veins imply a role for its DNA binding domain. *Hum Mutat* 34:801–811.
- Shelly M, Mosesson Y, Citri A, Lavi S, Zwang Y, Melamed-Book N, Aroeti B, Yarden Y. 2003. Polar expression of ErbB-2/HER2 in epithelia. Bimodal regulation by Lin-7. *Dev Cell* 5:475–486.
- Shriberg LD, Ballard KJ, Tomblin JB, Duffy JR, Odell KH, Williams CA. 2006. Speech, prosody, and voice characteristics of a mother and daughter with a 7;13 translocation affecting FOXP2. *J Speech Lang Hear Res* 49:500–525.
- Shu W, Lu MM, Zhang Y, Tucker PW, Zhou D, Morrisey EE. 2007. Foxp2 and Foxp1 cooperatively regulate lung and esophagus development. *Development* 134:1991–2000.
- Sifrim A, Hitz M-P, Wilsdon A, Breckpot J, Turki SHA, Thienpont B, McRae J, Fitzgerald TW, Singh T, Swaminathan GJ, Prigmore E, Rajan D, et al. 2016. Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nat Genet* 48:1060–1065.
- Simpson NH, Ceroni F, Reader RH, Covill LE, Knight JC; SLI Consortium, Hennessy ER, Bolton PF, Conti-Ramsden G, O'Hare A, Baird G, Fisher SE, et al. 2015. Genome-wide analysis identifies a role for common copy number variants in specific language impairment. *Eur J Hum Genet* 23:1370–1377.
- Sin C, Li H, Crawford DA. 2015. Transcriptional regulation by FOXP1, FOXP2, and FOXP4 dimerisation. *J Mol Neurosci* 55:437–448.
- Smith AW, Holden KR, Dwivedi A, Dupont BR, Lyons MJ. 2015. Deletion of 16q24.1 Supports a Role for the ATP2C2 Gene in Specific Language Impairment. *J Child Neurol*, 30:517–521.
- Snijders Blok L, Rousseau J, Twist J, Ehresmann S, Takaku M, Venselaar H, Rodan LH, Nowak CB, Douglas J, Swoboda KJ, Steeves MA, Sahai I, et al. 2018. CHD3 helicase domain mutations cause a neurodevelopmental syndrome with macrocephaly and impaired speech and language. *Nat Commun* 9:4619–4619.
- Sollis E, Deriziotis P, Saitsu H, Miyake N, Matsumoto N, Hoffer MJV, Ruivenkamp CAL, Alders M, Okamoto N, Bijlsma EK, Plomp AS, Fisher SE. 2017. Equivalent missense variant in the FOXP2 and FOXP1 transcription factors causes distinct neurodevelopmental disorders. *Hum Mutat* 38:1542–1554.

- Sollis E, Graham SA, Vino A, Froehlich H, Vreeburg M, Dimitropoulou D, Gilissen C, Pfundt R, Rappold GA, Brunner HG, Deriziotis P, Fisher SE. 2016. Identification and functional characterisation of de novo FOXP1 variants provides novel insights into the etiology of neurodevelopmental disorder. *Hum Mol Genet* 25:546–557.
- Song H, Makino Y, Noguchi E, Arinami T. 2015. A case report of de novo missense FOXP1 mutation in a non-Caucasian patient with global developmental delay and severe speech impairment. *Clin Case Rep* 3:110–113.
- Spector DL, Lamond AI. 2011. Nuclear speckles. *Cold Spring Harb Perspect Biol* 3.
- Spiteri E, Konopka G, Coppola G, Bomar J, Oldham M, Ou J, Vernes SC, Fisher SE, Ren B, Geschwind DH. 2007. Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. *Am J Hum Genet* 81:1144–1157.
- Srivastava S, Cohen JS, Vernon H, Barañano K, McClellan R, Jamal L, Naidu S, Fatemi A. 2014. Clinical whole exome sequencing in child neurology practice. *Ann Neurol* 76:473–483.
- St Pourcain B, Cents RAM, Whitehouse AJO, Haworth CMA, Davis OSP, O'Reilly PF, Roulstone S, Wren Y, Ang QW, Velders FP, Evans DM, Kemp JP, et al. 2014a. Common variation near ROBO2 is associated with expressive vocabulary in infancy. *Nat Commun* 5:4831.
- St Pourcain B, Skuse DH, Mandy WP, Wang K, Hakonarson H, Timpson NJ, Evans DM, Kemp JP, Ring SM, McArdle WL, Golding J, Smith GD. 2014b. Variability in the common genetic architecture of social-communication spectrum phenotypes during childhood and adolescence. *Mol Autism* 5:18.
- Stessman HAF, Willemsen MH, Fenckova M, Penn O, Hoischen A, Xiong B, Wang T, Hoekzema K, Vives L, Vogel I, Brunner HG, Burgt I van der, et al. 2016. Disruption of POGZ Is Associated with Intellectual Disability and Autism Spectrum Disorders. *Am J Hum Genet* 98:541–552.
- Stevenson RE, Bennett CW, Abidi F, Kleefstra T, Porteous M, Simensen RJ, Lubs HA, Hamel BCJ, Schwartz CE. 2005. Renpenning syndrome comes into focus. *Am J Med Genet A* 134:415–421.
- Stromswold K. 2006. Why aren't identical twins linguistically identical? Genetic, prenatal and postnatal factors. *Cognition* 101:333–384.
- Stroud JC, Wu Y, Bates DL, Han A, Nowick K, Paabo S, Tong H, Chen L. 2006. Structure of the forkhead domain of FOXP2 bound to DNA. *Structure* 14:159–166.
- Strumane K, Bonnomet A, Stove C, Vandenbroucke R, Nawrocki-Raby B, Bruyneel E, Mareel M, Birembaut P, Berx G, Roy F van. 2006. E-cadherin regulates human Nanos1, which interacts with p120ctn and induces tumor cell migration and invasion. *Cancer Res* 66:10007–10015.
- Sunyaev SR. 2012. Inferring causality and functional significance of human coding DNA variants. *Hum Mol Genet* 21:R10–7.
- Suter B, Kittanakom S, Stagljar I. 2008. Two-hybrid technologies in proteomics research. *Curr Opin Biotechnol* 19:316–323.
- Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, Mering C von. 2017. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* 45:D362–D368.
- Taipale M, Kaminen N, Nopola-Hemmi J, Haltia T, Myllyluoma B, Lyytinen H, Muller K, Kaaranen M, Lindsberg PJ, Hannula-Jouppi K, Kere J. 2003. A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. *Proc Natl Acad Sci USA* 100:11553–11558.

- Takahashi K, Liu F-C, Hirokawa K, Takahashi H. 2008. Expression of Foxp4 in the developing and adult rat forebrain. *J Neurosci Res* 86:3106–3116.
- Tam WY, Leung CKY, Tong KK, Kwan KM. 2011. Foxp4 is essential in maintenance of Purkinje cell dendritic arborisation in the mouse cerebellum. *Neuroscience* 172:562–571.
- Tang K, Rubenstein JLR, Tsai SY, Tsai M-J. 2012. COUP-TFII controls amygdala patterning by regulating neuropilin expression. *Development* 139:1630–1639.
- Teramitsu I, Kudo LC, London SE, Geschwind DH, White SA. 2004. Parallel FoxP1 and FoxP2 expression in songbird and human brain predicts functional interaction. *J Neurosci* 24:3152–3163.
- Thevenon J, Callier P, Andrieux J, Delobel B, David A, Sukno S, Minot D, Mosca Anne L, Marle N, Sanlaville D, Bonnet M, Masurel-Paulet A, et al. 2013. 12p13.33 microdeletion including ELKS/ERC1, a new locus associated with childhood apraxia of speech. *Eur J Hum Genet* 21:82–88.
- Thul PJ, Åkesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, Alm T, Asplund A, Björk L, Breckels LM, Bäckström A, Danielsson F, et al. 2017. A subcellular map of the human proteome. *Science* 356:.
- Tim-Aroon T, Jinawath N, Thammachote W, Sinpitak P, Limrungsikul A, Khongkhatithum C, Wattanasirichaigoon D. 2017. 1q21.3 deletion involving GATAD2B: An emerging recurrent microdeletion syndrome. *Am J Med Genet A* 173:766–770.
- Tomassy GS, De Leonibus E, Jabaudon D, Lodato S, Alfano C, Mele A, Macklis JD, Studer M. 2010. Area-specific temporal control of corticospinal motor neuron differentiation by COUP-TFI. *Proc Natl Acad Sci USA* 107:3576–3581.
- Tomblin B. 2011. Co-morbidity of autism and SLI: kinds, kin and complexity. *Int J Lang Commun Disord* 46:127–137.
- Tomblin JB, O'Brien M, Shriberg LD, Williams C, Murray J, Patil S, Bjork J, Anderson S, Ballard K. 2009. Language features in a mother and daughter of a chromosome 7;13 translocation involving FOXP2. *J Speech Lang Hear Res* 52:1157–1174.
- Turner SJ, Hildebrand MS, Block S, Damiano J, Fahey M, Reilly S, Bahlo M, Scheffer IE, Morgan AT. 2013. Small intragenic deletion in FOXP2 associated with childhood apraxia of speech and dysarthria. *Am J Med Genet A* 161A:2321–2326.
- Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, et al. 2015. Tissue-based map of the human proteome. *Science* 347:1260419.
- The UniProt Consortium. 2017. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 45D:158-169.
- Vandeweyer G, Helmsmoortel C, Van Dijck A, Vulto-van Silfhout AT, Coe BP, Bernier R, Gerds J, Rooms L, Ende J van den, Bakshi M, Wilson M, Nordgren A, et al. 2014. The transcriptional regulator ADNP links the BAF (SWI/SNF) complexes with autism. *Am J Med Genet C Semin Med Genet* 166C:315–326.
- Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. 2009. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* 10:252–263.
- Vargha-Khadem F, Watkins K, Alcock K, Fletcher P, Passingham R. 1995. Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Proc Natl Acad Sci USA* 92:930–933.
- Vargha-Khadem F, Watkins KE, Price CJ, Ashburner J, Alcock KJ, Connelly A, Frackowiak RS, Friston KJ, Pembrey ME, Mishkin M, Gadian DG, Passingham RE. 1998. Neural basis of an inherited speech and language disorder. *Proc Natl Acad Sci USA* 95:12695–12700.
- Veltman JA, Brunner HG. 2012. De novo mutations in human genetic disease. *Nat Rev Genet* 13:565–575.

