CHARACTERIZATION OF TRANSCRIPTION FACTORS IN MONOGENIC DISORDERS OF SPEECH AND LANGUAGE

SARA BUSQUETS ESTRUCH

ISBN: 978-90-76203-92-8
Printed and bound by Ipskamp Drukkers

Characterization of transcription factors in monogenic disorders of speech and language

Proefschriftter
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 maandag 11 juni 2018
om 14.30 uur precies

door
Sara Busquets Estruch
geboren op 16 maart 1988
te Barcelona (Spanje)

Promotor
Prof. dr. Simon E. Fisher

Copromotor
Dr. Sarah A. Graham (Birmingham Women's and Children's NHS Foundation Trust, Verenigd Koninkrijk)

Manuscriptcommissie
Prof. dr. Han G. Brunner
Prof. dr. Gudrun Rappold (UniversitätHeidelberg, Duitsland)
Prof. dr. Paul Coffer (UMC Utrecht)

Characterization of transcription factors in monogenic disorders of speech and language

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 Monday, June 11, 2018
at 14.30 hours

by
Sara Busquets Estruch
Born on March 16, 1988
in Barcelona (Spain)

Supervisor
Prof. dr. Simon E. Fisher

Copromotor
Dr. Sarah A. Graham (Birmingham Women's and Children's NHS Foundation Trust, United Kingdom)

Manuscriptcommissie
Prof. dr. Han G. Brunner
Prof. dr. Gudrun Rappold (University of Heidelberg, Germany)
Prof. dr. Paul Coffer (UMC Utrecht)

Aprendre

Caminar.
Caminar més de pressa.
Buscar.
Palpar.
Trobar.
Fugir.
Perdre's.
Tornar a caminar,
però més a poc a poc.

by Pep Mita

To Learn

To walk.
To walk faster.
To search.
To palpate.
To find.
To run away.
To get lost.
To walk again,
but slower.

by Pep Mita

CONTENTS

1 INTRODUCTION

1.1 CHARACTERIZING THE GENETIC BASIS OF HUMAN LANGUAGE

1.1.1 *Human language and language disorders*

Language is arguably the trait that most clearly distinguishes humans from other species. While other species do communicate and some of them do so using vocalizations, their communication systems are far simpler than human language (Levinson & Holler, 2014). Human language goes beyond transmitting simple and practical information about essential aspects of our lives, such as food or imminent danger, it also enables us to share our thoughts, express our feelings, and talk about even the most abstract matters we can imagine. Language allows us to exchange an unlimited number of ideas using a finite set of mental tools, thus it is not surprising that the mental processes underlying this trait are complex (Graham *et al.* , 2015). Yet, language comes naturally to us, and it is remarkable that, given the adequate environment, any normal child becomes a proficient speaker in their native language without consciously making any effort or needing formal tuition (Kuhl, 2004). Language is not only a uniquely human trait, but it also appears to be universal to all human beings (Fisher & Vernes, 2015). No human society that lacks complex language has yet been discovered (Hammarström, 2016). Moreover, language has been crucial to human cultural evolution, as it has enabled humans to cooperate and to share knowledge that has been transmitted over many generations (Smith & Kirby, 2008).

Our linguistic environment clearly plays a role in developing and shaping our language capacity, but there is also an innate component to these capabilities, as evident from heritability studies in twins (Bishop *et al.* , 1995; Lewis & Thompson, 1992). The first clues about the molecular basis of language came from looking at the genetic make-up of individuals with speech and language impairments. Individuals with speech and/or language impairment do not develop adequate skills in this area even if they are raised in a language-rich en-

vironment and receive appropriate input. Often in individuals with a generalized neurodevelopmental disorder, language is impaired as a secondary consequence of other cognitive deficits, but in some other cases the only cognitive aspect affected in that person is language itself, and those cases are defined as a specific language disorder (Graham & Fisher, 2015). There are several different kinds of developmental disorders that affect speech, language and/or reading (Table 1.1), all of which usually manifest in early childhood and continue into adulthood, affecting many aspects of life, such as educational achievement or social and job opportunities. Because language deficits are often a core feature of neuropsychiatric disorders such as autism-spectrum disorder (ASD) or intellectual disability (ID) (Bishop, 2010; Pal, 2011), these neurodevelopmental disorders are also useful conditions to investigate in order to disentangle the genetic basis of speech and language.

1.1.2 *The discovery of FOXP2-related language disorder*

The fact that language-related disorders typically cluster in families provided early indirect evidence that genetic factors may contribute to language (Neils & Aram, 1986; Tomblin, 1989; Barry *et al.* , 2007; Lewis *et al.* , 2007). Moreover, twin studies showed that monozygotic twins presented higher rates of concordance for language traits and disorders than dizygotic twins (Lewis & Thompson, 1992; Bishop *et al.* , 1995; Bishop & Hayiou-Thomas, 2008), highlighting the genetic contribution to linguistic abilities. Towards the end of the 1990s, with the rise of molecular genetic techniques, specific regions in the genome were identified to be involved in the etiology of language impairment by linkage studies (Fisher *et al.* , 2003), and followed up by targeted association studies of the specific linked regions and mutation screens of candidate genes (Newbury & Monaco, 2002).

A major breakthrough occurred in 2001, when geneticists identified the transcription factor *FOXP2* as the first gene implicated in a language disorder, through investigation of a large multigenerational family (the KE family) with severe communication problems (Lai *et al.* , 2001). Half of the members in the KE family suffer from a rare form of speech and language impairment characterized by difficulties in coordinating the orofacial movements required for speech (childhood apraxia of speech, CAS; also known as developmental verbal dyspraxia DVD). In addition to core deficits in orofacial motor control and

Table 1.1: Disorders of speech, language and/or reading mentioned in this thesis

Disorder	Definition
Childhood apraxia of speech (CAS)	Deficits in planning and coordination of speech and sound sequences necessary for fluid speech. Also known as developmental verbal dyspraxia (DVD)
Dysarthria	Speech that may be abnormally slow, fast, weak or imprecise caused by affections to muscles and nerves that control speech
Dyslexia	A difficulty with reading and spelling that cannot be explained by other causes such as low IQ, physical impairment or lack of opportunity to learn
Specific language impairment (SLI)	Unexplained impairment in acquisition of spoken language, affecting one or more of morphology, syntax, semantics and pragmatics. Also known as developmental language disorder (DLD)
Expressive language impairment	Type of SLI. Impaired ability to formulate ideas and messages using words and
Receptive language impairment	Type of SLI. Impaired ability to understand messages encoded in words and sentences
Stuttering	Involuntary repetitions, prolongations of syllables, and pauses during speech

spoken language production, verbal fluency and language comprehension were also impaired in the affected individuals (Vargha-Khadem *et al.* , 2005). The pedigree of the KE family suggested a simple autosomal dominant inheritance;

and linkage analysis identified a genomic region on chromosome 7 (7q31) that segregated with the disorder in the family (Fisher *et al.* , 1998). It was an independent clinical case with a similar phenotype to the KE family and a chromosomal translocation within the linked region that led researchers to the *FOXP2* gene (Lai *et al.* , 2000). Sequencing of *FOXP2* in the KE family revealed that a missense mutation affecting this gene (p.R553H) was present in the affected members of the family but not in the unaffected members, providing robust evidence for the implication of a single gene in a speech and language disorder (Lai *et al.* , 2001). Follow-up studies on the function of this transcription factor demonstrated that the p.R553H mutation severely disrupted the ability of the FOXP2 protein to bind to DNA and regulate gene expression, as well as affecting its subcellular localization (Vernes *et al.* , 2006). Since then, disruptions affecting *FOXP2* have been found in multiple additional cases of CAS (for further details see section 1.2.3.) (MacDermot *et al.* , 2005; Turner *et al.* , 2013; Laffin *et al.* , 2012; Shriberg *et al.* , 2006; Feuk *et al.* , 2006; Zeesman *et al.* , 2006; Reuter *et al.* , 2016). This gene has thus become a fundamental starting point to expand our knowledge on the genetic and biological basis of language impairment, by studying its function and molecular networks in the context of brain development. Note that throughout this dissertation symbols for genes are italicized (e.g., *FOXP2*), whereas symbols for proteins are not italicized (e.g., FOXP2); human protein and gene symbols are written all in uppercase letters; whereas when referring to murine orthologues only the first letter is uppercase and the rest are all lowercase (e.g., Foxp2); and orthologues in other species are written with the first and the last letters in uppercase and the rest in lowercase (e.g., FoxP2).

1.1.3 *The search for further language-related genes*

The discovery of *FOXP2* mutations causing a rare speech and language disorder was an exceptional finding, as it is very rare to find a well-defined behavioural phenotype with such a straightforward inheritance pattern. Identifying additional language-related genes remains challenging despite the technological advances that have emerged during the past decade. Language involves a highly complex set of skills, with a complex underlying genetic architecture, which is likely to be grounded in small contributions exerted by myriads of different genes. That is probably the main reason why it has been difficult to pinpoint

and disentangle the contributing genetic factors that shape this trait. Nevertheless, some notable progress has been made towards determining more of the genetic risk factors underpinning language and language-related disorders.

Several linkage studies have been carried out aiming at identifying more chromosomal regions co-inherited by individuals with a language disorder within a particular family (Newbury & Monaco, 2010; Kang & Drayna, 2011; Reader *et al.* , 2014; Graham & Fisher, 2015). However, none of these has yet led to the discovery of another simple monogenic form of impairment, like CAS in the KE family (Graham & Fisher, 2015). On the contrary, most of these studies have found linkage of multiple genomic regions to the language impairment, suggesting a heterogeneous genetic basis for these disorders. Follow-up genetic association studies focused on the identified linkage intervals and provided a number of candidate risk genes for different language impairments, all of which represented common variants with small effect sizes. Examples of this are *CMIP1* and *ATP2C2*, single nucleotide polymorphisms (SNPs) in which were associated with specific language impairment (SLI) risk; or *KIAA0319*, *DCDC2* and *MRPL19\C2ORF3*, associated with dyslexia (Anthoni *et al.* , 2007; Francks *et al.* , 2004; Meng *et al.* , 2005; Newbury *et al.* , 2009). *DYX1C1* and *ROBO1* are other candidate susceptibility genes for dyslexia, which were also identified by screening of mapped linkage regions (Hannula-Jouppi *et al.* , 2005; Taipale *et al.* , 2003). Like dyslexia, stuttering is also a very well-defined condition and analyses of a linkage region (12q) identified by studying various consanguineous families led to the discovery of a coding variant in the gene *GNPTAB* that imperfectly co-segregated with stuttering in one family (Kang *et al.* , 2010). The putative risk variant altered a conserved residue of the protein and was identified in a number of other cases, at a higher frequency than controls. Moreover, follow-up investigations screening for variants of *GNPTAB* as well as for closely related genes of the same pathway (*GNPTG* and *NAGPA*) in case-control studies identified several different rare coding variants in these genes that may be associated with stuttering (Kang *et al.* , 2010).

In the last few years the first genome-wide association studies (GWAS) for language-related disorders and language-related abilities in the general population have been carried out (Graham & Fisher, 2015). So far, this methodology has yielded only two genome-wide significant effects: one associated with social abilities near the *SCN11A* gene, and one associated with infant expressive vocabulary near the *ROBO2* gene (St Pourcain *et al.* , 2014a,b). The latter is

particularly intriguing given the association of its close paralogue *ROBO1* with dyslexia and the well-established functions of ROBO proteins in brain development (St Pourcain *et al.* , 2014a). However, in most language-related GWAS studies, no variant has reached the threshold for genome-wide significance, probably because these efforts have been relatively underpowered. Larger-scale GWASs and meta-analyses are promising strategies to circumvent this issue and they may be able to provide genome-wide significant hits in the future (Reader *et al.* , 2014; Graham & Fisher, 2015).

The development of cheaper, more efficient and faster next generation sequencing (NGS) technologies has enabled large-scale whole-exome sequencing (WES) and whole-genome sequencing (WGS), and consequently the genome-wide study of mutations has in recent years been extended to single-base resolution (Stranneheim & Wedell, 2016). This has had a revolutionary impact on the discovery of genes underlying developmental disorders, for which hundreds of causal variants have been discovered by focusing on rare *de novo* variants in severe, sporadic cases (Stranneheim & Wedell, 2016). Given that subjects with neurodevelopmental disorders typically present severe phenotypes in which the reproductive fitness is greatly reduced, it was long hypothesized that *de novo* variants might be particularly prevalent in this type of condition. That is unlikely to be the case for language-related disorders, which do not affect reproductive fitness as much as neurodevelopmental disorders such as severe autism or ID (Graham & Fisher, 2015). Thus, language-related disorders are more likely to be caused by inherited variants rather than by *de novo* variants. Nevertheless, a *de novo* frameshift mutation disrupting the *FOXP2* gene has been reported in a child with CAS (Turner *et al.* , 2013) and a *de novo* copy-number variant (CNV) affecting *ATP2C2*, previously associated with SLI, was reported in a case of language delay and expressive language impairment (Smith *et al.* , 2015).

WES and WGS have not only been useful to uncover *de novo* variation, but have also facilitated the identification of rare genetic variants in specific populations and families. An interesting example is the case of the Robinson Crusoe island, which has a population isolate in which the incidence of SLI is notably high. Initial linkage analyses had indicated that multiple genomic regions were implicated in the etiology of SLI in this population, but it was the use of exome sequencing to study five members of the population that pinpointed the puta-

tive transcription factor *NFXL1* as a novel candidate gene for involvement in SLI risk (Villanueva *et al.* , 2015).

In a recent study, WES of 43 unrelated individuals affected by severe SLI identified several potentially pathogenic inherited and *de novo* variants in several different genes (Chen *et al.* , 2017). They found variants in genes already implicated language-related disorders, such as *ERC1*, *GRIN2A*, *ATP2C2*, *CNTNAP2* or *ROBO1* (see sections below) but also some others are novel candidate genes for language-related disorders, such as *OXR1*, *SCN9A* and *KMT2D* (Chen *et al.* , 2017).

1.2 MONOGENIC LANGUAGE-RELATED DISORDERS AS A WINDOW INTO THE MOLECULAR NETWORKS UNDERLYING SPEECH AND LANGUAGE

1.2.1 *Insights into complex phenotypes from monogenic disorders*

Neurodevelopmental conditions such as language disorders typically have a highly complex genetic etiology that makes the understanding of their biology challenging. Rare monogenic cases may offer the best opportunities to further our insights into the molecular mechanisms underlying these disorders. In monogenic conditions the causality of any particular gene is of higher confidence than in oligogenic or polygenic phenotypes. The large effect that a genetic variant has on a monogenic disorder provides the possibility to efficiently investigate the biological and molecular impact that the variant might have using cell and animal models. This greatly helps to clarify the potential causal connections between mutation of that particular gene and the condition, strengthening (or not) the link between genotype and phenotype. By further investigating the functions of the gene of interest and its molecular networks, researchers can expand the knowledge of the biological foundations of complex neurodevelopmental disorders.

The disruption of *FOXP2* in CAS cases is a paradigm of a monogenic language disorder. As discussed in the following sections, the investigation of the function and dysfunction of FOXP2 in cell and animal models has provided ample insights into the biology underlying speech and language. To date, only a few other cases of monogenic language disorders have been described or suggested. For instance, the *ERC1* gene was found to be the smallest region of overlap in probands with deletions in the 12p13.33 locus that suffered from CAS

(Thevenon *et al.* , 2013). But the link between *ERC1* and language is not as clear as the one with *FOXP2*, as more genes are affected in the deletions and no functional studies have been performed to see the impact of the loss of function of *ERC1* in model systems.

Conversely, in other neurodevelopmental conditions such as ASD or ID, which can take advantage of much larger cohorts, advances in sequencing technologies such as NGS and WES have allowed the identification of hundreds of novel candidate variants (Vorstman *et al.* , 2017; Hoischen *et al.* , 2014). ASD genetic studies provide proof that the investigation of monogenic conditions can greatly help narrowing the neurobiological pathways underlying heterogenic neurodevelopmental disorders. Genes implicated in ASD converge on various biological processes such as chromatin remodeling and transcriptional regulation, cell growth and proliferation, and neuronal processes such as synaptic activity and organization, dendritic morphology and axonogenesis (Vorstman *et al.* , 2017).

Neurodevelopmental disorders frequently co-occur with speech and language difficulties, which suggests an overlapping etiology that probably includes shared genetic influences (Graham & Fisher, 2015). Therefore, investigating monogenic neurodevelopmental disorders in which language is severely impaired can also be of great help to unmask genetic factors underlying language impairment. Multiple language-related genes have been identified through studies involving individuals that suffered from additional cognitive deficits other than just language impairment. Relevant examples include mutations in *GRIN2A*, which have been linked to an epilepsy-aphasia disorder that includes speech apraxia and dysarthria (Turner *et al.* , 2015); and *de novo* mutations in *TBR1*, a transcription factor recurrently mutated in sporadic ASD cases that include language difficulties (Deriziotis *et al.* , 2014a). Initiatives like the Deciphering Developmental Disorders (DDD) study (`https://decipher.sanger.ac.uk/ddd#overview`) or the SFARI database (`https://gene.sfari.org`) are valuable resources of genetic variants found in neurodevelopmental disorders and autism, respectively. As of September 2017, more than 1500 genes have been implicated in developmental disorders by the DDD study and 900 genes have been implicated in ASD and are reported in the SFARI database. Although the confidence for these associations is variable since the mere occurrence of a rare mutation in a gene does not inevitably imply causality, a large number of them are supported by multiple and/or strong sources of evidence. As mentioned above, functional characterization of the putative risk variants can further support the role of the

variant in the disorder as well as reveal novel insights of the basic biology the gene that may be relevant for the phenotype. For instance, the pathogenic role of TBR1 variants in ASD was strongly confirmed with functional studies showing that patient mutations impaired core aspects of the biology of the encoded protein such as its transcriptional activity (Deriziotis *et al.* , 2014a). Moreover, these findings led researchers to investigate the physical interaction between the TBR1 and FOXP2 proteins, and to consequently discover that pathogenic mutations in both transcription factors disrupted the interaction. These findings revealed a relevant molecular link underlying distinct neurodevelopmental disorders that include impaired language ability, hence providing another piece in the intricate puzzle of genetics of language and brain development (Deriziotis *et al.* , 2014a).

Investigating the biology of language-related genes can uncover the networks in which they participate and can potentially feed back into the interpretation of NGS and GWAS studies. For instance, a number of FOXP2-related genes are also implicated in cases of neurodevelopmental disorders. As aforementioned, TBR1 mutations lead to ASD with language delay (Deriziotis *et al.* , 2014a). FOXP1 is the closest paralogue of FOXP2 and the encoded proteins form heterodimers to regulate gene expression; rare disruptions of FOXP1 result in a syndromic form of ID that includes language impairment (Bacon & Rappold, 2012; Sollis *et al.* , 2016). Various FOXP2 downstream target genes have also been connected to neurodevelopmental conditions such as ASD, schizophrenia or epilepsy (Song *et al.* , 2008; Roll *et al.* , 2006; Lambert *et al.* , 2014; Peng *et al.* , 2013; Mukamel *et al.* , 2011; Walker *et al.* , 2012; Roll *et al.* , 2010).

Transcription-related genes have emerged as key determinants of the biological processes that lead to neurodevelopmental disorders (De Rubeis *et al.* , 2014). Several transcription-related genes have been found in monogenic cases that presented an impaired language together with other cognitive deficits. Due to their roles in regulating the expression of large networks of genes, the neurodevelopmental conditions caused by mutations in transcription factors are often dosage sensitive. For instance, haploinsufficiency of the transcription factors FOXP1, TBR1, TCF4, FOXG1, SOX5 or SATB2 all lead to neurodevelopmental disorders that include language difficulties (Sollis *et al.* , 2016; Zarate & Fish, 2016; Lamb *et al.* , 2012; Peter *et al.* , 2014; Ma *et al.* , 2016; Nesbitt *et al.* , 2015; Lennertz *et al.* , 2011; Deriziotis *et al.* , 2014a). The way language is affected in each of these monogenic disorders is variable; probands carrying FOXP1 vari-

ants mostly show an affected expressive language, SOX5 and TBR1 mutations cause prominent language delays, and SATB2, FOXG1 and TCF4 mutation cases have absent or limited speech development.

The following sections will (1) outline how the investigation of FOXP2 function has opened up many paths into the molecular and neural basis of speech and language, and (2) introduce BCL11A, a transcription factor potentially relevant in neurodevelopmental language-related disorder.

1.2.2 *The FOXP family of proteins*

The forkhead (FOX) family of proteins includes over 40 members characterized by the presence of a forkhead box (FOX) domain, a highly conserved sequence that mediates binding of these proteins to DNA. FOX proteins play important roles during embryonic development in the regulation of genes important for orchestrating organogenesis (Hannenhalli & Kaestner, 2009; Jackson *et al.* , 2010; Golson & Kaestner, 2016). The FOX family is organized into 19 subfamilies based on phylogenetic hierarchy and sequence homology (Hannenhalli & Kaestner, 2009; Jackson *et al.* , 2010). FOXP2 belongs to the FOXP subfamily, which comprises FOXP1, FOXP2, FOXP3 and FOXP4, all of which generally act as transcriptional repressors (Golson & Kaestner, 2016; Li *et al.* , 2004; Shu *et al.* , 2001). FOXP1, FOXP2 and FOXP4 play overlapping yet distinct functions in neurodevelopment and cooperative roles in mouse lung and esophagus development (Bowers & Konopka, 2012b; Li *et al.* , 2004). The involvement of Foxp2 and Foxp4 in spinal cord development (Rousso *et al.* , 2012) has also been described. In contrast, FOXP3 is structurally and evolutionary divergent, and its function and expression is restricted to the immune system, where it regulates T-cell lymphocyte development (Fontenot *et al.* , 2003).

FOXP1, FOXP2 and FOXP4 are expressed in the brain, the heart and the lung (Shu *et al.* , 2001; Teufel *et al.* , 2003). They are all expressed in the cerebral cortex and in the striatum. FOXP1 and FOXP2 are also found in the thalamus and hypothalamus, and FOXP1 and FOXP4 overlap in the hippocampus. FOXP2 and FOXP4 are also specifically expressed in the Purkinje cells in the cerebellum. In the cerebral cortex, FOXP1 expression is confined to layers III-V; FOXP2, to layers V-VI; and FOXP4 is expressed through all the cortical layers (Takahashi *et al.* , 2003, 2008; Ferland *et al.* , 2003).

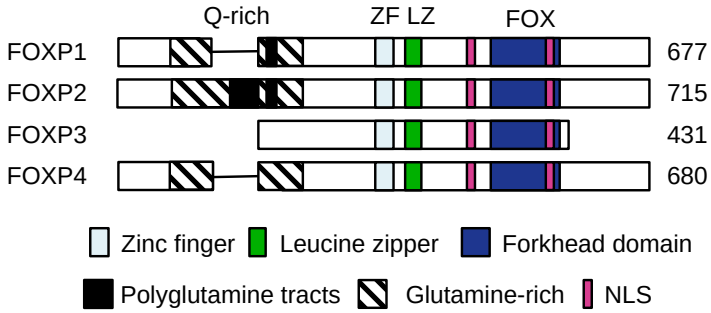


Figure 1.1: The FOXP family of proteins Schematic representation of the FOXP family of proteins. Known domains are shown: the glutamine-rich region (Q-rich) is shaded in black and white dashes; the polyglutamine tracts are shaded black; the zinc finger (ZF), in light blue; the leucine zipper (LZ), in green; the forkhead domain (FOX), in dark blue; and the nuclear localization signals (NLS), in pink. The number of amino acid residues is indicated to the right of the schematic

The FOXP proteins share a conserved leucine zipper domain, a zinc finger motif, a long glutamine-rich region (not present in FOXP3), and the FOX DNA-binding domain, which is the most evolutionary conserved part of all FOXP members (Li *et al.* , 2004; Wang *et al.* , 2003) (Figure 1.1). The FOX domain includes two nuclear localization signals (NLS) (Mizutani *et al.* , 2007; Vernes *et al.* , 2007). The FOXP proteins bind to the specific nucleotide sequence in the DNA 5'-TRTTKRY through the FOX domain, but sequence variation among the family members is thought to confer specificity to each protein in the recognition of different targets (Vernes *et al.* , 2007; Stroud *et al.* , 2006; Wang *et al.* , 2003). The crystal structure of the FOX domain of FOXP2 bound to DNA further revealed key residues in the recognition helix (helix 3) that mediate this interaction (Stroud *et al.* , 2006).

The leucine zipper mediates the homo- and hetero-dimerization of the FOXP proteins, which is a unique characteristic of this subfamily. This property is needed for them to bind DNA and regulate gene expression (Li *et al.* , 2004). A mutation in the leucine zipper of FOXP3 that disrupts its dimerization ability results in an X-linked autoimmunity and allergic dysregulation syndrome (IPEX) (Chatila *et al.* , 2000). This mutation involves the deletion of a conserved glutamic amino acid in the leucine zipper (Chae *et al.* , 2006). When this con-

served residue was deleted in FOXP1, FOXP2 or FOXP4, it also impaired their ability to homo- and hetero-dimerize (Li *et al.* , 2004). FOXP family hetero-dimerization may offer another level of DNA site recognition, elegantly controlled by the expression or availability of different FoxP partners.

All members of the FOXP subfamily also have a C2H2 zinc finger domain. In some proteins this kind of domain is involved in binding DNA (Wolfe *et al.* , 2003), but it may also mediate protein-RNA and protein-protein interactions. The function of the zinc finger domain in FOXP proteins has not been determined yet. All FOXP proteins (except FOXP3) have a glutamine-rich region. FOXP2 has two polyglutamine tracts (polyQ) of 40 and 10 glutamines, neither of which are present in human FOXP1 or FOXP4 (Estruch *et al.* , 2016a). Rare polymorphisms of the polyQ tract have been identified in FOXP2 in individuals with neurodevelopmental disorders; for instance, an addition of 4 glutamines was found in a case of CAS (MacDermot *et al.* , 2005). However, the causal role of these variants in the disorder is unclear, as variations in the number of glutamines in FOXP2 are also found in the general population (Exome Aggregation Consortium (ExAC), Cambridge, MA (`http://exac.broadinstitute.org`)). Unlike polyQs in triplet repeat diseases such as Huntington disease which are encoded by long CAG repeats, FOXP2 polyQ tracts are encoded by a mixture of CAAs and CAGs, which probably makes them more stable and less prone to expansion (Butland *et al.* , 2007; Bruce & Margolis, 2002). One study suggested a role in transcriptional repression for this region of the FOXP proteins, based on the finding that mouse Foxp1 isoforms lacking the glutamine-rich region showed stronger repression than the isoforms that have it (Wang *et al.* , 2003). However, an effect of this kind was not observed in studies of FOXP2, which found that the protein repressed transcription at similar levels, regardless of presence or absence of the polyQ tract (Estruch *et al.* , 2016a).

1.2.3 *FOXP2 mutations in childhood apraxia of speech*

As introduced in section 1.1.2, heterozygous disruptions of the FOXP2 gene cause a rare speech and language disorder, in which the most prominent feature is childhood apraxia of speech (CAS), which is a motor speech disorder characterized by difficulties in coordinating and articulating the rapid orofacial muscle movements necessary for fluent speech (Lai *et al.* , 2001). Together with CAS, subjects with FOXP2 haploinsufficiency also present receptive and expres-

sive language difficulties, IQ is usually (but not always) within the normal range (Morgan *et al.* , 2016), and fine motor skills may be impaired but gross motor skills are typically unaltered.

The first DNA sequence variant that was implicated in this speech and language disorder was a missense mutation in one copy of the FOXP2 gene which yields an arginine-to-histidine substitution (p.R553H) in the highly conserved DNA-binding domain of the encoded transcription factor. The speech and motor deficits were accompanied by other expressive and receptive problems in both oral and written language, as well as poor processing and production of grammatical structures. Some of the affected KE family members also exhibited lower IQ scores than unaffected members, but this did not entirely co-segregate with the language disorder indicating that effects on non-verbal cognition were unlikely to be central to the phenotype.

The underlying neuropathology of the disorder was studied using structural magnetic resonance imaging (MRI) and functional MRI (fMRI) techniques. Together, these approaches revealed that the affected KE family members presented bilateral structural and functional abnormalities in several motor-related regions of the brain (Vargha-Khadem *et al.* , 1998; Watkins *et al.* , 2002; Liégeois *et al.* , 2003, 2011). Affected KE family members exhibited significant bilateral reductions in grey matter density in the caudate nucleus of the basal ganglia, the ventral cerebellum and the inferior frontal gyrus of the cortex, which contains Broca's area. Conversely, the putamen of the basal ganglia and several other cortical regions including the Wernicke's area presented increased levels of grey matter. Interestingly, the caudate nucleus was smaller in individuals with CAS and its volume correlated significantly with the performance on a test of oral praxis (Vargha-Khadem *et al.* , 1998; Watkins *et al.* , 2002; Belton *et al.* , 2003). Consistent with the morphological findings, fMRI during the performance of semantic language tasks revealed an under-activation of Broca's area and the putamen in the affected family KE members (Liégeois *et al.* , 2003, 2011). All these data indicate that the frontostriatal and frontocerebellar networks involved in learning, planning and execution of speech motor sequences could be key circuits affected in impaired KE family members and FOXP2 might play an important role in the development of such circuits.

Since the discoveries of the mutation in the KE family and the translocation affecting FOXP2 in a child with CAS, several etiological inherited and *de novo* FOXP2 variants in CAS have been identified, although overall disruption of this

gene remains a rare cause of language disorder. Various large 7q31 deletions that span FOXP2 have also been reported in individuals with CAS (Feuk *et al.* , 2006; Lennon *et al.* , 2007; Palka *et al.* , 2012; Rice *et al.* , 2012; Zeesman *et al.* , 2006; Zilina *et al.* , 2012). The deletions encompassed several genes other than FOXP2, such that the phenotype of some these individuals is broader and more severe than that of the KE family, but it is likely that disruption of FOXP2 is the cause of the speech and language problems of these individuals. Chromosomal rearrangements directly disrupting the FOXP2 gene in subjects with CAS further support its involvement in the etiology of the disorder (Feuk *et al.* , 2006; Lai *et al.* , 2000, 2001; Shriberg *et al.* , 2006; Tomblin *et al.* , 2009).

Most interestingly, several intragenic FOXP2 mutations affecting the amino-acid sequence of the encoded protein have been reported in cases of speech-related disorders (Figure 1.2). Based on the inheritance patterns, some of the variants are clearly pathogenic, such as the variant of the KE family, p.R553H (Lai *et al.* , 2001). Other likely pathogenic variants include the nonsense variants p.R328*, identified in two independent families with CAS, and p.R478*, found *de novo* in one proband with delayed speech development and also to co-segregate in another family with speech delay (MacDermot *et al.* , 2005; Reuter *et al.* , 2016). Mutations that are predicted to truncate FOXP2 protein, such as nonsense and frameshift variants, are usually etiological since they usually lead to a complete loss of protein function. Examples of this are FOXP2 variants p.R328* and p.R478* or *de novo* variants identified in a proband with CAS (p.Q390Vfs*7) and in individuals with mild ASD and speech impairment (p.R564* and p.F538Vfs*28) (Turner *et al.* , 2013; Reuter *et al.* , 2016). Of note, inherited and *de novo* intragenic mutations in *FOXP2*, which include some of the above described (p.R478* and p.F538Vfs*28), have been recently identified in 14 individuals from eight unrelated families. The subjects present variable degrees and types of speech and language impairment and most of them also exhibited mild cognitive deficits, broadening the phenotypical and clinical spectrum associated with *FOXP2* disruptions (Reuter *et al.* , 2016).

Several reported missense variants of FOXP2 are of uncertain significance, since they either do not fully segregate with the typical disease phenotype in the respective families in which they were reported or their inheritance pattern is unknown (Graham & Fisher, 2015). For example, variant p.Q17L was identified in a subject with CAS but not in his affected sibling, and variant p.M406T was carried by a proband that suffered from epilepsy-aphasia spectrum disorder

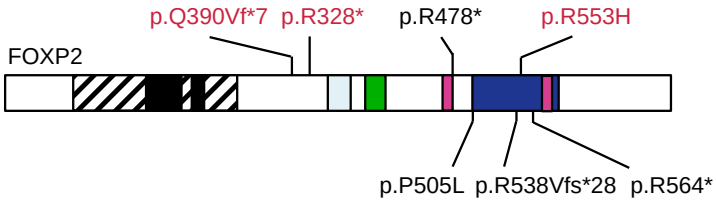


Figure 1.2: FOXP2 causal variants Schematic representation of FOXP2 variants probably pathogenic identified in cases of speech impairment. Known domains are shown: the glutamine-rich region (Q-rich) is shaded in black and white dashes; the polyglutamine tracts are shaded black; the zinc finger (ZF), in light blue; the leucine zipper (LZ), in green; the forkhead domain (FOX), in dark blue; and the nuclear localization signals (NLS), in pink. Variants that have been functionally characterized in cell models are in red.

but also by his unaffected siblings and father (MacDermot *et al.* , 2005; Roll *et al.* , 2010). Variant p.N597H was identified in a proband with CAS but the parental genotypes were not determined (Laffin *et al.* , 2012) and variant p.R536P, found in a child with severe speech delay and autistic traits, was inherited from his unaffected father, who probably presented mosaicism for this variant, but the affected brother was not sequenced (Reuter *et al.* , 2016). In the cases where the information on the pedigree is incomplete or where the variant does not clearly co-segregate with the phenotype it is essential to determine with experimental methods whether the variant has an impact on protein function in order to clarify putative causality. Bioinformatics algorithms, which are based on known information on protein sequence and evolutionary history, can predict whether an amino acid change may be damaging, but often provide false positives or negatives. Functional characterization assays have proven to be fundamental in determining the causality of FOXP2 variants in the disorder (Vernes *et al.* , 2006; Estruch *et al.* , 2016a). These have been done extensively for the mutation of the KE family p.R553H, which causes abnormal subcellular localization of the FOXP2 protein and, most notably, impairs its ability to bind DNA and regulate gene expression (Vernes *et al.* , 2006; Estruch *et al.* , 2016a). Similarly, p.R328* and p.Q390Vfs*7 mutations impair subcellular localization and transcriptional activity of FOXP2, which is expected since these mutations result in a truncated form of the protein that lacks the DNA-binding domain as well as the NLSs

(Vernes *et al.* , 2006; Estruch *et al.* , 2016a). In both cases, the functional assays provided further support for the pathogenic role of the variants in the etiology of CAS. Conversely, the FOXP2 missense variants p.Q17L and p.M406T, which did not co-segregate with disorder in the respective families but were predicted to be damaging variants, displayed normal nuclear localization and DNA binding, which suggests that they are incidental to the phenotype (Estruch *et al.* , 2016a). Similarly, FOXP2 variant p.N597H, for which the genotype of the parents of the affected proband was unknown, did not exhibit any altered protein function in cellular assays (Estruch *et al.* , 2016a). These are examples that show how functional characterization studies are critical to ascertain the significance of FOXP2 variants for speech and language disorder.

1.2.4 *FOXP1 variants in neurodevelopmental disorders*

FOXP1 is the closest paraloguous protein to FOXP2, sharing 68% amino acid identity. The expression of FOXP1 significantly overlaps with FOXP2 in several brain structures (Teramitsu *et al.* , 2004; Ferland *et al.* , 2003). Moreover, the two proteins are known to functionally interact by forming hetero-dimers to cooperatively regulate shared downstream target genes (Li *et al.* , 2004; Shu *et al.* , 2001). The close relationship of FOXP1 with FOXP2 suggested that mutations affecting FOXP1 might also be related to speech and language disorders.