- Verger A, Quinlan KGR, Crofts LA, Spanò S, Corda D, Kable EPW, Braet F, Crossley M. 2006. Mechanisms directing the nuclear localisation of the CtBP family proteins. *Mol Cell Biol* 26:4882–4894.
- Vernes SC, MacDermot KD, Monaco AP, Fisher SE. 2009. Assessing the impact of FOXP1 mutations on developmental verbal dyspraxia. *Eur J Hum Genet* 17:1354–1358.
- Vernes SC, Newbury DF, Abrahams BS, Winchester L, Nicod J, Groszer M, Alarcón M, Oliver PL, Davies KE, Geschwind DH, Monaco AP, Fisher SE. 2008. A functional genetic link between distinct developmental language disorders. *N Engl J Med* 359:2337–2345.
- Vernes SC, Nicod J, Elahi FM, Coventry JA, Kenny N, Coupe A-M, Bird LE, Davies KE, Fisher SE. 2006. Functional genetic analysis of mutations implicated in a human speech and language disorder. *Hum Mol Genet* 15:3154–3167.
- Vernes SC, Oliver PL, Spiteri E, Lockstone HE, Puliyadi R, Taylor JM, Ho J, Mombereau C, Brewer A, Lowy E, Nicod J, Groszer M, et al. 2011. Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. *PLoS Genet* 7:e1002145.
- Vernes SC, Spiteri E, Nicod J, Groszer M, Taylor JM, Davies KE, Geschwind DH, Fisher SE. 2007. High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders. *Am J Hum Genet* 81:1232–1250.
- Villanueva P, Nudel R, Hoischen A, Fernández MA, Simpson NH, Gilissen C, Reader RH, Jara L, Echeverry MM, Echeverry MM, Francks C, Baird G, et al. 2015. Exome sequencing in an admixed isolated population indicates NFXL1 variants confer a risk for specific language impairment. *PLoS Genet* 11:e1004925.
- Vissers LELM, Ligt J de, Gilissen C, Janssen I, Steehouwer M, Vries P de, Lier B van, Arts P, Wieskamp N, Rosario M del, Bon BWM van, Hoischen A, et al. 2010. A de novo paradigm for mental retardation. *Nat Genet* 42:1109–1112.
- Vissers LELM, Ravenswaaij CMA van, Admiraal R, Hurst JA, Vries BBA de, Janssen IM, Vliet WA van der, Huys EHLPG, Jong PJ de, Hamel BCJ, Schoenmakers EFPM, Brunner HG, et al. 2004. Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat Genet* 36:955–957.
- Waldron L, Steimle JD, Greco TM, Gomez NC, Dorr KM, Kweon J, Temple B, Yang XH, Wilczewski CM, Davis IJ, Cristea IM, Moskowitz IP, et al. 2016. The Cardiac TBX5 Interactome Reveals a Chromatin Remodeling Network Essential for Cardiac Septation. *Dev Cell* 36:262–275.
- Wang B, Lin D, Li C, Tucker P. 2003. Multiple domains define the expression and regulatory properties of Foxp1 forkhead transcriptional repressors. *J Biol Chem* 278:24259–24268.
- Wang B, Weidenfeld J, Lu MM, Maika S, Kuziel WA, Morrissey EE, Tucker PW. 2004a. Foxp1 regulates cardiac outflow tract, endocardial cushion morphogenesis and myocyte proliferation and maturation. *Development* 131:4477–4487.
- Wang T-F, Ding C-N, Wang G-S, Luo S-C, Lin Y-L, Ruan Y, Hevner R, Rubenstein JLR, Hsueh Y-P. 2004b. Identification of Tbr-1/CASK complex target genes in neurons. *J Neurochem* 91:1483–1492.
- Willemsen MH, Nijhof B, Fenckova M, Nillesen WM, Bongers EM, Castells-Nobau A, Asztalos L, Viragh E, Van BB, Tezel E, Veltman JA, Brunner HG, et al. 2013. GATAD2B loss-of-function mutations cause a recognisable syndrome with intellectual disability and are associated with learning deficits and synaptic undergrowth in Drosophila. *J Med Genet* 50:507–514.
- Worthey EA, Raca G, Laffin JJ, Wilk BM, Harris JM, Jakielski KJ, Dimmock DP, Strand EA, Shriberg LD. 2013. Whole-exome sequencing supports genetic heterogeneity in childhood apraxia of speech. *J Neurodev Disord* 5:29.

- Xu J, Bauer DE, Kerenyi MA, Vo TD, Hou S, Hsu Y-J, Yao H, Trowbridge JJ, Mandel G, Orkin SH. 2013. Corepressor-dependent silencing of fetal hemoglobin expression by BCL11A. *Proc Natl Acad Sci USA* 110:6518–6523.
- Yamada T, Yang Y, Hemberg M, Yoshida T, Cho HY, Murphy JP, Fioravante D, Regehr WG, Gygi SP, Georgopoulos K, Bonni A. 2014. Promoter decommissioning by the NuRD chromatin remodeling complex triggers synaptic connectivity in the mammalian brain. *Neuron* 83:122–134.
- Yamazaki T, Yoshimatsu Y, Morishita Y, Miyazono K, Watabe T. 2009. COUP-TFII regulates the functions of Prox1 in lymphatic endothelial cells through direct interaction. *Genes Cells* 14:425–434.
- Zeesman S, Nowaczyk MJM, Teshima I, Roberts W, Cardy JO, Brian J, Senman L, Feuk L, Osborne LR, Scherer SW. 2006. Speech and language impairment and oromotor dyspraxia due to deletion of 7q31 that involves FOXP2. *Am J Med Genet A* 140:509–514.
- Zentner GE, Layman WS, Martin DM, Scacheri PC. 2010. Molecular and phenotypic aspects of CHD7 mutation in CHARGE syndrome. *Am J Med Genet A* 152A:674–686.
- Zhang X-M, Chang Q, Zeng L, Gu J, Brown S, Basch RS. 2006. TBLR1 regulates the expression of nuclear hormone receptor co-repressors. *BMC Cell Biol* 7:31.
- Zhou C, Tsai SY, Tsai MJ. 2001. COUP-TFI: an intrinsic factor for early regionalisation of the neocortex. *Genes Dev* 15:2054–2059.
- Zilina O, Reimand T, Zjablovskaja P, Männik K, Männamaa M, Traat A, Puusepp-Benazzouz H, Kurg A, Ounap K. 2012. Maternally and paternally inherited deletion of 7q31 involving the FOXP2 gene in two families. *Am J Med Genet A* 158A:254–256.
- Zimmerman E, Maron JL. 2016. FOXP2 gene deletion and infant feeding difficulties: a case report. *Cold Spring Harb Mol Case Stud* 2:a000547.
- Zollino M, Orteschi D, Murdolo M, Lattante S, Battaglia D, Stefanini C, Mercuri E, Chiurazzi P, Neri G, Marangi G. 2012. Mutations in KANSL1 cause the 17q21.31 microdeletion syndrome phenotype. *Nat Genet* 44:636–638.