The first evidence implicating *FOXP1* in speech and language impairment came from patients that presented large deletions of chromosome 3p, which spanned this gene, as well as other genes (Petek *et al.* , 2003; Pariani *et al.* , 2009). Although the phenotypes were understandably varied probably due to involvement of multiple genes, most of the subjects presented language difficulties. Subsequent studies of individuals with ID reported missense and truncating variants or smaller deletions encompassing *FOXP1*, demonstrating that the disruption of a single copy of the gene was sufficient to cause a neurodevelopmental disorder characterized by global developmental delay and ID together with moderate to severe speech impairment (Hamdan *et al.* , 2010; O'Roak *et al.* , 2011; Srivastava *et al.* , 2014; Lozano *et al.* , 2015; Sollis *et al.* , 2016; Carr *et al.* , 2010; Le Fevre *et al.* , 2013; Siper *et al.* , 2017; Meerschaut *et al.* , 2017). Autistic behaviours frequently accompany the *FOXP1*-related disorder (Siper *et al.* , 2017; Meerschaut *et al.* , 2017). Recent large-scale exome sequencing efforts have identified *FOXP1* as a gene with recurrent *de novo* mutations associated

with ASD (Iossifov *et al.* , 2014). Also, distinctive facial features such as a high broad forehead and a short nose are associated with *FOXP1* mutations (Siper *et al.* , 2017; Meerschaut *et al.* , 2017). The language impairment seen in patients with *FOXP1* mutations consistently affects expressive skills more severely than receptive skills. Although articulation problems were reported for a large number of cases, formal diagnoses of CAS have not been described (Bacon & Rappold, 2012). Moreover, no causative *FOXP1* mutations were identified in an early study that screened a cohort of 49 patients with CAS, which is perhaps not surprising given that it is now known that the phenotype resulting from haploinsufficiency of this gene is typically more severe than that due to *FOXP2* mutations (Vernes *et al.* , 2009).

Notably, all of the causative *FOXP1* mutations reported to date in which the inheritance could be determined have occurred *de novo*, consistent with the fact that this type of mutation represents the major genetic cause of severe forms of sporadic neurodevelopmental disorders, such as ID, epilepsy and autism (Siper *et al.* , 2017; Veltman & Brunner, 2012). The disrupting effect of missense and truncating variants on FOXP1 protein function was demonstrated in cell models, which showed that etiological mutations disrupt nuclear subcellular localization and transcriptional repression activity (Sollis *et al.* , 2016). Interestingly, whilst most of the assessed variants showed impaired dimerization with FOXP1 or FOXP2, some others retained the interaction, suggesting the possibility of a dominant-negative effect, for example by preventing wild-type FOXP protein from binding to DNA and regulating gene expression (Sollis *et al.* , 2016).

Given that the phenotype caused by *FOXP1* mutations is more severe than that due to FOXP2 mutations, it may be that the former gene has broader effects on neurodevelopment than the latter. All patients carrying heterozygous *FOXP1* mutations present with problems in the language domain, but it is not known to what extent these language deficits share molecular pathophysiology with *FOXP2*-related language disorder. It is interesting to speculate that the speech and language deficits in some of the reported *FOXP1* mutation cases could be exacerbated through a potential dominant-negative effect on wild-type FOXP2 protein (Bacon & Rappold, 2012; Sollis *et al.* , 2016).

1.2.5 *Functions of FOXP2 in the developing and adult brain*

The discovery of FOXP2 mutations in speech and language disorder put this transcription factor in the spotlight and promoted many investigations to explore its functions in the brain. FOXP2 is expressed in several brain areas, including the basal ganglia, the deep cortical layers, the thalamus and the cerebellum (Lai *et al.* , 2003), sites in which structural and functional abnormalities have been reported in the KE family (Vargha-Khadem *et al.* , 1998; Watkins *et al.* , 2002; Liégeois *et al.* , 2003, 2011), as described in section 1.2.3. The involvement of these FOXP2-expressing regions in neural processes such as learning, planning and executing movement sequences suggests a role for this gene in the circuitry underlying speech and language.

Animal models, including knockouts as well as cell lines carrying etiological variants, have been essential to further the knowledge on FOXP2 functions in the developing brain. Human FOXP2 and mouse Foxp2 are expressed in the same brain regions, including the cerebral cortex, the striatum and the cerebellum (French & Fisher, 2014). Homozygous disruption of murine Foxp2 results in mice with smaller cerebellums that die 3 to 4 weeks after birth (French & Fisher, 2014). Heterozygous knockouts of Foxp2 in mice have also been studied to model the haploinsufficiency of humans with *FOXP2*-related CAS. Some studies report that Foxp2 haploinsufficiency leads to striatal and cerebellar structural defects together with motor impairments (Shu *et al.* , 2005). Like homozygous knockouts, mice carrying two copies of Foxp2 p.R552H, the mouse variant analogous to the mutation found in the KE family, die soon after birth and have a cerebellum of reduced size with decreased foliation (Groszer *et al.* , 2008). Foxp2 p.R552H heterozygote mice (matching the heterozygous status observed in humans with CAS) are viable and healthy. Despite exhibiting a normal baseline motor performance, experiments using a voluntary running-wheel system and accelerating rotarods showed that these heterozygous mice have impaired motor-skill learning (Groszer *et al.* , 2008). For example, Foxp2 p.R552H heterozygotes had significant delays in learning the necessary motor skills to use the wheel (Groszer *et al.* , 2008). This is in line with the problems in fine motor control observed in some subjects with *FOXP2* disruptions; such underlying difficulties might be particularly apparent for learning the motor sequences that are crucial for speech articulation (Alcock *et al.* , 2000; Palka *et al.* , 2012; Turner *et al.* , 2013). Furthermore, electrophysiology experiments

revealed abnormal synaptic plasticity for Foxp2 p.R552H heterozygous mice in the striatum and cerebellum, brain regions implicated in motor learning and control (Groszer *et al.* , 2008).

FOXP2 is among the most conserved genes in humans. Aside from variation in polyQ tracts, the mouse and human proteins differ in only three amino acids and show conserved expression patterns. Two of the relevant substitutions occurred on the human lineage after splitting from chimpanzees. Researchers inserted these two human substitutions into the endogenous *Foxp2* gene in mice, creating a partially humanized version of the protein, and observed that striatal, cortical and thalamic neurons had significantly longer dendrites and increased synaptic plasticity (Enard *et al.* , 2009; Reimers-Kipping *et al.* , 2011). Notably, mice carrying one copy of the Foxp2 p.R552H mutation showed the opposite effects (Groszer *et al.* , 2008; Enard *et al.* , 2009). All together, this points towards an important role for Foxp2 in cortico-striatal circuits that regulate motor-skill learning and function, and also suggests that this particular function might have been important for the evolution of speech and language in humans (Enard *et al.* , 2009; Reimers-Kipping *et al.* , 2011). Further investigations indicated that Foxp2 controls the development of cortico-striatal circuits through directly repressing the expression of Mef2c, which is a regulator of cortico-striatal synapse formation and striatal spinogenesis (Chen *et al.* , 2016b).

Independent investigations of downstream target genes of FOXP2 both in cultured neurons and human fetal brain further highlighted its role in regulating multiple aspects of neural plasticity. FOXP2 directly regulates genes involved in synaptic plasticity, neurite outgrowth and axon guidance (Spiteri *et al.* , 2007; Vernes *et al.* , 2007). Follow-up functional experiments demonstrated that in the developing brain the KE mutation significantly impaired normal neurite outgrowth and branching of medium spiny neurons, a type of neuron that represent 85% of all striatal neurons (Vernes *et al.* , 2011). Gain- and loss-of-function experiments indicated that FOXP2 promotes neuronal differentiation in neuronal-like cell lines and medium spiny neurons (Devanna *et al.* , 2014; Chiu *et al.* , 2014). Interestingly, this effect seems to be mediated by interaction of FOXP2 with retinoic acid signaling pathways, some members of which are regulated by FOXP2 (Devanna *et al.* , 2014).

Animal models have also been used to investigate the role of this transcription factor in the production of vocalizations. Wild-type mouse pups generate ultrasonic vocalizations when separated from their mother, known as isolation

calls. These vocalizations are absent in Foxp2 complete knockouts and reduced in the heterozygote knockouts (Shu *et al.* , 2005). However, pups carrying Foxp2 p.R552H heterozygotes produce similar numbers of isolation calls to wild-type littermates (Groszer *et al.* , 2008). Ultrasonic vocalizations of mouse pups are entirely innate and cannot be considered a proxy to human speech or language. Conversely, the vocalizations of zebra finches have to be learned from an adult tutor. Juvenile male zebra finches learn complex vocalizations by imitation, which makes this animal a valuable model to study roles of FoxP2 in vocal learning. Additionally, the neural circuits that underlie vocal motor learning in songbirds show interesting parallels to the ones in humans (Fisher & Scharff, 2009). In zebra finches, FoxP2 is expressed in a striatal brain structure of the song system called Area X (Teramitsu *et al.* , 2004). Interestingly, the expression of FoxP2 in Area X is increased during the period of song learning. Moreover, viral knockdown of FoxP2 in this structure results in an incomplete and inaccurate vocal imitation, which alters the ability of juvenile birds to learn songs from other individuals (Murugan *et al.* , 2013; Haesler *et al.* , 2007). Follow-up studies found that knocking down FoxP2 affected the signaling between area X and the connected cortical structures, in part through a decrease of dopamine levels in these circuits (Murugan *et al.* , 2013). Together with the fact that overexpression of FoxP2 in this area also results in inaccurate song imitation (Haesler *et al.* , 2007), these studies indicate that a tightly regulated FoxP2 dosage is key for normal vocal learning in zebra finches.

1.2.6 *Functions of FOXP1 in the developing brain*

It has been known for many years that FoxP1 is expressed in the developing brain (Shu *et al.* , 2001; Ferland *et al.* , 2003) but its specific roles there were not explored until several independent studies consistently implicated this gene in neurodevelopmental disorder (Hamdan *et al.* , 2010; O'Roak *et al.* , 2011; Srivastava *et al.* , 2014; Lozano *et al.* , 2015; Sollis *et al.* , 2016; Carr *et al.* , 2010; Le Fevre *et al.* , 2013).

FOXP1 is widely expressed in human and murine tissues and is involved in the development of multiple organs, such as brain, heart, lung, esophagus and the immune system (Shu *et al.* , 2001; Shi *et al.* , 2004; Wang *et al.* , 2004; Shu *et al.* , 2007; Hu *et al.* , 2006). Within the central nervous system, key sites of expression include the cerebral cortex, striatum, the hippocampus and the spinal

cord (Shu *et al.* , 2001; Morikawa *et al.* , 2009; Ferland *et al.* , 2003; Teramitsu *et al.* , 2004). During development, it regulates cortical migration, midbrain dopamine neuron differentiation and also plays a role in the development of the striatum (Palmesino *et al.* , 2010; Konstantoulas *et al.* , 2010; Araujo *et al.* , 2015; Bacon *et al.* , 2015; Precious *et al.* , 2016).

Foxp1 knockout mice die at embryonic day 14.5 (E14.5) due to cardiac defects, which indicates that this gene is also important for non-neural developmental processes (Wang *et al.* , 2004; Shu *et al.* , 2007). Mice with conditional brain-specific loss of Foxp1 show altered hippocampal electrophysiology and striatal morphology (Bacon *et al.* , 2015). These animals also exhibited memory deficits, repetitive behaviours as well as reduced social interests (Bacon *et al.* , 2015). It has also been shown that Foxp1 knockdown specifically in the mouse cerebral cortex leads to an abnormal neuronal migration (Li *et al.* , 2015b). Like Foxp2, FoxP1 is also involved in the development and function of striatal medium spiny neurons. Reduction of Foxp1 levels results in abnormalities in the excitability and differentiation of medium spiny neurons (Araujo *et al.* , 2015). Also, it has been suggested that Foxp1 regulates pathways involved in striatal development and function both in mice and human neuronal cell models (Araujo *et al.* , 2015). The implication of this gene in striatal and cortical development indicates that, similarly to Foxp2, it may also be important for cortico-striatal circuits that underlie motor function. Furthermore, Foxp1 is also expressed in the motor neuronal region of the spinal cord (Tamura *et al.* , 2003; Shu *et al.* , 2001), where it plays a critical role regulating the formation and connection of motor neurons to target muscles by acting as a downstream effector of Hox gene pathways (Rousso *et al.* , 2012; Arber, 2008; Dasen *et al.* , 2008). So far, findings from mouse models are in line with the phenotypes observed in patients with *FOXP1* mutations and therefore shed light on the molecular mechanisms that may be involved in the etiology of related neurodevelopmental disorders.

1.2.7 *FOXP4*

FOXP4 is the least studied of the FOXP subfamily members. FOXP4 expression has been reported in several tissues including brain, lung, heart, spleen, liver, testes and kidney (Teufel *et al.* , 2003; Lu *et al.* , 2002). While heterozygous Foxp4 knockout mice do not exhibit any obvious defects, homozygous knock-

outs die at E12.5 due to cardiac malformations, similarly to Foxp1 knockout mice. Further examination of Foxp4 mutant mice revealed the development of two complete hearts (Li *et al.* , 2004). A neural role for Foxp4 has been reported in the mouse cerebellum (Tam *et al.* , 2011). FoxP4 is expressed in Purkinje cells of the cerebellum from embryonic stages to adulthood. Knocking down the expression of Foxp4 at postnatal day 10 (P10), but not at P5, resulted in defects of Purkinje cell dendritic arborization (Tam *et al.* , 2011).

The only FOXP4 disruption reported in humans is a homozygous frameshift mutation found in a child with neurodevelopmental delay and malformations of the larynx and the heart (Charng *et al.* , 2016). Given the known expression pattern and data from mouse knockouts, it is plausible that this phenotype might be due to *FOXP4* mutation; however further observations of disruptions of the gene are necessary to confirm the existence of a *FOXP4*-related syndrome. Although *FOXP4* was the strongest candidate gene in this case, variants in two other genes were also found in a homozygous state in the child and heterozygous state in the parents (Charng *et al.* , 2016).

1.2.8 *FOXP protein-protein interactions*

The fine control that the FOXP proteins exert spatially and temporally in the regulation of gene expression appears to be crucial for the development of the language and motor circuitry. Transcription factors, such as the FOXPs, are precisely regulated via their associations with other proteins (Smith & Matthews, 2016). Identifying the proteins that interact with the FOXP proteins can shed light on their functions in normal brain development, and also help to explain the different phenotypes resulting from mutations affecting these proteins. Moreover, finding proteins that bind these transcription factors can provide new candidate genes that can be examined for variants in patient cohorts. Studies of the key interactions can facilitate interpretation of the biological relevance of rare variants identified by exome/genome sequencing and of associations found in genome-wide association studies of language impairment. Understanding the FOXP interactome will extend our knowledge of the roles of these transcription factors in the brain and consequently, help further elucidate the genetic basis of speech and language.

Two prominent experimental approaches that are typically used to screen for protein-protein interactions are yeast two-hybrid screens (Y2H) and affinity

purification followed by mass spectrometry-based proteomics (AP/MS). Both these approaches have been used to try identifying putative FOXP interactors. In the Y2H assay, the two proteins of interest are fused to the DNA binding domain (DBD) and the transcriptional activation domain (AD) of a transcription factor, respectively. The protein fused to the DBD is known as the bait, and the protein fused to the AD as the prey. Both fusion proteins are expressed together in a genetically modified yeast strain in which the specific DNA binding site of the DBD is upstream of a reporter gene. Upon interaction between the bait and the prey, the DBD and AD are brought in close proximity and a functional transcription factor is reconstituted, consequently activating the expression of the reporter gene. This leads to a specific phenotype that can be selected usually by growing the yeast in selective media for that phenotype. For instance, the reporter gene may be an essential amino acid (e.g. leucine) that will allow only the yeast cells that express it to grow in plates lacking that specific amino acid. To perform a genome-wide Y2H screen for interactors of a given bait, a library of cDNAs fused to ADs is used as the prey. In an AP/MS screen, a protein usually fused to an epitope-tag is either purified by affinity columns recognizing the tag or immuno-precipitated using a specific antibody. This results in the purification of multi-protein complexes containing the tagged protein. The complexes are digested and the masses of the resulting peptides are subsequently analyzed with MS to identify the different proteins that were present in the pulled-down complex.

Screening techniques are highly sensitive but can be susceptible to false positives. In AP/MS, common false positives include molecules that are difficult to remove with the washing steps, such as abundant proteins (e.g. actin, tubulin or ribosomal proteins), proteins that bind unfolded peptides (heat-shock proteins) and proteins that bind non-specifically to the affinity matrix. In Y2H, false positives may be of diverse origins. For instance, unfolded peptides can show non-specific interactions, while proteins that activate the expression of the reporter proteins or proteins that interfere with the selection method can also result in false positive signals. Thus, to fully ensure the veracity of an interaction found with these techniques, it is imperative to confirm the results obtained with a different method.

Pull-downs are a standard commonly used strategy for validating interactions, but it can be challenging to develop reliable and efficient assays, since the conditions have to be optimized for each protein pair. Techniques such as

Bioluminescence Resonance Energy Transfer (BRET) or Mammalian two-hybrid (M2H) are advantageous alternatives because they are performed in living mammalian cells and can therefore better mimic real in vivo interactions. In contrast to pull-down strategies, BRET and M2H methods do not require cell lysis, which can disrupt interactions, and additionally they offer a higher throughput. However, techniques like these don't have the power to detect all kinds of interactions. A certain technique might be well suited for one type of interaction, but not appropriate to detect a different one. It is very difficult to predict which method will work best for detecting any particular interaction; therefore it is ideal to assess and confirm protein interactions using several complementary approaches.

A number of putative FOXP-interacting proteins have been identified in large-scale high-throughput protein-protein interaction screening surveys (Corominas *et al.* , 2014; Sakai *et al.* , 2011; Li *et al.* , 2015a; Ravasi *et al.* , 2010). Most of these have not yet been confirmed with an alternative method, but a small number have been validated and further characterized (Figure 1.3). The FOXP proteins homo and hetero-dimerize with each other through their leucine zipper. (Li *et al.* , 2004; Deriziotis *et al.* , 2014b; Estruch *et al.* , 2016a). As explained above, the dimerization ability of these proteins is needed for them to bind DNA and regulate transcription, and it may play a role in modulating their DNA-binding specificity (Sin *et al.* , 2015; Li *et al.* , 2004).

The most well described FOXP interactors are the C-terminal binding proteins (CTBP) 1 and 2, which function as co-repressors for several transcription factors (Li *et al.* , 2004; Estruch *et al.* , 2016a). CTBP1 was identified in a Y2H screen using a fragment of FOXP2 encompassing residues 269 to 500 as bait. The interaction was validated with coIPs in HEK293 cells, which also showed that CTBP1 was able to interact with FOXP1 and FOXP2 but not with FOXP4. Accordingly, in luciferase assays CTBP1 enhanced the transcriptional repression activity of FOXP1 and FOXP2 but not FOXP4. A putative CTBP1 binding site (PLNLV) is conserved in FOXP1 and FOXP2 (Li *et al.* , 2004). FOXP2 truncations lacking the PLNLV site led to a loss of repression, however mutations of the CTBP1 binding site in both FOXP1 and FOXP2 did not significantly alter their ability to repress transcription. These findings suggest that CTBP1 may be important but is not essential for transcriptional repression mediated by FOXP1 and FOXP2. Evidence of the interaction between CTBP2 and FOXP2 has come from multiple independent Y2H screens (Sakai *et al.* , 2011; Rolland *et al.* , 2014;

Corominas *et al.* , 2014). Recently, a study by our group confirmed FOXP2 interaction with CTBP1 and CTBP2 using a BRET assay (Estruch *et al.* , 2016a). In line with the previously reported findings for CTBP1, CTBP2 also interacted with FOXP1 but not with FOXP4 (Estruch *et al.* , 2016a). However, two etiological FOXP2 truncations that lack the putative CTBP binding site proposed by Li and colleagues (Li *et al.* , 2004) partially retained the interaction with the CTBPs, suggesting that the PLNLV motif in FOXP2 is not essential for the interaction. Using a series of FOXP2 C-terminal truncations, the minimal CTBP binding region was narrowed down to residues 259 to 329. This region includes both of the amino acid substitutions that occurred in FOXP2 since the divergence of the human and the chimpanzee lineages, but these substitutions did not affect the CTBP interaction (Estruch *et al.* , 2016a).

Another co-factor that has been shown to interact with FOXP1, FOXP2 and FOXP4 is GATAD2B, a component of the NuRD chromatin- remodeling complex (Chokas *et al.* , 2010). This interaction was also identified using a Y2H screen using a FOXP2 fragment containing residues 250 to 500 as bait, which includes the leucine zipper and the zinc finger domains. The interaction was validated for FOXP1, FOXP2 and FOXP4 using coIP assays. To assess whether GATAD2B acted as a co-repressor of the FOXPs, a luciferase assay was carried out in cultured lung cells using a construct containing the lung-specific PDPN (also known as T1-alpha) promoter. Overexpression of GATAD2B with FOXP1 or FOXP4 significantly reduced the expression of the luciferase gene, suggesting that GATAD2B cooperates with FOXP1 and FOXP4 to repress the PDPN promoter. This effect was not observed for GATAD2B and FOXP2. Additional experiments corroborated the ability of FOXP1 and GATAD2B to synergistically repress expression of the PDPN endogenous gene. Further interactions with other components of the NuRD complex were shown for FOXP1 and FOXP4; MTA1 interacted with FOXP1 and HDAC1/2 interacted with FOXP1 and FOXP4. Strikingly, by cooperatively regulating lung-specific gene expression, HDAC2 and FOXP1 interaction was shown to play a role in the modulation of lung epithelium immunological responses to injury (Chokas *et al.* , 2010). Interestingly, GATAD2B mutations have been associated with ID (Willemsen *et al.* , 2013), but the relevance of the FOXP-GATAD2B interaction in the brain has not been studied.

FOXP1 also interacts with the co-repressor NCOR2 (Jepsen *et al.* , 2008). This interaction was first hypothesized based on the fact that knockout mice of NCOR2

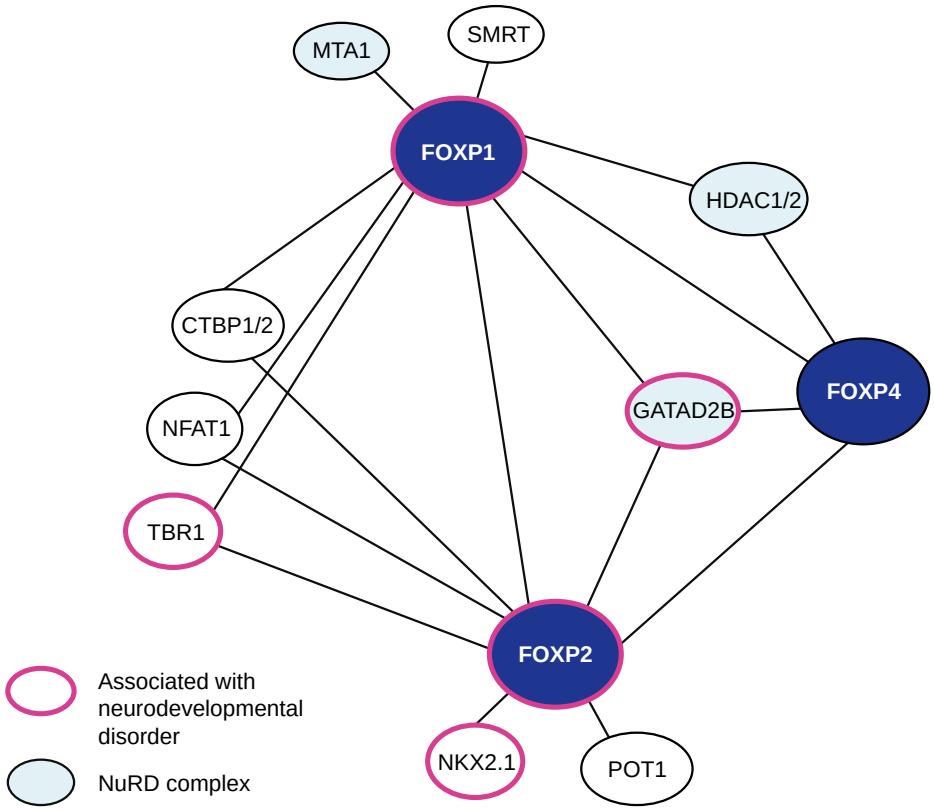


Figure 1.3: FOXP interacting-proteins. Network diagram of FOXP protein-protein interactions at least confirmed with two different experimental approaches. Lines indicate protein-protein interactions. A pink border indicates that the genes encoding for these proteins have been associated with neurodevelopmental disorders. Shaded in light blue are the members of the NuRD complex.

or FOXP1 exhibit similar cardiac malformation phenotypes. NCOR2-FOXP1 interaction was detected by co-IP experiments performed in HEK293 cells overexpressing FOXP1 and NCOR2, and also in dissected mice hearts, which express both proteins endogenously. Furthermore, it was demonstrated that FOXP1 and NCOR2 cooperatively repress the expression of common target genes to promote cardiac growth and regulate macrophage differentiation (Jepsen *et al.* , 2008).

An association between FOXP2 and POT1, a nuclear protein involved in telomere maintenance and DNA repair mechanisms has also been reported (Tanabe *et al.* , 2011). POT1 was identified in a Y2H screen using the C-terminal part of FOXP2 (484-715) as bait. Further Y2H assays narrowed down the POT1-interacting region in FOXP2 to residues 484-598, which include the FOX DNA-binding domain. It was observed that POT1 localized in the cytoplasm when transfected alone, but co-transfection with FOXP2 promoted its translocation to the cell nucleus. The etiological variant FOXP2 p.R553H did not change the subcellular localization of POT1. These findings contradict other studies that report that POT1 protein is localized in the cell nucleus, specifically in the telomeric foci (Calvete *et al.* , 2015; Ramsay *et al.* , 2013). Although this potential interaction is intriguing since it could reveal an additional non-canonical function for FOXP2, further evidence will be needed to support these findings.

In addition to the co-repressors described above, a few transcription factors have also been shown to interact with FOXP2. NKX2-1 is a transcription factor with important roles in lung, brain and thyroid development (Kimura, 2003; Sussel *et al.* , 1999). Nkx2-1 interaction with Foxp2 was identified in mouse lung cell lines using a mammalian two-hybrid assay and confirmed with CoIP experiments (Zhou *et al.* , 2008). In vitro GST pull-down assays revealed the Nkx2-1 DNA-binding domain mediated the interaction with Foxp2, which suggested that this interaction might prevent Nxk2-1 from binding to DNA. Subsequently, this was confirmed by competitive EMSA assays, in which increasing amounts of FOXP2 inhibited the formation of NKX2-1-DNA complexes in a dose-dependent manner. In line with this finding, Foxp2 inhibited Nkx2-1 ability to regulate the transcription of the lung-specific gene SP-C. Taken together, these data indicate that the interaction between Foxp2 and NKx2-1 acts to prevent Nkx2-1 from binding DNA and regulating lung-specific target genes (Zhou *et al.* , 2008).

Another transcription factor known to interact with FOXP2 is NFATC2. This was discovered in a study in which the FOXP2 forkhead domain was co-crystalized with NFATC2 to investigate the interaction between FOXP3 and NFATC2 (Wu *et al.* , 2006). Chemical crosslinking of the FOX domain of FOXP2 and NFATC2 revealed that they only form stable complexes in the presence of a specific DNA sequence. NFATC2 is expressed in several organs including the brain (Vihma *et al.* , 2016), hence it is possible that it interacts with the FOXPs in neural sites of co-expression to regulate brain development. However, to date the functional consequences of the interaction FOXP-NFATC2 have only been studied for FOXP3 in the context of the immune system (Wu *et al.* , 2006).
A large-scale Y2H screen first suggested the neuron-specific transcription factor TBR1 as a putative interactor of FOXP2 (Sakai *et al.* , 2011). Interaction of TBR1 with FOXP2 interaction was later confirmed using a BRET assay (Deriziotis *et al.* , 2014a). The interaction involves the FOXP2 region encompassing amino acid residues 122-258, and the DNA-binding domain (T-box) of TBR1 (Deriziotis *et al.* , 2014a). Interestingly, *de novo* heterozygous mutations in TBR1 have been found in subjects that exhibit ASD together with language delay. Etiological mutations in both FOXP2 and TBR1 disrupted the interaction between these transcription factors (Deriziotis *et al.* , 2014a), pointing towards a role for this interaction in the pathophysiology of the distinct yet overlapping neurodevelopmental disorders caused by mutations in these different genes. Together with the fact that the two encoded proteins are expressed in overlapping areas of the brain, these findings suggest that the interaction FOXP2-TBR1 may be important for brain development.

A large Y2H screen identified several interactors of murine FoxP1: Tle6, Tcf3, Phf2, Per2, Per1, Smad4 (Ravasi *et al.* , 2010). These are all proteins involved in transcriptional regulation, however the biological or physiological function of these putative interactions remains to be explored.

Most of the known FOXP-interacting proteins are transcription-related proteins such as members of the chromatin remodeling complex NuRD or the transcription factors TBR1, NKX2-1, and NFATC2 (Chokas *et al.* , 2010; Zhou *et al.* , 2008; Wu *et al.* , 2006; Deriziotis *et al.* , 2014a). But over their lifespan the FOXPs also have to interact with other classes of proteins, such as chaperones to be properly folded, or nuclear import proteins. Another class of proteins that are likely to interact with the FOXPs are post-translational modification (PTM) enzymes. PTMs tightly regulate many aspects of transcription factor function,

such as protein stability, subcellular localization, transcriptional activity, DNA-binding, and protein-interactions (Filtz *et al.* , 2014). Interestingly, large high-throughput screens identified MAPK3 and PIAS3, phosphorylation and sumoylation enzymes, respectively, as FOXP2 interacting-proteins (Corominas *et al.* , 2014; Sakai *et al.* , 2011). However, at the start of this PhD project, these interactions had not been confirmed and their functional roles had not been characterized. Thus, the post-translational mechanisms that may regulate FOXP function remained elusive.

1.2.9 *Transcriptional regulatory networks of FOXP1 and FOXP2*

Mutations in *FOXP1* and *FOXP2* lead to monogenic rare language-related disorders. However, the genetic factors underlying speech and language are multiple and probably interconnected forming complex molecular networks of genes. Therefore, exploring upstream and downstream molecular pathways that regulate and are regulated by these transcription factors, together with their protein-protein interactions provides a sensible entry point to uncover the neurogenetic networks that may contribute to the formation of brain circuits that shape speech and language. Figures 1.4 and 1.5 depict FOXP2 and FOXP1 known molecular networks, respectively.

There have been multiple genome-wide studies of FOXP2 transcriptional targets, which have revealed a large number of potential candidate target genes (Vernes *et al.* , 2007, 2011; Spiteri *et al.* , 2007; Nelson *et al.* , 2013). Many of the genes regulated by FOXP2 belong to networks implicated in neuronal-specific processes such as neurite outgrowth and synaptic plasticity (Vernes *et al.* , 2011; Spiteri *et al.* , 2007), but for most transcriptional targets the physiological relevance is unclear and only a subset have been further confirmed and characterized.

Perhaps the most well-established FOXP2 target gene reported to date is *CNTNAP2*, which encodes a transmembrane protein expressed in the brain that mediates cell-cell interactions, with important roles in the regulation of the strength and plasticity of synapses (Rodenas-Cuadrado *et al.* , 2014). Multiple heterozygous mutations in *CNTNAP2* have been reported in patients with a range of complex phenotypes including ID, autism and schizophrenia, which can include language difficulties (Poot, 2015; Rodenas-Cuadrado *et al.* , 2014). However, numerous heterozygous genetic disruptions in CNTNAP2 are also

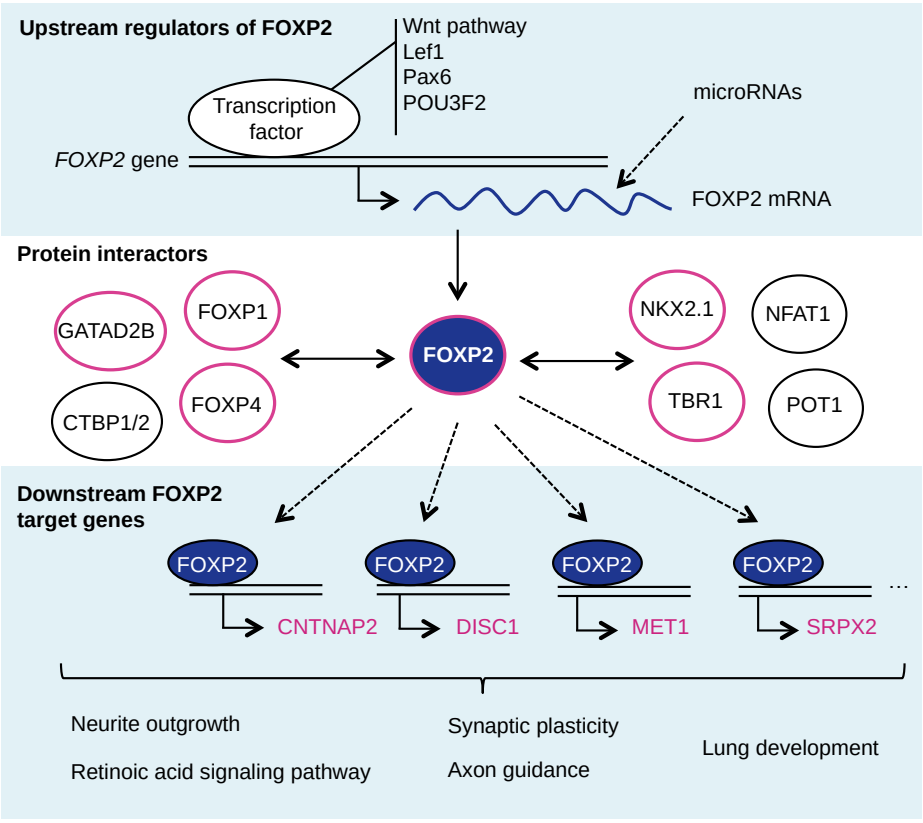


Figure 1.4: FOXP2 molecular networks. Diagram depicting upstream FOXP2 regulators (top part), FOXP2 protein interactors (middle part), and downstream FOXP2 target genes and biological processes in which they are involved (bottom part). Genes in pink or circled in pink have been associated with language-related or neurodevelopmental disorders.

found in the normal population and, therefore, it is not clear whether haploinsufficiency of this gene confers an increased risk for neurodevelopmental disorders. In contrast, homozygous mutations of *CNTNAP2* have been identified in a few sporadic cases of neurodevelopmental disorders that present more severe phenotypes such as cortical dysplasia, focal epilepsy, language regression and severe cognitive impairment (Rodenas-Cuadrado *et al.* , 2016). Further molecular studies of *CNTNAP2* gene function both in humans and mice have indicated that it is involved in neural processes such as synaptic transmission, dendritic arborization or neuronal migration, suggesting important roles in cortical development (Anderson *et al.* , 2012; Peñagarikano *et al.* , 2011).

Other FOXP2 candidate target genes have also been linked to neurodevelopmental disorders, such as *SRPX2*, *uPAR*, *MET* and *DISC1* (Mukamel *et al.* , 2011; Roll *et al.* , 2010; Walker *et al.* , 2012). It would be interesting to further confirm these as FOXP2 target genes and explore whether they may be relevant in the etiology of FOXP2-related CAS cases. Most of the investigations of FOXP1 target genes have focused on non-neuronal tissues. FOXP1 directly regulates the transcription of genes involved in B cell development, monocyte differentiation and macrophage function (Shi *et al.* , 2004, 2008; Hu *et al.* , 2006; Jepsen *et al.* , 2008). In lung tissue, FoxP1 and FoxP2 regulate *PDPN* as well as *SCGB1A1* gene expression (Shi *et al.* , 2004; Shu *et al.* , 2007). FoxP1 also functions as a transcriptional repressor of *Nkx2-5*, a key regulator of cardiomyocyte proliferation during cardiac development (Zhang *et al.* , 2010).

In the brain, FoxP1 regulates Pitx3, a transcription factor required for midbrain dopaminergic neuron differentiation and survival during development (Konstantoulas *et al.* , 2010). Araujo and colleagues performed a broad characterization of FOXP1 downstream target neuronal pathways by coupling RNA-seq and ChIP-seq methodologies using both mice striatal tissue and human neuronal progenitor cells. They found that FOXP1 target genes significantly overlapped with FOXP2-regulated genes in the striatum and, notably, that they were enriched for ASD-related genes (Araujo *et al.* , 2015).