SAMENVATTING

Taal is een zeer complexe vaardigheid die mensen toch gemakkelijk en zonder formele instructie onder de knie krijgen. Kinderen leren een taal door gewoon te observeren en door interactie met de taalgebruikers om ze heen. Daarbij maken ze gebruik van hulpmiddelen in het menselijk brein die vastgelegd zijn in onze genen. Genen bevatten de instructies om eiwitten te maken. Deze eiwitten zijn de bouwstenen van elke cel in ons lichaam en de bewegende stukjes die onze ontwikkeling aansturen. Hiertoe behoren ook de hersencircuits die worden gebruikt voor taal. Specifieke genen worden in verschillende cellen aan en uit gezet op specifieke momenten om zo de hoeveelheid eiwit die wordt geproduceerd te veranderen en om heel precies te controleren hoe deze circuits zich ontwikkelen. Maar hoe kunnen we er achter komen welke genen betrokken zijn bij de ontwikkeling van taal in de hersenen?

Een van de mogelijkheden is om te kijken wat er gebeurt als er iets mis gaat. Soms ontstaat er een toevallige mutatie in een gen die leidt tot teveel of te weinig productie van een specifiek eiwit, of tot abnormale eigenschappen van het eiwit. In de hersenen kan dit complexe ontwikkelingsprocessen verstoren, waardoor de taalcircuits niet werken zoals ze zouden moeten werken. Dit kan leiden tot een taalstoornis bij mensen die deze mutaties hebben. Onderzoekers kunnen het DNA van mensen met taalstoornissen onderzoeken om zo uit te vinden welke genen gemuteerd zijn. Daaruit kunnen we afleiden dat deze genen belangrijk zijn voor de ontwikkeling van taal in de hersenen.

Deze strategie heeft geleid tot de ontdekking van *FOXP2*, het eerste gen dat een duidelijke relatie met taal heeft. Onderzoekers vonden een grote familie met meerdere generaties waarin ongeveer de helft van de familieleden een ernstige spraak- en taalstoornis had. Ze kwamen er achter dat de aangedane familieleden een mutatie in *FOXP2* hadden die niet aanwezig was in de gezonde familieleden. Andere *FOXP2* mutaties werden later gevonden in andere, niet-verwante personen met een vergelijkbare aandoening; dit versterkte de eerder gevonden link tussen *FOXP2* en taalstoornissen. Het *FOXP2* gen staat aan in specifieke delen van de hersenen en vormt de code voor het FOXP2 eiwit (volgens internationale afspraken worden genen schuingedrukt, terwijl de bijbehorende eiwitten worden geschreven als normale tekst). FOXP2 is onderdeel van een groep eiwitten die transcriptiefactoren worden genoemd. Transcriptiefactoren bepalen wanneer andere genen worden aan en uit gezet. Ze kunnen daarom een belangrijke rol spelen in het reguleren van de ontwikkelingsprocessen die nodig zijn voor het ontwikkelen van taal in de hersenen. *FOXP2* mutaties beschadigen het vermogen van het FOXP2 eiwit om genen aan en uit te zetten, en dit verstoort de hersenontwikkeling en leidt tot een spraak- en taalstoornis. Maar *FOXP2* mutaties spelen slechts een rol bij een klein deel van de personen met een spraak- en taalstoornis. Omdat taal een zeer complex cognitief proces is, is het waarschijnlijk dat veel andere genen ook betrokken zijn. Helaas is *FOXP2* een vrij uitzonderlijk geval en zijn er weinig andere genen die met een vergelijkbaar niveau van betrouwbaarheid met taal geassocieerd zijn. Gelukkig zijn er andere vormen van bewijs, naast het directe verband

tussen een taalstoornis en een mutatie, die we kunnen gebruiken om genen te identificeren die mogelijk met taal te maken hebben.

In dit proefschrift heb ik de zoektocht uitgebreid door te kijken naar andere genen die een functionele relatie hebben met *FOXP2*. Genen werken zelden onafhankelijk van andere genen. Eiwitten – de producten van deze genen – werken meestal samen met andere eiwitten in verschillende combinaties om structuren te vormen of deel te nemen aan chemische reacties in de cel. Bij complexe processen, zoals de ontwikkeling van hersencircuits die nodig zijn voor taal, zijn meestal veel verschillende eiwitten betrokken die met elkaar samenwerken als onderdeel van een dynamisch netwerk. Eiwit-eiwit interacties kunnen worden opgespoord in het laboratorium, dus als we één lid van het netwerk kennen (bijvoorbeeld *FOXP2*) kunnen we op zoek gaan naar de andere leden, zelfs als deze nog niet geassocieerd zijn met taalstoornissen.

Ik heb in dit proefschrift ook breder gekeken, naar genen die betrokken zijn bij andere aspecten van de ontwikkeling dan taal. Genen en eiwitten spelen vaak een rol in meer dan één biologisch proces. Dit betekent dat een mutatie die een enkel eiwit beïnvloedt soms een aantal verschillende symptomen kan veroorzaken. In tegenstelling tot *FOXP2* mutaties, die uitsluitend invloed lijken te hebben op taal, veroorzaken mutaties die invloed hebben op andere taalgerelateerde eiwitten niet alleen taalstoornissen maar ook bredere cognitieve problemen (zoals een verstandelijke beperking of autismspectrumstoornis) of zelfs problemen in geheel andere delen van het lichaam. Toch kunnen ook deze mutaties een cruciale rol spelen in de ontwikkeling van taal in de hersenen. Door ons alleen te concentreren op zeer specifieke taalstoornissen kunnen we andere belangrijke elementen van het netwerk over het hoofd zien. We kunnen daarom baat hebben bij het onderzoeken van bredere aandoeningen waarbij taal een belangrijke rol speelt.

In dit proefschrift heb ik de taalgerelateerde eiwitten die we al kennen (zoals *FOXP2*) als een startpunt gebruikt om een groter netwerk van eiwitten te ontdekken die een rol kunnen spelen bij het vormen van de taalmechanismen van de hersenen. Ik ben begonnen met eiwitten waarvan bekend is dat ze samenwerken met *FOXP2* en werkte zo naar buiten in het netwerk, hierbij gefocust op eiwitten betrokken bij aandoeningen waarbij taalstoornissen tenminste één van de belangrijkste klinische kenmerken waren. Mijn doelen waren a) het vinden van nieuwe interacties tussen taalgerelateerde eiwitten en b) ontdekken wat er gebeurt als een mutatie een van deze eiwitten verstoort. Welke kenmerken hebben de aangedane individuen? Wat gebeurt er met de locatie en functie van dat eiwit in de cel? En hoe beïnvloedt dit de samenwerking van verschillende eiwitten?

Ik ben begonnen door te kijken naar een bekende samenwerkingspartner van *FOXP2*, een sterk gerelateerde en zeer vergelijkbare transcriptiefactor genaamd *FOXP1*. De *FOXP*-eiwitten komen vaak voor als paar – ofwel twee kopieën van hetzelfde eiwit (bijvoorbeeld *FOXP1* met *FOXP1*), ofwel twee verschillende eiwitten (bijvoorbeeld *FOXP1* met *FOXP2*) – en deze koppeling zou nodig kunnen zijn voor deze eiwitten om genen ‘aan’ of ‘uit’ te kunnen zetten. Er is zelfs bewijs dat *FOXP1* en *FOXP2* hun functie soms op dezelfde genen kunnen uitoefenen. Dit alles maakte *FOXP1* een goede kandidaat voor het reguleren van

taalgerelateerde processen in de hersenen. Mutaties in FOXP1 zijn ook gevonden bij personen met taalproblemen, hoewel zij ook andere symptomen hadden, zoals bijvoorbeeld een verstandelijke beperking, autistische kenmerken en ontwikkelingsproblemen in andere delen van het lichaam zoals de spieren, ogen en urinewegen.

In **Hoofdstuk 2** heb ik gewerkt met andere onderzoekers om drie nieuwe *FOXP1* mutaties te rapporteren. We beschreven de klinische kenmerken van de patiënten die een dergelijke mutatie hadden en kwamen er achter dat die sterk overeenkwamen met eerdere beschrijvingen van *FOXP1*-gerelateerde aandoeningen. Daarna heb ik de effecten van deze *FOXP1* mutaties en drie aanvullende mutaties die eerder waren beschreven onderzocht. Ik wilde ontdekken welk effect deze mutaties zouden kunnen hebben op de locatie van het FOXP1 eiwit in de cel en op het vermogen van het eiwit om als transcriptiefactor te functioneren. Ik heb eerst naar cellen gekeken onder de microscoop en zag dat het FOXP1 eiwit zich normaal in de celkern bevindt, waar ook het DNA zich bevindt. Maar de *FOXP1* mutaties verstoorden dit normale patroon. Sommige mutaties zorgden ervoor dat het FOXP1 eiwit zich verspreidde buiten de celkern, terwijl andere mutaties leidden tot abnormale klontering van het eiwit – beide patronen suggereerden dat het eiwit zich niet langer bond aan DNA zoals normaal. Een tweede set experimenten liet zien dat FOXP1 met deze mutaties niet langer genen ‘uit’ kon zetten. Als laatste keek ik naar hoe de mutaties de koppeling van FOXP1 met een andere kopie van FOXP1 of met FOXP2 beïnvloedden. Sommige mutaties verstoorden deze interacties, terwijl bij andere mutaties FOXP1 eiwitten nog steeds konden koppelen met FOXP1 en FOXP2, maar normale eiwitten in de abnormale klontering leken te worden getrokken. Dit betekent dat FOXP1 mutaties niet alleen de functie van FOXP1 kunnen beïnvloeden, maar ook de functie van FOXP2. Dit doorstroomeffect van een FOXP1 mutatie op de functie van FOXP2 zou mogelijk een verklaring kunnen zijn voor sommige overeenkomsten die worden gezien bij taalstoornissen ten gevolge van FOXP1 en FOXP2 mutaties.

In **Hoofdstuk 3** heb ik opnieuw samengewerkt met onderzoekers om drie aanvullende patiënten met mutaties in FOXP1 te beschrijven. Deze patiënten hadden alle drie dezelfde mutatie. Deze mutatie was het directe equivalent van een zeer bekende mutatie in *FOXP2* – de mutatie die ontdekt was in die eerste grote familie met taalproblemen. Ik heb deze bij elkaar passende mutaties van FOXP1 en FOXP2 vergeleken en kwam er achter dat de twee eiwitten op een zeer vergelijkbare manier waren veranderd. Beide mutaties leidden tot abnormale klontering van de aangedane eiwitten en verstoorden het vermogen om andere genen uit te zetten. Deze mutaties verstoorden niet de interactie met normale versies van FOXP1 of FOXP2, maar konden deze normale eiwitten wel meetrekken in de klontering, en daarmee kunnen deze mutaties mogelijk een negatief effect hebben op de functie van de normale versies van FOXP1 en FOXP2. De *FOXP1* en *FOXP2* mutaties waren geassocieerd met enkele vergelijkbare klinische kenmerken, met name spraakproblemen, die zouden kunnen worden beïnvloed door de manier waarop de twee eiwitten samenwerken. Toch waren de twee aandoeningen opvallend verschillend op andere vlakken: de *FOXP2* mutatie had een zeer specifiek effect op spraak, terwijl aangedane personen een normale intelligentie

hebben en weinig problemen in andere delen van het lichaam; de *FOXP1* mutatie veroorzaakte daarentegen bredere cognitieve problemen en andere ontwikkelingsproblemen. Ik concludeerde dat hoewel de twee mutaties zeer vergelijkbare effecten hebben op eiwitniveau, andere verschillen tussen de FOXP1 en FOXP2 eiwitten (bijvoorbeeld waar in de hersenen en op welk moment van de ontwikkeling ze voorkomen) groot genoeg zijn om behoorlijk verschillende aandoeningen te veroorzaken.

In **Hoofdstuk 4** begon ik met het uitbreiden van het netwerk van taalgerelateerde eiwitten door te kijken naar nieuwe eiwitten die mogelijk samenwerken met de FOXP transcriptiefactoren. Eerdere studies hadden drie eiwitten (GATAD2B, KANSL1 en CHD3) geïdentificeerd die worden geclassificeerd als ‘chromatine remodelerende factoren’. Deze eiwitten werken vaak samen met transcriptiefactoren en zijn verantwoordelijk voor het aanpassen van de DNA-structuur als onderdeel van het proces van het aan en uit zetten van genen. Alle drie de eiwitten hadden mutaties die werden gevonden in personen met een verstandelijke beperking en duidelijke taalproblemen, waardoor dit goede kandidaten waren voor verder onderzoek. Eerder werden al interacties beschreven voor de ‘muizenversie’ van GATAD2B en FOXP1/2, maar niet voor de humane versies van dit eiwit. Ik bevestigde dat de humane versie van GATAD2B en FOXP1/2 ook konden interacteren en ontdekte welke delen van elk eiwit betrokken waren bij deze interactie. Ik kwam er achter dat mutaties in GATAD2B of FOXP1/2 deze interacties konden verstoren, wat enkele van de gelijkenissen tussen deze aandoeningen zou kunnen verklaren. Ik kon de interactie tussen FOXP1/2 en CHD3 of tussen FOXP1/2 en KANSL1 niet bevestigen, maar toch wijzen de connecties van CHD3 en KANSL1 met taalstoornissen er op dat deze eiwitten een belangrijke rol kunnen spelen in de ontwikkeling van taal in de hersenen.

In **Hoofdstuk 5** heb ik de zoektocht verder uitgebreid door te kijken naar TBR1, een andere transcriptiefactor die ook samenwerkt met FOXP1 en FOXP2. Mutaties in TBR1 waren al gevonden in patiënten met een autismespectrumstoornis en verstandelijke beperking; veel van hen hadden opvallende taalachterstanden. Dit maakte TBR1 een veelbelovende kandidaat voor betrokkenheid bij taalgerelateerde processen. Ik besloot om uit te vinden welke andere eiwitten zouden kunnen samenwerken met TBR1 en te kijken of deze ook geassocieerd waren met ontwikkelingsstoornissen. Mijn collega’s en ik hebben een methode gebruikt die ‘affiniteitspurificatie’ wordt genoemd om zo TBR1 uit cellen te extraheren, samen met andere eiwitten die hier aan vastzaten, en deze andere eiwitten te identificeren door massa-spectrometrie. Na enkele kwaliteitscontrolestappen kwam ik uit op een definitieve lijst van bijna 250 mogelijke TBR1-interactoren. Door gebruik te maken van databases en eerder gepubliceerde wetenschappelijke literatuur kon ik bepaalde patronen ontdekken tussen deze eiwitten. De meeste eiwitten bevonden zich in grote hoeveelheden in de hersenen en waren met name gelokaliseerd in de celkern, dichtbij het DNA. Veel eiwitten waren ‘chromatine remodelerende factoren’, die samen zouden kunnen werken met TBR1 om genen ‘aan’ en ‘uit’ te zetten. In een aanzienlijk deel van deze eiwitten werden ook mutaties gevonden die een verstandelijke beperking of autismespectrumstoornis kunnen veroorzaken, net als TBR1. Deze resultaten gaven de eerste uitgebreide kijk op eiwitten die

samenwerken met TBR1 en laten zo zien hoe TBR1 zou kunnen werken. Daarnaast leverde het nieuwe kandidaten op die mogelijk een rol spelen bij taalgerelateerde stoornissen en andere ontwikkelingsstoornissen.

In **Hoofdstuk 6** heb ik enkele van de meest veelbelovende kandidaten uit Hoofdstuk 5 geselecteerd om te proberen de interacties in levende cellen te valideren. Ik heb me met name gericht op eiwitten met bekende associaties met verstandelijke beperking en autismespectrumstoornissen. Veel van deze eiwitten waren ook geassocieerd met duidelijke taalstoornissen. Ik heb van vijf nieuwe eiwitten bevestigd dat ze met TBR1 interacteren (GATAD2B, BCOR, ADNP, NR2F1 en NR2F2). Van sommige van deze eiwitten was al bekend dat zij interacteren met de FOXP transcriptiefactoren (bijvoorbeeld van GATAD2B, zoals te zien in Hoofdstuk 4), wat ze nog verder verbindt met het netwerk van taalgerelateerde eiwitten. Mutaties in *TBR1* verstoorden verschillende van deze interacties – waarbij mutaties in twee verschillende regio's van het TBR1 eiwit meer of juist minder beschadigend kunnen zijn voor verschillende interacties. Ik liet zien dat mutaties in GATAD2B en ADNP ook de interactie met TBR1 konden verstoren. Het verlies van deze interacties zou een bijdragende factor kunnen zijn in abnormale hersenontwikkeling, cognitieve stoornissen en taalproblemen.