Investigations of factors that regulate FOXP2 and FOXP1 expression and function are scarce. In the zebra fish, FoxP2 expression is regulated by the transcription factors pax6 (Coutinho *et al.* , 2011) and lef1, which is involved in the Wnt signaling pathway (Bonkowsky *et al.* , 2008). Also, manipulation of the Wnt signaling pathway in embryonic chicken retinas affected the expression of FoxP2 (Trimarchi *et al.* , 2009). Another study showed that the neural tran-

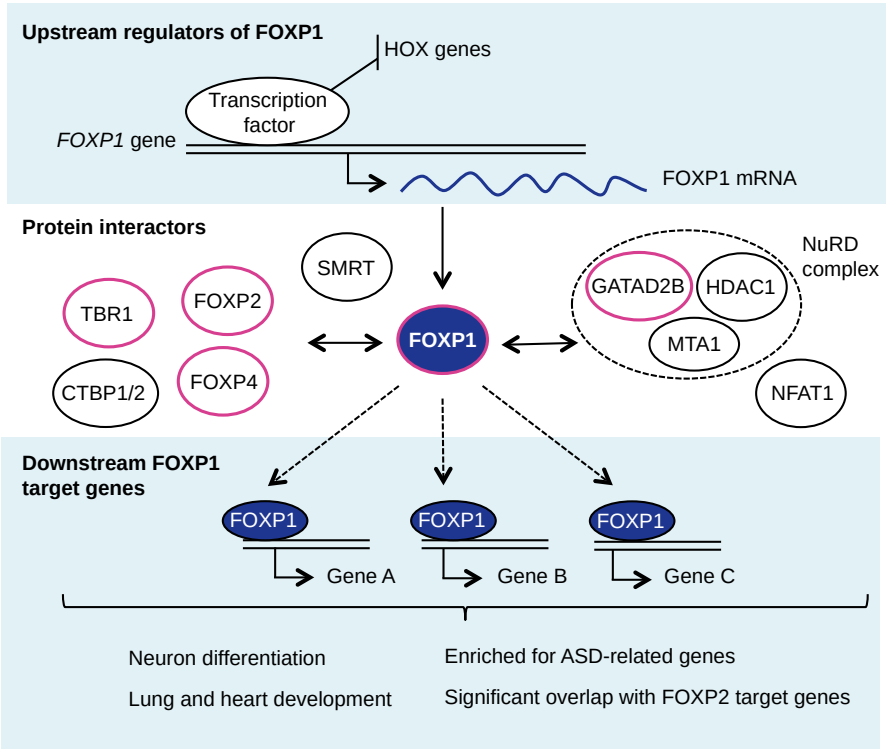


Figure 1.5: FOXP1 molecular networks. Diagram depicting upstream FOXP1 regulators (top part), FOXP1 protein interactors (middle part), and information on downstream FOXP1 target genes (bottom part). Genes in pink or circled in pink have been associated with language-related or neurodevelopmental disorders.

scription factor POU3F2 regulates *FOXP2* in human cell lines (Maricic *et al.* , 2013). Interestingly, the binding site of POU3F2 is affected by a substitution shared by nearly all humans but absent or polymorphic in Neanderthals (Maricic *et al.* , 2013). The amount of FOXP2 protein can also be regulated by post-transcriptional mechanisms. A study by Clovis et al. found that the microRNAs miR-9 and miR-132 repressed the expression of Foxp2 protein by targeting its 3'UTR region. Furthermore, removing the 3'UTR of *Foxp2*, which includes regulatory elements, resulted in an ectopic expression of Foxp2 and led to a significant delay of neurite outgrowth (Clovis *et al.* , 2012). Although all these studies are interesting and potentially relevant to understanding how FOXP2 is regulated, additional replication studies and further investigations are needed to confidently confirm the findings.

1.2.10 *BCL11A mutations in neurodevelopmental disorder*

As noted above during the last two decades researchers have been successfully exploiting *FOXP2* and its molecular networks to gain insights into the molecular basis of speech and language. The vast majority of this knowledge has come from studying the functions of this gene in cellular and animal models. With the rise of next generation sequencing technologies, more monogenic cases of language-related disorders are being uncovered, providing novel candidate genes that, like *FOXP2*, can be further investigated to provide fresh perspectives into the biology of speech and language.

One novel putative language-related gene is *BCL11A*. A microdeletion that solely affects this gene was reported in a child with CAS, the same phenotype that is caused by mutations in *FOXP2* (Peter *et al.* , 2014). In this case, though, the proband also presented additional signs such as mild cognitive deficits and hypotonia. *BCL11A* had previously been associated with neurodevelopmental disorder in multiple cases of individuals suffering from the 2p15p16.1 microdeletion syndrome (OMIM #612513), a rare genetic neurodevelopmental syndrome characterized by ID, dysmorphic features and microcephaly (Balci *et al.* , 2015; Peter *et al.* , 2014; Rajcan-Separovic *et al.* , 2007; Chabchoub *et al.* , 2008; de Leeuw *et al.* , 2008; Fannemel *et al.* , 2014; Félix *et al.* , 2010; Florisson *et al.* , 2013; Hancarova *et al.* , 2013; Hucthagowder *et al.* , 2012; Liang *et al.* , 2009; Piccione *et al.* , 2012; Prontera *et al.* , 2011; Basak *et al.* , 2015). Patients with the 2p16.1-p15 microdeletion syndrome present varying degrees of neu-

rodevelopmental delay and ID together with motor delay, growth retardation, microcephaly and other cranial and skeletal anomalies. Of note, the majority of the patients with this syndrome also show delays in the acquisition of language abilities. *BCL11A* is a reasonable candidate for contributing to the key features of this syndrome because it is a transcription factor expressed in the brain, but the contribution of other genes within the deleted genomic region cannot be discarded. Moreover, there are two non-overlapping critical regions for this syndrome and only one includes *BCL11A* (Lévy *et al.* , 2017).

BCL11A, also known as *CTIP1*, encodes a C2H2 zinc finger transcription factor expressed both in the blood and in the brain, with several splicing isoforms (Satterwhite *et al.* , 2001; Kuo & Hsueh, 2007; Liu *et al.* , 2006), which are described in chapter 5. Its functions in the hematopoietic system have been extensively studied, where it plays a central role in hematopoiesis and in regulating the switch from fetal to adult hemoglobin by acting as a transcriptional repressor of fetal hemoglobin (HbF) (Satterwhite *et al.* , 2001; Sankaran *et al.* , 2010). The presence of chromosomal microdeletions at 2p15-p16.1 encompassing the gene in individuals with speech sound disorder and with more severe ID, as well as the fact that it has been suggested as a recurrently mutated gene in ASD (Peter *et al.* , 2014; Iossifov *et al.* , 2014; De Rubeis *et al.* , 2014; Balci *et al.* , 2015), indicate that this protein can also have a crucial role in the brain.

However, BCL11A functions in the brain have not been as thoroughly investigated as its role as a key hematopoietic transcription factor. BCL11A is expressed in the cortex and the hippocampus of mouse and rat brain (Kuo & Hsueh, 2007; Leid *et al.* , 2004). BCL11A molecular links are depicted in Figure 1.6. This protein interacts and acts as a co-repressor with the neural nuclear receptors: NR2F1, NR2F2 and NR2E1; but it can also independently bind DNA as a transcription factor through its C2H2 zinc finger domains (Avram *et al.* , 2000; Estruch *et al.* , 2012; Chan *et al.* , 2013). At the cellular level, BCL11A is known to regulate dendrite outgrowth and axon branching in cultured hippocampal neurons (Kuo *et al.* , 2009), which is controlled in part by its physical interaction with the ID-related transcription factor CASK (Kuo *et al.* , 2010). In the cortex, BCL11A helps orchestrate the migration and specification of cortical projection neurons by regulating the expression of *SEMA3C* and *TBR1* (Woodworth *et al.* , 2016; Cánovas *et al.* , 2015; Wiegreffe *et al.* , 2015). Moreover, BCL11A directs the acquisition of sensory areas in the developing neocortex and represses motor area identity (Greig *et al.* , 2016).

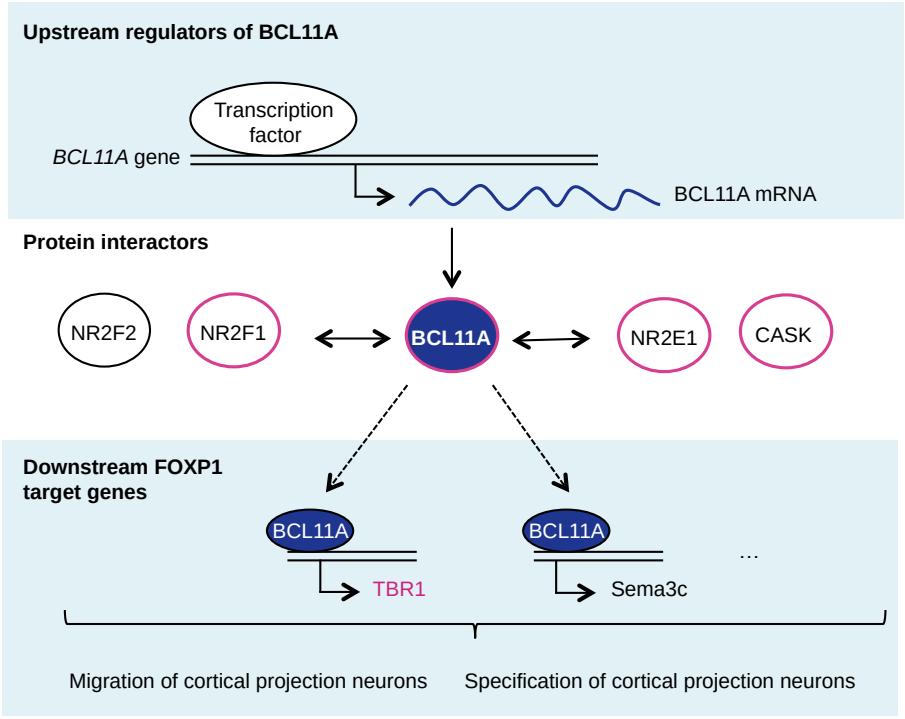


Figure 1.6: BCL11A molecular networks. Diagram depicting upstream BCL11A regulation (top part), BCL11A protein interactors (middle part), and downstream BCL11A target genes and biological the processes in which they are involved (bottom part). Genes in pink or circled in pink have been associated with language-related or neurodevelopmental disorders.

Clearly, BCL11A is an intriguing language-related gene with important roles in the cerebral cortex and therefore has the potential to provide fresh insights into the biology of speech and language.

1.3 AIMS AND SCOPE OF THE THESIS

Language is a complex trait with heterogeneous multifactorial genetic foundations; hence it is challenging to unravel the individual contributions of each single gene to its development and functions. A sensible entry point to start dissecting the crucial molecular pathways is to look at monogenic language-related disorders.

In this dissertation, my main aim was to shed light on the molecular biology underlying language by investigating the molecular functions of transcription factors that have been implicated in monogenic language-related disorders. To do so, I use two main approaches: (1) The identification and investigation of FOXP protein interaction partners and (2) the functional characterization of *BCL11A* mutations found in cases of neurodevelopmental disorders that include language impairments.

Deciphering the protein-protein interactions of a particular protein helps understand the molecular and physiological function of that protein and at the same time, uncovers the molecular networks underlying particular biological aspects that involve that protein. The major part of this thesis is concerned with finding and characterizing protein-protein interactions of three FOXP proteins: FOXP1, FOXP2 and FOXP4, with a special focus on FOXP2. In **Chapter 2**, I validate the interaction between FOXP2 and the post-translational enzyme PIAS1, which was identified in multiple prior independent Y2H screens but never further investigated. The investigation of this interaction leads to the discovery that FOXP2 is subject to SUMOylation, the post-translational modification of proteins with small ubiquitin-like modifiers (SUMO), which regulates many aspects of protein function. I find that PIAS1, as well as other members of the PIAS family, promotes FOXP2 SUMOylation. Moreover, I show that the most well-studied etiological FOXP2 mutation implicated in speech and language disorder leads to substantially reduced SUMOylation of the encoded protein.

In **Chapter 3**, I seek to validate SFPQ and NONO as interaction partners of FOXP2. These proteins were identified in mass spectrometry screens for FOXP2 interactors previously performed in our group. I find that the interaction of

FOXP2 with SFPQ and NONO is consistently not validated in all the different complementary protein-protein interaction methods employed, suggesting that these putative interactors were false positives in the initial mass spectrometry screens.

The results obtained in Chapter 3 indicate the need for new mass spectrometry screens to search for novel FOXP protein interaction partners. Therefore, in **Chapter 4**, I use affinity purification of FOXP1, FOXP2 and FOXP4 complexes followed by mass spectrometry analysis to generate a list of several new putative FOXP-interacting proteins. I use validation methods to confirm seven FOXP-interacting transcription factors from this screen: SOX5, NR2F1, NR2F2, SATB1, SATB2, YY1 and ZMYM2. I also map the binding regions of these co-factors within FOXP2. Strikingly, most of the newly identified FOXP-interacting partners have well-established roles in brain development and/or are implicated in neurodevelopmental disorders with symptoms that overlap with those resulting from mutations in *FOXP1* and/or *FOXP2*. I confirm that the majority of these transcription factors are co-expressed with *FOXP2* in neuronal subpopulations, suggesting that the interactions may occur in vivo and have physiological relevance. Furthermore, I find that these interactions are disrupted by etiological mutations of *FOXP1* and *FOXP2*. Thus, the findings in this chapter expand the FOXP interactome, revealing that these proteins are part of a broader transcription factor network that underlies brain development and neurodevelopmental disorders.

The functional characterization of gene variants identified in people with a certain disorder is crucial to confidently determine the etiological role of that particular gene. Additionally, it provides new insights into the function of that gene in both health and disease. In **Chapter 5**, I perform functional characterization of several *de novo* missense mutations in BCL11A that were found in individuals with neurodevelopmental delay and impaired language function. By using a wide range of molecular assays, I assess the impact of these mutations on protein function in cellular models and find that they all lead to a loss of protein function thus confirming their causal role in the disorder and contributing to the delineation of a novel monogenic ID syndrome. The results in this chapter also highlight the importance of the N-terminus region of BCL11A in core protein functions such as protein-protein interactions and gene transcription and suggest that transcriptional dysregulation underlies *BCL11A*-associated syndrome.

Finally, in **Chapter 6** I summarize the work of the thesis, discussing how the findings from the four experimental chapters have increased our understanding of the molecular and genetic basis of language-related disorders, and considering the future of this growing field.

2

THE LANGUAGE-RELATED TRANSCRIPTION FACTOR FOXP2 IS POST-TRANSLATIONALLY MODIFIED WITH SMALL UBIQUITIN-LIKE MODIFIERS[1]

Abstract: *Mutations affecting the transcription factor FOXP2 cause a rare form of severe speech and language disorder. Although it is clear that sufficient FOXP2 expression is crucial for normal brain development, little is known about how this transcription factor is regulated. To investigate post-translational mechanisms for FOXP2 regulation, we searched for protein interaction partners of FOXP2, and identified members of the PIAS family as novel FOXP2 interactors. PIAS proteins mediate post-translational modification of a range of target proteins with small ubiquitin-like modifiers (SUMOs). We found that FOXP2 can be modified with all three human SUMO proteins and that PIAS1 promotes this process. An aetiological FOXP2 mutation found in a family with speech and language disorder markedly reduced FOXP2 SUMOylation. We demonstrate that FOXP2 is SUMOylated at a single major site, which is conserved in all FOXP2 vertebrate orthologues and in the paralogues FOXP1 and FOXP4. Abolishing this site did not lead to detectable changes in FOXP2 subcellular localization, stability, dimerization or transcriptional repression in cellular assays, but the conservation of this site suggests a potential role for SUMOylation in regulating FOXP2 activity in vivo.*

1 This chapter has been published as: Estruch, S.B., Graham S.A., Deriziotis P., Fisher S.E. (2016). The language-related transcription factor FOXP2 is post-translationally modified with small ubiquitin-like modifiers. Scientific reports, 6:20911. 10.1038/srep20911.

2.1 INTRODUCTION

Heterozygous disruption of the *FOXP2* gene, which encodes a member of the forkhead box (FOX) family of transcription factors, leads to a rare and severe form of speech and language disorder (MIM 605317 (gene), 602081 (disorder)). This developmental disorder was first described in a three-generation pedigree (the KE family), in which half of the family members have difficulties with learning to make coordinated orofacial movements underlying speech (childhood apraxia of speech), together with wide-ranging impairments in comprehension and production of spoken and written language, but without serious impact on other aspects of cognitive functioning (Lai *et al.* , 2001). All affected members of the KE family carry a missense mutation (R553H) within the FOX DNA-binding domain, which abolishes DNA binding and transcriptional repression by FOXP2 (Lai *et al.* , 2001; Nelson *et al.* , 2013; Vernes *et al.* , 2006). Around twenty further cases of speech/language disorder resulting from *FOXP2* haploinsufficiency have since been reported, including nonsense and frameshift point mutations, as well as chromosomal rearrangements disturbing the locus (Adegbola *et al.* , 2015; Nazaryan *et al.* , 2014; Turner *et al.* , 2013; Utine *et al.* , 2014). Thus, adequate FOXP2 expression appears to be essential for normal development of language-related brain circuits, presumably in order to establish correct expression levels of crucial downstream target genes involved in processes such as neurite outgrowth and synaptic plasticity (Groszer *et al.* , 2008; Vernes *et al.* , 2011).

FOXP2 shows evolutionarily-conserved expression in brain structures including the cortex, basal ganglia, thalamus and cerebellum (Ferland *et al.* , 2003; Lai *et al.* , 2003). Studies in animal models further support the notion that precisely controlled levels of FOXP2 are necessary for normal brain development (as reviewed in (Fisher & Scharff, 2009; French & Fisher, 2014)). Mice in which both copies of the *Foxp2* gene have been disrupted show severe motor impairments and developmental delay, and die 3-4 weeks after birth (French & Fisher, 2014; Groszer *et al.* , 2008) (Note that the murine orthologue of *FOXP2* is designated as *Foxp2* and orthologues in other species as *FoxP2*). When mice are heterozygous for a *Foxp2* mutation equivalent to that found in the KE family, they are overtly normal, but exhibit deficits in motor skill learning and abnormal electrophysiology within cortico-striatal circuits (French & Fisher, 2014; Groszer *et al.* , 2008; French *et al.* , 2012). Transient manipulation of *Foxp2* levels also has deleterious

consequences in the developing mouse brain: both knock-down and overexpression of *Foxp2* have been reported to affect neurogenesis, neuronal morphology and migration (Garcia-Calero *et al.* , 2016; Tsui *et al.* , 2013). The effect of manipulating *FoxP2* levels has also been investigated in the zebra finch, a species which, like humans, has the unusual ability to learn vocalizations from other individuals (Fisher & Scharff, 2009). Both knock-down and overexpression of *FoxP2* in key parts of the brains of juvenile zebra finches disrupts the normal process of song learning (Haesler *et al.* , 2007; Heston & White, 2015), indicating that precise control of FoxP2 levels is necessary for normal vocal learning behaviour in this species.

Studies in songbirds indicate that, in addition to the spatial regulation of expression, dynamic temporal regulation of activity of this transcription factor is important for its functions in the developing and adult brain. Zebra finch *FoxP2* expression is elevated in a specific song-related brain region (Area X) during the critical period in which juvenile birds learn their song (Haesler *et al.* , 2007; Rochefort *et al.* , 2007). Furthermore, in adult birds, *FoxP2* levels in Area X decrease when males practice songs alone, but not during performance of songs to females, which may contribute to the increased variability in song output during solo practice compared to female-directed singing (Miller *et al.* , 2008; Murugan *et al.* , 2013; Teramitsu & White, 2006; Thompson *et al.* , 2013). Similar dynamic regulation of the human orthologue could potentially play a role in vocal learning during speech acquisition.

While several studies have examined patterns of FOXP2 protein expression in the brains of different species (Ferland *et al.* , 2003; Lai *et al.* , 2003; Mendoza *et al.* , 2015), few investigations have addressed potential mechanisms for regulation of FOXP2 activity (Fu *et al.* , 2014). Transcription factor activity is often regulated via interaction with other transcription factors, co-repressors/co-activators, chromatin-modifiers, and post-translational modification enzymes. Such interactions can alter protein turnover, increase or decrease transcriptional activation/repression activity, or influence selection of downstream targets. A small number of FOXP2-interacting proteins have been described, notably the paralogues FOXP1 and FOXP4, the transcription factor TBR1, and the co-repressor CtBP1 (Deriziotis *et al.* , 2014b,a; Li *et al.* , 2004; Lozano *et al.* , 2015). To uncover additional mechanisms for regulation of FOXP2 activity we sought to identify novel protein interaction partners. We found that members of the PIAS family of proteins interact with FOXP2, and also with the paralogue FOXP1.

PIAS proteins mediate post-translational modification of nuclear proteins with small ubiquitin-like modifiers (SUMOs) (Meulmeester & Melchior, 2008; Rytinki *et al.* , 2009). SUMOs are ubiquitously-expressed polypeptides that are reversibly coupled to many different proteins with a variety of functional outcomes. We show that FOXP2 is modified with SUMOs at a single major evolutionarily-conserved site, and that PIAS1 promotes this modification. SUMOylation of FOXP2 is an excellent candidate mechanism for dynamic regulation of FOXP2 activity in vivo.

2.2 MATERIAL AND METHODS

2.2.1 *Yeast two-hybrid assay*

The yeast two-hybrid assay was performed by Dualsystems Biotech AG (Switzerland). The bait construct was produced by fusing the coding sequence of full-length human FOXP2 (Uniprot accession O15409) to the DNA-binding domain of the bacterial transcription factor LexA. Preys consisted of a human foetal brain cDNA library fused to the activation domain of yeast Gal4. LacZ was used as the reporter gene and interactions identified by the presence of blue colour. It was confirmed that transfection of the FOXP2 bait construct alone did not activate transcription of the reporter gene. False positive interactors were removed using the bait dependency test to identify prey constructs which activated transcription without co-transfection with the FOXP2 bait construct.

2.2.2 *DNA constructs*

The coding sequences of PIAS1 (NM_016166.1), PIAS2 (NM_004671.3), PIAS3 (NM_006099.3), PIAS4 (NM_015897.1), SUMO1 (NM_003352.4), SUMO2 (NM_006937.3), and SUMO3 (NM_001286416.1), and a 1146 bp region of the promoter of SRPX2, were amplified from human foetal brain cDNA using the primers listed in Supplementary Table S2.1. The cloning of wild-type FOXP2, FOXP1 and CtBP1, and of synthetic truncated forms of FOXP2, has been described previously (Deriziotis *et al.* , 2014a,b; Lozano *et al.* , 2015). For expression of fusion proteins with *Renilla* luciferase, YFP and mCherry, cDNAs were subcloned into the pLuc, pYFP and pmCherry expression vectors, respec-

tively, which have been described previously (Deriziotis *et al.* , 2014b,a). For expression of proteins with three tandem N-terminal Myc tags or an N-terminal V5 tag, cDNAs were subcloned into vectors which were created by modification of the vector pEGFP-C2 (Clontech), and have an identical backbone to the pLuc, pYFP and pmCherry vectors, with the exception of the N-terminal tag and polylinker. To generate the FOXP2-UBC9 fusion protein, the UBC9 coding sequence (NM_194260.2) plus 58 upstream nucleotides were fused to the 3' end of the FOXP2 coding sequence in the V5-tag vector, removing the FOXP2 stop codon. The FOXP2 K674R and R553H mutants, PIAS1 C350S mutant, and SUMO alanine mutants were generated by site-directed mutagenesis using the Quick-Change Site-Directed Mutagenesis kit (Stratagene) following the manufacturers protocol. Primers used in site-directed mutagenesis are listed in Supplementary Table S2.2. The SRPX2 luciferase reporter plasmid was generated by subcloning a 1146 bp region of the SRPX2 promoter into the promoterless firefly luciferase vector pGL4.23 (Promega). All constructs were verified by Sanger sequencing. Plasmid sequences are available upon request.

2.2.3 *Cell culture and transfection*

HEK293 were obtained from ECACC (cat. no. 85120602) and cultured in DMEM supplemented with 10% foetal bovine serum. HeLa-SUMO3 cells, which stably express SUMO3 with an N-terminal hexahistidine tag (Tatham *et al.* , 2009), and the parental HeLa cell line, were kindly provided by Professor Ronald Hay and Dr Michael Tatham, and were cultured in DMEM supplemented with 10% foetal bovine serum (with the addition of 5 μ M puromycin for the HeLa-SUMO3 cell line). Transfections were performed using GeneJuice (Merck-Millipore) according to the manufacturer's instructions.

2.2.4 *BRET assay*

BRET assays were performed as described (Deriziotis *et al.* , 2014b).

2.2.5 *Fluorescence microscopy*

HEK293 cells were seeded on coverslips coated with poly-L-lysine. Cells were cultured for 30 h post-transfection, and then fixed with methanol and nuclei were stained with Hoechst 33342. Fluorescence images were acquired using a Zeiss LSM510 confocal microscope with LSM Image Software or a Zeiss Axio Imager 2 upright fluorescence microscope with ApoTome.2 using ZEN Image software.

2.2.6 *Gel shift assay*

HEK293 cells were transfected in 6-well plates and cultured for 48 h. Cells were lysed in 300 μ L of Laemmli sample buffer containing 10% Tris(2-carboxyethyl)-phosphine hydrochloride (TCEP) and incubated for 10 min at 95° C. Proteins were resolved on 10% SDS-polyacrylamide gels and transferred to PVDF membranes using a TransBlot Turbo Blotting apparatus (Bio-Rad). Membranes were blocked in PBS containing 0.5% milk and 0.1% Tween-20 and incubated overnight at 4° C with primary antibody. The following antibodies were used: anti-GFP (Clontech cat. no. 632380, 1:8000, for YFP constructs); anti-mCherry (Novus cat. no. NBP1-96751, 1:1000); anti-V5 tag (Genetex cat. no. GTX42525, 1:3000); anti-Myc tag (Abcam cat. no. ab9106, 1:1000); anti-β -actin (Sigma cat. no. A5441, 1:10,000). After washing, membranes were incubated with horseradish peroxidase-conjugated goat anti-mouse or anti-rabbit IgG for 45 min at room temperature. Proteins were visualized using Novex ECL Chemiluminescent Substrate Reagent Kit (Life Technologies) and a ChemiDoc XRS+ imaging system (Bio-Rad). Densitometry was performed using the Chemidoc XRS+ System image analysis software (Bio-Rad).

2.2.7 *Pull-down assay*

Purification of SUMO3 conjugates from HeLa cells stably expressing His-tagged SUMO3 was performed as previously described (Tatham *et al.* , 2009). Briefly, cells were seeded in 6-well plates and transfected with YFP-tagged FOXP2 variants or YFP alone. After 48 h cells were lysed in 6 M Guanidinium-HCl, 10 mM Tris, 100 mM Sodium phosphate buffer pH 8.0, 5 mM β -mercaptoethanol

and 5 mM imidazole. An aliquot of the lysate (10%) was retained as the input sample and the remainder was incubated with His-tag Dynabeads (Life Technologies) overnight at 4° C with rotation. Beads were washed with 8 M Urea, 10 mM Tris pH 6.3, 100 mM sodium phosphate buffer, 0.1 % Triton x-1000 and 5 mM β-mercaptoethanol. SUMO3 conjugates were eluted by incubation at room temperature for 20 minutes in 200 mM imidazole, 150 mM Tris pH 8.0, 5% SDS, 30% glycerol, 720 mM β-mercaptoethanol and 0.0025% bromophenol blue. Western blotting was performed as described above; His-tagged SUMO3 conjugates were detected using an anti-His tag antibody (Abgent cat. no. AM1010a, 1:1000).

2.2.8 *Protein degradation assay*

HEK293 cells were transfected in 6-well plates and cultured for 48 h. Cycloheximide was added for the indicated times at a final concentration of 50 μ g/ml. Cells were lysed for 10 min at 4° C with 100 mM Tris pH 7.5, 150 mM NaCl, 10 mM EDTA, 0.2% Triton X-100, 1% PMSF, and protease inhibitor cocktail. Cell lysates were cleared by centrifugation at 10,000x*g* for 3 min at 4° C. Gel electrophoresis, western blotting and densitometry were performed as described above.

2.2.9 *Fluorescence measures of protein expression levels*

HEK293 cells were transfected with YFP-FOXP2 and mCherry in clear-bottomed black 96-well plates in triplicate. Cells were cultured in a TECAN M200PRO microplate reader at 37° C with 5% CO_2. Fluorescence intensity measurements were taken at multiple time points. For each well and time point, the background-subtracted YFP intensity was divided by the background-subtracted mCherry intensity. Triplicate conditions were averaged.

2.2.10 *Luciferase assays*

HEK293 cells were seeded in clear-bottomed white 96-well plates and transfected in triplicate. For the SV40 assay, cells were transfected with 12 ng of pGL3-promoter firefly luciferase reporter construct (Promega), 5 ng of pRL-TK

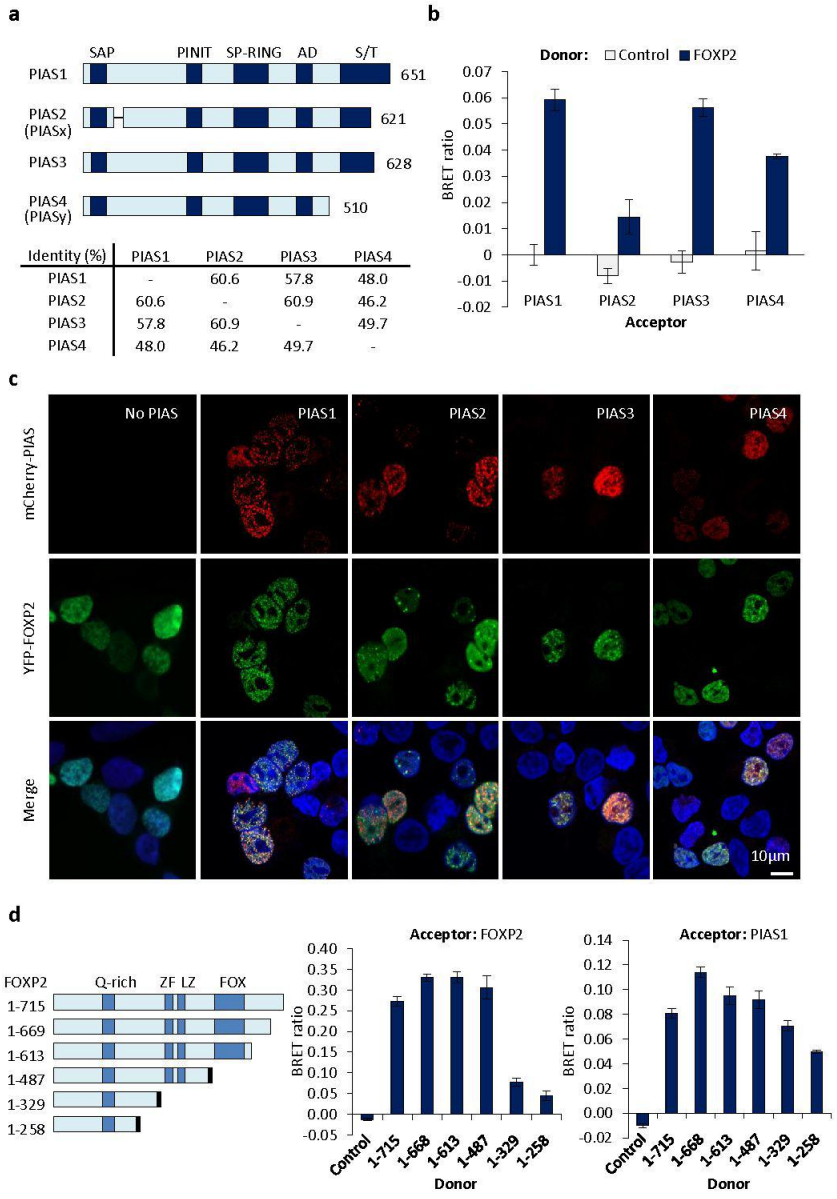
Renilla luciferase normalization control (Promega), and 16 ng of YFP-FOXP2 (wild-type or K674R or R553H mutant) or YFP control construct. For the SRPX2 assay, cells were transfected with 4.3 ng of SRPX2 luciferase reporter construct, 5 ng of pGL4.74 *Renilla* luciferase normalization control (Promega), and 45 ng of YFP-FOXP2 (wild-type or K674R or R553H mutant) or YFP control construct. After 48 h, luciferase activity was measured in a TECAN F200PRO microplate reader using the Dual-Luciferase Reporter Assay system (Promega).

2.3 RESULTS

2.3.1 *FOXP2 interacts with members of the PIAS family of proteins*

To identify candidate interaction partners of FOXP2, a screen of a human foetal brain yeast two-hybrid library was conducted using the full-length human protein as bait. The most frequently observed prey in this screen was PIAS1 (Supplementary Table S2.3). The vertebrate PIAS family includes four proteins with conserved domain architecture and 45-60% sequence identity (Fi gure 2.1a). Interestingly, PIAS3 was one of four proteins identified as candidate FOXP2 interactors in an independent screen of a human foetal brain yeast two-hybrid library, also using full-length human FOXP2 as the bait (Sakai *et al.* , 2011).PIAS proteins are known to interact with and modulate the activity of a range of transcription factors (Kotaja *et al.* , 2002; Rytinki *et al.* , 2009). Members of the PIAS family therefore appeared to be strong candidates for FOXP2 interaction partners.

In order to confirm the interaction of FOXP2 with PIAS proteins, we used a Bioluminescence Resonance Energy Transfer (BRET) assay, which allows protein-protein interactions to be monitored in live mammalian cells in culture (Deriziotis *et al.* , 2014b). In the BRET assay, a protein of interest is expressed as a fusion with *Renilla* luciferase (Luc) and a candidate interaction partner is expressed as a fusion with yellow fluorescent protein (YFP). An interaction between the two proteins brings the Luc and YFP moieties into sufficient proximity to allow resonance energy transfer to occur upon addition of a luciferase substrate, shifting the wavelength of the emitted light from 480 nm to 530 nm. Using Luc-FOXP2 and YFP-PIAS fusion proteins, we confirmed that FOXP2 interacts with PIAS1 and PIAS3, and in addition demonstrated interaction with PIAS4, and a poten-



Identity (%)	PIAS1	PIAS2	PIAS3	PIAS4
PIAS1	-	60.6	57.8	48.0
PIAS2	60.6	-	60.9	46.2
PIAS3	57.8	60.9	-	49.7
PIAS4	48.0	46.2	49.7	-

Figure 2.1

Figure 2.1: FOXP2 interacts with members of the PIAS family of SUMO E3 ligases(a) Top: schematic representation of human PIAS proteins. Domains are shaded in dark blue: SAP domain (SAP); PINIT domain (PINIT); SP-RING domain (SP-RING); acidic domain (AD); serine/threonine- rich domain (S/T). The number of amino acid residues is shown to the right of the schematic. Bottom: identity matrix for PIAS proteins. **(b)** BRET assay for interaction between FOXP2 and PIAS proteins. HEK293 cells were transfected with luciferase-FOXP2 (donor) and YFP-PIAS (acceptor). The control donor protein is a nuclear-targeted luciferase. Values are mean corrected BRET ratios ± S.E.M. (n = 3) **(c)** Fluorescence micrographs of HEK293 cells transfected with mCherry-PIAS (red) and YFP-FOXP2 (green). Nuclei were stained with Hoechst 33342 (blue). **(d)** Left panel: Schematic representation of synthetic truncated forms of FOXP2. The number of amino acid residues is shown on the left; 1-715 represents the full-length protein. Known domains are shown in dark blue: glutamine-rich region (Q-rich); zinc finger (ZF); leucine zipper (LZ); forkhead domain (FOX). A nuclear-targeting signal (shown in black) was appended to the C-terminus of variants 1-487, 1-329, and 1-258 because these variants lack endogenous nuclear targeting signals. Centre panel: BRET assay for interaction between synthetic FOXP2 truncations and wild-type FOXP2. HEK293 cells were transfected with luciferase-FOXP2 truncations (donor) and YFP-FOXP2 (acceptor) Right panel: BRET assay for the interaction between synthetic FOXP2 truncations and PIAS1. HEK293 cells were transfected with luciferase-FOXP2 truncations (donor) and YFP-PIAS1 (acceptor). The control donor protein is a nuclear-targeted luciferase. Values are mean corrected BRET ratios ± S.E.M. (n = 3).

tial weaker interaction with PIAS2 (Figure 2.1b). Note that all experiments were performed using HEK293 cells, unless indicated otherwise.