Samengevat heb ik verschillende nieuwe interacties geïdentificeerd waarbij de FOXP en TBR1 transcriptiefactoren betrokken zijn en op die manier bijgedragen aan het inzicht in een netwerk van eiwitten die de ontwikkeling van taalmechanismen in de hersenen kunnen beïnvloeden. Mutaties die deze eiwitten beïnvloeden veroorzaakten verschillende aandoeningen; sommige ernstiger dan andere, soms met behoorlijk verschillende en zelfs contrasterende symptomen, maar vaak met een vergelijkbare uitwerking op taalstoornissen. Ik kwam er achter dat verschillende mutaties de locatie en functie van de aangedane eiwitten in de cel negatief beïnvloedden. Ik ontdekte ook dat verschillende mutaties interacties tussen de eiwitten verstoorden en dat een mutatie die één eiwit verstoort ook een negatieve impact kan hebben op het vermogen van andere eiwitten in het netwerk om hun normale functie uit te oefenen. Deze resultaten dragen bij aan onze kennis over hoe genetische factoren taal beïnvloeden en onderstrepen de complexe rol van eiwit-eiwit interacties in ontwikkeling en aandoeningen van de mens.

SUMMARY

Language is a highly complex ability that humans nonetheless acquire readily and without formal teaching. Children learn a language simply by observing and interacting with the language users around them, using tools that are built into the human brain and encoded in our genes. Genes contain the instructions to produce proteins, which are the building blocks of every cell in our bodies and the moving pieces that drive our development, including the formation of brain circuits used for language. Certain genes are switched on or off in different cells at specific times to change the amount of each protein produced and to precisely control how these circuits develop. But how can we find out which genes are involved in the development of language in the brain?

One way is to look at what happens when something goes wrong. Sometimes a random mutation appears in a gene that leads to the production of too much or not enough of a particular protein or that gives the protein abnormal properties. In the brain, this can throw the complex developmental processes out of balance, so that the circuits used for language don't work the way they should. This leads to language impairment in people who carry these mutations. Researchers can study the DNA of people with language impairment in order to find which genes are mutated – we can then infer that these genes must be important for the development of language in the brain.

This approach led to the discovery of *FOXP2*, the first gene with a firm link to language. Researchers identified a large multigenerational family in which about half of the members had a severe speech and language disorder. They found that the affected family members had a mutation in *FOXP2* that was not present in their unaffected relatives. Other *FOXP2* mutations were later found in unrelated people with a similar disorder, which solidified the link between *FOXP2* and language disorders. The *FOXP2* gene is switched on in specific parts of the brain and encodes the FOXP2 protein (by convention, genes are italicised, while their protein products are written in normal text). FOXP2 is a member of a class of proteins called transcription factors, which control when other genes are switched on or off. Therefore, it potentially plays a very important role in regulating the developmental processes required for language in the brain. *FOXP2* mutations damage the protein's ability to switch genes on or off, disrupting brain development and leading to language impairment. However, *FOXP2* mutations only account for a small proportion of language impairment cases. Given the complexity of language as a cognitive process, it's likely that many other genes are also involved. Unfortunately, *FOXP2* has been a rather exceptional case and few other genes have been linked to language with the same degree of confidence. Thankfully, there are other lines of evidence, besides the direct link between language impairment and a mutation, that we can take into account to identify probable language-related genes.

In this thesis, I expanded the search by looking at other genes that share a functional relationship with *FOXP2*. Genes rarely work independently. Proteins – the functional products of genes – very often interact (that is, join together) with other proteins in various combinations to form structures or to participate in chemical reactions in the cell. Complex

processes, like the development of brain circuits required for language, are likely to involve many different proteins interacting with each other as part of a dynamic network. Protein-protein interactions can be detected in the laboratory, so if we know one member of the network (for example, FOXP2), we can start to search for others, even if they haven't yet been linked to a particular language disorder.

I also looked more widely at genes involved in other aspects of development in addition to language. Genes and proteins often play a part in more than one biological process. This means that a mutation affecting a single protein can sometimes cause a number of different symptoms. Unlike FOXP2 mutations, which seem to exclusively affect language, mutations affecting other language-related proteins might cause not only language impairment, but also broader cognitive difficulties (such as intellectual disability or autism spectrum disorders) or even problems in completely different parts of the body. But they can still be crucial elements in the development of language in the brain. By concentrating only on very specific language disorders, we might be missing some important elements of the network. We could therefore benefit from looking at broader disorders that have a strong language component.

In this thesis, I used the language-related proteins that we already know (such as FOXP2) as a starting point to explore a larger network of proteins that may be involved in forming the brain's language machinery. I started with known interaction partners of FOXP2 and worked outwards through the network, focussing on proteins connected to disorders where language impairment was at least one of the major clinical features. My aims were to a) identify new interactions between language-related proteins and b) to find out what happens when a mutation disrupts one of these proteins. What characteristics do the affected individuals have? What happens to the location and function of that protein in the cell? And how does it affect the interactions between the different proteins?

I began by looking at a known interaction partner of FOXP2, a closely related and very similar transcription factor called FOXP1. The FOXP proteins often occur in pairs – either two copies of the same protein (for example, FOXP1 with FOXP1), or two different proteins (for example, FOXP1 with FOXP2) – and this pairing may actually be necessary for them to be able to switch genes on or off. There is even evidence that FOXP1 and FOXP2 act on some of the same genes. All of this made FOXP1 a good candidate for regulating language-related processes in the brain. Indeed, mutations in FOXP1 had also been identified in people with language difficulties, although their symptoms were broader and included intellectual disability, autistic features and developmental problems in other parts of the body such as the muscles, eyes and urinary tract.

In **Chapter 2**, I worked with collaborators to report three new *FOXP1* mutations. We described the clinical features of the patients who carried each mutation and found that these closely matched previous reports of *FOXP1*-related disorder. I then investigated the effects of these *FOXP1* mutations, as well as three others that had been reported before. I wanted to find out what effect these might have on the location of the FOXP1 protein in the cell and on its ability to act as a transcription factor. First, I looked at cells under a microscope and saw that the FOXP1 protein was typically found throughout the nucleus of the cell, where

the DNA is located. But the *FOXP1* mutations disrupted this normal pattern. Some mutations made the FOXP1 protein spread outside of the nucleus, while others made the protein collect in abnormal clumps – both of which suggested it was no longer binding to the DNA normally. A second set of experiments showed that with these mutations, FOXP1 could no longer switch genes off. Finally, I looked at how the mutations affected the pairing of FOXP1 – either with another copy of FOXP1, or with FOXP2. Some mutations stopped the interactions, while others could still interact with normal FOXP1 or FOXP2 but appeared to pull the normal proteins into the abnormal clumps. This meant that FOXP1 mutations might not only affect the function of FOXP1, but also of FOXP2. This flow-on effect of a *FOXP1* mutation on the function of FOXP2 could potentially explain some of the similarities in language impairment between *FOXP1* and *FOXP2* mutations.

In **Chapter 3**, I worked with collaborators again to describe three more patients with mutations in *FOXP1*. But this time, all three patients had an identical mutation. This mutation was directly equivalent to a well-known mutation in *FOXP2* – the one identified in that first large family with language difficulties. I directly compared these matching mutations affecting FOXP1 and FOXP2 and found that the two proteins were altered in very similar ways. Both mutations led to abnormal clumping of the affected proteins and disrupted their abilities to switch off other genes. They did not stop the mutated proteins from interacting with normal FOXP1 or FOXP2, but drew them into the abnormal clumps, perhaps negatively influencing the function of the normal proteins as well. The *FOXP1* and *FOXP2* mutations were associated with some similar clinical features, especially speech difficulties, which might be influenced by the way the two proteins interact. However, the two disorders were markedly different in other ways: the *FOXP2* mutation had a very specific effect on speech, with affected people having normal intelligence and few problems in other parts of the body, while the *FOXP1* mutation caused broader cognitive and other developmental delays. I concluded that while the two mutations had very similar effects at the protein level, other differences between the FOXP1 and FOXP2 proteins (for example, where in the brain and at which developmental stages they occur) are great enough to cause quite distinct disorders.

In **Chapter 4**, I started to expand the network of language-related proteins by looking for new proteins that might interact with the FOXP transcription factors. Previous studies had identified three proteins (GATAD2B, KANSL1 and CHD3) that are classified as “chromatin remodelling factors”. These proteins often interact with transcription factors and are responsible for adjusting the structure of the DNA as part of the process of turning genes on or off. All three had mutations that were found in people with intellectual disability and prominent language deficits, making them good candidates for further investigation. Interactions had previously been reported for the mouse versions of GATAD2B and FOXP1/2, but not for the human versions. I confirmed that the human GATAD2B and FOXP1/2 proteins also interacted and found which parts of each protein were involved in the interaction. I found that mutations affecting GATAD2B or FOXP1/2 could prevent them from interacting, which might explain some of the similarities between these disorders. I

wasn't able to confirm the interactions between FOXP1/2 and CHD3 or between FOXP1/2 and KANSL1. However, the connections of CHD3 and KANSL1 to language impairment mean that they may still play important roles in the development of language in the brain.