We noted that PIAS proteins exhibited nuclear localization with a distinctive speckled appearance, as has been reported previously (Kotaja *et al.* , 2002). We therefore examined if co-expression of PIASs with FOXP2 would cause redistribution of FOXP2, which normally exhibits a diffuse localization within nuclei. Expression of PIAS1 together with FOXP2 caused a dramatic change, involving extensive co-localization of FOXP2 with PIAS1 in nuclear speckles (Figure 2.1c). A similar effect was observed upon expression of PIAS3 and PIAS4, consistent with the interaction observed in the BRET assay. Upon overexpression of PIAS2, FOXP2 retained a largely diffuse nuclear distribution, with little FOXP2 exhibiting co-localization with PIAS2 within speckles, consistent with the lower level of interaction observed between FOXP2 and PIAS2 in BRET experiments. The interaction between FOXP2 and PIASs may therefore draw FOXP2 into nuclear speckles.

To try to identify the region of FOXP2 involved in binding to PIAS proteins, we performed BRET assays using a series of synthetic, truncated versions of FOXP2 (Deriziotis *et al.* , 2014a) (Figure 2.1d). These truncations appear to be effective in mapping interaction sites because deletion of the region containing residues 330-487, which contains the leucine zipper dimerization domain (Li *et al.* , 2004), results in a substantial reduction in interaction with full-length FOXP2 (Figure 2.1, centre). Notably, even the shortest FOXP2 truncation tested (residues 1-258) retained the ability to interact with PIAS1, though perhaps to a slightly lesser degree than the full-length protein (Figure 2.1d, right). These results suggest that some key determinants of PIAS binding reside within the N-terminal region of FOXP2. Apart from a polyglutamine tract, this region does not contain any known domains, but it does include regions of polypeptide that are highly conserved in FOXP1 and FOXP4, and has also been identified as the region interacting with the autism-related transcription factor TBR1, suggesting that this region may coordinate multiple protein-protein interactions (Deriziotis *et al.* , 2014a).

2.3.2 *FOXP2 is SUMOylated*

PIAS proteins function as SUMO E3 ligases, promoting the transfer of SUMO from the SUMO-conjugating enzyme UBC9 to an acceptor lysine residue in a target protein, in a manner analogous to the transfer of ubiquitin to proteins by ubiquitin E3 ligases (Meulmeester & Melchior, 2008; Rytinki *et al.* , 2009). There are three SUMO proteins in vertebrates, SUMO1, SUMO2 and SUMO3, all of which have a molecular weight of 11 kDa. SUMO2 and SUMO3 have ~ 95% amino acid sequence identity and are thought to be functionally very similar, whereas SUMO1 has only ~ 50% amino acid sequence identity with SUMO2/3 and is not functionally redundant with these proteins (Meulmeester & Melchior, 2008) The SUMOylation of specific proteins is typically difficult to detect due to the dynamic nature of the modification, which is readily removed by SUMO-specific proteases of the SENP family, and the fact that only a minor proportion of target protein molecules carry a SUMO moiety at any one time (Meulmeester & Melchior, 2008; Rytinki *et al.* , 2009). To facilitate detection of protein SUMOylation, the 18 kDa SUMO-conjugating enzyme UBC9 can be fused to a target protein of interest (Jakobs *et al.* , 2007). We therefore generated

a FOXP2-UBC9 fusion construct, which also carries a V5 epitope tag to enable the fusion to be detected independently from endogenous FOXP2 (Figure 2.2a).

We transfected cells with FOXP2-UBC9 together with a YFP-fusion of SUMO1, SUMO2 or SUMO3. The use of YFP-tagged SUMOs allows discrimination between proteins modified with endogenous and exogenous SUMO. Lysates of transfected cells were probed by western blotting with anti-V5 antibody to detect any shift in the migration of FOXP2-UBC9 resulting from SUMOylation (Figure 2.2b). All samples contained a FOXP2-UBC9 species that migrated at ~ 110 kDa, representing unSUMOylated protein. Cells transfected with FOXP2-UBC9 and a YFP-SUMO contained a ~ 170 kDa FOXP2-UBC9 species that was not present in control cells transfected with FOXP2-UBC9 and YFP, indicating that FOXP2 can be SUMOylated with all three SUMOs (Figure 2.2b). To confirm that the ~ 170 kDa species represents YFP-SUMO conjugated to FOXP2-UBC9, we generated mutant forms of YFP-SUMO, in which the two C-terminal glycine residues required for conjugation to target proteins (and to UBC9) were mutated to alanine. As expected, cells transfected with FOXP2-UBC9 together with an alanine mutant YFP-SUMO did not contain the ~ 170 kDa species (Figure 2.2b). In cells transfected with FOXP2-UBC9 together with a mutant YFP-SUMO or YFP alone, a ~ 130 kDa FOXP2-UBC9 species was observed, which may represent FOXP2-UBC9 modified with endogenous SUMO (Figure 2.2b). Note that the observed molecular weights of the different FOXP2-UBC9 species do not necessarily correspond with theoretical values because SUMOylated proteins are branched polypeptides and exhibit anomalous migration.

In order to confirm the results of the gel shift assay, we also examined FOXP2 SUMOylation in a BRET assay using Luc-FOXP2 with YFP-SUMO fusion proteins. Interaction of FOXP2 was observed with SUMO1, SUMO2 and SUMO3, but not with their respective alanine mutants, in agreement with the gel shift assay (Figure 2.2c). Notably, the BRET assay readily allows detection of FOXP2 SUMOylation without the need to fuse FOXP2 to UBC9, indicating that the technique is highly sensitive for monitoring SUMOylation. BRET has rarely been used in studies of SUMOylation, but may be a widely applicable technique for examining this modification, with the advantage that its use in live cells overcomes the difficulties in maintaining protein SUMOylation encountered in most experimental procedures.

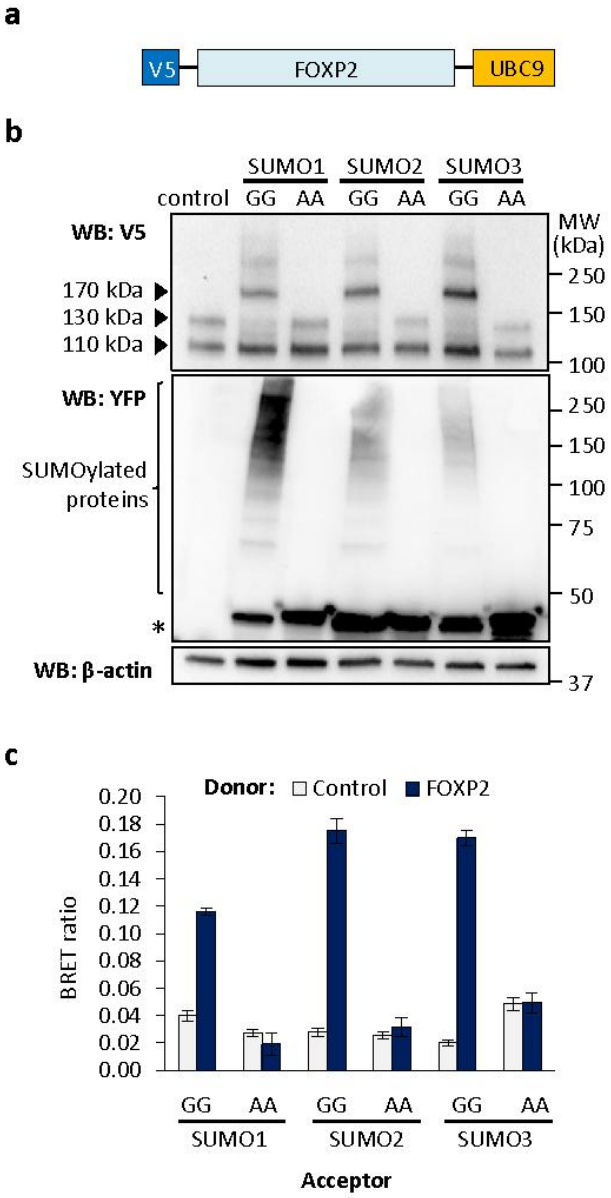

a
V5
FOXP2
UBC9
b
SUMO1
SUMO2
SUMO3
control
GG
AA
GG
AA
GG
AA
WB: V5
MW (kDa)
170 kDa
130 kDa
110 kDa
250
150
100
WB: YFP
SUMOylated proteins
250
150
100
75
50
*
WB: β-actin
37
c
Donor: Control FOXP2
BRET ratio
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
GG
AA
GG
AA
GG
AA
SUMO1
SUMO2
SUMO3
Acceptor

Figure 2.2

Figure 2.2: FOXP2 can be SUMOylated with SUMO1, 2 and 3. (a) Schematic representation of the FOXP2-UBC9 fusion protein with an N-terminal V5 epitope tag. **(b)** Gel shift assay for FOXP2 SUMOylation. FOXP2-UBC9 was expressed in HEK293 cells together with a YFP-fusion of either wild-type SUMO (GG), or mutant SUMO in which the two C-terminal glycine residues required for conjugation to the target protein were mutated to alanine (AA), or with YFP alone (control). Top panel: western blot probed with anti-V5 antibody to detect FOXP2-UBC9. The 110 kDa species is unmodified FOXP2-UBC9. The 130 kDa species is FOXP2-UBC9 modified with endogenous SUMO. The 170 kDa species is FOXP2-UBC9 modified with YFP-SUMO. Middle panel: western blot probed with anti-YFP antibody. The asterisk indicates unconjugated YFP-SUMO. Higher molecular weight species are cellular proteins modified with YFP-SUMO. Bottom panel: western blot probed with anti-β-actin to confirm equal loading. **(c)** BRET assay for interaction between FOXP2 and SUMO. HEK293 cells were transfected with luciferase-FOXP2 (donor) and YFP-SUMO (acceptor), using either wild-type SUMO (GG) or alanine mutants (AA). The control donor protein is a nuclear-targeted luciferase. Values are mean corrected BRET ratios ± S.E.M. (n = 3).

2.3.3 *PIAS1 promotes SUMOylation of FOXP2*

To determine if PIAS proteins are involved in the SUMOylation of FOXP2, we focused on PIAS1, the PIAS family member identified in the original yeast two-hybrid screen. We tested if overexpression of PIAS1, together with SUMO, would allow SUMOylation of FOXP2 to be detected in a gel shift assay without the need to fuse FOXP2 to UBC9. Overexpression of myc-tagged PIAS1 together with any of the three SUMO proteins (fused to mCherry) gave rise to a new FOXP2 species of ~ 140 kDa, suggesting that PIAS1 is able to stimulate SUMOylation of FOXP2 (Figure 2.3a). Importantly, the observation of a new high molecular weight FOXP2 species in this experiment shows that FOXP2 can be SUMOylated without being fused to UBC9.

To confirm that the increase in SUMOylation was mediated directly by PIAS1, we generated a catalytically inactive version of PIAS1 that has a point mutation (C350S) within the SP-RING domain, which is involved in the recognition of target proteins (Figure 2.3b) (Kahyo *et al.* , 2001; Rytinki *et al.* , 2009). The C350S PIAS1 mutant was unable to interact with FOXP2 in a BRET assay, indicating that the SP-RING domain may be involved in recognition of FOXP2 as a SUMOylation target (Figure 2.3b). The C350S mutant displayed a more

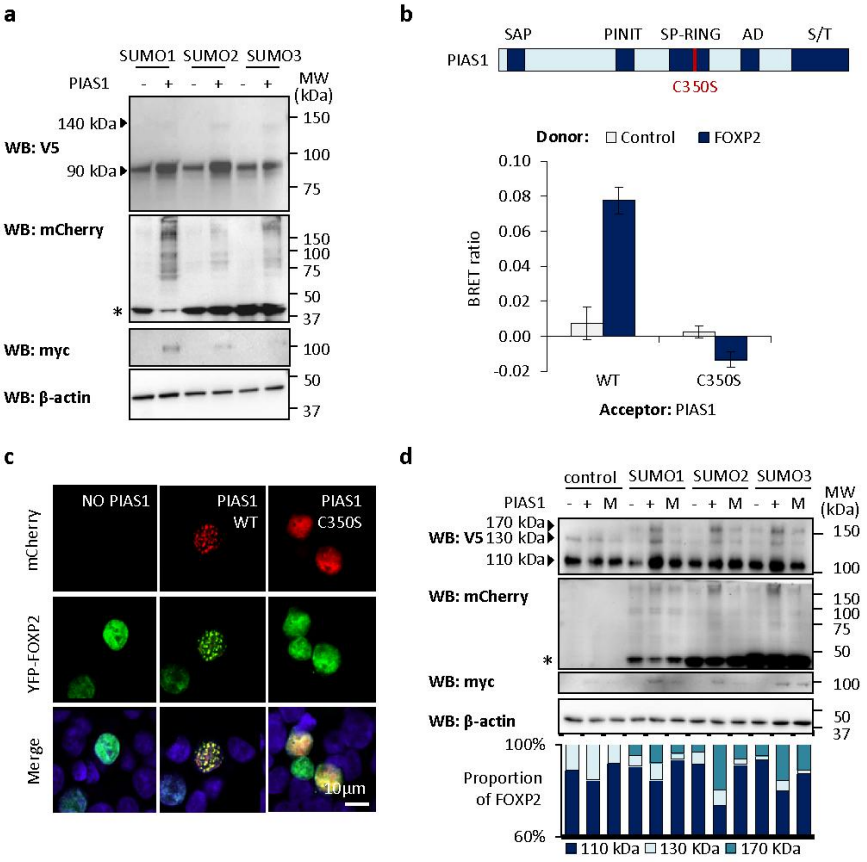


Figure 2.3: PIAS1 promotes FOXP2 SUMOylation. (a) Gel shift assay for SUMOylation of FOXP2. HEK293 cells were transfected with V5-tagged FOXP2 and mCherry-SUMO, together with myc- tagged PIAS1 (+) or an empty vector (-). Top panel: western blot probed with anti-V5 antibody. The 90 kDa species is unmodified FOXP2. The 140 kDa species is FOXP2 modified with mCherry- SUMO. Second panel: western blot probed with anti-mCherry. The asterisk indicates unconjugated mCherry-SUMO. Higher molecular weight species are cellular proteins modified with mCherry- SUMO. Third panel: western blot probed with anti-myc tag to detect PIAS1. Bottom panel: western blot probed with anti-β-actin to confirm equal loading. **(b)** Top: schematic representation of PIAS1 C350S mutant. Bottom: BRET assay for interaction of FOXP2 with PIAS1. HEK293 cells were transfected with luciferase-FOXP2 (donor) and YFP-PIAS (wild-type (WT) or C350S mutant, acceptor). The control donor protein is a nuclear-targeted luciferase. Values are mean corrected BRET ratios ± S.E.M. (n = 3).

Figure 2.3: (Cont.)**(c)** Fluorescence micrographs of HEK293 cells transfected with mCherry-tagged wild-type PIAS1 (WT) or C350S mutant (red) and YFP-FOXP2 (green). Nuclei were stained with Hoechst 33342 (blue). **(d)** Gel shift assay for SUMOylation of FOXP2. HEK293 cells were transfected with FOXP2-UBC9 together with mCherry-SUMO or mCherry alone (control) and myc-tagged wild-type PIAS1 (+), C350S mutant (M) or empty vector (-). Top panel: western blot probed with anti-V5 antibody to detect FOXP2-UBC9. The 110 kDa species is unmodified FOXP2-UBC9. The 130 kDa species is FOXP2-UBC9 modified with endogenous SUMO. The 170 kDa species is FOXP2-UBC9 modified with mCherry-SUMO. Second panel: western blot probed with anti-mCherry antibody. The asterisk indicates unconjugated mCherry- SUMO. Higher molecular weight species are cellular proteins modified with mCherry-SUMO. Third panel: western blot probed with anti-myc to detect PIAS1. Fourth panel: western blot probed with anti-β-actin to confirm equal loading. Bottom panel: densitometry analysis of FOXP2-UBC9 species.

diffuse nuclear localization than wild-type PIAS1, suggesting that the localization of PIAS1 within nuclear speckles is connected to its activity as a SUMO E3 ligase (Figure 2.3c). Consistent with this, the C350S mutant did not induce redistribution of FOXP2 into nuclear speckles (Figure 2.3c). In a gel shift assay, wild-type PIAS1 promoted the modification of FOXP2-UBC9 with both endogenous SUMO (~ 130 kDa species) and mCherry-tagged SUMO (~ 170 kDa species) (Figure 2.3d). The C350S mutant was unable to promote SUMOylation of FOXP2-UBC9 (Figure 2.3d), indicating that the increase in FOXP2 SUMOylation observed upon PIAS1 overexpression is due to the SUMO E3 ligase activity of PIAS1.

2.3.4 *K674 is the major SUMOylation site in FOXP2*

SUMOs are conjugated to target proteins via an isopeptide bond formed by the C-terminal carboxyl group of SUMO and the amino group of a lysine side chain in the target protein. Lysine residues that are subject to SUMOylation are often found within the consensus sequence Ψ KX(D/E), where Ψ is a hydrophobic amino acid and X is any amino acid, although many SUMOylation sites do not conform to this pattern (Geiss-Friedlander & Melchior, 2007). We used three prediction algorithms to identify potential SUMOylation sites in FOXP2: SUMOplot (`www.abgent.com/sumoplot`), GPS-SUMO (`sumosp.biocuckoo.org`)

(Zhao *et al.* , 2014), and JASSA (www.jassa.fr) (Beauclair *et al.* , 2015). All three algorithms identified K674 as a high-confidence potential SUMOylation site (Supplementary Table S2.4). No other lysine residues in FOXP2 were identified as potential SUMOylation sites by more than one algorithm. Residue K674 lies in the C-terminal region of FOXP2, which does not contain any previously described functional domains, consistent with the typical localization of SUMOylation sites within structurally disordered regions of polypeptide (Figure 2.4a) (Diella *et al.* , 2008). This residue is within a VKEE sequence that matches the consensus Ψ KX(D/E) motif (Figure 2.4b). In addition, the putative SUMOylation site at K674 belongs to a class of predicted SUMOylation sites termed KEPE motifs, which have the consensus sequence Ψ KX(D/E)PXXX(D/E) (Diella *et al.* , 2009). KEPE motifs are found in over 130 human proteins, and are enriched among proteins involved in transcription (Diella *et al.* , 2009). The critical residues of the KEPE motif are conserved in vertebrate FOXP2 proteins, supporting a functional role for the motif (Figure 2.4b).

To assess if K674 functions as a SUMOylation site, we mutated this residue to arginine, thus removing the amino group required for SUMO conjugation but preserving the positive charge at this position of the polypeptide. We performed a gel shift assay using wild-type or mutant FOXP2-UBC9 co-transfected with PIAS1 and each of the three SUMOs. In this assay, the ~ 170 kDa SUMOylated FOXP2-UBC9 species observed in cells transfected with wild-type FOXP2 was not present in cells transfected with the K674R mutant (Figure 2.4c). Thus K674 is the major site in FOXP2 that is subject to modification by SUMO1, SUMO2 and SUMO3.

The dramatic reduction in SUMOylation resulting from mutation of residue K674 suggests that this residue may be the most important SUMOylation site in vivo. To detect if a proportion of FOXP2 might be SUMOylated at alternative sites, we employed a HeLa cell line stably expressing His-tagged SUMO3 (Tatham *et al.* , 2009). Using cobalt affinity purification under denaturing conditions, SUMOylated species were purified from HeLa-SUMO3 cells transfected with YFP-tagged wild-type FOXP2, YFP-tagged K674R mutant, or YFP alone (Figure 2.4d). YFP-FOXP2 was detected in the affinity-purified protein fraction from the HeLa-SUMO3 cells, confirming that FOXP2 can be SUMOylated without being fused to UBC9 (Figure 2.4d). Note that in this key experiment we did not detect an observable shift in the molecular weight of FOXP2 as a result of SUMOylation, in contrast to the clear size shifts observed in our earlier exper-

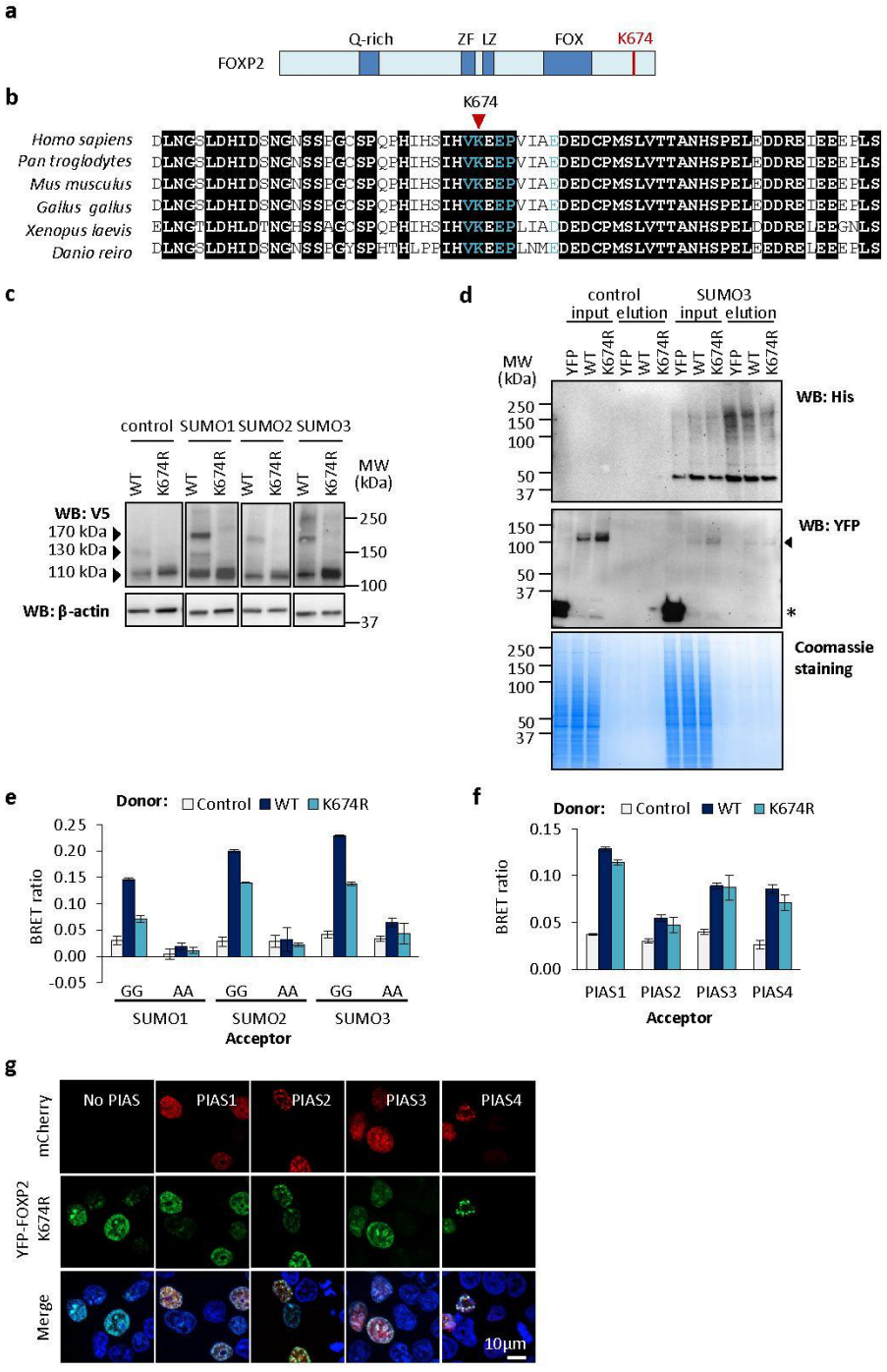

a
Q-rich
ZF
LZ
FOX
K674
FOXP2
b
K674
Homo sapiens
Pan troglodytes
Mus musculus
Gallus gallus
Xenopus laevis
Danio reiro
c
control
SUMO1
SUMO2
SUMO3
WT
K674R
MW (kDa)
WB: V5
170 kDa
130 kDa
110 kDa
WB: β-actin
250
150
100
37
d
control
SUMO3
input
elution
YFP
MW (kDa)
WB: His
WB: YFP
Coomassie staining
e
Donor:
Control
WT
K674R
BRET ratio
GG
AA
SUMO1
SUMO2
SUMO3
Acceptor
f
PIAS1
PIAS2
PIAS3
PIAS4
g
No PIAS
mCherry
YFP-FOXP2
Merge
10µm

Figure 2.4

Figure 2.4: (Cont.) **K674 is the major SUMOylation site in FOXP2. (a)** Schematic representation of FOXP2 showing the predicted SUMOylation site. Known domains are shown in dark blue: glutamine-rich region (Q-rich); zinc finger (ZF); leucine zipper (LZ); forkhead domain (FOX). **(b)** Sequence alignment of the region surrounding the putative SUMOylation site in FOXP2 orthologues. Conserved residues are shown on a black background. Critical residues of the KEPE-type SUMOylation site motif are shown in turquoise. UniProt accession numbers: *Homo sapiens* O15409; *Pan troglodytes* Q8MJ98; *Mus musculus* P58463; Gallus gallus Q5IHK1; Xenopus laevis Q4VYS1; *Danio rerio* Q4JNX5. **(c)** Gel shift assay for FOXP2 SUMOylation. HEK293 cells were transfected with FOXP2-UBC9 (wild-type (WT) or K674R mutant) together with YFP-SUMOs or YFP alone (control). Top panel: western blot probed with anti-V5. The 110 kDa species is unmodified FOXP2-UBC9. The 130 kDa species is FOXP2-UBC9 modified with endogenous SUMO. The 170 kDa species is FOXP2-UBC9 modified with YFP-SUMO. Bottom panel: western blot probed with anti-β-actin. **(d)** Pull-down assay for FOXP2 SUMOylation. HeLa cells stably expressing His-tagged SUMO3, or the parental HeLa cell line (control), were transfected with YFP-FOXP2 (wild-type (WT) or K674R mutant) or YFP alone. His-tagged species were isolated under denaturing conditions using cobalt affinity purification. Western blots of total lysate (input) and affinity-purified material (elution) were probed with anti-His tag antibody to visualize SUMO-conjugated proteins (top panel). YFP and YFP-FOXP2 were visualized using anti-GFP antibody: FOXP2 is indicated by an arrowhead and YFP with an asterisk (middle panel). Total protein was visualized by Coomassie blue staining (bottom panel). **(e)** BRET assay for interaction of FOXP2 with SUMO. HEK293 cells were transfected with luciferase-FOXP2 (wild-type (WT) or K674R mutant, donor) and YFP-SUMO (wild-type (GG) or alanine mutant (AA), acceptor). The control is a nuclear-targeted luciferase. Values are mean corrected BRET ratios ± S.E.M. (n = 3). **(f)** BRET assay for interaction of FOXP2 and PIAS proteins. Cells were transfected with luciferase-FOXP2 (wild-type (WT) or K674R mutant, donor) and YFP-PIAS (acceptor). **(g)** Fluorescence micrographs of HEK293 cells transfected with mCherry-PIAS (red) and YFP-FOXP2 K674R mutant (green). Nuclei were stained with Hoechst 33342 (blue).

iments. One potential explanation of the discrepancy is that this experiment employed His-tagged SUMO, with a molecular weight of only ~ 10 kDa, comparable to that of endogenous SUMO, while our prior experiments used YFP- and mCherry-tagged SUMO proteins, which have substantially higher molecular weights of >35 kDa. Importantly, YFP-FOXP2 was not detectable among proteins eluted from resin incubated with lysate from the parental HeLa cell line, demonstrating that FOXP2 does not bind non-specifically to the affinity resin (Figure 2.4d). Moreover, the YFP control protein was not detectable in the

affinity-purified fraction from the HeLa-SUMO3 cells, demonstrating that only SUMOylated proteins are purified using this procedure (Figure 2.4d).

The K674R mutant was also present in the affinity-purified material from the HeLa-SUMO3 cell line, indicating that specific enrichment of SUMOylated proteins allows detection of rarer forms of FOXP2 that are SUMOylated at one or more alternative sites (Figure 2.4d). Different SUMOylation site prediction tools variously identify potential additional SUMOylation sites at K74, K285, K417 and K560 (Supplementary Table S2.4). However, none of these predictions are consistent across two or more prediction tools, and none lie within a typical consensus SUMOylation motif. While it is possible that several additional sites in FOXP2 may occasionally be SUMOylated, modification at these sites may not serve a critical biological function.

In a BRET assay, the K647R mutant displayed consistently reduced, but not abolished, interaction with wild-type SUMO1, SUMO2 and SUMO3, and no interaction with the respective alanine mutants (Figure 2.4e). The residual interaction between the K674R mutant and SUMOs in the BRET assay might be accounted for by the presence of minor secondary SUMOylation sites. Given that the reduction in the BRET signal resulting from the K674R mutation is modest, there may also be a contribution to this signal from non-covalent association of the mutant with SUMO. Low-affinity, non-covalent interactions between SUMOs and other proteins are mediated by SUMO-interaction motifs (SIMs) (Meulmeester & Melchior, 2008). We employed the JASSA (Beauclair et al., 2015) and GPS-SUMO (Zhao *et al.* , 2014) algorithms to identify potential SIMs in FOXP2, but no high-confidence SIMs were found (Supplementary Table S2.5).

SUMO may instead associate with the K674R mutant as part of a complex with UBC9 and PIAS. In support of this model, the K674R mutant exhibited a similar degree of interaction with PIASs as wild-type FOXP2 in a BRET assay (Figure 2.4f). We also observed co-localization of the K674R mutant with PIASs in nuclear speckles (Figure 2.4g). Therefore K674 is not required for interaction of FOXP2 with PIAS proteins, consistent with the mapping of the PIAS binding site to the N-terminal region of FOXP2 (Figure 2.1d). Furthermore, the relocalization of FOXP2 to nuclear speckles that is observed upon overexpression of PIASs appears to be due to the interaction between PIAS and FOXP2, and not to the SUMOylation of FOXP2.

2.3.5 *Functional consequences of FOXP2 SUMOylation*

SUMOylation can affect the function of transcription factors in several ways (Meulmeester & Melchior, 2008). To identify potential effects of SUMOylation on FOXP2 function, we assessed if the K647R SUMOylation site mutant displayed any altered properties in cellular assays. The mutant did not exhibit any differences in subcellular localization, retaining a diffuse nuclear distribution (Figure 2.5a). To assess differences in protein expression level, we transfected cells with YFP fusions of wild-type FOXP2 or the K674R mutant, and measured fluorescence intensity over time, relative to the fluorescence intensity of co-transfected mCherry. No differences were observed in the ultimate expression level of the wild-type and mutant proteins, or in the time course of induction of expression (Figure 2.5b). To test for differences in protein degradation, cycloheximide was added to cells expressing wild-type FOXP2 or the K674R mutant to arrest protein synthesis, and the decrease in FOXP2 protein over time was monitored by western blotting. For both wild-type and mutant FOXP2, the amount of protein had dropped to approximately 25% of starting levels after 6 h incubation with cycloheximide, and no difference in the rate of degradation was observed between the wild-type and mutant proteins (Figure 2.5c). Abolishing the major SUMOylation site therefore does not have a substantial effect on FOXP2 turnover in cultured cells. It is possible that differences in stability might be evident after longer incubations with cycloheximide, but reliable quantification of FOXP2 is precluded by the low levels of remaining protein.

To determine if SUMOylation might affect the transcriptional regulatory activity of FOXP2, we employed a luciferase reporter assay in which luciferase expression is driven by the SV40 viral promoter (Vernes *et al.* , 2006). As previously reported, wild-type FOXP2 repressed luciferase activity by around 60% , whereas the mutant FOXP2 found in the KE family (R553H), which is unable to bind DNA, did not repress luciferase activity (Figure 2.5d) (Vernes *et al.* , 2006). The K674R mutant did not differ significantly in its repressive capability from the wild-type protein (Figure 2.5d). A luciferase assay was also performed using the human SRPX2 promoter. It has previously been reported that FOXP2 represses transcription from this promoter (Roll *et al.* , 2010). In our assay, wild-type FOXP2 substantially repressed luciferase activity, and the R553H mutant showed loss of repression, but the K674R mutant again did not differ signifi-

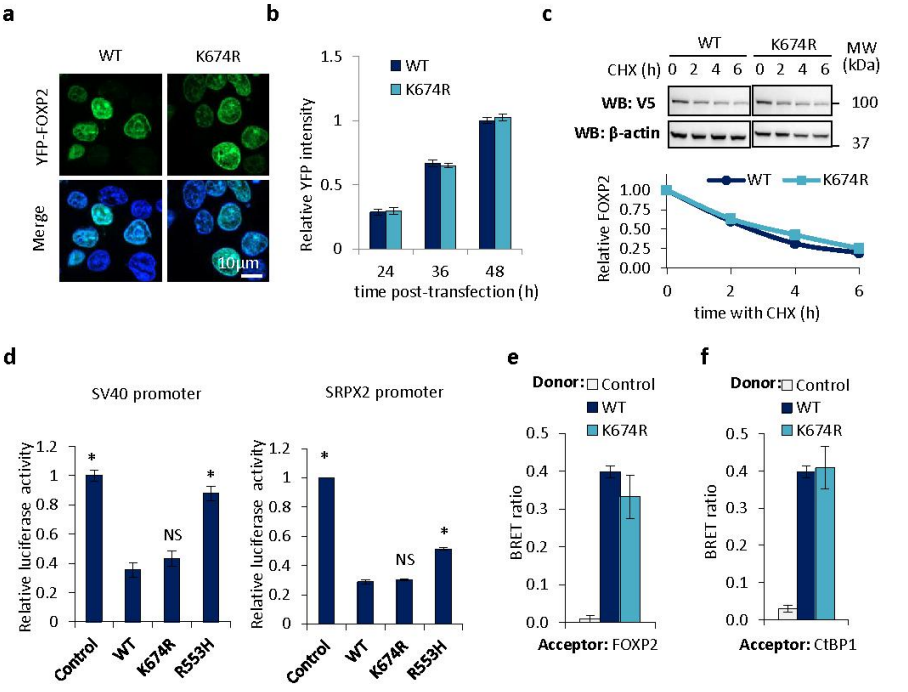


Figure 2.5: Abolishing the K674R SUMOylation site has no effect in cellular assays of FOXP2 function. (a) Fluorescence micrographs of cells transfected with YFP-tagged wild-type (WT) and K674R mutant forms of FOXP2. Nuclei were stained with Hoechst 33342 (blue). **(b)** Fluorescence- based measurement of FOXP2 expression level. HEK293 cells were transfected with YFP-FOXP2 (wild-type (WT) or K674R mutant), together with mCherry for normalization. Fluorescence intensity was measured 24, 36 and 48 h post-transfection. Values are mean YFP/mCherry fluorescence ratios ± S.E.M. (n = 3), relative to the value for wild-type FOXP2 at 48 h.**(c)** Western blot assay for FOXP2 degradation. HEK293 cells were transfected with V5-tagged FOXP2 (wild- type (WT) or K674R mutant). Cycloheximide (CHX) was added to cells 48 h after transfection for varying amounts of time. Top: western blots of whole cell extracts probed with anti-V5 and anti-β- actin antibodies. Bottom: densitometry quantification of FOXP2. Values are normalized to β-actin and plotted relative to the 0 h time point. **(d)** Luciferase reporter assays for transcriptional regulatory activity of FOXP2. HEK293 cells were transfected with a luciferase reporter vector containing the SV40 promoter (left) or the human SRPX2 promoter (right), together with YFP- FOXP2 (wild-type (WT), K674R or R553H mutants), or YFP alone (control). Values are mean relative luciferase activity ± S.E.M. (n = 3),

Figure 2.5: (Cont.) expressed relative to the control. Asterisks indicate significant differences compared to wild-type FOXP2 ($p<0.05$, one-tailed student's t-test). NS, not significant. Exact p-values for the SV40 assay are 0.0043 for the control, 0.0048 for R553H, 0.1598 for K674R. Exact p-values for the SRPX2 assay are 0.0009 for the control, 0.0017 for R553H, and 0.2566 for K674R.**(e)** BRET assay for FOXP2 dimerization. HEK293 cells were transfected with luciferase- FOXP2 (wild-type (WT) or K674R mutant, donor) and YFP-FOXP2 (acceptor). The control is a nuclear-targeted luciferase. Values are mean corrected BRET ratios ± S.E.M. (n = 3). **(f)** BRET assay for interaction of FOXP2 with CtBP1. HEK293 cells were transfected with luciferase-FOXP2 (wild- type (WT) or K674R mutant, donor) and YFP-CtBP1 (acceptor). The control is a nuclear-targeted luciferase. Values are mean corrected BRET ratios ± S.E.M. (n = 3)

cantly in its repressive capability from the wild-type protein (Figure 2.5d). Thus the loss of the SUMOylation site does not have a generalized effect on the repressive capability of FOXP2.

We then looked to see if abolishing the SUMOylation site affects the ability of FOXP2 to form homodimers (Li *et al.* , 2004). The mutant displayed normal dimerization ability in a BRET assay (Figure 2.5e), as might be expected given that FOXP2 dimerization is mediated by the leucine zipper domain, which is not located near to the SUMOylation site (Figure 2.4a) (Li *et al.* , 2004; Deriziotis *et al.* , 2014b). Finally we employed a BRET assay to assess if the K674R mutant differed in its ability to bind to the co-repressor CtBP1 (Li *et al.* , 2004; Deriziotis *et al.* , 2014b). Again, no differences were observed between the wild-type and mutant proteins (Figure 2.5f). Thus, in our cell-based assays, abolishing the major SUMOylation site in FOXP2 does not have a substantial impact on the behaviour of the protein.

2.3.6 *A FOXP2 mutant that causes speech/language disorder shows reduced SUMOylation*

Members of the KE family who are affected by speech and language disorder all carry a heterozygous mutation, R553H, within the DNA-binding domain of FOXP2 (Figure 2.6a) (Lai *et al.* , 2001). This mutation affects a critical residue within the DNA-recognition helix of the FOX domain and abolishes DNA binding (Lai *et al.* , 2001; Nelson *et al.* , 2013; Stroud *et al.* , 2006; Vernes *et al.* , 2006).

Unexpectedly, a gel shift assay in which cells were transfected with wild-type or mutant FOXP2, together with PIAS1 and SUMO, showed a clear reduction in SUMOylation of the R553H mutant, though it was still modified to a greater extent than the K674R mutant (Figure 2.6b). A BRET assay also showed a near total loss of interaction between SUMOs and the R553H mutant in comparison to wild-type FOXP2 (Figure 2.6c), in contrast to the partially retained interaction between SUMOs and the K674R mutant (Figure 2.4d).

Given that the major SUMOylation site is intact in the R553H mutant, the reduction in SUMOylation may be a consequence of reduced interaction between the mutant and components of the SUMOylation machinery, such as PIASs or UBC9. Consistent with this possibility, the R553H mutant showed reduced or abolished interaction with PIASs in a BRET assay (Figure 2.6d). The R553H mutant continued to show some co-localization with PIASs in transfected cells, however the tendency of this mutant to form aggregates makes it unclear if the punctae containing both proteins are nuclear speckles or protein aggregates (Figure 2.6e) (Vernes *et al.* , 2006). Decreased interaction with PIAS1 could account for the relatively higher level of R553H SUMOylation in the gel shift assay compared to the BRET assay, because the overexpression of PIAS1 in the gel shift assay might have a compensatory effect on SUMOylation. Interestingly, the R553H mutant was SUMOylated to a similar extent as wild-type FOXP2 when these proteins were fused to UBC9 (Figure 2.6f), suggesting that fusion to UBC9 might rescue a loss of interaction with the SUMOylation machinery in the R553H mutant.

Thus in contrast to the K674R mutant, which is able to interact with PIASs but cannot be SUMOylated, the R553H mutant has reduced interaction with PIASs, but can still be SUMOylated by employing overexpression of PIAS or fusion to UBC9. It is unexpected that the R553H mutation should reduce interaction with PIASs, because truncated forms of FOXP2 that lack the entire FOX domain, and are thus also unable to bind DNA, are still able to interact with PIAS1 (Figure 2.1d). The partial mislocalization and increased propensity for protein aggregation resulting from the R553H mutation may contribute to the reduction in interaction, although the majority of mutant protein still displays a normal diffuse nuclear localization, and retains the ability to interact with wild-type FOXP2, indicating that the mutation does not cause gross misfolding of the entire population of molecules (Vernes *et al.* , 2006; Deriziotis *et al.* , 2014a). Potentially the R553H mutation causes a conformational change that

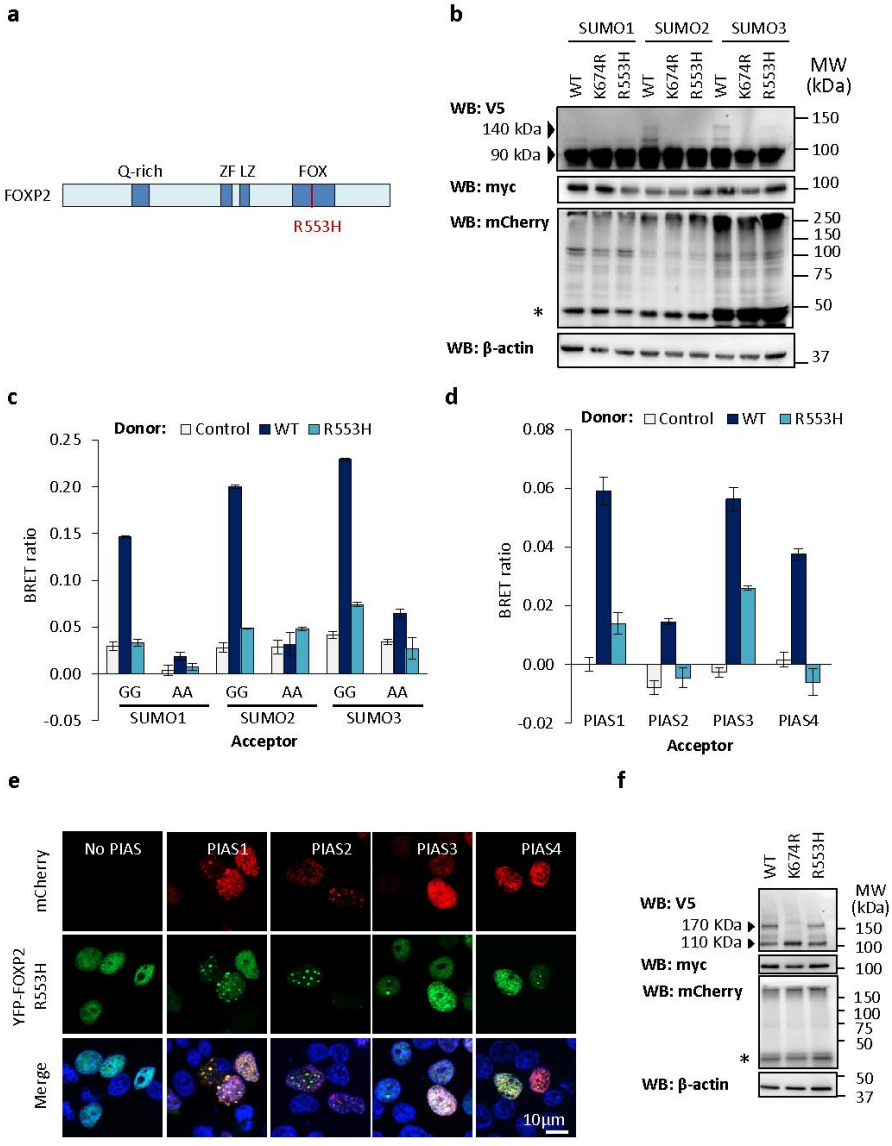


Figure 2.6: The R553H FOXP2 mutant which causes speech/language disorder exhibits reduced SUMOylation. (a) Schematic representation of the FOXP2 R553H mutant. Known domains are shown in dark blue: glutamine-rich region (Q-rich); zinc finger (ZF); leucine zipper (LZ); forkhead domain (FOX).

Figure 2.6: (Cont.) **(b)** Gel shift assay for FOXP2 SUMOylation. HEK293 cells were transfected with V5-tagged FOXP2 (wild-type (WT), K674R or R553H mutant) together with mCherry-SUMO and myc-tagged PIAS1. Top panel: western blot probed with anti-V5 antibody. The 90 kDa species is unmodified FOXP2. The 140 kDa species is FOXP2 modified with mCherry-SUMO. Second panel: western blot probed with anti-myc tag antibody to detect PIAS1. Third panel: western blot probed with anti-mCherry. The asterisk indicates unconjugated mCherry-SUMO. Higher molecular weight species are cellular proteins modified with mCherry-SUMO. Bottom panel: western blot probed with anti-β-actin to confirm equal loading. **(c)** BRET assay for interaction of FOXP2 with SUMO. HEK293 cells were transfected with luciferase-FOXP2 (wild-type (WT) or R553H mutant, donor) and YFP-SUMO (wild-type (GG) or alanine mutant (AA), acceptor). The control donor protein is a nuclear-targeted luciferase. Values are mean corrected BRET ratios ± S.E.M. (n = 3). **(d)** BRET assay for interaction of FOXP2 with PIAS proteins. HEK293 cells were transfected with luciferase-FOXP2 (wild-type (WT) or R553H mutant, donor) and YFP-PIAS (acceptor). The control donor protein is a nuclear-targeted luciferase. Values are mean corrected BRET ratios ± S.E.M. (n = 3). **(e)** Fluorescence micrographs of HEK293 cells transfected with mCherry-PIAS (red) and YFP-FOXP2 R553H mutant (green). Nuclei were stained with Hoechst 33342 (blue). **(f)** Gel shift assay for FOXP2 SUMOylation. HEK293 cells were transfected with FOXP2-UBC9 (wild-type (WT), K674R or R553H mutant) together with mCherry-SUMO3 and myc-tagged PIAS1. Top panel: western blot probed with anti-V5. The 110 kDa species is unmodifed FOXP2-UBC9. The 170 kDa species is FOXP2-UBC9 modified with mCherry-SUMO3. Second panel: western blot probed with anti-myc tag antibody to detect PIAS1. Third panel: western blot probed with anti-mCherry. The asterisk indicates unconjugated mCherry-SUMO. Higher molecular weight species are cellular proteins modified with mCherry-SUMO. Bottom panel: western blot probed with anti-β-actin to confirm equal loading.

blocks the PIAS binding site. Alternatively, the loss of DNA-binding capacity and/or destabilization of the FOX domain resulting from the R553H mutation may permit interactions with other cellular proteins, such as those involved in protein degradation, that in turn interfere with PIAS binding. Although the mechanism by which the R553H mutation reduces interaction with PIAS is unclear, it seems likely that a functional FOX domain is needed in addition to the K674 SUMOylation site to permit normal levels of FOXP2 SUMOylation in cells.

2.3.7 *SUMOylation of other FOXP proteins*

FOXP2 has three mammalian paralogues, FOXP1, FOXP3 and FOXP4 (Figure 2.7a). FOXP1, FOXP2 and FOXP4 exhibit 55-65% sequence identity, are able to form heterodimers, and are expressed in overlapping cell populations in the brain and other organs, suggesting that they may co-operate in the regulation of certain subsets of target genes (Ferland *et al.* , 2003; Li *et al.* , 2004; Lu *et al.* , 2002). FOXP3 is structurally divergent, and its expression is restricted to regulatory T lymphocytes (Fontenot *et al.* , 2003). The critical residues of the KEPE SUMOylation motif in FOXP2 are conserved in FOXP1 and FOXP4, which is particularly striking because the C-terminal regions of these proteins generally exhibit a low level of similarity (the region is absent in FOXP3) (Figure 2.7a,b). This low level of sequence conservation in the C-terminal region of the FOXP proteins is consistent with the polypeptide in this region being structurally disordered, in order to maintain accessibility of the SUMOylation site (Diella *et al.* , 2008).

The conservation of the SUMOylation site in FOXP1 prompted us to assess if this protein may also be subject to PIAS-mediated SUMOylation. The essential role of FOXP1 in brain development has recently come to light since it was discovered that haploinsufficiency of *FOXP1* is associated with intellectual disability, autistic features, expressive speech deficits and dysmorphic features (Hamdan *et al.* , 2010; Horn *et al.* , 2010; Le Fevre *et al.* , 2013; Lozano *et al.* , 2015; O'Roak *et al.* , 2011). The FOXP1-related disorder is more severe than that resulting from haploinsufficiency of *FOXP2*, indicating that the two proteins have non-redundant functions in human brain development (Bacon & Rappold, 2012).Furthermore, and in contrast to aetiological *FOXP2* variants, all aetiological *FOXP1* variants reported to date have occurred de novo (Le Fevre *et al.* , 2013; Lozano *et al.* , 2015). In a BRET assay, FOXP1 and FOXP2 exhibited similar levels of interaction with PIAS1, PIAS2 and PIAS4, but unlike FOXP2, FOXP1 showed little or no interaction with PIAS3 (Fig 7c). BRET assays also showed clear interaction of FOXP1 with all three SUMOs (Fig 7d). It is therefore likely that FOXP1 is also subject to modification by SUMO1, SUMO2 and SUMO3, at an equivalent site to that in FOXP2 (K636), but that the members of the PIAS family have differing levels of importance in the SUMOylation of FOXP1 and FOXP2. Three proteome-wide studies of SUMOylation in human cell lines have also identified FOXP1 and FOXP4 as substrates for SUMOyla-

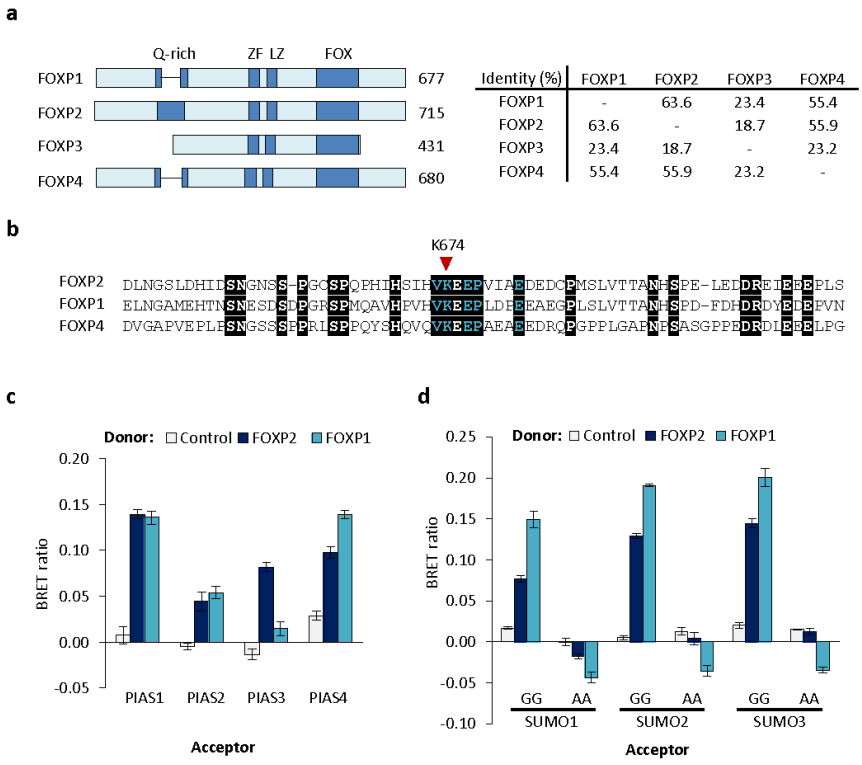


Figure 2.7: The FOXP2 paralogue FOXP1 is also SUMOylated. (a) Left: Schematic representation of the FOXP family of proteins. Known domains are shown in dark blue: glutamine-rich region (Q-rich); zinc finger (ZF); leucine zipper (LZ); forkhead domain (FOX). The number of amino acid residues is indicated to the right of the schematic. Right: identity matrix for FOXP proteins. **(b)** Sequence alignment of the region surrounding the SUMOylation site in FOXP proteins. Conserved residues are shown on a black background. Critical residues of the KEPE-type SUMOylation site motif are shown in turquoise and the SUMO conjugation site is labeled. UniProt accession numbers: FOXP2 O15409; FOXP1 Q9H334; FOXP4 Q8IVH2. **(c)** BRET assay for interaction of FOXP1 and FOXP2 with PIAS proteins. HEK 293 cells were transfected with luciferase-FOXP1 or luciferase-FOXP2 (donor) and YFP-PIAS (acceptor). The control donor protein is a nuclear-targeted luciferase. Values are mean corrected BRET ratios ± S.E.M. (n = 3). **(d)** BRET assay for interaction of FOXP1 and FOXP2 with SUMO. HEK293 cells were transfected with luciferase-FOXP1 or luciferase-FOXP2 (donor) and YFP-SUMO (wild-type (GG) or alanine mutant (AA), acceptor). The control donor protein is a nuclear-targeted luciferase. Values are mean corrected BRET ratios ± S.E.M. (n = 3).

tion, supporting a conserved role for SUMOylation in the regulation of FOXP transcription factors (Golebiowski *et al.* , 2009; Tatham *et al.* , 2011; Wen *et al.* , 2014).

2.4 DISCUSSION

In this study we have shown that FOXP2 has a single major SUMOylation site at K674, which can be modified by SUMO1, SUMO2 and SUMO3. This site is fully conserved in orthologues of FOXP2, and lies within the C-terminal region of the protein, which previously had no known function. We have demonstrated that FOXP2 interacts with members of the PIAS family of E3 SUMO ligases, and that this interaction probably involves the SP-RING domain of PIAS and the N-terminal region of FOXP2. The interaction between FOXP2 and PIAS proteins causes relocalization of FOXP2 to nuclear speckles, and promotes SUMOylation of FOXP2. The FOXP2 paralogues FOXP1 and FOXP4 probably also undergo PIAS-mediated SUMOylation at equivalent sites.

SUMOylation is believed to occur in all cell types and across all developmental stages, and thousands of nuclear proteins are thought to be modified in this way (Meulmeester & Melchior, 2008). The essential role of SUMOylation in development is evidenced by the early embryonic lethality resulting from *Ubc9* knockout in mice (Nacerddine *et al.* , 2005). Global changes in SUMOylation levels during brain development have been documented in mouse and rat, but the functional impact of these changes is uncertain (Hasegawa *et al.* , 2014; Loriol *et al.* , 2012). Effects of SUMOylation on several neural proteins have been reported, with impacts on neuronal specification and dendritic and synaptic morphogenesis, but for most proteins the function of SUMOylation remains unclear (Gwizdek *et al.* , 2013; Meulmeester & Melchior, 2008).

In our cellular assays, abolishing the major SUMOylation site in FOXP2 did not produce changes in subcellular localization, stability, transcriptional regulation, dimerization with wild-type FOXP2, or interaction with the co-repressor CtBP1. Abolition of the SUMOylation site may have effects on unknown protein interactions, or on regulation of a particular subset of target genes. It would therefore be of interest in future to perform proteomic and transcriptomic studies to search for protein-protein interactions and target genes that are affected by loss of the FOXP2 SUMOylation site. At the same time, the failure to observe differences between wild-type FOXP2 and the K647R mutant is not necessarily

surprising given the small proportion of wild-type protein molecules modified by SUMOylation at any one time. Like several other post-translational modifications, SUMOylation is a dynamic, reversible process which allows protein activity to be regulated on short time-scales in response to external signals that may change over the course of development. The consequences of loss of SUMOylation in FOXP2 may therefore only be apparent in the context of a developing organism, and SUMOylation of FOXP2 is potentially essential in vivo but not in cultured cells. It would therefore be interesting to abolish the FOXP2 SUMOylation site in an animal model, in order to assess the effect of loss of SUMOylation on developmental regulation of gene expression. There are still very few examples of animal models in which a SUMOylation site in a specific protein has been ablated. However, abolishing the SUMOylation sites of the transcription factor NR5A1 in the mouse resulted in aberrant regulation of target genes and prominent endocrine abnormalities, without affecting protein stability or localization (Lee *et al.* , 2011), consistent with the suggestion that the consequences of loss of SUMOylation may only be apparent in an organismal context. To our knowledge there has been no systematic survey of rare variants in human developmental disorders for changes likely to affect SUMOylation. Thus, the importance of SUMOylation of specific proteins to normal development is not yet fully appreciated.

The rapid and dynamic nature of SUMOylation makes it well suited as a mechanism for modifying the activity of proteins such as FOXP2 in response to activity within neural circuits. Such mechanisms may be important for supporting neural plasticity, a process in which FOXP2 orthologues have been shown to play a role in animal models (French *et al.* , 2012; Groszer *et al.* , 2008; Murugan *et al.* , 2013). The absolute conservation of the SUMOylation site and surrounding KEPE motif in FOXP2 orthologues and in the paralogues FOXP1 and FOXP4 suggests that SUMOylation may be an evolutionarily ancient and conserved mechanism for regulating the activity of these transcription factors in the brain and elsewhere, both during development and in the adult organism.

The selection of target proteins for SUMOylation is mediated by SUMO E3 ligases, such as members of the PIAS family. Our results indicate that multiple PIAS proteins may participate in FOXP SUMOylation in vivo, and that different PIASs may be involved in the SUMOylation of different FOXPs. RNA expression data from the Human Protein Atlas (www.proteinatlas.org) suggests that the four PIAS proteins have ubiquitous expression, with PIAS1 and PIAS3 showing

moderate expression levels across all tissues tested, whereas PIAS2 and PIAS4 have lower average expression levels with high expression in testis. However the PIAS proteins are not functionally redundant, because *Pias1* knockout mice exhibit perinatal lethality, whereas *Pias2* and *Pias4* knockouts have no obvious phenotype (*Pias3* knockouts have not yet been reported) (Liu *et al.* , 2004; Santti *et al.* , 2005; Wong *et al.* , 2004). There may be some temporal or spatial specificity in the interaction of PIASs with FOXPs, and we also do not exclude the involvement of other kinds of SUMO E3 ligase in the SUMOylation of FOXPs.

The reduced level of SUMOylation of the R553H mutant, and the reduced interaction of the mutant with PIAS proteins, indicate that mutations within the FOX domain of FOXP2 can interfere with PIAS-mediated SUMOylation. Interestingly, PIAS1-mediated SUMOylation has also been reported for FOXL2 and FOXA2, which belong to different subfamilies of the FOX transcription factor family (Belaguli *et al.* , 2012; Marongiu *et al.* , 2010). The different subfamilies of FOX proteins show little similarity outside the FOX domain, raising the possibility that the FOX domain may act in concert with other subfamily-specific protein regions to promote SUMOylation of FOX transcription factors. Aetiological mutations have been reported in the FOX domains of several other FOX transcription factors (Benayoun *et al.* , 2011), but the effects of these mutations on protein SUMOylation have not been investigated. We predict that disorder-related FOX domain mutations would disrupt SUMOylation of other FOX transcription factors, including FOXP1.

Concurrently with the submission of this manuscript, SUMOylation of FOXP2 by SUMO1 and SUMO3 was reported by an independent research group (Meredith *et al.* , 2015). In agreement with our findings, this report identified K674 as the major SUMOylation site in FOXP2, and found that the R553H variant displays reduced SUMOylation. Furthermore, no effect on protein stability or localization was observed when mutating the SUMOylation site, in line with our data. Some potential small effects on transcriptional regulation in luciferase reporter assays were observed, including for SRPX2, using a similar reporter construct to that employed in our experiments, emphasizing that the effects of SUMOylation on FOXP2-mediated transcriptional regulation warrant further investigation, ideally in a more biologically-relevant model, and that the effects of SUMOylation may be promoter-dependent.

SUMOylation is currently the only confirmed post-translational modification of FOXP2. A key part of the way SUMOylation affects protein function is through

interaction with other post-translational modifications, for example by competition with ubiquitination and acetylation for modification of specific lysine side chains. Our findings thus emphasize the need to investigate further the post-translational modifications of FOXP2 in order to understand how the activity of this transcription factor may be dynamically regulated in the developing and adult brain.

2.5 SUPPLEMENTARY MATERIAL

Supplementary Table S2.1: Cloning primer sequences
Restriction sites are underlined.

Gene	Forward primer	Site	Reverse primer	Site
PIAS1	AAGATCTCGGCGGACAGT GCGGAACTAAAGC	BglII	CGCTAGCTTAGTCCAATGA AATAATGTCTGGT	NheI
PIAS2	GGATCCTGGCGGATTTCG AAGAGTTG	BamHI	TCTAGATTAGTCCAATGAG ATGATGTCAGG	XbaI
PIAS3	AGATCTTGGCGGAGCTGG GCGAATTAAA	BglII	TCTAGATCAGTCCAGGGAA ATGATGTC	XbaI
PIAS4	GGATCCTGGCGGCGGAG CTGGTG	BamHI	TCTAGATCAGCAGGCCGGC ACCAGGCCCTT	XbaI
SUMO1	GGATCCTGTCTGACCAGG AGGCAAAACCTT	BamHI	TCTAGACTAAACTGTTGAAT GACCCCC	XbaI
SUMO2	GGATCCTGGCCGACGAAA AGCCCAAGGAAG	BamHI	TCTAGATTAGTAGACACCT CCCGTCTG	XbaI
SUMO3	GGATCCTCTCCGAGGAGA AGCCCAAGGAG	BamHI	TCTAGACTAGAAACTGTGC CCTGCCAG	XbaI
UBC9	AGATCTGGAACACCTGTC CGCTACGCTC	BglII	TCTAGATTATGAGGGCGCA AACTTCTT	XbaI
SRPX2 pro-moter	GGTACCCTCTGCCTCCTG GGTTCAAG	KpnI	AAGCTTGATGGGGGAGAAG GAACACA	HindIII

Supplementary Table S2.2: Site-directed mutagenesis primer sequences

Mutant	Primer 1	Primer 2
FOXP2 K674R	CATACATTCAATCCACGTC AGGGAAGAGCCAGTGATTG	CAATCACTGGCTCTTCCCTGA CGTGGATTGAATGTATG
PIAS1 C350S	ACATTGTAGATGAGAACTT GTAAGGGCCCGACAC	GTGTCGGGCCCTTACAAGTTC TCATCTACAATGT

SUMO1 AA	AGACTAAACTGTTGAATGAGC CGCCGTTTGTTCCTGATAAAC	GTTTATCAGGAACAAACGGC GGCTCATTCAACAGTTTAGTCT
SUMO2 AA	TTCTAGATCAGTAGACAGCTGC CGTCTGCTGTTGGAAC	GTTCCAACAGCAGACGGCAGCT GTCTACTGATCTAGAA
SUMO3 AA	CTCTCCGGCACAGCTGCCGTCT GCTGCT	AGCAGCAGACGGCAGCTGTG CCGGAGAG

Supplementary Table S2.3: Summary of yeast two-hybrid results

DNA was isolated from positive colonies, the cDNAs of the prey constructs were sequenced, and BLAST search was used to identify the proteins encoded by positive clones. Proteins represented by two or more clones are listed.

HGNC symbol	Number of preys
PIAS1	7
CHD3	4
FXYD6	4
NREP	3
FKBP1A	2
NRGN	2

Supplementary Table S2.4: Prediction of putative FOXP2 SUMOylation sites

The table lists all the lysine residues in FOXP2 (Uniprot O15409). SUMOylation sites were predicted using three web-based algorithms: Joined Advanced SUMOylation Site and SIM Analyser (JASSA, www.jassa.fr), SUMOplot analysis Program (www.abgent.

com/SUMOplot), and GPS-SUMO (SUMOsp.biocuckoo.org/). JASSA uses a scoring system based on a Position Frequency Matrix derived from the alignment of experimental SUMOylation sites. One of the two sites identified by JASSA corresponds to an inverted consensus SUMOylation site D/E-X-K-ψ, where ψ = A/F/I/L/M/P/V/W; X = any amino acid. The second is a standard consensus SUMOylation site in the context of a negatively charged amino acid-dependent SUMOylation motif (NDSM): ψ-K-X-D/E-X-α, where 2 out of 6 α must be D/E. SUMOplot predicts the probability (0-1) for the major SUMO canonical consensus sequence ψ-K-X-D/E where ψ =I/L/V. The GPS-SUMO score is based on 983 SUMOylation sites in 545 proteins collected from the literature.

Residue	Sequence context	JASSA	SUMOplot	GPS-SUMO
K74	LLLQQQTSGLKSPKSSDKQRP		0.80	
K77	QQQTSGLKSPKSSDKQRPLQV			
K81	SGLKSPKSSDKQRPLQVPVSV			
K140	LQQQQLQEFYKKQQEQLHLQL			
K141	QQQQLQEFYKKQQEQLHLQLL			
K195	QQQQQQQHPGKQAKEQQQQQQ			
K198	QQQQHPGKQAKEQQQQQQQQQ			
K271	LSPAEIQQLWKEVTGVHSMED			
K285	GVHSMEDNGIKHGGLDLTTNN		0.77	
K306	SSSTTSSNTSKASPPITHHSI			
K349	SHTLYGHGVCKWPGCESICED			
K365	SICEDFGQFLKHLNNEHALDD			
K397	VVQQLEIQLSKERERLQAMMT			
K417	THLHMRPSEPKPSPKPLNLVS	High (Inverted)		
K421	MRPSEPKPSPKPLNLVSSVTM			
K433	LNLVSSVTMSKNMLETSPQSL			
K482	VGAIRRRHSDKYNIPMSSEIA			
K499	SEIAPNYEFYKNADVRPPFTY			
K549	AYFRRNAATWKNAVRHNLSLH			

K560	NAVRHNLSLHKCFVRVENVKG		0.17	
K569	HKCFVRVENVKGAVWTVDEVE			
K582	VWTVDEVEYQKRRSQKITGSP			
K587	EVEYQKRRSQKITGSPTLVKN			
K596	QKITGSPTLVKNIPTSLGYGA			
K674	PQPHIHSIHVKEEPVIAEDED	High (NDSM)	0.93	24.943

Supplementary Table S2.5: Prediction of putative FOXP2 SUMO-interacting motifs (SIMs)

SUMO-interacting motifs (SIMs) were predicted using two web-based algorithms: Joined Advanced SUMOylation Site and SIM Analyser (JASSA, www.jassa.fr/), and GPS-SUMO (SUMOsp.biocuckoo.org/). JASSA uses a scoring system based on a Position Frequency Matrix derived from the alignment of experimental SIMs and the maximum predictive score is 38. The GPS-SUMO score is based on 151 SIMs in 80 proteins collected from the literature. In both cases, putative SIMs were only identifiable by using the lowest possible detection threshold. Potential SIMs are indicated in bold type.

Residues	Sequence context	JASSA	GPS-SUMO
47-50	DTSSEVST**VELL**HLQQQQAL	0.275	
49-52	SSEVSTVE**LLHL**QQQQALQA	0	
63-66	QQALQAAR**QLLL**QQQTSGLK	0	
148-151	KKQQEQLHL**QLL**QQQQQQQQ	0.053	
233-236	QQLQQQQH**LLSL**QRQGLISI	0.018	
241-244	LLSLQRQG**LISIP**PGQAALP	0.097	28.42
423-426	EPKPSPKP**LNLV**SSVTMSKN	0.01	
618-624	QAALA**ESSLPLL**SNPGLINN	0.036	

3

ASSESSMENT OF FOXP2-CANDIDATE PROTEIN INTERACTIONS IDENTIFIED IN PREVIOUS MASS SPECTROMETRY SCREENS

Abstract: *Heterozygous disruptions in the FOXP2 gene cause a rare speech and language disorder characterized by difficulty in coordinating the oral motor sequences necessary for fluent speech. The FOXP2 gene encodes a transcription factor that is expressed in a subset of brain structures such as the cortex, striatum and cerebellum, where it regulates aspects of neural development and function. The activity of transcription factors is dependent on and regulated by multiple protein-protein interactions. Therefore, determining which co-factors FOXP2 interacts with can provide insights into the molecular functions of this protein in the context of speech and language development. Although a number of studies have investigated roles of FOXP2 in the brain and sought to identify downstream target genes, knowledge regarding its protein-protein interactions remains scarce. Prior research in our group sought to identify FOXP2-interacting proteins by means of immunoprecipitation followed by mass spectrometry. In this study, I re-evaluated the dataset from this earlier screen, selecting two promising candidate FOXP2 interactors, NONO and SFPQ, for targeted investigation. I aimed to validate these interactions using multiple complementary methods: bioluminescence resonance energy transfer assay, cobalt-affinity purification, and mammalian two-hybrid assay. The interaction between FOXP2 and NONO/SFPQ failed to validate using any of these techniques, suggesting that these proteins are unlikely to be physiological interaction partners. Consequently, the methodology used in the previous mass spectrometry screen may have been vulnerable to false positive results, indicating the need for new screens that search for FOXP2-interacting proteins.*

3.1 INTRODUCTION

The transcription factor *FOXP2* was the first gene to be conclusively linked with speech and language development (Lai *et al.* , 2001). In particular, a heterozygous etiological mutation in this gene was identified in a large, multigenerational family (the KE family) in which half of the members suffered from childhood apraxia of speech (CAS), a disorder characterized by difficulties coordinating the precise oral motor sequences required for speech production. This mutation yields a substitution in the DNA-binding domain of the FOXP2 protein and abolishes its ability to bind its usual target sequences and regulate gene expression (Stroud *et al.* , 2006; Vernes *et al.* , 2006). Since then, a number of other alterations affecting the gene have been found in individuals with similar speech and language deficits (Laffin *et al.* , 2012; MacDermot *et al.* , 2005; Reuter *et al.* , 2016; Turner *et al.* , 2013), firmly establishing FOXP2 as a critical genetic factor underlying language development and function.

FOXP2 belongs to the FOXP family of transcription factors, which also includes FOXP1, FOXP3 and FOXP4. FOXP2, FOXP1 and FOXP4 are expressed in the brain where they play overlapping but distinct roles in its development (Bowers & Konopka, 2012a; Shu *et al.* , 2001). These three proteins share the same domain organization: a glutamine-rich region, a zinc finger domain, a leucine zipper and a forkhead DNA-binding domain. FOXP3 is divergent from the other FOXPs in sequence and structure, and its expression is restricted to the immune system (Fontenot *et al.* , 2003). FOXP2 DNA-binding sites and target genes have been characterized in several studies, which show that FOXP2 regulates genes implicated in neuronal processes such as synaptic plasticity and neurite outgrowth (Spiteri *et al.* , 2007; Vernes *et al.* , 2011, 2007). In contrast, knowledge of FOXP2 protein interactions remains scarce. In order to regulate gene expression, transcription factors need to physically interact with transcription-related proteins, such as chromatin remodelling enzymes or other transcription factors (Smith & Matthews, 2016). Protein-protein interactions regulate numerous aspects of a transcription factor function, such as DNA-binding affinity and specificity, transcriptional activation/repression activity, subcellular localization, and protein turnover.

FOXP2 forms homo-dimers and also hetero-dimerizes with FOXP1 and FOXP4 through the leucine zipper domain (Li *et al.* , 2004). This dimerization is essential for transcriptional regulatory activity, and heterodimerization among FOXP

proteins may generate a repertoire of dimers with different DNA-binding specificities (Sin *et al.* , 2015).

A small number of other proteins have also been reported to interact with FOXP2, including the transcriptional co-repressors CTBP1, CTBP2 and GATAD2B (Chokas *et al.* , 2010; Estruch *et al.* , 2016a). Large-scale yeast two-hybrid (Y2H) screens have provided only a small number of candidate FOXP2 interactors (Corominas *et al.* , 2014; Rolland *et al.* , 2014; Sakai *et al.* , 2011). Most of the candidates identified in these screens have not been confirmed, but previous studies from our group have successfully validated and further investigated two of these proteins, namely, the transcription factor TBR1 and the post-translational modification enzyme PIAS1 (Estruch *et al.* , 2016b; Deriziotis *et al.* , 2014a). Rare mutations in TBR1 cause autism spectrum disorder (ASD) together with language deficits (Deriziotis *et al.* , 2014a). Strikingly, in cellular assays the interaction between FOXP2 and TBR1 could be disrupted by known mutations in either transcription factor, suggesting a shared molecular etiology for distinct neurodevelopmental disorders that include language impairment (Deriziotis *et al.* , 2014a). As described in Chapter 2, investigations of the interaction between FOXP2 and PIAS1 led to the discovery that FOXP2 is SUMOylated (Estruch *et al.* , 2016b). Evidence of FOXP2 SUMOylation has also been found in two independent studies which additionally suggested that this post-translational modification is involved in regulating FOXP2 activity in the cerebellum, where it may help modulate motor function and vocal communication (Meredith *et al.* , 2015; Usui *et al.* , 2016). These studies represent clear examples of how identifying and understanding FOXP2 physical interactions with other proteins can reveal valuable knowledge about the transcription factor and its roles in brain development.