In **Chapter 5**, I expanded the search further by looking at TBR1 - another transcription factor, which also interacts with FOXP1 and FOXP2. Mutations in TBR1 had been discovered in patients with autism spectrum disorder and intellectual disability, many of whom had noticeable language delays. This made TBR1 a promising candidate for involvement in language-related processes. I decided to find out which other proteins might interact with TBR1 and see if they were also linked to developmental disorders. My collaborators and I used a method called affinity purification to extract TBR1 from cells, along with any other proteins that were attached to it, and identified them using mass spectrometry. After some quality control steps, I arrived at a final list of nearly 250 potential TBR1-interactors. Drawing on databases and previous scientific literature, I identified certain patterns amongst these proteins. Most could be found at high levels in the brain and were predominantly located in the nucleus of the cell, close to the DNA. Many were chromatin remodelling factors, which might work together with TBR1 to switch genes on and off. A significant proportion were also affected by mutations that cause intellectual disability or autism spectrum disorders, just like TBR1. These results provided the first comprehensive look at the proteins that interact with TBR1, shedding light on how TBR1 might work, as well as pinpointing possible new candidates for language-related and other developmental disorders.

In **Chapter 6**, I selected some of the most promising candidates from Chapter 5 to try to validate the interactions in living cells. I particularly focussed on those proteins with known links to intellectual disability and autism spectrum disorders, many of which also included significant language impairment. I confirmed five new TBR1-interacting proteins (GATAD2B, BCOR, ADNP, NR2F1 and NR2F2). Some of these proteins were already known to interact with the FOXP transcription factors (for example, GATAD2B, as shown in Chapter 4), which connects them even more with the network of language-related proteins. Mutations in *TBR1* disrupted several of the interactions – with mutations affecting two distinct regions of the TBR1 protein being more or less damaging for different interactions. I showed that mutations affecting GATAD2B and ADNP could also disrupt interaction with TBR1. The loss of these interactions might be a contributing factor in abnormal brain development, cognitive disorders and language impairment.

In summary, I identified several new interactions involving the FOXP and TBR1 transcription factors, adding to our understanding of a network of proteins that may influence the development of language mechanisms in the brain. Mutations affecting these proteins caused distinct disorders of differing severity, sometimes with quite varied and contrasting symptoms, but often with similar effects on language impairment. I found that multiple mutations adversely influenced the location and function of the affected proteins in the cell. I also found that several mutations disrupted interactions between the proteins, and that a mutation affecting one protein could also negatively impact the ability of other proteins in

the network to perform their usual functions. These results add to our knowledge of how genetics influences language and highlight the complex role of protein-protein interactions in human development and disease.

CURRICULUM VITAE

Elliot Sollis was born in Perth, Western Australia. He obtained a Bachelor of Science, majoring in Genetics and Anatomy & Human Biology, from the University of Western Australia in 2009. He then completed a Bachelor of Science (Honours) degree in 2010, also at the University of Western Australia, with a thesis on the role of actin mutations in genetic muscle disorders. He then worked as a research assistant studying the molecular biology of mesothelioma, before moving overseas for his PhD.

At the Max Planck Institute for Psycholinguistics in Nijmegen, Elliot found the perfect opportunity to combine his genetics background with a lifelong interest in language and linguistics. He began his PhD research in the Language and Genetics Department in 2013, characterising genetic variants identified in people with language impairment, leading to the work presented in this thesis.

Elliot continues to work in genetics, now as a scientific curator for the Genome-wide Association Studies Catalog, at the European Bioinformatics Institute in Hinxton, U.K.

PUBLICATIONS

Ravenscroft G, **Sollis E**, Charles AK, North KN, Baynam G, Laing NG. 2011. Fetal akinesia: review of the genetics of the neuromuscular causes. *J Med Genet* 48:793-801.

Sollis E, Graham SA, Vino A, Froehlich H, Vreeburg M, Dimitropoulou D, Gilissen C, Pfundt R, Rappold GA, Brunner HG, Deriziotis P, Fisher SE. 2016. Identification and functional characterisation of *de novo* *FOXP1* variants provides novel insights into the etiology of neurodevelopmental disorder. *Hum Mol Genet* 25:546–557.

Sollis E, Deriziotis P, Saitsu H, Miyake N, Matsumoto N, Hoffer MJV, Ruivenkamp CAL, Alders M, Okamoto N, Bijlsma EK, Plomp AS, Fisher SE. 2017. Equivalent missense variant in the *FOXP2* and *FOXP1* transcription factors causes distinct neurodevelopmental disorders. *Hum Mutat* 38:1542–1554.

Estruch SB, Graham SA, Quevedo M, Vino A, Dekkers DHW, Deriziotis P, **Sollis E**, Demmers J, Poot RA, Fisher SE. 2018. Proteomic analysis of FOXP proteins reveals interactions between cortical transcription factors associated with neurodevelopmental disorders. *Hum Mol Genet* 27:1212-1227.

Ravenscroft G, Pannell S, O'Grady G, Ong R, Ee HC, Faiz F, Marns L, Goel H, Kumarasinghe P, **Sollis E**, Sivadorai P, Wilson M, Magoffin A, Nightingale S, Freckmann ML, Kirk EP, Sachdev R, Lemberg DA, Delatycki MB, Kamm MA, Basnayake C, Lamont PJ, Amor DJ, Jones K, Schilperoort J, Davis MR, Laing NG. 2018. Variants in *ACTG2* underlie a substantial number of Australasian patients with primary chronic intestinal pseudo-obstruction. *Neurogastroenterol Motil* 30:e13371.

Hoed J den, **Sollis E**, Venselaar H, Estruch SB, Deriziotis P, Fisher SE. 2018. Functional characterisation of *TBR1* variants in neurodevelopmental disorder. *Sci Rep* 8:14279–14279.

Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, **Sollis E**, Suveges D, Vrousseau O, Whetzel PL, Amode R, Guillen JA, Riat HS, Trevanion SJ, Hall P, Junkins H, Flicek P, Burdett T, Hindorff LA, Cunningham F, Parkinson H. 2019. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 47:D1005-1012.

ACKNOWLEDGEMENTS

At the risk of stating the obvious, completing a PhD is really hard... Fortunately, I've had a great group of colleagues, friends and family to guide, encourage and support me. You've all helped me to get to this point and made it an enjoyable and rewarding experience.

Firstly, I would like to thank my supervisors, Simon Fisher and Pelagia Derizioti. Simon, thanks for building a great department around such a unique and fascinating research area and for giving me the opportunity to work on a really interesting project. Your input was invaluable and there were many times when your astute suggestions were just what I needed to get to the next stage. Pela, I couldn't have hoped for a better guide over the last few years. Thanks for sharing your scientific expertise, for celebrating the successes and for supporting me through the difficult moments with the perfect mix of humour and sympathetic ire. *Και... πόσα χρυσά αστέρια κερδίζω που τελείωσα το διδακτορικό;*

Many thanks also to my collaborators. For the *FOXP1* project, I am indebted to the clinical collaborators who identified the variants and provided clinical reports, including Nobuhiko Okamoto, whose visit to my poster presentation in Baltimore provided the initial impetus for Chapter 3. Sincere thanks also to the patients and their families, without whom this research would not be possible. For the *TBR1* project, thank you to Raymond Poot and Martí Quevedo for welcoming me into their lab, helping me to conduct the affinity purifications and providing valuable comments on my analyses. While unfortunately not all of my projects came to fruition, I'm very grateful to those who helped: thanks to Sonja, Kai and Pedro for generous assistance with mouse experiments, as well as Rob Woestenenk for help at the flow cytometry facility.

A lot of this thesis builds on the everyday teamwork of my colleagues and friends in the Protein Group. Our weekly discussions helped immensely, and I know that behind the experiments in this thesis, there are flasks of cells and vials of plasmids with all of your initials on them! Sara, *la meva germana*, thanks for being a great friend and for always being happy to help out and share what you learnt on the PhD journey before me. Sarah, thanks for your biochemistry brilliance and for lots of help troubleshooting tricky experiments. Arianna, thanks for your cryosectioning expertise and lots of help in the lab. Lot, thanks for being my inside source on anything clinical, and for all your help as a paranymph (including great translation work on the samenvatting). Joery, thanks for your great work on the *TBR1* project – I really enjoyed working together on that – and thanks for helping out with some of my final experiments. Thank you to Fabian, Moritz, Swathi and Rocío for valuable discussions and for helping to make the group what it was. And thanks to all of you for lots of fun and ridiculousness in and out of the lab!