Further characterizing the FOXP2 interactome is critical to fully comprehend its molecular functions. In this chapter I aimed to expand the known FOXP2 interactome by validating putative binding partners suggested by mass spectrometry analysis of immunoprecipitated FOXP2 complexes in a previous unpublished investigation by our group. From this dataset I selected the transcriptional regulators SFPQ and NONO as promising candidates, and attempted to validate their interaction with FOXP2 using a range of complementary protein-protein interaction techniques. The results consistently demonstrate a lack of interaction between FOXP2 and NONO/SFPQ with all methods employed, which suggests that these proteins are not physiological interaction partners. I con-

clude that the experimental and/or analytical methods used in the previous mass spectrometry study to create a list of candidate FOXP2-interacting proteins were likely suboptimal and vulnerable to false positives. This chapter thus highlights the need for fresh mass spectrometry screens using improved methodology to deliver further insights into the FOXP2 interactome.

3.2 MATERIAL AND METHODS

DNA constructs

The cloning of human FOXP2 (NM_014491) has been described previously (Estruch et al., 2016b). The coding sequences of NONO (NM_007363) and SFPQ (NM_005066) were amplified from human foetal brain cDNA using the primers listed in Table 1 and then cloned into pCR2.1-TOPO (Invitrogen). For expression of fusion proteins with *Renilla* luciferase, yellow fluorescent protein (YFP), mCherry, HisV5 and myc tags, the coding sequences of FOXP2, NONO and SFPQ were subcloned into the pLuc, pYFP, pmCherry, pHisV5 and pMyc expression vectors, respectively, which have been described previously (Estruch *et al.* , 2016b).

Table 3.1: Cloning primers for NONO and SFPQ.
∗ BclI generates sticky ends that are complementary to BamHI.

	Forward primer	Reverse primer	Restriction sites
NONO	AGGATCCAGAGTA ATAAAACTTTTAA CTTGG	CTCTAGATTAGTATCG GCGACGTTTGTTTGGG	BamHI + XbaI
SFPQ	GATGATCAGCTCTCG GGATCGGTTCCG GAGTC	CGCTAGCCTAAAAT CGGGGTTTTTTGT TTGGG	BclI* + XbaI

For the mammalian two-hybrid assay, modified versions of the vectors pBIND and pACT (Promega) were created. The pBIND and pACT vectors contain the yeast GAL4 DNA-binding domain and the herpes simplex virus VP16 activation domain, respectively, upstream of a multiple cloning region. The *Renilla* luciferase gene was removed from pBIND, together with its promoter and polyA

signal, by excising a ClaI fragment and religating the backbone, creating the modified construct pBIND2 (Supplementary Figure S2.1). The construct pACT2 was generated by inserting the functional region of the original pACT vector into the pBIND2 backbone using the NdeI and XbaI sites (Supplementary Figure S2.2). The cDNAs of FOXP2, NONO and SFPQ were subcloned into the pBIND2 and pACT2 vectors using the BamHI and XbaI sites. A firefly luciferase reporter vector for the mammalian two-hybrid assay (pGL4-luc2-GAL4UAS-AP) was engineered by inserting a KpnI-NcoI fragment of pGL5 (Promega), containing the yeast GAL4 upstream activating sequence (GAL4UAS) directly upstream of the adenovirus major late promoter, into the backbone of pGL4.23 (Promega), replacing the minimal promoter in this vector (Supplementary Figure S2.3). A negative control reporter vector, which lacks the GAL4UAS (pGL4-luc2-AP) was engineered by inserting a NheI-AgeI fragment of pGL4-luc2-GAL4UAS-AP, containing the adenovirus promoter and partial luciferase coding sequence, into pGL4.23 (Promega) (Supplementary Figure S2.4). All constructs were verified by Sanger sequencing.

Cell culture and transfection

HEK293 cells were obtained from ECACC (cat. no. 85120602) and cultured in DMEM supplemented with 10% foetal bovine serum. Transfections were performed using GeneJuice (Merck-Millipore) according to the manufacturer?s instructions.

BRET assay

BRET assays were performed as described (Deriziotis et al., 2014a). Briefly, HEK293 cells were transfected in 96-well plates with plasmids encoding YFP- and luciferase-fusion proteins. After 36-48 h, Enduren live cell luciferase substrate (Promega) was added at a final concentration of 60 μ M. Cells were cultured for a further 4 h, and emission readings (integrated over 10 s) were taken using a TECAN F200PRO microplate reader using the Blue1 and Green1 filter sets. Expression levels of the YFP-fusion proteins were measured by taking fluorescence intensity readings using the filter set and dichroic mirror suitable for green fluorescent protein (excitation 480 nm, emission 535 nm). The corrected BRET ratio

was calculated with the following formula: [Green1(experimental condition)/Blue1(experimental condition)]-[Green1(control condition)/Blue1(control condition)]. The control conditions used luciferase or YFP fused to a C-terminal nuclear localization signal.

Fluorescence microscopy

HEK293 cells were seeded on coverslips coated with poly-L-lysine. Cells were cultured for 30 h post-transfection, and then fixed with 4% paraformaldehyde. Nuclei were stained with Hoechst 33342. Fluorescence images were acquired using an Axiovert A-1 fluorescent microscope with ZEN Image software (Zeiss).

Pull-down assays

Cells were cultured in 6-well plates and transfected with HisV5-tagged FOXP2, NONO, or SFPQ, or empty pHisV5 vector, together with myc-tagged FOXP2, NONO, or SFPQ, or pMyc vector. After 48 h cells were lysed in 25 mM sodium phosphate pH 8.0, 250 mM NaCl, 1% Triton X-100, 20 mM imidazole, 1% v/v PMSF and 0.5% v/v EDTA-free protease inhibitor cocktail set III (Sigma-Aldrich). An aliquot of the lysate (10%) was retained as the input sample and the remainder was incubated with His-tag Dynabeads (Life Technologies) overnight at 4° C with rotation. Beads were washed with 25 mM sodium phosphate pH 8.0, 250 mM NaCl, 0.1% Triton X-100, 20 mM imidazole. His-V5-tagged proteins were eluted in 25 mM sodium phosphate pH 8.0, 250 mM NaCl, 0,1% Triton X-100, 400 mM imidazole. Proteins were resolved on 10% SDS-polyacrylamide gels and transferred to PVDF membranes using a TransBlot Turbo Blotting apparatus (Bio-Rad). Membranes were blocked in PBS containing 5% milk and 0.1% Tween-20, and were then incubated overnight at 4° C with primary antibody in PBS containing 5% milk. The following antibodies were used: anti-V5 tag (Genetex cat. no. GTX42525, 1:3000); anti-Myc tag (Abcam cat. no. ab9106, 1:1000). After washing, membranes were incubated with horseradish peroxidase-conjugated goat anti-mouse or anti-rabbit IgG for 45 min at room temperature. Proteins were visualized using Novex ECL Chemiluminescent Substrate Reagent Kit (Life Technologies) and a ChemiDoc XRS+ imaging system (Bio-Rad).

Mammalian two-hybrid assay

Protein interactions were assayed using the CheckMate Mammalian Two-Hybrid System (Promega). In brief, HEK293T cells were seeded in clear-bottomed white 96-well plates and cultured for 24 h. Cells were transfected in triplicate with an equimolar mixture of pACT2 and pBIND2 constructs (5 mol each per well), together with the firefly luciferase reporter construct pGL4-luc2-GAL4UAS-AP or the negative control reporter pGL4-luc2-AP (5 fmol), and the pGL4.74 *Renilla* luciferase normalization control (2 fmol). The total mass of transfected DNA was adjusted to 6 ng with filler plasmid. After 48 h, cells were lysed with passive lysis buffer and firefly and *Renilla* luciferase activities were measured using the Dual Luciferase Reporter Assay System (Promega) with a TECAN F200PRO microplate reader. In order to correct for variation in transfection efficiency, each firefly luciferase activity value was normalized to the respective *Renilla* luciferase activity value.

3.3 RESULTS

3.3.1 *Evaluation of candidate FOXP2-interacting proteins*

As part of previous work in our group to identify potential interacting proteins, FOXP2 protein complexes immunoprecipitated using a FOXP2-specific antibody were analysed by mass spectrometry (Elahi, 2008). FOXP2 protein complexes were purified from four different sources: (1) SH-SY5Y (human neuroblastoma) cells stably expressing wild-type FOXP2; (2) SH-SY5Y cells stably expressing FOXP2.pR553H, the variant found in the KE family; (3) HEK293 cells, which endogenously express FOXP2; and (4) embryonic mouse brain at day E16.5, a time point of high Foxp2 expression. The proteins identified in the immunoprecipitated complexes are summarized in Table 3.2.

To prioritize potential FOXP2-interacting proteins, I first excluded previously reported FOXP2 interactors (i.e. FOXP1, FOXP2 and FOXP4) and probable non-specific interactors (histones, ribosomal subunits). I then selected proteins that (1) are localized to the cell nucleus, (2) are expressed in the brain, and (3) were identified in two or more experiments. This filtering process resulted in a short-

list of 6 putative FOXP2 interactors: FOXR1, NCOR1, NONO, PRKRIR, SBF1 and SFPQ.

Table 3.2: Candidate FOXP2-interacting proteins previously suggested by immunoprecipitation and mass spectrometry.

Immunoprecipitation was performed on 4 different sources: SH-SY5Y cell line stably expressing wild-type FOXP2 (a); SH-SY5Y cell line stably expressing FOXP2.pR533H (b); HEK293 cells (c); and mouse brain at embryonic day E16.5 (d). Proteins identified with $p<0.05$ are indicated with an "X". Italicized words highlight biological information suggesting that the interaction is not physiologically relevant.

Gene name	a	b	c	d	Description	Expression	
						Tissue	Cell compartment
ACTB	X				Beta-actin	Ubiquitous	*Cytoplasm*
ACTL6B	X				Chromatin remodeller	Brain	Nucleus
ARMC3			X		Beta-catenin-like protein	Ubiquitous	
BPTF	X				Nucleosome-remodelling factor	Ubiquitous	*Cytoplasm/ nucleus*
CRB1			X		Calcium receptor	Brain	*Membrane/ secreted*
CTSD	X				Protease	Brain	*Secreted*
EEF1A2	X				Protein elongation factor	Brain	Nucleus
EGF			X		Growth factor	Brain	*Membrane*
EIF4B	X	X			Protein translation initiation factor	Ubiquitous	*Cytoplasm*
FN1	X				Extracellular matrix	Ubiquitous	*Secreted*
FOXP1		X			Transcription factor	Brain	Nucleus
FOXP2	X	X			Transcription factor	Brain	Nucleus
FOXP4	X				Transcription factor	Brain	Nucleus
FOXR1	X		X		Transcription factor	Brain	Nucleus
GABRA3			X		GABA receptor subunit	Brain	*Membrane*
GFAP	X				Intermediate filament	Brain	*Cytoplasm*
GRIN3A		X			Glutamate receptor ionotropic	Brain	*Membrane*
H2AZ	X		X		Histone	Ubiquitous	Nucleus
HIST1H2AB	X				Histone	Ubiquitous	Nucleus

HIST1H2AG	X				Histone	Ubiquitous	Nucleus
HIST1H2BK	X				Histone	Ubiquitous	Nucleus
HIST1H3A	X				Histone	Ubiquitous	Nucleus
HIST1H4A	X				Histone	Ubiquitous	Nucleus
HIST2H2BF	X		X		Histone	Ubiquitous	Nucleus
HNRPK	X				Ribonucleoprotein	Brain	Cytoplasm/ nucleus
HNRPL	X				Ribonucleoprotein	Brain	Cytoplasm/ nucleus
HSP90AA1				X	Chaperone	Ubiquitous	*Cytoplasm*
HSP90AB1		X			Chaperone	Ubiquitous	*Cytoplasm*
IFIH1			X		Helicase	Ubiquitous	Cytoplasm/ nucleus
KIF21A	X				Microtubule-binding motor protein	Ubiquitous	*Cytoplasm*
LRSAM1		X	X		E3 ubiquitin-protein ligase	Spinal cord	*Cytoplasm*
NACA				X	Protein translation complex subunit	Ubiquitous	*Cytoplasm*
NCOR1	X	X			Trancriptional co-repressor	Ubiquitous	Nucleus
NEB		X			Structural protein	*Muscle*	*Cytoplasm*
NONO	X	X			DNA- and RNA binding protein	Brain	Nucleus
NPM1	X				Nucleolar phosphoprotein	Ubiquitous	Cytoplasm/ nucleus
PKM	X				Pyruvate kinase isozyme	*Proliferating cells*	Cytoplasm/ nucleus
RNF149			X		E3 ubiquitin-protein ligase	Brain	*Membrane*
RPL10	X				Large ribosomal subunit	Ubiquitous	*Cytoplasm*
RPL6	X	X			Large ribosomal subunit	Ubiquitous	*Cytoplasm*
RPL7	X				Large ribosomal subunit	Ubiquitous	*Cytoplasm*
RPL9	X				Large ribosomal subunit	Ubiquitous	*Cytoplasm*
RPS3		X			Large ribosomal subunit	Ubiquitous	*Cytoplasm*
RPS4X	X				Large ribosomal subunit	Ubiquitous	*Cytoplasm*
RPS6	X				Large ribosomal subunit	Ubiquitous	*Cytoplasm*
SBF1	X	X			Pseudophosphatase	Ubiquitous	Nucleus
SERPINB7	X				Inhibitor of lys-specific proteases	Ubiquitous	*Cytoplasm*
SFPQ	X		X		DNA- and RNA binding protein	Ubiquitous	Nucleus
SHROOM3	X				Structural protein	Brain	*Cytoplasm*
SYNCRIP	X				Ribonucleoprotein	Ubiquitous	Cytoplasm/ nucleus
THAP12	X		X		Kinase modulator	Ubiquitous	Nucleus

TMEFF1		X		Egf-like domains containing protein	Brain	*Membrane*
TRIM21	X	X		E3 ubiquitin-protein ligase	Heart and lung	Cytoplasm/ nucleus
VIM		X	X	Class-iii intermediate filaments	*Fibroblasts*	*Cytoplasm*
WNT7A	X			Frizzled receptor ligand	Brain	*Secreted*
WWP2	X			E3 ubiquitin-protein ligase	Brain	Nucleus

From this list, NONO and SFPQ were chosen as a primary focus because they are transcription factors, and are also known to interact with each other (Fox et al., 2005), which is consistent with both proteins being pulled-down in the same protein complexes. NONO and SFPQ, together with a third protein, PSPC1, form the *Drosophila* behaviour/human splicing (DBHS) protein family (Knott *et al.* , 2016). DBHS proteins form homo- and heterodimers (Fox *et al.* , 2005) and have roles in transcriptional activation and repression, splicing, pre-mRNA processing and RNA transport (Amelio *et al.* , 2007; Dong *et al.* , 2007; Izumi *et al.* , 2014; Kanai *et al.* , 2004; Kaneko *et al.* , 2007; Kim *et al.* , 2011; Park *et al.* , 2013).Although NONO and SFPQ are expressed in the brain, little is known regarding their roles in neural processes. NONO is implicated in the regulation of the circadian clock function, acting as a transcriptional cofactor of PER proteins (Kowalska *et al.* , 2012, 2013) and SFPQ is involved in the activation of neural-specific alternative splicing (Kim *et al.* , 2011). NONO and SFPQ therefore represented particularly interesting candidates for interaction with FOXP2.

3.3.2 *Bioluminescence Resonance Energy Transfer assay*

To try to confirm the interaction between FOXP2 and NONO/SFPQ, I first used the Bioluminescence Resonance Energy Transfer (BRET) assay, a method which allows the monitoring of protein-protein interactions in living cells, and which has been used successfully in prior studies of FOXP2 (Deriziotis *et al.* , 2014b). To test for an interaction between two proteins using the BRET assay, the two proteins of interest are expressed as fusion proteins in cultured cells: one protein (the donor) is fused to *Renilla* luciferase (RLuc), and the second (the acceptor) is fused to yellow fluorescent protein (YFP). An interaction between the

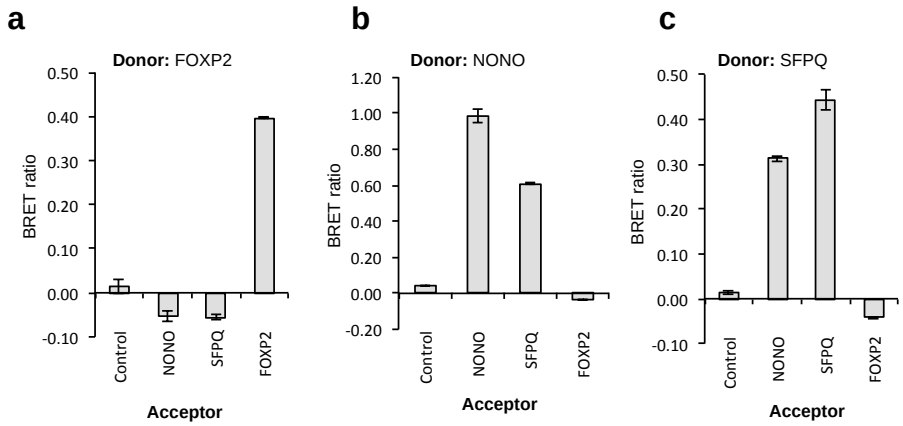


Figure 3.1: BRET assay for testing interaction between FOXP2 and NONO/SFPQ. **(a)** HEK293 cells were transfected with FOXP2 fused to Renilla luciferase, and FOXP2, NONO or SFPQ fused to YFP. **(b)** HEK293 cells were transfected with NONO fused to Renilla luciferase, and FOXP2, NONO and SFPQ fused to YFP. **(c)** HEK293 cells were transfected with SFPQ fused to Renilla luciferase, and FOXP2, NONO and SFPQ fused to YFP. Values are mean corrected BRET ratio± SEM (n = 3).

two proteins may bring the RLuc and YFP moieties in sufficient proximity to allow resonance energy transfer to occur from RLuc to YFP when a luciferase substrate is supplied, which leads to a shift in the wavelength of the emitted light (the BRET signal).

When RLuc-FOXP2 (donor) was co-expressed with YFP-NONO or YFP-SFPQ (acceptor), there was no increase in the BRET signal compared with the control (Figure 3.1a). Similar results were obtained when the assay was performed with FOXP2 as the acceptor and NONO or SFPQ as the donor (Figure 3.1b,c). In contrast, both the homodimerization of FOXP2 and the interaction between NONO and SFPQ were efficiently detected (Figure 3.1), confirming the suitability of the assay to detect protein-protein interactions involving these transcription factors. The results of the BRET assay therefore do not support the existence of an interaction between FOXP2 and NONO/SFPQ.

In parallel to the BRET experiments, the subcellular localization of the FOXP2, NONO and SFPQ proteins was examined by transfecting cells with fusions of

these proteins to the fluorescent proteins YFP and mCherry. FOXP2, NONO and SFPQ all localize to the cell nucleus (Figure 3.2a), confirming that the lack of interaction between these proteins in the BRET assay is not due to mislocalization of one or more proteins. FOXP2 presents a diffuse distribution in the nucleus, whereas NONO and SFPQ exhibit a speckled pattern, as reported in previous studies (Fox *et al.* , 2005; Shav-Tal & Zipori, 2002) (Figure 3.2a). Co-transfections of YFP-tagged SFPQ/NONO and mCherry-tagged FOXP2 suggest that FOXP2 is not incorporated into the SFPQ/NONO-containing nuclear substructures (Figure 3.2b), consistent with a lack of interaction between these proteins.

3.3.3 *Pull-down assay*

The BRET assay is a sensitive and versatile technique for detecting protein-protein interactions because it can detect weak and transient interactions, and it is applied to proteins in their native environment within living cells. However for some pairs of interacting proteins, the BRET assay may give a false negative result because the luciferase and YFP moieties are not sufficiently close, or not appropriately oriented, to allow resonance energy transfer to occur. Because the putative interaction between FOXP2 and NONO/SFPQ was initially detected by immunoprecipitation, I next attempted to confirm the interaction using a pull-down assay. Cells were transiently transfected with a bait protein carrying an N-terminal hexahistidine and V5 epitope (HisV5) tag, and a prey protein carrying an N-terminal Myc epitope tag. Bait-containing complexes were pulled down from cell lysates using magnetic beads coated with cobalt, which has a high affinity for histidine. The bait and prey proteins were detected by immunoblotting. The His-tag/cobalt pull-down system offers an advantage over co-immunoprecipitation strategies because the latter require an antibody with high affinity and specificity for the protein of interest, as well as optimization of binding and elution conditions for each antibody-antigen pair.

In cells co-transfected with HisV5-FOXP2 and Myc-FOXP2, Myc-tagged FOXP2 was detected in the affinity-purified (elution) fraction, demonstrating the efficacy of the pull-down (Fig. 3a). However, HisV5-FOXP2 did not pull down Myc-tagged SFPQ, suggesting that SFPQ does not interact with FOXP2 in these cells (Figure 3.3a). In cells transfected with HisV5-tagged FOXP2, NONO, or empty vector together with Myc-FOXP2, an enrichment of Myc-FOXP2 in the

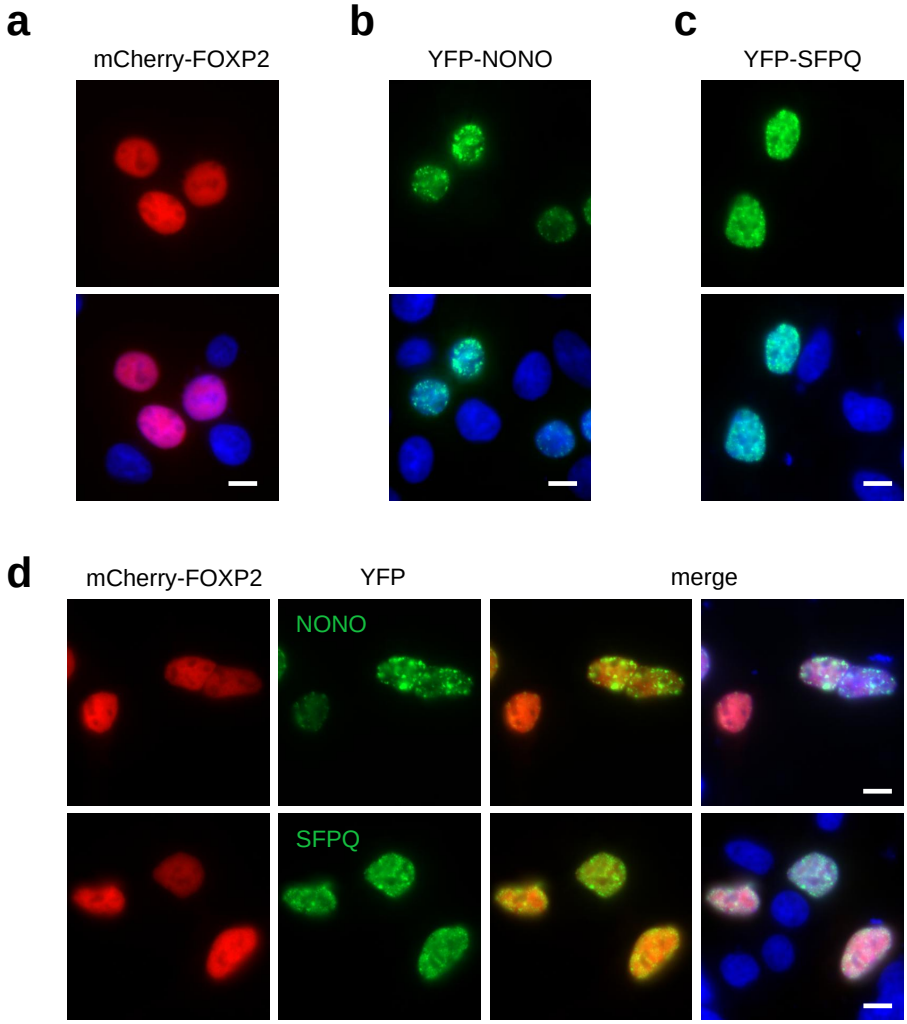


Figure 3.2: Co-localization of FOXP2 with NONO and SFPQ in HEK293 cells. (a-c) Fluorescence micrographs of cells transfected with FOXP2 fused to mCherry **(a)**, NONO fused to YFP **(b)** and SFPQ fused to YFP **(c)**. **(d)** Fluorescence micrographs of cells transfected with FOXP2 fused to mCherry (red) and NONO or SFPQ fused to YFP (green). Nuclei were stained with Hoechst 33342 (blue). Scale bars represent 10 μm.

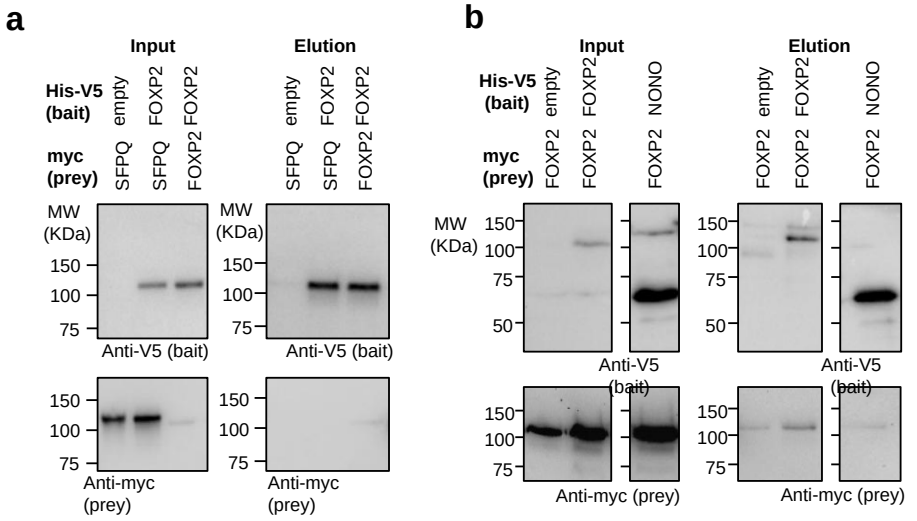


Figure 3.3: Cobalt-affinity pull-down assay for interaction between FOXP2 and NONO/SFPQ. (a) Pull- down assay to test for interaction of FOXP2 with SFPQ. HEK293 cells were transfected with HisV5- tagged FOXP2 (bait) or empty vector together with myc-SFPQ or myc-FOXP2 (prey). His-tagged species were isolated using cobalt affinity purification. Western blots of total lysate (input) and affinity-purified material (elution) were probed with anti-V5 tag antibody to visualize FOXP2 (top panels). Myc-SFPQ and myc-FOXP2 were visualized using anti-myc antibody (bottom panels). **(b)** Pull-down assay to test for interaction of FOXP2 with NONO. HEK293 cells were transfected with HisV5-tagged FOXP2, NONO or empty vector (baits) together with myc-FOXP2 (prey). His-tagged species were isolated using cobalt affinity purification. Western blots of total lysate (input) and affinity-purified material (elution) were probed with anti-V5 tag antibody to visualize NONO and FOXP2 (top panels). Myc-FOXP2 and myc-FOXP2 was visualized using anti-myc antibody (bottom panels).

eluted material was observed only when HisV5-FOXP2 was used as the bait (Figure 3.3b). No enrichment of Myc-FOXP2 compared to the control condition was observed when HisV5-NONO was used as the bait, suggesting that the presence of FOXP2 in the affinity-purified fraction was due to non-specific binding (Figure 3.3b). Thus, the pull-down assay also fails to support the existence of an interaction between FOXP2 and NONO/SFPQ.

3.3.4 *Mammalian Two-Hybrid assay*

Affinity purification can be a reliable technique for detecting protein-protein interactions. However the appropriate experimental conditions must be identified for each specific pair of proteins, because it is critical that the interaction of interest is preserved after the cells have been lysed. Excessively stringent washing steps can result in the loss of specific interactors. Conversely, non-specific binding of the prey protein to the affinity matrix can result in contamination of the eluted material if binding conditions or washing steps are not stringent enough. In some cases it may not be possible to achieve ideal balance of conditions that retain specific interactions while eliminating non-specific ones. Therefore, I employed a third technique, the mammalian two-hybrid (M2H) assay, to assess the putative interaction of FOXP2 with NONO and SFPQ.

In the M2H assay, bait and prey proteins are fused to a GAL4 DNA-binding domain (encoded by the pBIND2 vector) and a VP16 transcriptional activation domain (encoded by the pACT2 vector), respectively. Cells are transfected with the two vectors coding for the fusion proteins, together with a reporter vector that contains tandem GAL4 DNA binding sites upstream of a basic promoter and the firefly luciferase gene. An interaction between the bait and prey proteins physically connects the GAL4 and VP16 domains, creating a functional transcriptional activator that will drive transcription of the reporter firefly luciferase gene (Figure 3.4a, Supplementary Figure S2.5). The sensitivity with which luciferase activity can be detected makes the M2H assay potentially able to pick up weak and transient interactions. In addition, the interaction is not detected in real time, as in the case of the BRET assay; rather it is captured over a longer period in the expression of the luciferase protein, which may also increase sensitivity. The M2H assay must however be used with caution when investigating protein-protein interactions of transcription factors. For example, when fused to the GAL4 DNA-binding domain, a transcription factor may acti-

vate transcription in the absence of any interaction with the VP16 transactivation domain fusion protein, leading to a high background signal that can make the results challenging to interpret.
I first verified that the M2H assay was effective, by using the known interaction between the proteins Id and MyoD as a positive control, and empty bait/prey vectors as a negative control (Finkel *et al.* , 1993). A reporter vector that lacks the GAL4 binding sites served as an additional negative control. These experiments confirmed that an increase in luciferase activity occurred only when both the MyoD and Id fusion proteins, and the reporter construct containing the GAL4 binding sites, were present (Figure 3.4b).

No interaction was observed between FOXP2 and NONO/SFPQ, when FOXP2 was expressed as the bait protein and NONO/SFPQ as the prey (Figure 3.4c). The assay could not be performed in the reverse configuration, with FOXP2 as the prey, because FOXP2 activated luciferase transcription when fused to the VP16 activation domain, suggesting that the reporter plasmid contains binding sites for FOXP2 (data not shown). In contrast, the interaction between SFPQ and NONO was readily detected (Figure 3.4d) (Fox *et al.* , 2005; Knott *et al.* , 2016). Thus, the mammalian two-hybrid assay also provides little support for the existence of an interaction between FOXP2 and NONO/SFPQ.

3.4 DISCUSSION

In the work described in this chapter I sought to define novel FOXP2-interacting proteins by evaluating an existing dataset of putative interactors generated by a previous mass spectrometry screen. The most promising of these putative FOXP2 interactors, NONO and SFPQ, were selected for in-depth follow-up. However, the interaction between FOXP2 and SFPQ/NONO consistently failed to validate, whether using BRET assays, pull-down methods or a mammalian two-hybrid system. The use of these three complementary methodologies, each of which optimally detects protein-protein interactions with different characteristics, reduces the probability that the lack of interaction observed in our validations is a false negative result. Moreover, these assays were able to detect the known interactions of the proteins under investigation, such as the homodimerization of FOXP2 and the heterodimerization of NONO and SFPQ, indicating that the methods used are suitable for detecting interactions of these transcrip-

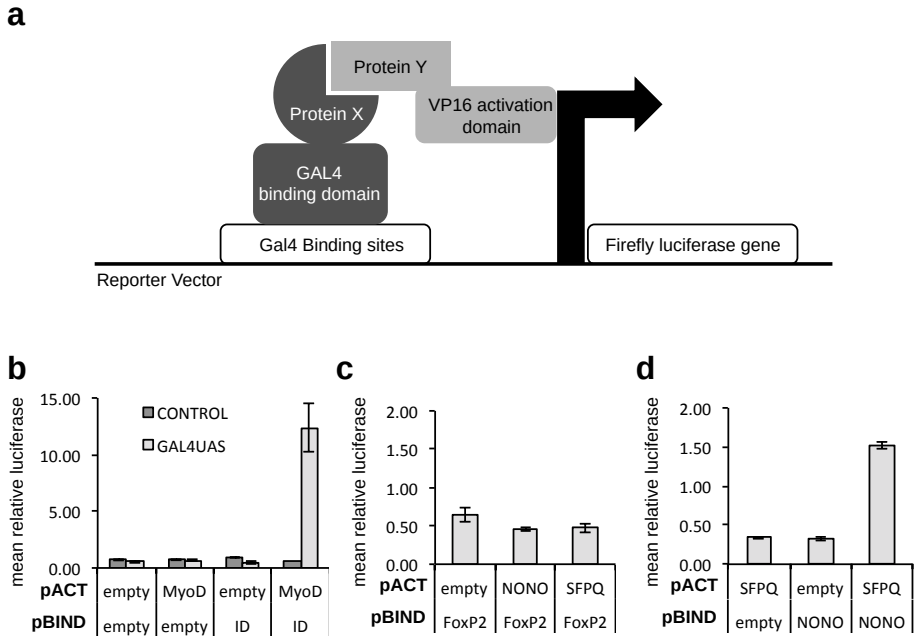


Figure 3.4: Mammalian two-hybrid assay for interaction of FOXP2 with NONO / SFPQ
(a) Schematic representation of the mammalian two-hybrid assay. Cells are transfected with plasmids coding for Proteins "X" and "Y" fused to the GAL4 DNA binding DNA binding domain (pBIND2 vector) and to the VP16 transcriptional domain (pACT2 vector), respectively; and a reporter vector that contains GAL4 DNA binding sites upstream of the firefly luciferase gene. An interaction between proteins "X" and "Y" physically connects the GAL4 DNA-binding domain and the VP16 transactivation domain, reconstructing a functional transcriptional factor that will activate expression of the firefly luciferase reporter gene.**(b)** Validation of the mammalian two-hybrid assay. HEK293 cells were transfected with pBIND2-MyoD or empty pBIND2; pACT2-ID or empty pACT2; the GAL4UAS firefly reporter plasmid or a control reporter lacking the GAL4UAS binding sites; and a normalizer vector that expresses Renilla luciferase. **c** M2H assay for FOXP2 interaction with NONO and SFPQ. HEK293 cells were transfected with pBIND2-FOXP2; pACT2-NONO, pACT2-SFPQ or empty pACT2; the GAL4UAS firefly reporter plasmid; and a normalizer vector that expresses Renilla luciferase. **d** M2H assay for the interaction between NONO and SFPQ. HEK293 cells were transfected with pBIND2-NONO or empty pBIND2; pACT2-SFPQ or empty pACT2; the GAL4UAS firefly reporter; and a normalizer vector that expresses Renilla luciferase. In (b)-(d), values are mean relative luciferase activity ± S.E.M. (n = 3)

tion factors. FOXP2 and SFPQ/NONO also showed different patterns of localization within the nucleus, which is not suggestive of a functional interaction.

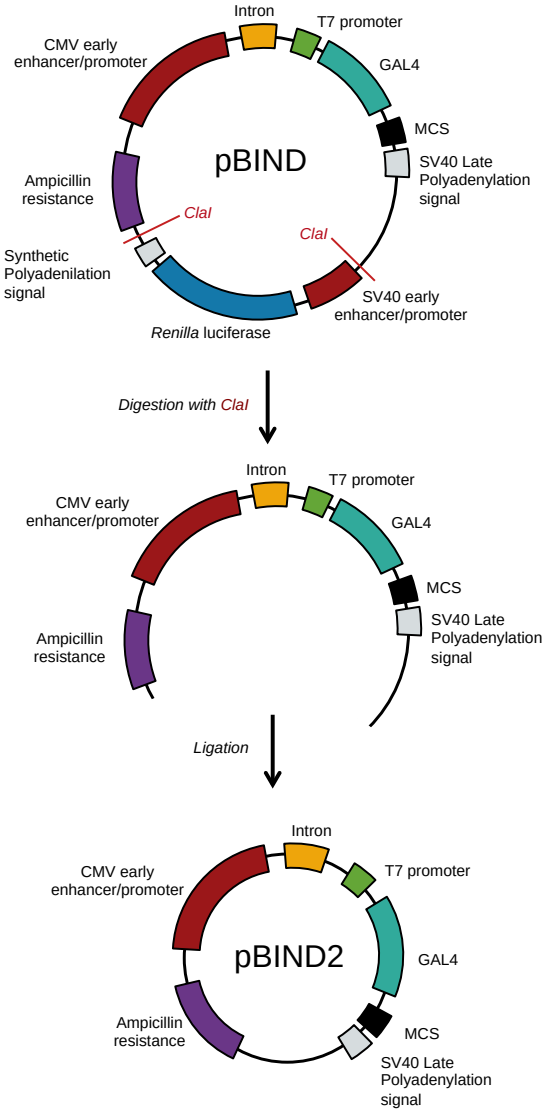
It is therefore likely that NONO and SFPQ were false positive results in the mass spectrometry experiments carried out previously. The immunoprecipitation that was performed to isolate FOXP2-containing protein complexes in these experiments may not have been effective due to the sub-optimal affinity and specificity of the antibody used, or due to the composition of the lysis and wash buffers. Of note, the band representing FOXP2 in the Coomassie-stained polyacrylamide gels of precipitated proteins was not stronger following pull-down with anti-FOXP2 compared to the control condition, casting doubt on the efficiency of the pull-down (Elahi, 2008). In those prior experiments, FOXP2 itself was not identified in the mass spectrometry analysis of FOXP2-affinity purifications from HEK293 cells or mouse brain tissue, indicating inefficient precipitation in these samples. In addition, the known FOXP2 interactors FOXP1 and FOXP4, which would be expected to strongly co-precipitate with FOXP2 in all the conditions, were only identified in one condition (SH-SY5Y cells overexpressing wild-type FOXP2) (Table 3.2). There were also insufficient replicates included for each protein source to test the reproducibility of the results (Elahi, 2008).

The selection of NONO and SFPQ as putative FOXP2 interactors to further investigate was based on multiple factors, including their biological function, since both proteins are neural transcription factors and they are already known to interact with one another (Fox *et al.* , 2005; Knott *et al.* , 2016), and the relative strength of the data, since both proteins had been detected in experiments using two different cell lines (Table 3.2). Following completion of the work described in this chapter, mutations in NONO were reported to lead to neurodevelopmental disorders including intellectual disability, confirming the importance of this transcription factor in brain development (Mircsof *et al.* , 2015; Reinstein *et al.* , 2016). However, the phenotype associated with NONO mutations is not especially similar to that caused by FOXP2 haploinsufficiency. Because NONO and SFPQ were among the most biologically-plausible and highest-confidence candidates from the previous mass spectrometry effort, it appears unlikely that other candidates identified in this screen would prove to be true FOXP2 interactors upon further investigation. Thus, this chapter highlights the need for fresh mass spectrometry-based screens to identify further candidate FOXP2 interactors. Optimization of the pull-down methodology and data analysis pipeline,

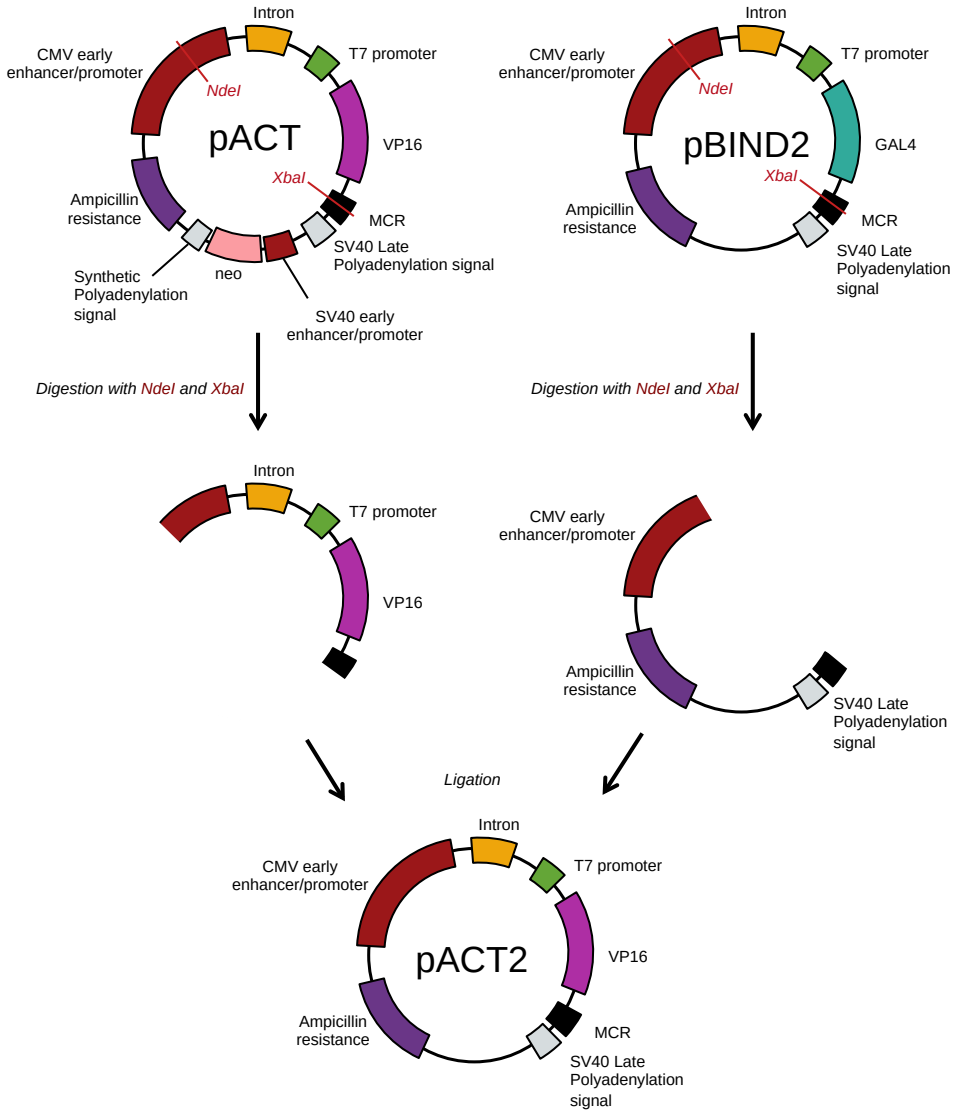
together with use of adequate replicates, represent important factors for improving the sensitivity and specificity of such a screen.

3.5 SUPPLEMENTARY MATERIAL

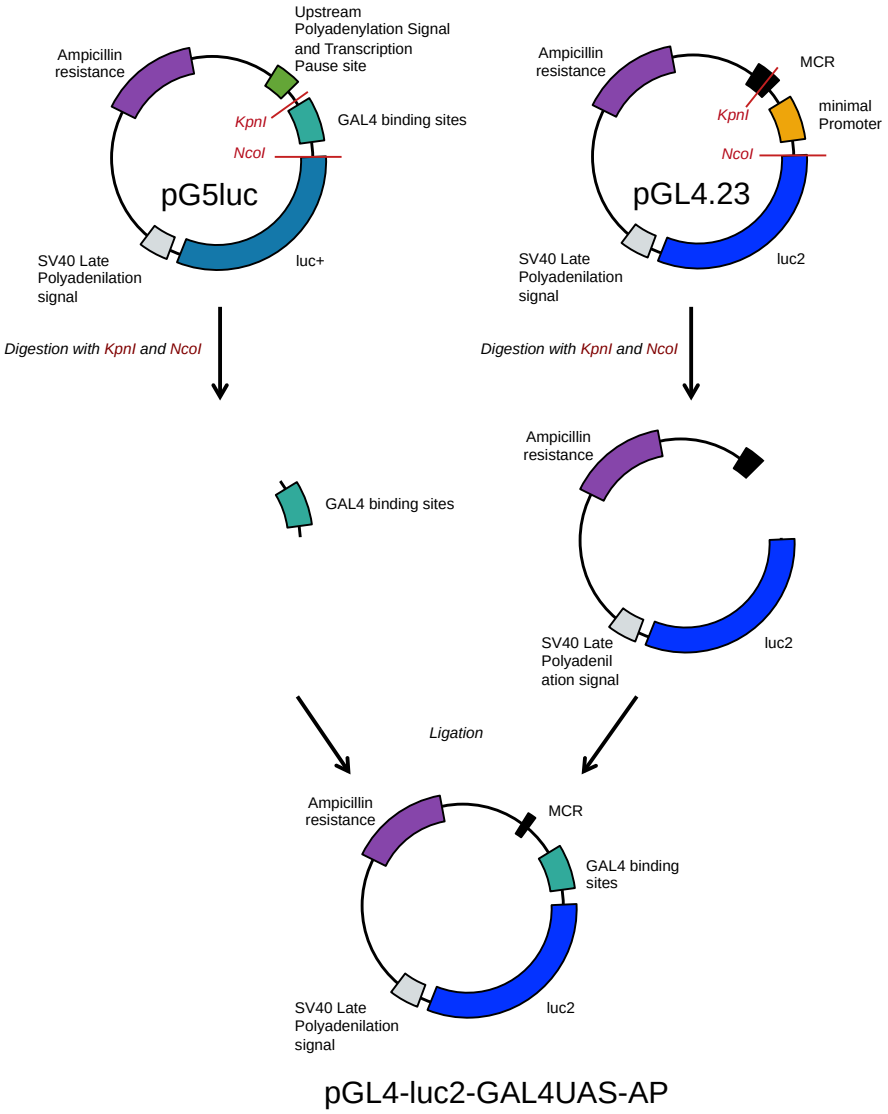
Supplementary Figure S3.1. Cloning strategy to generate M2H plasmid pBIND2.



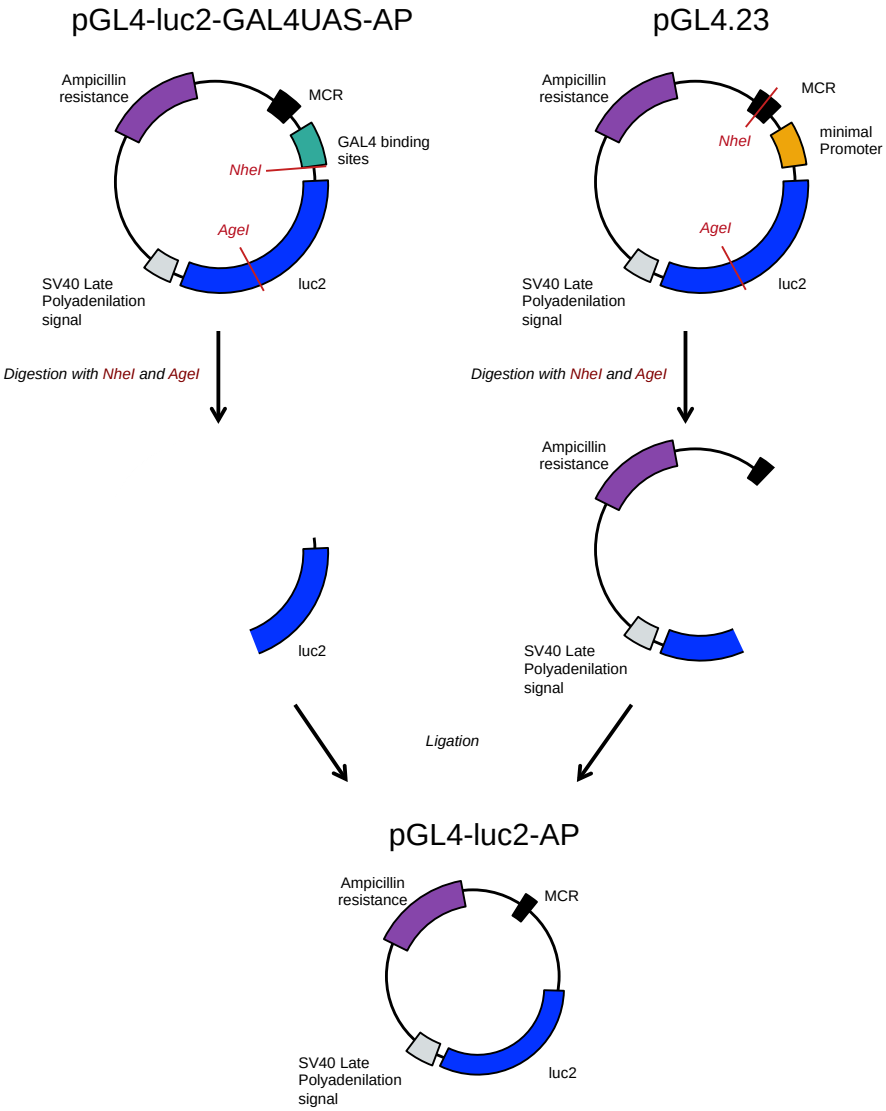
Supplementary Figure S3.2. Cloning strategy to generate M2H plasmid pACT2.



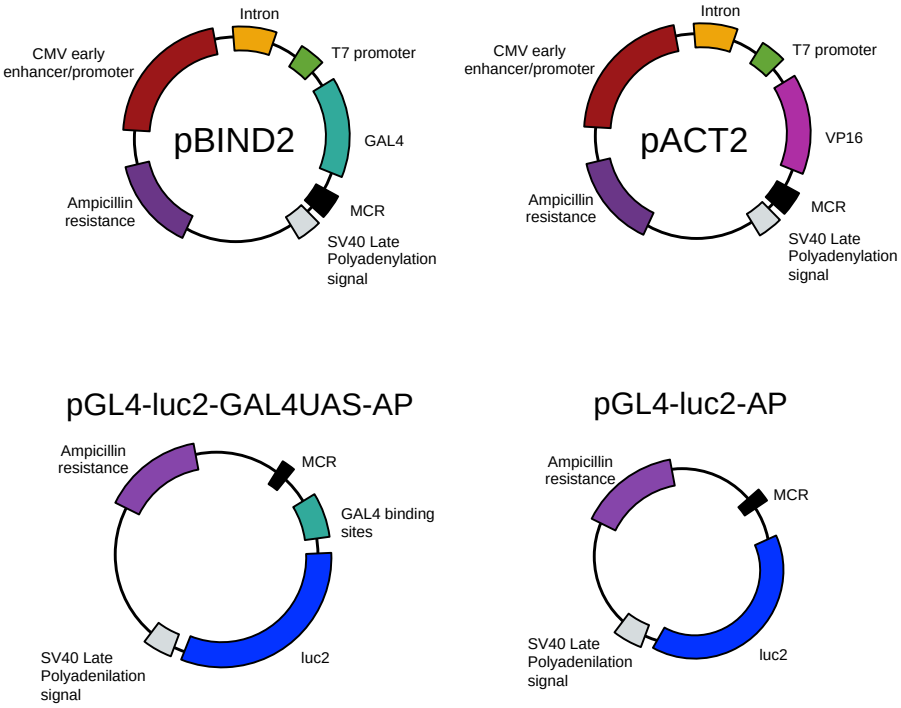
Supplementary Figure S3.3. Cloning strategy to generate M2H plasmid pGL4-luc2-GAL4UAS-AP.



Supplementary Figure S3.4. Cloning strategy to generate M2H plasmid pGL4-luc2-AP.



Supplementary Figure S3.5. Diagram of the plasmids used in the M2H assay.



4

PROTEOMIC ANALYSIS OF FOXP PROTEINS REVEALS INTERACTIONS BETWEEN CORTICAL TRANSCRIPTION FACTORS ASSOCIATED WITH NEURODEVELOPMENTAL DISORDERS[1]

Abstract: *FOXP transcription factors play important roles in neurodevelopment, but little is known about how their transcriptional activity is regulated. FOXP proteins cooperatively regulate gene expression by forming homo- and hetero-dimers with each other. Physical associations with other transcription factors might also modulate the functions of FOXP proteins. However, few FOXP-interacting transcription factors have been identified so far. Therefore, we sought to discover additional transcription factors that interact with the brain-expressed FOXP proteins, FOXP1, FOXP2 and FOXP4, through affinity-purifications of protein complexes followed by mass spectrometry. We identified seven novel FOXP-interacting transcription factors (NR2F1, NR2F2, SATB1, SATB2, SOX5, YY1 and ZMYM2), five of which have well-established roles in cortical development. Accordingly, we found that these transcription factors are co-expressed with FoxP2 in the deep layers of the cerebral cortex and also in the Purkinje cells of the cerebellum, suggesting that they may cooperate with the FoxPs to regulate neural gene expression in vivo. Moreover, we demonstrated that etiological mutations of FOXP1 and FOXP2, known to cause neurodevelopmental disorders, severely disrupted the interactions with FOXP-interacting transcription factors. Additionally, we pinpointed specific regions within FOXP2 sequence involved in mediating these interactions. Thus, by expanding the FOXP interactome we have uncovered part of a broader neural transcription factor network involved in cortical development, providing novel molecular insights into the transcriptional architecture underlying brain development and neurodevelopmental disorders.*

1 This chapter has been published as: Estruch S.B., Quevedo-Calero M., Graham S.A., Vino A., Sollis E., Deriziotis P., Fisher S.E. (2018). Proteomic analysis of FOXP proteins reveals interactions between cortical transcription factors associated with neurodevelopmental disorders. Human Molecular Genetics. doi:10.1093/hmg/ddy035.

4.1 INTRODUCTION

Transcription factors have emerged as a key class of genes disrupted in monogenic forms of neurodevelopmental disorders such as intellectual disability (ID) and autism spectrum disorder (ASD), consistent with the precise temporal and spatial control of gene expression that underpins neurodevelopmental processes (Silbereis *et al.* , 2016). Large-scale next-generation DNA sequencing studies have been particularly successful in identifying monogenic ID/ASD-related disorders caused by high-penetrance *de novo* mutations in transcription factor genes (De Rubeis *et al.* , 2014; Kwan, 2013; Peter *et al.* , 2014).

Among the many families of human transcription factors, the FOXP subfamily of forkhead box proteins is notable for the neurodevelopmental phenotypes which have been associated with their disruption (Bowers & Konopka, 2012b; Charng *et al.* , 2016). The FOXP family includes four proteins: FOXP1, FOXP2, FOXP3 and FOXP4 (Shu *et al.* , 2001). FOXP1, FOXP2 and FOXP4 exhibit 55–65% sequence identity and show overlapping expression in the developing brain, as well as in other organs (Shu *et al.* , 2001; Teufel *et al.* , 2003). FOXP3 is structurally divergent and its expression is limited to T lymphocytes (Fontenot *et al.* , 2003). FOXP3 disruption causes an immunological disorder, IPEX (immunodysregulation polyendocrinopathy enteropathy X-linked, OMIM # 304790) syndrome, while mutations of FOXP1, FOXP2 and FOXP4 have each been linked to distinct neurodevelopmental disorders. Heterozygous disruptions of FOXP1 cause a broad neurodevelopmental syndrome, which includes global developmental delay and ID, frequently accompanied by features of autism and impaired speech and language abilities (OMIM # 613670) (Hamdan *et al.* , 2010; Lozano *et al.* , 2015; O'Roak *et al.* , 2011; Sollis *et al.* , 2016; Srivastava *et al.* , 2014). All pathogenic variants observed to date have occurred *de novo*, consistent with the severe phenotype of the disorder (Sollis *et al.* , 2016). Heterozygous disruptions of FOXP2 cause a rare monogenic form of speech and language impairment including childhood apraxia of speech as a core feature (OMIM # 602081) (Estruch *et al.* , 2016a; Laffin *et al.* , 2012; Lai *et al.* , 2001; MacDermot *et al.* , 2005; Reuter *et al.* , 2016; Roll *et al.* , 2010; Turner *et al.* , 2013). Unlike FOXP1-associated disorder, general cognitive functions may be within the normal range, and both inherited and *de novo* cases have been reported (Graham & Fisher, 2015). Although the FOXP2-related disorder is less severe than that resulting from FOXP1 disruption, individuals with FOXP2 and FOXP1 mu-

tations display overlapping features such as language impairment, suggesting that the syndromes may involve disruption of similar molecular and cellular networks (Bacon & Rappold, 2012).

Disorders relating to FOXP4 disruption had not been reported until recently, when Charng et al. described a homozygous frameshift variant in a child from a consanguineous family affected by developmental delay and malformations in the larynx and the heart (Charng *et al.* , 2016). It is interesting that this, the only currently suspected case of FOXP4-related disorder, involves a recessive mutation, in contrast to the dominant disorders described for FOXP1 and FOXP2 disruptions. The potential complete absence of FOXP4 protein in this child may also explain why the disorder appears to have broader effects in organs other than the brain. Indeed, all three neurally-expressed FOXP proteins also have roles in the development of other organs, but their neurodevelopmental functions appear to show more dosage sensitivity than roles in non-neural tissues (Li, 2004; Li *et al.* , 2015b; Shu *et al.* , 2005, 2007). Although FOXP4 is the strongest candidate causal gene in the case identified by Charng and colleagues, additional observations of FOXP4 disruption are necessary to confirm its etiological role, particularly since variants in two other genes were also found in a homozygous state in the child and heterozygous state in the parents (Charng *et al.* , 2016).

The precise mechanisms by which FOXP family members regulate transcription are just beginning to be explored. Several studies have investigated FOXP2 and FOXP1 target genes in the brain (Araujo *et al.* , 2015; Spiteri *et al.* , 2007; Vernes *et al.* , 2007, 2011) but genes regulated by FOXP4 remain to be discovered. An important mechanism in regulation of gene expression is that transcription factors act in a combinatorial fashion, and thus can generate the complex patterns of gene expression underlying development of the brain and other systems using a limited set of transcription factors (Silbereis *et al.* , 2016; Smith & Matthews, 2016). The genes regulated by the FOXP proteins in specific cell types are therefore likely to depend on the co-expression of other transcription factors. Indeed, the transcriptional activity of FOXP1, FOXP2 and FOXP4 is regulated by their ability to form homo- and hetero-dimers with each other through their leucine zipper domain (Li *et al.* , 2004). The three proteins show partially overlapping patterns of expression in the brain, such that different combinations of FOXP homo and hetero-dimers may regulate distinct target

genes (Ferland *et al.* , 2003; Lai *et al.* , 2003; Sin *et al.* , 2015; Takahashi *et al.* , 2008).

In addition to dimerization between FOXP family members, a small number of interactions between FOXP proteins and other transcription factors have been reported. FOXP1 and FOXP2 interact with the neural transcription factor TBR1, rare mutations of which cause ASD accompanied by language deficits (Deriziotis *et al.* , 2014a). Notably, this interaction can be disrupted by etiological mutations in either FOXP2 or TBR1 (Deriziotis *et al.* , 2014a). Other reported FOXP-interacting transcription factors include GATAD2B and NKX2.1 (Chokas *et al.* , 2010; Zhou *et al.* , 2008). We hypothesized that there may be further interactions between FOXPs and other transcription factors with relevance to physiological developmental processes and neurodevelopmental disorders. We therefore sought to identify transcription factors which may cooperate with FOXP proteins in regulating gene expression during neurodevelopment, by applying a mass spectrometry approach. We identified seven novel FOXP-interacting transcription factors: NR2F1, NR2F2, SATB1, SATB2, SOX5, YY1 and ZMYM2. Several of these novel interactors have well-established roles in the development of the nervous system and/or are associated with neurodevelopmental disorders. The interactions with these binding partners involve different regions of the FOXP2 polypeptide, and are variably affected by different pathogenic variants in FOXP1 and FOXP2. Thus our findings provide new clues to the molecular function of the FOXP proteins, and point to a network of transcription factors involved in cortical development, disruption of which manifests as neurodevelopmental disorder.

4.2 MATERIALS AND METHODS

DNA constructs

The cloning of human FOXP2 (NM_ 014491), FOXP1 (NM_ 032682), FOXP4 (NM_ 001012426), has been described previously (Estruch et al., 2016b). The coding sequences of TFDP1 (NM_ 007111), NR2F1 (NM_ 005654), NR2F2 (NM_ 021005), SATB1 (NM_ 002971), SATB2 (NM_ 001172509), ZBTB2 (NM_020861), TP53 (NM_000546), ZNF687 (NM_001304763), ZMYM2 (NM_003453), NFAT5 (NM_ 138714), YY1 (NM_ 003403) and SOX5 (NM_ 001330785) were amplified using the primers listed in Supplementary Table S4.1 and cloned into pCR2.1-

TOPO. Human fetal brain cDNA was used as template for PCR, except in the cases of NR2F2 and NFAT5, which were amplified from the plasmids pAAV-hNR2F2-RFP (Addgene, # 22926) and pEGFP-NFAT5 (Addgene, # 13627), respectively. The cloning of FOXP2 variants p.Q17L, p.M406T, p.P416T, p.R553H, p.N597H, p.R328* , P.Q390Vfs*7, of synthetic truncated forms of FOXP2, and of FOXP1 variants p.A339Sfs*4, p.V423Hfs*37, p.Y439* , p.R465G, p.R514C, p.R252* , and p.W534R has been described previously (Estruch *et al.* , 2016a; Sollis *et al.* , 2016). For expression of FLAG-tagged protein for affinity purification, the coding regions of FOXP1, FOXP2 and FOXP4 were subcloned into the pEF1α FLAGbio expression vector, provided by Dr. J.W. Wang (Kim *et al.* , 2009b). For expression of fusion proteins with *Renilla* luciferase, yellow fluorescent protein (YFP) and mCherry tags, cDNAs were subcloned into the pLuc, pYFP, and pmCherry expression vectors, respectively, which have been described previously (Deriziotis *et al.* , 2014a; Estruch *et al.* , 2016a). All constructs were verified by Sanger sequencing.

Cell culture and transfection

HEK293 cells were obtained from ECACC (cat. No. 85120602) and cultured in DMEM supplemented with 10% fetal bovine serum. Transient transfections were performed using GeneJuice (Merck-Millipore) according to the manufacturer's instructions. To generate stable cell lines expressing FLAG-tagged FOXP proteins, HEK293 cells were transfected with pEF1αFLAGbio containing the coding sequence of FOXP1, FOXP2, or FOXP4 using GeneJuice (Merck-Millipore) according to the manufacturer's instructions. Single clones were isolated following selection with culture medium containing 10 μM puromycin. Expression of the tagged proteins in selected clones was confirmed by western blotting of cell lysates using anti-FLAG antibody (Sigma). Selected stable cell lines were maintained in DMEM supplemented with 10% fetal bovine serum and 5 μM puromycin.

Affinity Purification and Mass Spectrometry (AP/MS)

HEK293 cell lines stably expressing FLAG-tagged FOXP1, FOXP2 or FOXP4, or control cell lines, were expanded to twenty 15 cm dishes, washed with PBS, and

harvested by scraping. The affinity purification 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, 0.2 mM EDTA, 1.5 mM $MgCl_2$, 100 mM KCl, 20% glycerol) and then incubated with anti-FLAG M2 agarose beads (Sigma) and benzonase (Novagen) for 3 h at 4° C. Beads were washed 5 times for 5 min with buffer C-100 containing 0.02% NP-40 (C-100∗) and bound proteins were eluted 4 times for 15 min at 4° C with buffer C-100∗ containing 0.2 mg/ml FLAG-tripeptide (Sigma). The purification of FOXP1, FOXP2 or FOXP4 was checked by western blot using anti-FLAG antibody. The elution fractions were pooled and proteins were TCA-precipitated, resolved by SDS-polyacrylamide gel electrophoresis, stained with Colloidal Blue Staining Kit (Invitrogen) and analyzed by mass spectrometry as described previously (Engelen *et al.* , 2011; Gagliardi *et al.* , 2013; Ninkovic *et al.* , 2013; van den Berg *et al.* , 2010).

Mass-spectrometry analyses were performed on three independent purifications from FOXP1- and FOXP2-expressing cell lines, and two independent purifications from FOXP4-expressing cell lines. Each purification was performed in parallel with stable cell lines transfected with empty control vector. To remove non-specific interactors, peptide lists were filtered using peptides found in the control conditions or in the CRAPome database, a collection of common contaminants in AP/MS data (Mellacheruvu *et al.* , 2013). Proteins with a Mascot score lower than 45 were also excluded.

Western blotting

Proteins from cell lysates were resolved on 10% SDS-polyacrylamide gels and transferred to PVDF membranes using a TransBlot Turbo Blotting apparatus (Bio-Rad). Membranes were blocked in PBS containing 5% milk and 0.1% Tween-20 and incubated overnight at 4° C with primary antibody. After washing, membranes were incubated with horseradish peroxidase-conjugated goat anti-mouse or anti-rabbit IgG for 45 min at room temperature. Proteins were visualized using Novex ECL Chemiluminescent Substrate Reagent Kit (Invitrogen) and a ChemiDoc XRS+ imaging system (Bio-Rad).

BRET assay

BRET assays were performed as described previously (Deriziotis et al., 2014a). Briefly, HEK293 cells were transfected in 96-well plates with plasmids encoding YFP- and luciferase-fusion proteins. After 36-48 h, Enduren live cell luciferase substrate (Promega) was added at a final concentration of 60 μ M. Cells were cultured for a further 4 h, and emission readings (integrated over 10 s) were taken using a TECAN F200PRO microplate reader using the Blue1 and Green1 filter sets. Expression levels of the YFP-fusion proteins were measured by taking fluorescent readings using the filter set and dichroic mirror suitable for green fluorescent protein (excitation 480 nm, emission 535 nm). The corrected BRET ratio was calculated with the following formula: $[\text{Green1}_{\text{(experimental condition)}}/\text{Blue1}_{\text{(experimental condition)}}] - [\text{Green1}_{\text{(control condition)}}/\text{Blue1}_{\text{(control condition)}}]$. The control conditions used luciferase or YFP fused to a C-terminal nuclear localization signal (Deriziotis et al., 2014a).

Fluorescence microscopy

HEK293 cells were seeded on coverslips coated with poly-L-lysine. Cells were transfected with plasmids encoding proteins of interest fused to YFP or mCherry. Cells were cultured for 30 h post-transfection, and then fixed with 4 % paraformaldehyde. Nuclei were stained with Hoechst 33342. Fluorescence images were acquired using an Axiovert A-1 fluorescent microscope with ZEN Image software (Zeiss).

Immunofluorescence

Immunofluorescence experiments were performed on mouse brain sections at postnatal day 3 (P3). Brains were harvested and embedded in optimal cutting temperature (OCT) compound on dry ice. Sagittal sections were prepared at a thickness of 4 μ m using a Leica CM1950 cryostat and then preserved at -20° C. Tissue sections were fixed in ice-cold acetone for 10 minutes at -20° C, blocked using 10% donkey serum in PBS for 1 h at room temperature, and incubated overnight at 4° C with primary antibody diluted in 2% donkey serum. The following antibodies were used: goat anti-FOXP2 N-16 antibody (sc-21069,

Santa Cruz), rabbit anti-FOXP2 antibody (ab16046, Abcam), rabbit anti-SOX5 (ab94396, Abcam), goat anti-SATB1 E-15 antibody (sc-5990, Santa Cruz), rabbit anti-SATB2 antibody (ab34735, Abcam), rabbit anti-NR2F1 (ab181137, Abcam) and rabbit anti-NR2F2 (ab42672, Abcam). After washing, tissue sections were incubated with fluorescently-labeled secondary antibodies diluted in 2% donkey serum: donkey anti-rabbit IgG Alexa488 (a21206, Invitrogen) and donkey anti-goat IgG Alexa594 (a11058, Invitrogen). Slides were mounted with VectaShield Antifade Mounting Medium with DAPI (Vector Labs) and then imaged using Axiovert A-1 fluorescent microscope with ZEN Image software (Zeiss).

4.3 RESULTS

4.3.1 *Identification of FOXP-interacting transcription factors*

To identify transcription factors that may cooperate with FOXP proteins to regulate gene expression, we generated HEK293 cell lines stably expressing FOXP1, FOXP2 or FOXP4 fused to an N-terminal FLAG tag. HEK293 cells endogenously express all three FOXP proteins as well as many other neural genes (Shaw *et al.* , 2002; Stepanenko & Dmitrenko, 2015). FLAG affinity-purified protein complexes from these cell lines were analyzed by mass spectrometry (Figure 4.1a,b). Three independent experiments were performed for FOXP1 and FOXP2, and two experiments for FOXP4. After filtering out non-specific interactors, a total of 381 putative FOXP-interacting proteins were identified, with substantial overlap between the three FOXP proteins (Figure 4.1c, Supplementary Table S4.2). For each FOXP protein, the other two FOXPs were among the interacting proteins identified in each experiment, reflecting heterodimerization between FOXP family members (Li *et al.* , 2004) (Supplementary Table S4.2). Moreover, the set of putative FOXP-interacting proteins included previously reported interaction partners such as CTBP1, CTBP2 and GATAD2B, confirming that the affinity purification procedure retrieved physiologically relevant FOXP-interacting proteins (Chokas *et al.* , 2010; Estruch *et al.* , 2016a) (Supplementary Table S4.2).

The transcription factors among the putative FOXP-interacting proteins were identified, resulting in a set of 28 proteins (Figure 4.1c, Table 4.1, Supplementary Table S4.2). To prioritize potential interacting transcription factors of most interest for further investigation, we excluded previously reported FOXP inter-

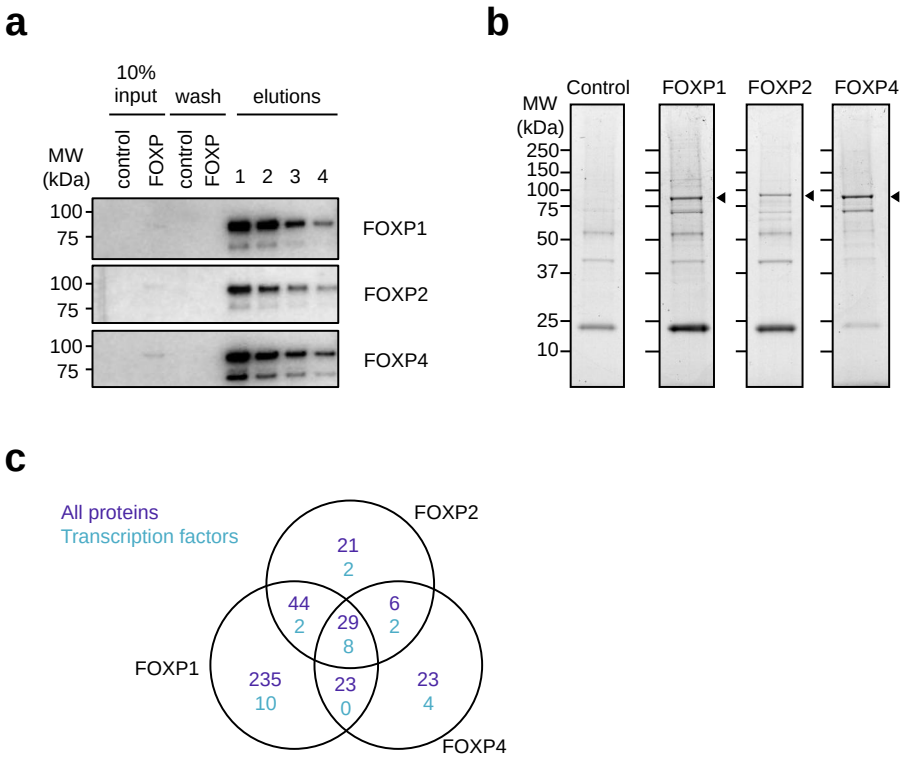


Figure 4.1: Identification of FOXP1-, FOXP2- and FOXP4-interacting proteins. (a) Affinity purifications of FOXP protein complexes using HEK293 cell lines stably expressing FLAG- FOXP1 (top panel), -FOXP2 (middle panel) and -FOXP4 (bottom panel) protein complexes. FLAG-tagged species were isolated using a FLAG affinity purification strategy. Western blots of total lysate (10% Input), washed proteins (wash), and affinity-purified material (elutions) were probed with an anti-FLAG antibody. **(b)** Coomassie-stained SDS- polyacrylamide gels of the affinity purifications of FOXP1, FOXP2, FOXP4 and control purification. Arrows indicate bands corresponding to FOXP1, FOXP2 or FOXP4 for each IP. **(c)** Venn diagram showing the overlaps between the proteins (purple) or transcription factors (light blue) identified in each FOXP purification.

actors, and then selected proteins that were present in two or more replicates, or that had a known involvement in neurodevelopmental disorder. Because the FOXP proteins have high sequence similarity and are likely to have shared in-

teractors, the appearance of a protein in two experiments with different FOXPs was treated as a replicate. We also included transcription factors if two or more members of the same family of transcription factors were observed (Supplementary Table S4.3). This selection resulted in a filtered list of 12 putative FOXP interactors for follow up.