Thanks to everyone else in the lab – Pedro, Paolo, Kai, Martin, Midas, Jasper, Jurgen, Janine, Joses, Ine, Laura and anyone else I may have missed (sorry!) – for the lunchtime antics, eclectic playlists and general good times. Thanks to the rest of the Language and Genetics department for great discussions and for expanding my scientific horizons in department meetings – I now read GWAS papers for a living, so some of that polygenic trait

stuff must have stuck. Particular thanks to Amaia and Amanda for bioinformatics advice and helpful scripts that made the *TBRI* analyses much easier. And thanks to Martina for keeping everything in the department running smoothly and for helping to keep me organised during my time there and beyond.

Thanks to the MPI for providing a great working environment and for allowing me to ~~indulge secret interests in topics barely related to my thesis~~ develop valuable interdisciplinary connections across departments. Thanks also to Els, Dirkje and Kevin at the IMPRS for support throughout my PhD and for helping to keep me on track after I left the institute.

There are lots of other people in Nijmegen who each added something special to my time there. Thanks to Gwilym, for conquering the Eregilde at de Deut with me, and for all our pub conversations ranging from the serious to the beef-themed. Couldn't be happier to have you as my paranymp! To Lisa, for being my Dutch-class buddy and for giving me plenty of opportunities to dress up in ridiculous costumes. To Elliott, for podcast recommendations that got me through long sessions in the lab (causing only occasional laughter-induced chemical spills). To Amie, for lots of laughs, for country music, monkeys and that nail polish puzzle – which I never solved! And thanks to Suzanne, Ashley, Johanne, Will, Richard, Rick, Franziska, Evelien, Emma, Sean and all the other fellow Max Planckers, party-goers and -throwers and Cultuurcafé regulars who added so much to my time in Nijmegen. Thanks to the PoL gang for having me as an honorary member in the latter years, and particularly to Jeroen for proofreading my samenvatting, and Joe and Mischa for a generous metaphysical tenancy arrangement at a crucial time.

Music was a big part of my time in Nijmegen and one of the best things I did was to join an amazing band, Cherry and the Cakes. I think improvising raucous jazz in a basement every Tuesday (or in a beer festival in a church!) was the perfect antidote to the rigours of a PhD and I loved it! Thank you so much to Tim, Ilse, Martha, Christian, Giel, Sara, René, Tom, Dan, Greta, Chris, Paul and anyone else who played with us or came to one of our gigs. Thanks to Tulya for a foray into theatrical music. And thanks to Nietzsche, Maarten, Yves, Varun and anyone else who was up for a jam over the years.

Thanks also to Leny, Tommy, Dorien and family, who were not just landlords but *mijn Nijmeegse familie*. *Bedankt voor de steun toen ik ziek was, dat jullie er altijd waren voor een praatje (en een kans om mijn Nederlands te oefenen!), en dat jullie me altijd naar huis hebben gestuurd met een berg eten!*

Thanks to my friends back home and in the global Perth diaspora – Sam, Lee, Kalina, Luke, Tim, Stephen, Jasmine, Elaine, Kat, Thom, Nadia, Andrew... I love that no matter how long it's been, it's always the same when we see each other.

Thanks also to my former supervisors, Nigel, Gina and Cleo, and many others in Perth for their encouragement. And thanks to my new colleagues at the GWAS Catalog, for making the move out of the lab such a smooth transition and for the gentle encouragement of an occasional “so how's your thesis coming along?” to keep me going in the final year.

To my wonderful girlfriend Eirini, thank you for making the last few years so special. Thank you for helping and supporting me at every step of the way and for giving me much needed reminders that I'm almost there. You've been amazing! *Ευχαριστώ πολύ αγάπη μου!*

And finally, thank you so much to my family. Mum and Dad (Liz and James), you've always been there for me and I couldn't have done this without you. Thanks for the weekend Skype sessions, for all the long trips north and for your constant love and encouragement. Adam, thanks for being a great brother and for being interested in what I'm up to. Although you're also a very good actor, so who knows... nah but seriously, thanks! Thanks to all the Sollises for their support and for catching up around Europe and at home. To my Auntie Linda, I'm proud to be joining you as the second doctor in the family – thanks for leading the way! And thanks to my generous and inspiring grandparents – Granny, Grandda, Nana and Opa – sadly not all of you could be here to see it, but this is for you.

MPI SERIES IN PSYCHOLINGUISTICS

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing. *Miranda van Turenhout*
2. The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography. *Niels O. Schiller*
3. Lexical access in the production of ellipsis and pronouns. *Bernadette M. Schmitt*
4. The open-/closed-class distinction in spoken-word recognition. *Alette Haveman*
5. The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach. *Kay Behnke*
6. Gesture and speech production. *Jan-Peter de Ruiter*
7. Comparative intonational phonology: English and German. *Esther Grabe*
8. Finiteness in adult and child German. *Ingeborg Lasser*
9. Language input for word discovery. *Joost van de Weijer*
10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe. *James Essegbey*
11. Producing past and plural inflections. *Dirk Janssen*
12. Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea. *Anna Margetts*
13. From speech to words. *Arie van der Lugt*
14. Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language. *Eva Schultze-Berndt*
15. Interpreting indefinites: An experimental study of children's language comprehension. *Irene Krämer*
16. Language-specific listening: The case of phonetic sequences. *Andrea Weber*
17. Moving eyes and naming objects. *Femke van der Meulen*
18. Analogy in morphology: The selection of linking elements in Dutch compounds. *Andrea Krott*
19. Morphology in speech comprehension. *Kerstin Mauth*
20. Morphological families in the mental lexicon. *Nivja H. de Jong*
21. Fixed expressions and the production of idioms. *Simone A. Sprenger*
22. The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria). *Birgit Hellwig*
23. Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies. *Fermín Moscoso del Prado Martín*
24. Contextual influences on spoken-word processing: An electrophysiological approach. *Daniëlle van den Brink*
25. Perceptual relevance of prevoicing in Dutch. *Petra M. van Alphen*
26. Syllables in speech production: Effects of syllable preparation and syllable frequency. *Joana Cholin*

27. Producing complex spoken numerals for time and space. *Marjolein Meeuwissen*
28. Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction. *Rachèl J. J. K. Kemps*
29. At the same time...: The expression of simultaneity in learner varieties. *Barbara Schmiedtová*
30. A grammar of Jalonke argument structure. *Friederike Lüpke*
31. Agrammatic comprehension: An electrophysiological approach. *Marlies Wassenaar*
32. The structure and use of shape-based noun classes in Miraña (North West Amazon). *Frank Seifart*
33. Prosodically-conditioned detail in the recognition of spoken words. *Anne Pier Salverda*
34. Phonetic and lexical processing in a second language. *Mirjam Broersma*
35. Retrieving semantic and syntactic word properties. *Oliver Müller*
36. Lexically-guided perceptual learning in speech processing. *Frank Eisner*
37. Sensitivity to detailed acoustic information in word recognition. *Keren B. Shatzman*
38. The relationship between spoken word production and comprehension. *Rebecca Özdemir*
39. Disfluency: Interrupting speech and gesture. *Mandana Seyfeddinipur*
40. The acquisition of phonological structure: Distinguishing contrastive from non-contrastive variation. *Christiane Dietrich*
41. Cognitive cladistics and the relativity of spatial cognition. *Daniel B.M. Haun*
42. The acquisition of auditory categories. *Martijn Goudbeek*
43. Affix reduction in spoken Dutch. *Mark Pluymaekers*
44. Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence. *Valesca Kooijman*
45. Space and iconicity in German Sign Language (DGS). *Pamela Perniss*
46. On the production of morphologically complex words with special attention to effects of frequency. *Heidrun Bien*
47. Crosslinguistic influence in first and second languages: Convergence in speech and gesture. *Amanda Brown*
48. The acquisition of verb compounding in Mandarin Chinese. *Jidong Chen*
49. Phoneme inventories and patterns of speech sound perception. *Anita Wagner*
50. Lexical processing of morphologically complex words: An information-theoretical perspective. *Victor Kuperman*
51. A grammar of Savosavo, a Papuan language of the Solomon Islands. *Claudia Wegener*
52. Prosodic structure in speech production and perception. *Claudia Kuzla*
53. The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production and comprehension. *Sarah Schimke*
54. Studies on intonation and information structure in child and adult German. *Laura de Ruiter*
55. Processing the fine temporal structure of spoken words. *Eva Reinisch*
56. Semantics and (ir)regular inflection in morphological processing. *Wieke Tabak*
57. Processing strongly reduced forms in casual speech. *Susanne Brouwer*

58. Ambiguous pronoun resolution in L1 and L2 German and Dutch. *Miriam Ellert*
59. Lexical interactions in non-native speech comprehension: Evidence from electroencephalography, eye-tracking, and functional magnetic resonance imaging. *Ian FitzPatrick*
60. Processing casual speech in native and non-native language. *Annelie Tuinman*
61. Split intransitivity in Rotokas, a Papuan language of Bougainville. *Stuart Robinson*
62. Evidentiality and intersubjectivity in Yurakaré: An interactional account. *Sonja Gipper*
63. The influence of information structure on language comprehension: A neurocognitive perspective. *Lin Wang*
64. The meaning and use of ideophones in Siwu. *Mark Dingemans*
65. The role of acoustic detail and context in the comprehension of reduced pronunciation variants. *Marco van de Ven*
66. Speech reduction in spontaneous French and Spanish. *Francisco Torreira*
67. The relevance of early word recognition: Insights from the infant brain. *Caroline Junge*
68. Adjusting to different speakers: Extrinsic normalization in vowel perception. *Matthias J. Sjerps*
69. Structuring language. Contributions to the neurocognition of syntax. *Katrien R. Segaert*
70. Infants' appreciation of others' mental states in prelinguistic communication: A second person approach to mindreading. *Birgit Knudsen*
71. Gaze behavior in face-to-face interaction. *Federico Rossano*
72. Sign-spatiality in Kata Kolok: how a village sign language of Bali inscribes its signing space. *Conny de Vos*
73. Who is talking? Behavioural and neural evidence for norm-based coding in voice identity learning. *Attila Andics*
74. Lexical processing of foreign-accented speech: Rapid and flexible adaptation. *Marijt Witteman*
75. The use of deictic versus representational gestures in infancy. *Daniel Puccini*
76. Territories of knowledge in Japanese conversation. *Kaoru Hayano*
77. Family and neighbourhood relations in the mental lexicon: A cross-language perspective. *Kimberley Mulder*
78. Contributions of executive control to individual differences in word production. *Zeshu Shao*
79. Hearing speech and seeing speech: Perceptual adjustments in auditory-visual processing. *Patrick van der Zande*
80. High pitches and thick voices: The role of language in space-pitch associations. *Sarah Dolscheid*
81. Seeing what's next: Processing and anticipating language referring to objects. *Joost Rommers*
82. Mental representation and processing of reduced words in casual speech. *Iris Hanique*
83. The many ways listeners adapt to reductions in casual speech. *Katja Poellmann*
84. Contrasting opposite polarity in Germanic and Romance languages: Verum Focus and affirmative particles in native speakers and advanced L2 learners. *Giuseppina Turco*

85. Morphological processing in younger and older people: Evidence for flexible dual-route access. *Jana Reifegerste*
86. Semantic and syntactic constraints on the production of subject-verb agreement. *Alma Veenstra*
87. The acquisition of morphophonological alternations across languages. *Helen Buckler*
88. The evolutionary dynamics of motion event encoding. *Annemarie Verkerk*
89. Rediscovering a forgotten language. *Jiyoun Choi*
90. The road to native listening: Language-general perception, language-specific input. *Sho Tsuji*
91. Infants' understanding of communication as participants and observers. *Gudmundur Bjarki Thorgrímsson*
92. Information structure in Avatime. *Saskia van Putten*
93. Switch reference in Whitesands. *Jeremy Hammond*
94. Machine learning for gesture recognition from videos. *Binyam Gebrekidan Gebre*
95. Acquisition of spatial language by signing and speaking children: a comparison of Turkish sign language (TID) and Turkish. *Beyza Sümer*
96. An ear for pitch: on the effects of experience and aptitude in processing pitch in language and music. *Salomi Savvatia Asaridou*
97. Incrementality and Flexibility in Sentence Production. *Maartje van de Velde*
98. Social learning dynamics in chimpanzees: Reflections on (nonhuman) animal culture. *Edwin van Leeuwen*
99. The request system in Italian interaction. *Giovanni Rossi*
100. Timing turns in conversation: A temporal preparation account. *Lilla Magyari*
101. Assessing birth language memory in young adoptees. *Wencui Zhou*
102. A social and neurobiological approach to pointing in speech and gesture. *David Peeters*
103. Investigating the genetic basis of reading and language skills. *Alessandro Gialluisi*
104. Conversation Electrified: The Electrophysiology of Spoken Speech Act Recognition. *Rósa Signý Gísladóttir*
105. Modelling Multimodal Language Processing. *Alastair Smith*
106. Predicting language in different contexts: The nature and limits of mechanisms in anticipatory language processing. *Florian Hintz*
107. Situational variation in non-native communication. *Huib Kouwenhoven*
108. Sustained attention in language production. *Suzanne Jongman*
109. Acoustic reduction in spoken-word processing: Distributional, syntactic, morphosyntactic, and orthographic effects. *Malte Viebahn*
110. Nativeness, dominance, and the flexibility of listening to spoken language. *Laurence Bruggeman*
111. Semantic specificity of perception verbs in Maniq. *Ewelina Wnuk*
112. On the identification of FOXP2 gene enhancers and their role in brain development. *Martin Becker*

113. Events in language and thought: The case of serial verb constructions in Avatime. *Rebecca Defina*
114. Deciphering common and rare genetic effects on reading ability. *Amaia Carrión Castillo*
115. Music and language comprehension in the brain. *Richard Kunert*
116. Comprehending Comprehension: Insights from neuronal oscillations on the neuronal basis of language. *Nietzsche H.L. Lam*
117. The biology of variation in anatomical brain asymmetries. *Tulio Guadalupe*
118. Language processing in a conversation context. *Lotte Schoot*
119. Achieving mutual understanding in Argentine Sign Language. *Elizabeth Manrique*
120. Talking Sense: the behavioural and neural correlates of sound symbolism. *Gwilym Lockwood*
121. Getting under your skin: The role of perspective and simulation of experience in narrative comprehension. *Franziska Hartung*
122. Sensorimotor experience in speech perception. *Will Schuerman*
123. Explorations of beta-band neural oscillations during language comprehension: Sentence processing and beyond. *Ashley Lewis*
124. Influences on the magnitude of syntactic priming. *Evelien Heyselaar*
125. Lapse organization in interaction. *Elliott Hoey*
126. The processing of reduced word pronunciation variants by natives and foreign language learners: Evidence from French casual speech. *Sophie Brand*
127. The neighbors will tell you what to expect: Effects of aging and predictability on language processing. *Cornelia Moers*
128. The role of voice and word order in incremental sentence processing. *Sebastian Sauppe*
129. Learning from the (un)expected: Age and individual differences in statistical learning and perceptual learning in speech. *Thordis Neger*
130. Mental representations of Dutch regular morphologically complex neologisms. *Laura de Vaan*
131. Speech production, perception, and input of simultaneous bilingual preschoolers: Evidence from voice onset time. *Antje Stoehr*
132. A holistic approach to understanding pre-history. *Vishnupriya Kolipakam*
133. Characterization of transcription factors in monogenic disorders of speech and language. *Sara Busquets Estruch*
134. Indirect request comprehension in different contexts. *Johanne Tromp*
135. Envisioning Language - An Exploration of Perceptual Processes in Language Comprehension. *Markus Ostarek*
136. Listening for the WHAT and the HOW: Older adults' processing of semantic and affective information in speech. *Juliane Kirsch*
137. Let the agents do the talking: on the influence of vocal tract anatomy on speech during ontogeny and glossogeny. *Rick Janssen*
138. Age and hearing loss effects on speech processing. *Xaver Koch*
139. Vocabulary knowledge and learning: Individual differences in adult native speakers. *Nina Mainz*

140. The face in face-to-face communication: Signals of understanding and non-understanding. *Paul Hömke*
141. Person reference and interaction in Umpila/Kuuku Ya'u narrative. *Clair Hill*
142. Beyond the language given: The neurobiological infrastructure for pragmatic inferencing. *Jana Bašňáková*
143. From Kawapanan to Shawi: Topics in language variation and change. *Luis Miguel Rojas-Berscia*
144. On the oscillatory dynamics underlying speech-gesture integration in clear and adverse listening conditions. *Linda Drijvers*
145. Understanding temporal overlap between production and comprehension. *Amie Fairs*
146. The role of exemplars in speech comprehension. *Annika Nijveld*
147. A network of interacting proteins disrupted in language-related disorders. *Elliot Sollis*

**M A X
P L A
N C K**

MAX PLANCK INSTITUTE
FOR PSYCHOLINGUISTICS

VISITING ADDRESS

Wundtlaan 1
6525 XD Nijmegen
The Netherlands

POSTAL ADDRESS

P.O. Box 310
6500 AH Nijmegen
The Netherlands

CONTACT

T +31(0)24 3521 911
F +31(0)24 3521 213
E info@mpi.nl
Twitter [@MPI_NL](https://twitter.com/MPI_NL)
www.mpi.nl