4.3.2 *Validation of interactions between FOXPs and transcription factors*

To validate the selected putative FOXP-interacting transcription factors, we used a Bioluminescence Resonance Energy Transfer (BRET) assay. The BRET assay allows protein-protein interactions to be observed in live cells, and its effectiveness has previously been shown for successfully confirming interactions between FOXPs and other proteins (Deriziotis *et al.* , 2014b,a; Estruch *et al.* , 2016b,a; Sollis *et al.* , 2016). In the BRET assay, cells are transfected with a protein of interest fused to *Renilla* luciferase, and a candidate interaction partner fused to yellow fluorescent protein (YFP). An interaction between the two proteins under investigation can bring the luciferase and YFP moieties into sufficient proximity for resonance energy transfer to occur, causing a shift in the wavelength of the emitted light.

We generated YFP-fusion proteins for the 12 putative FOXP interactors, and tested them for interaction with luciferase-fusions of FOXP1, FOXP2 and FOXP4 using the BRET assay. We detected interactions between FOXP proteins and seven of the 12 putative interactors: SOX5, SATB1, SATB2, NR2F1, NR2F2, YY1 and ZMYM2 (Figure 4.2a,b). In some cases a specific FOXP-cofactor interaction was detected in the BRET assay that was not identified in the mass spectrometry screen. For example, NR2F1, NR2F2 and SATB1 (identified through FOXP4 interactome screening) interacted with FOXP1 and FOXP2 in the BRET but had not been detected by mass spectrometry for those two FOXP proteins (Figure 4.2b). This is likely due to the greater sensitivity of the targeted BRET investigations; these proteins may not have been detected among the complex mixture of proteins isolated by affinity purification. Transfection of cells with YFP-fusions of the interactors validated using the BRET assay and mCherry-FOXP2 showed that all seven interactors exhibited nuclear localization that completely overlapped with FOXP2 (Figure 4.2c). ZMYM2 and SOX5 formed nuclear speckles when expressed as YFP-fusions, which caused partial re-distribution of co-

Table 4.1: List of candidate FOXP-interacting transcription factors identified in the mass spectrometry screens
* Tentative association

Gene Symbol	Function	Associated Phenotype
BNC2	Regulation of skin pigmentation	
CDC5L	Regulation of cell cycle	
EMSY	DNA repair	
FOXC1	Embryonic and ocular development	Iris hypoplasia and glaucoma
FOXP1	Brain, heart and lung development	ID with language impairment
FOXP2	Brain, heart and lung development	Severe language impairment
FOXP4	Brain, heart and lung development	Developmental delay*
FUBP1	Regulation of cell cycle	
GATAD2B	Brain development	ID
HIC2	Unknown function	
NFAT5	Regulation of inflammatory response	
NR2F1	Brain and heart development	ID with optic atrophy
NR2F2	Brain and heart development	Congenital heart defects
RFX1	Unknown function	
SATB1	Regulation of immune system and brain development	
SATB2	Brain and bone development	ID with language impairment
SOX13	T-cell development	
SOX5	Brain and bone development	ID with language impairment
TFDP1	Regulation of cell cycle	ID with language impairment*
TP53	Regulation of cell cycle	Li-Fraumeni syndrome
YY1	Embryonic development	ID
ZBTB2	Regulation of cell cycle	
ZBTB39	Unknown function	
ZFHX4	Brain and muscle development	Ptosis
ZMYM2	Immune system regulation	
ZNF148	Unknown function	ID with language impairment
ZNF687	Bone development	Paget disease of bone

expressed FOXP2, consistent with the presence of a physical interaction between ZMYM2/SOX5 and FOXP2 (Figure 4.2c).

Five candidate interactors (NFAT5, TFDP1, TP53, ZBTB2 and ZNF687) did not show evidence of interaction in the BRET assay. The lack of interaction in this assay was not due to the candidate interactors being localized outside the nucleus in live cells, as all these proteins showed total or partial nuclear localization, overlapping with that of FOXP2 (Figure 4.2c). NFAT5, ZBTB2 and TFDP1 were present in the cytoplasm, as has been reported previously, but still showed overlapping expression with FOXP2 in the nucleus (Ishida *et al.* , 2005; Jeon *et al.* , 2009) (Figure 4.2c). The failure to observe an interaction with these proteins in the BRET assay does not exclude that these proteins might be true FOXP interactors. For a signal to be detected in the BRET assay, the luciferase and YFP peptides need to be in close proximity and also must be oriented correctly so that resonance energy transfer may occur. Therefore certain pairs of interactors may not allow for efficiency resonance transfer (Sun *et al.* , 2016). In addition, the affinity purification procedure can purify large complexes of proteins, some of which may interact only indirectly with the bait protein (Budayeva & Cristea, 2014; Hayes *et al.* , 2016). Pairs of proteins that interact indirectly are less likely to produce a signal in the BRET assay because the interaction-mediating proteins are not overexpressed, and the distance between the protein pair may also be too great for efficient energy transfer. Therefore, NFAT5, TFDP1, TP53, ZBTB2 and ZNF687 may have been isolated by affinity purification due to indirect interaction with the FOXP bait protein.

4.3.3 *Neural transcription factors interact with distinct sites within FOXP proteins*

We next sought to determine which FOXP regions are involved in the interactions with each of the transcription factors validated using the BRET assay. Here, we focused on FOXP2, building on our prior experience with mapping interaction sites within this particular protein. Specifically, we performed further BRET assays using a series of synthetic variants of FOXP2 isoform I (NM_014491.3, NP_055306.1) truncated at the N- or C-terminus, which have been employed previously to map the binding sites of FOXP2 interaction partners (Figure 4.3a) (Deriziotis *et al.* , 2014a; Estruch *et al.* , 2016a,b). For SOX5 and ZMYM2, an N-terminal FOXP2 fragment containing residues 1-258 of isoform I retained the ability to interact, while the counterpart C-terminal fragment

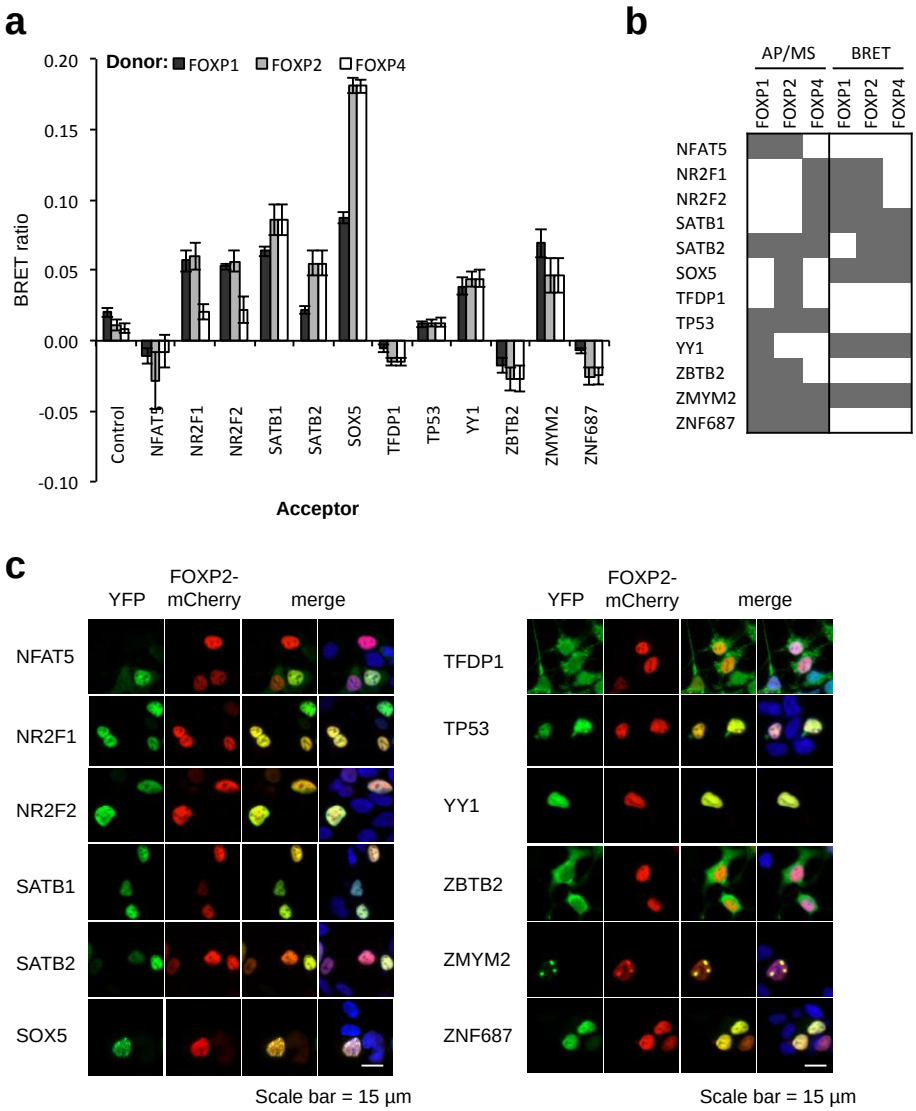


Figure 4.2: Validation of identified FOXP-interacting proteins. (a) Bioluminescence resonance energy transfer (BRET) assay for interaction of FOXP1, FOXP2 and FOXP4 variants with the putative FOXP-interacting transcription factors. HEK293 cells were transfected with Renilla luciferase-FOXP1, -FOXP2 or –FOXP4 (donor) and YFP (acceptor) fusion proteins of the putative FOXP-interacting transcription factors. The control acceptor protein is a nuclear-targeted YFP. Values are mean corrected BRET ratios ± S.E.M (n = 3).

Figure 4.2: (Cont.) **(b)** Heatmap summarizing the results of the affinity purification followed by mass spectrometry (AP/MS) and the BRET assays. Grey shading indicates interaction. **(c)**Fluorescence micrographs of HEK293 cells transfected with YFP-fusions of the putative FOXP-interacting transcription factors, together with FOXP2 fused to mCherry. Nuclei were stained with Hoechst 33342 (Scale bar = 15 μm).

(residues 259-715) did not, indicating that the critical binding determinants for SOX5 and ZMYM2 lie within the N-terminal 258 residues of FOXP2 (Figure 4.3b,c). In addition, SOX5 and ZMYM2 interacted with a naturally-occurring alternative FOXP2 isoform (isoform III) that lacks the N-terminal 92 amino acids of other isoforms (Supplementary Figure S4.1), narrowing down the critical region to residues 93 to 258.The N-terminal region of FOXP2 also mediates interaction with the transcription factor TBR1, and with members of the PIAS family of proteins which are involved in FOXP2 post-translational modification, which suggests that this region may coordinate multiple protein-protein interactions (Deriziotis *et al.* , 2014a; Estruch *et al.* , 2016b). The region includes two polyglutamine tracts of unknown function that are expanded in FOXP2 relative to other FOXP family members. Shortening of these tracts to bring them into line with the sequence of FOXP1 does not affect interaction with TBR1 or PIAS proteins in cellular assays (Deriziotis *et al.* , 2014a; Estruch *et al.* , 2016b). Tract shortening also did not affect interaction with SOX5 and ZMYM2 in the current study, as expected given that these proteins also interact with FOXP1 (Supplementary Fig. S4.2). Thus, the binding sites for these proteins may lie in the glutamine-rich regions flanking the polyglutamine tracts

For the paralogous transcription factors SATB1 and SATB2, only C-terminally truncated FOXP2 forms containing an intact forkhead DNA-binding domain retained a full ability to interact with SATB1/2, suggesting that the interaction requires an intact forkhead domain together with additional elements from the N-terminal region of FOXP2 (Figure 4.3d). Interestingly, a previous mass spectrometry study reported SATB1/2 in affinity purifications of FOXP1 and FOXP3 as well as of FOX proteins from several other subfamilies (Li *et al.* , 2015a). Given that SATBs interact with diverse FOX proteins that have substantial structural divergence, it is highly likely that the forkhead DNA-binding domain itself plays an important role in mediating the interaction.

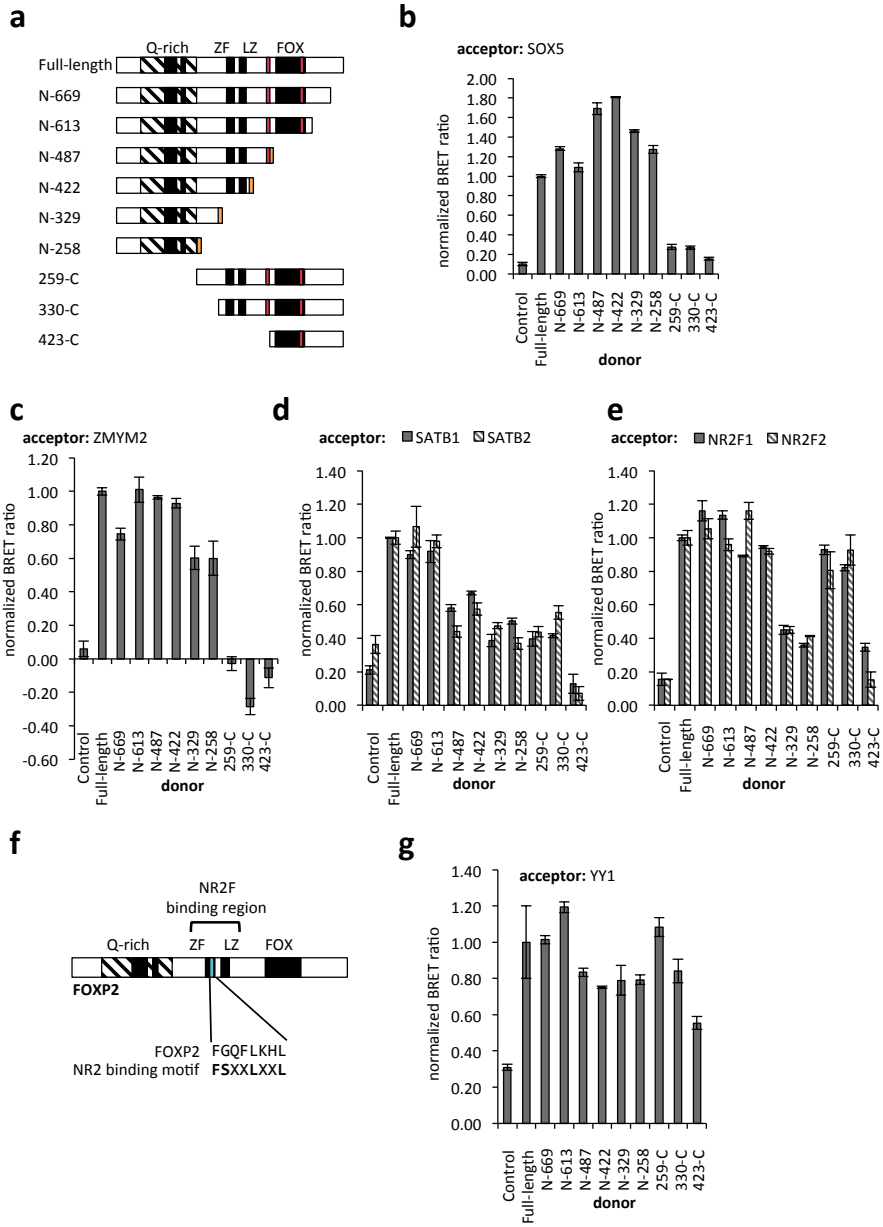

a
Q-rich
ZF
LZ
FOX
Full-length
N-669
N-613
N-487
N-422
N-329
N-258
259-C
330-C
423-C
b
acceptor: SOX5
normalized BRET ratio
donor
c
acceptor: ZMYM2
d
acceptor: SATB1 SATB2
e
acceptor: NR2F1 NR2F2
f
NR2F
binding region
FOXP2
FOXP2 FGQFLKHL
NR2 binding motif FSXXLXXL
g
acceptor: YY1
Control

Figure 4.3

Figure 4.3: Mapping of the regions in FOXP2 that mediate protein-protein interactions (a) Schematic representation of synthetic truncated forms of FOXP2 isoform I (NM_014491.3, NP_055306.1). Known domains are labeled: glutamine-rich (Q-rich) region (hatched shading) including long and short poly-glutamine tracts, zinc finger (ZF), leucine zipper (LZ), and forkhead domain (FOX). Nuclear localization signals are indicated with red bars. A synthetic nine-residue nuclear targeting sequence (orange bars) was appended to the C-terminus of variants that lack one or both of the endogenous nuclear localization signals. **(b-e,g)** BRET assay for interaction of synthetic truncated FOXP2 variants with FOXP-interacting transcription factors. HEK293 cells were transfected with truncated FOXP2 variants fused to Renilla luciferase (donor) and SOX5 **(b)**, ZMYM2 **(c)**, SATB1 **(d)**, SATB2 **(d)**, NR2F1 **(e)**, NR2F2 **(e)** or YY1 **(g)** fused to YFP (acceptor). The control donor protein is a nuclear-targeted luciferase. Values are mean corrected BRET ratios normalized to full-length FOXP2 ± S.E.M. (n = 3). **(f)** Schematic representation of FOXP2 showing the putative NRF2 binding motif (light blue bar). The sequence alignment of the region shows the putative NRF2 binding motif in FOXP2 paralogues FOXP1 and FOXP4.

The second pair of paralogous transcription factors in our screen, NR2F1 and NR2F2, displayed normal interaction with a FOXP2 form truncated beyond residue 422, but reduced interaction with more severely truncated forms, suggesting a role for the region encompassing residues 330-422 (Figure 4.3e). Accordingly, normal interaction was observed with a C-terminal fragment of FOXP2 beginning at residue 330, whereas a fragment beginning at residue 423 showed little or no binding (Figure 4.3e). Residues 330-422 include a leucine zipper domain which mediates dimerization, and a zinc finger domain with unknown function.

FOXP2 dimerization does not appear to be essential for interaction with NR2F1/2, because we found that a synthetic dimerization-deficient FOXP2 variant retained the ability to interact with NR2F1 (Supplementary Figure S4.2d) (Li *et al.* , 2004). Interestingly, the zinc finger domain in FOXP2 contains a short amino acid sequence (residues 360-367) with high similarity to the NR2 binding motif F/YSXXLXXL/Y, which has been shown to mediate the interaction of NR2F1/2 with other transcription factors and co-repressors (Chan *et al.* , 2013) (Figure 4.3f). This motif is conserved in FOXP1 but not FOXP4, which could explain why NR2F1/2 interacted with FOXP1 and FOXP2, but not FOXP4, in our BRET assays (Figure 4.2a,b) – the presence of NR2F1/2 in FOXP4 complexes analyzed

by mass spectrometry may then be due to an indirect interaction mediated by FOXP1/2.

Finally, in the case of YY1, all of the truncated FOXP2 variants tested retained some degree of interaction, suggesting that the interaction between FOXP2 and YY1 may involve multiple binding sites (Figure 4.3g).

4.3.4 *Disorder-related FOXP variants disrupt interactions with transcription factors*

We next investigated etiological variants of FOXP proteins related to neurodevelopmental disorders, assessing their effects on the transcription factor interactions that we identified. Heterozygous disruptions of the *FOXP2* gene cause a rare form of speech and language disorder (Graham & Fisher, 2015). The disorder was first reported in a three-generation family, in which all of the affected members carry the FOXP2 variant p.R553H, which is unable to bind to its typical DNA target sequence (Lai *et al.* , 2001; Vernes *et al.* , 2006). Several other variants in FOXP2 have since been reported in individuals with neurodevelopmental disorders; functional studies indicate that some of these are etiological variants whereas others are benign rare variants (Figure 4.4a) (Estruch *et al.* , 2016a; Laffin *et al.* , 2012; MacDermot *et al.* , 2005; Reuter *et al.* , 2016; Roll *et al.* , 2006; Turner *et al.* , 2013). Etiological variants had been shown to severely disrupt core molecular aspects of FOXP2 protein function, such as transcriptional activity and subcellular localization (Estruch *et al.* , 2016a). Using the BRET assay, we examined the effects of seven rare variants on the interaction of FOXP2 with the transcription factors identified in our interactor screen (Figure 4.4b-h). Four of these seven rare variants had been shown to be benign in prior cellular assays (Estruch *et al.* , 2016a). All four of these benign rare variants displayed normal interaction with the novel interaction partners, whereas almost all interactions involving one of the three pathogenic FOXP2 variants were either abolished or disrupted (Figure 4.4b-i).

Heterozygous mutations of the *FOXP1* gene result in a neurodevelopmental syndrome characterized by mild to moderate ID with features of ASD and speech and language impairments (Lozano *et al.* , 2015; Sollis *et al.* , 2016). We examined the effects of seven etiological variants of FOXP1 on interactions with the six FOXP1-interacting transcription factors identified in our screen. The etiological FOXP1 variants comprised four truncated and three missense variants.

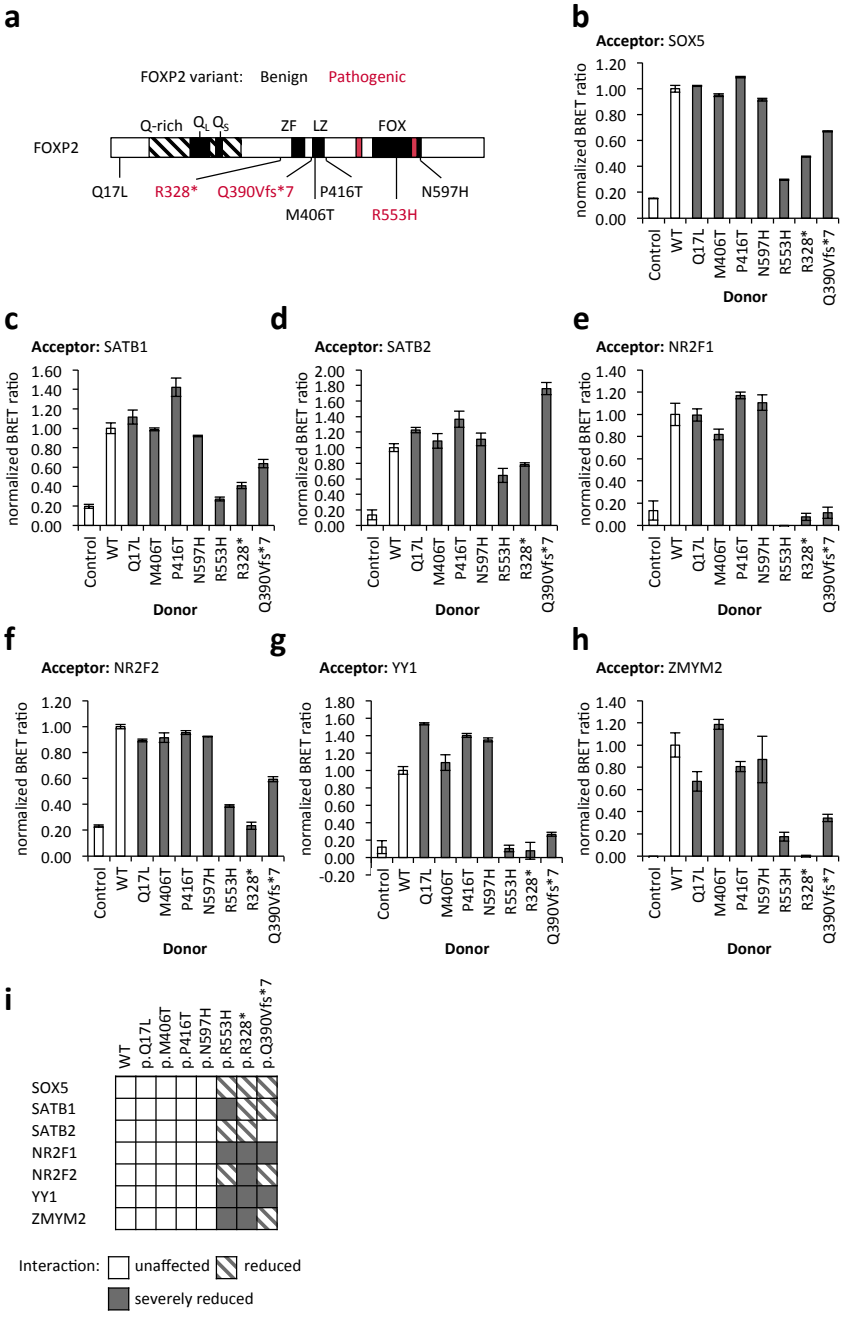

a
FOXP2 variant: Benign Pathogenic
FOXP2
Q-rich
ZF
LZ
FOX
Q17L
R328*
Q390Vfs*7
P416T
M406T
R553H
N597H
b
Acceptor: SOX5
c
Acceptor: SATB1
d
Acceptor: SATB2
e
Acceptor: NR2F1
f
Acceptor: NR2F2
g
Acceptor: YY1
h
Acceptor: ZMYM2
normalized BRET ratio
Control
WT
Q17L
M406T
P416T
N597H
R553H
R328*
Q390Vfs*7
Donor
i
SOX5
SATB1
SATB2
NR2F1
NR2F2
YY1
ZMYM2
Interaction: unaffected reduced
severely reduced

Figure 4.4

Figure 4.4: Etiological mutations in FOXP2 disrupt protein interactions (a) Schematic representation of the FOXP2 protein isoform I (NM_014491.3, NP_055306.1) showing rare variants found in individuals with neurodevelopmental disorders. Pathogenic variants are shown in red and benign variants in black. Known domains are labeled: glutamine-rich (Q-rich) region (hatched shading), long (Q_L) and short (Q_S) polyglutamine tracts, zinc finger (ZF), leucine zipper (LZ) and forkhead domain (FOX). Nuclear localization signals are indicated with red bars. **(b-h)** BRET assays for interaction of FOXP2 variants with FOXP-interacting proteins. HEK293 cells were transfected with Renilla luciferase-FOXP2 wild type (WT) or variants (donor) and YFP (acceptor) fusions of SOX5 **(b)**, SATB1 **(c)**, SATB2 **(d)**, NR2F1 **(e)**, NR2F2 **(f)**, YY1 **(g)** or ZMYM2 **(h)**. The control acceptor protein is a nuclear-targeted YFP. Values are mean corrected BRET ratios normalized to wild-type FOXP2 ± S.E.M. (n = 3). **(i)** Heatmap summarizing the results of the BRET assays between FOXP2 variants and the FOXP-interacting transcription factors. White squares indicate an unaffected interaction; a hatched pattern, a reduced interaction; and grey shading, a severely reduced interaction.

The truncated variants did not interact with any of the transcription factors tested in our assays, whereas two of the three missense variants retained some degree of interaction with one or more interaction partner (Figure 4.5b-h). The FOXP1 missense variants p.R465G and p.R514C retained the ability to interact with SOX5 (Figure 4.5b). Accordingly, co-expression of these variants with SOX5 led to mislocalization of SOX5 into the nuclear aggregates formed by the abnormal FOXP1 proteins, an effect not observed when SOX5 was co-expressed with other FOXP1 variants (Figure 4.5i). Thus, most etiological FOXP1 and FOXP2 variants represent a loss of function with respect to interaction with other neural transcription factors. However the FOXP1 p.R465G and p.R514C variants may exert dominant negative effects *in vivo* by interfering with the function of interacting transcription factors such as SOX5. Such differences in the behaviour of different protein variants has the potential to contribute to phenotypic variation observed between individuals with FOXP1-related disorder (Sollis *et al.* , 2016).

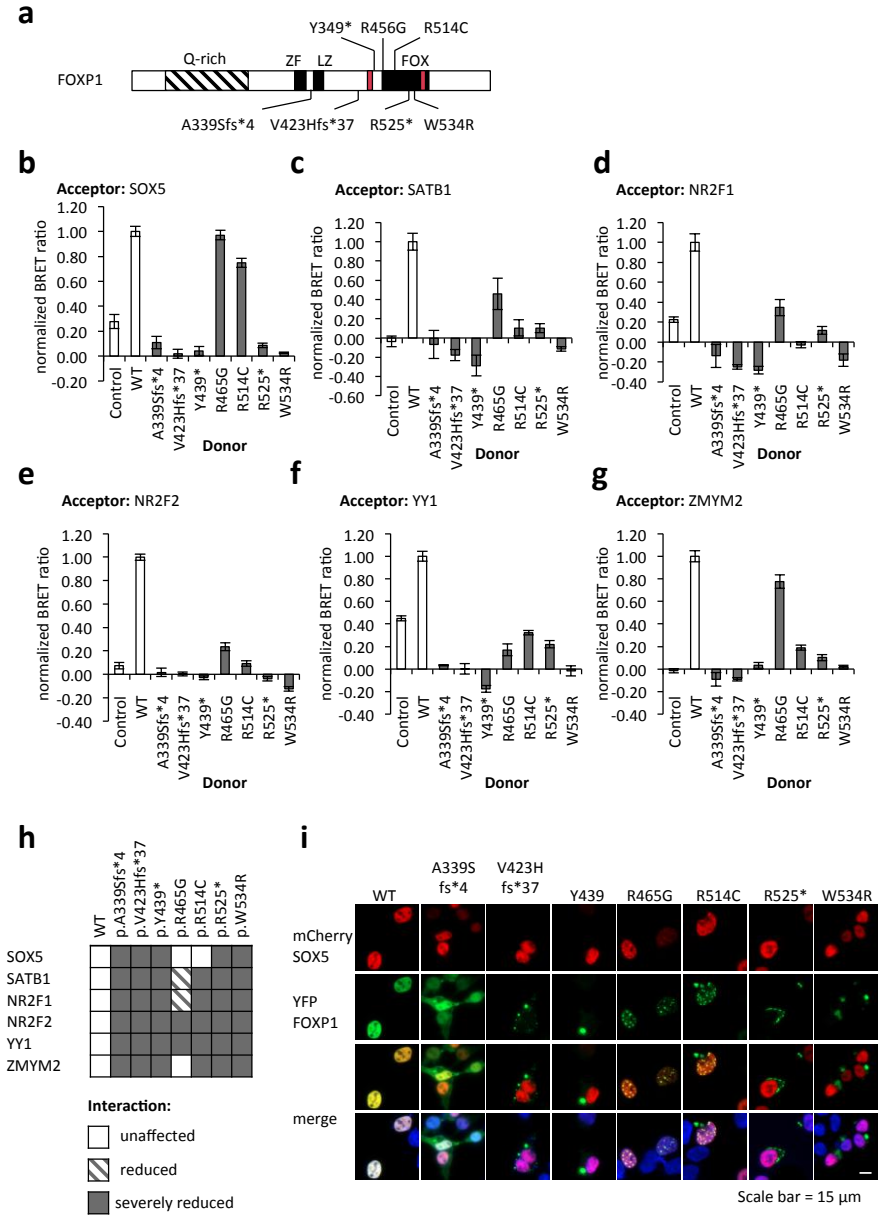

a
Y349* R456G R514C
Q-rich ZF LZ FOX
FOXP1
A339Sfs*4 V423Hfs*37 R525* W534R
b
Acceptor: SOX5
c
Acceptor: SATB1
d
Acceptor: NR2F1
e
Acceptor: NR2F2
f
Acceptor: YY1
g
Acceptor: ZMYM2
normalized BRET ratio
Control WT A339Sfs*4 V423Hfs*37 Y439* R465G R514C R525* W534R
Donor
h
WT p.A339Sfs*4 p.V423Hfs*37 p.Y439* p.R465G p.R514C p.R525* p.W534R
SOX5
SATB1
NR2F1
NR2F2
YY1
ZMYM2
Interaction:
unaffected
reduced
severely reduced
i
WT A339S fs*4 V423H fs*37 Y439 R465G R514C R525* W534R
mCherry SOX5
YFP FOXP1
merge
Scale bar = 15 µm

Figure 4.5

Figure 4.5: Etiological mutations in FOXP1 disrupt protein interactions. (a) Schematic representation of the FOXP1 protein (NM_032682, NP_116071) showing rare variants found in individuals with neurodevelopmental disorders. Known domains are labeled: glutamine-rich (Q-rich) region (hatched shading), zinc finger (ZF), leucine zipper (LZ) and forkhead domain (FOX). Nuclear localization signals are indicated with red bars. **(b-g)** BRET assays for interaction of FOXP1 variants with the FOXP-interacting proteins. HEK293 cells were transfected with Renilla luciferase-FOXP1 wild type (WT) or variants (donor) and YFP (acceptor) fusions of SOX5 **(b)**, SATB1 **(c)**, NR2F1 **(d)**, NR2F2 **(e)**, YY1 **(f)** or ZMYM2 **(g)**. The control acceptor protein is a nuclear-targeted YFP. Values are mean corrected BRET ratios normalized to wild-type FOXP1 ± S.E.M. (n = 3). **(h)** Heatmap summarizing the results of the BRET assays between FOXP2 variants and the FOXP-interacting transcription factors. White squares indicate an unaffected interaction; a hatched pattern, a reduced interaction; and grey shading, a severely reduced interaction. **(i)** Fluorescence micrographs of HEK293 cells transfected with YFP-fusions of FOXP1 wild-type or etiological variants, together with SOX5 fused to mCherry. Nuclei were stained with Hoechst 33342. (Scale bar = 15 μm).

4.3.5 *FOXP2 is co-expressed with interacting transcription factors in specific neuronal subpopulations*

We sought to identify the cell types in which the newly identified interactions between FOXP2 and transcription factors may be physiologically relevant. Of the seven FOXP-interacting transcription factors validated using the BRET assay, ZMYM2 and YY1 are ubiquitously expressed, while SOX5, SATB1, SATB2, NR2F1 and NR2F2 are all neural transcription factors that are expressed in subsets of neurons within the cerebral cortex (Huang *et al.* , 2011; Kim *et al.* , 2009a; Lai *et al.* , 2008; Tripodi *et al.* , 2004). We therefore examined the co-localization of Foxp2 with the interacting neural transcription factors in mouse brain by immunostaining (Figure 4.6). Brains from mice at postnatal day 3 were used because by this point in development the cortical layers have formed and layer-specific Foxp2 expression can readily be detected (Ferland *et al.* , 2003). Foxp2 was observed in the deep layers of the cortex, as well as in the striatum, as previously reported (Figure 4.6) (Ferland *et al.* , 2003) Each of the five neural transcription factors showed a different pattern of expression in the cortex. Sox5 exhibited extensive co-expression with Foxp2 in deep layer cortical neurons (Figure 4.6). Satb2 expression was found throughout all the six layers of

the cortex and its expression coincided with Foxp2 in a subpopulation of neurons of layer V (Figure 4.6). Satb1 was detected in layers IV to VI, including a subpopulation of Foxp2-expressing cells (Figure 4.6). Nr2f1 was expressed in all the layers of the cortex, including the vast majority of Foxp2-positive cells in layers V and VI (Figure 4.6). In contrast, and despite being expressed in a few scattered cells in the cortex layers I, IV and V, Nr2f2 was not co-expressed with Foxp2 in any cortical neurons (Figure 4.6). However, Nr2f2 was found to be extensively co-expressed with Foxp2 in the Purkinje cells of the cerebellum, another key neuronal subtype with high Foxp2 expression (Figure 4.7). These data are consistent with the prevailing view that expression of different combinations of transcription factors is essential to determining and maintaining cell identity and function.

4.4 DISCUSSION

In this study, through mass spectrometry-based proteomics followed by targeted analysis using BRET, we identified and characterized the interactions between FOXP family proteins and seven neurally-expressed transcription factors. We demonstrated co-expression of the interactors with Foxp2 in functionally-relevant neuronal subtypes in the early post-natal mouse brain, confirming that the reported interactions are of potential physiological significance *in vivo*. We found that different interaction partners have distinct binding sites within the FOXP2 protein, and that disorder-associated variants in FOXP1 and FOXP2 generally cause broad disruption of protein-protein interactions, but in some cases may lead to a dominant-negative effect in which abnormal FOXP protein may interfere with the functions of a normal interactor.

The interaction partners we identified have important, yet diverse, roles in neurodevelopment. Sox5 is expressed in deep cortical layers during embryonic and early post-natal stages and controls timing of the generation of distinct corticofugal neuron subtypes (Lai *et al.* , 2008). Loss of Sox5 results in aberrant differentiation and abnormal migration of cortical projection neurons (Lai *et al.* , 2008). Satb2 is a key regulator of cortical development and is expressed throughout all cortical layers (Huang *et al.* , 2011; Szemes *et al.* , 2006). The neural functions of Satb1 have not been as extensively investigated as its roles in the immune system (Alvarez *et al.* , 2000), but it is reported to regulate development of cortical interneurons and facilitate cortical neuron plasticity by

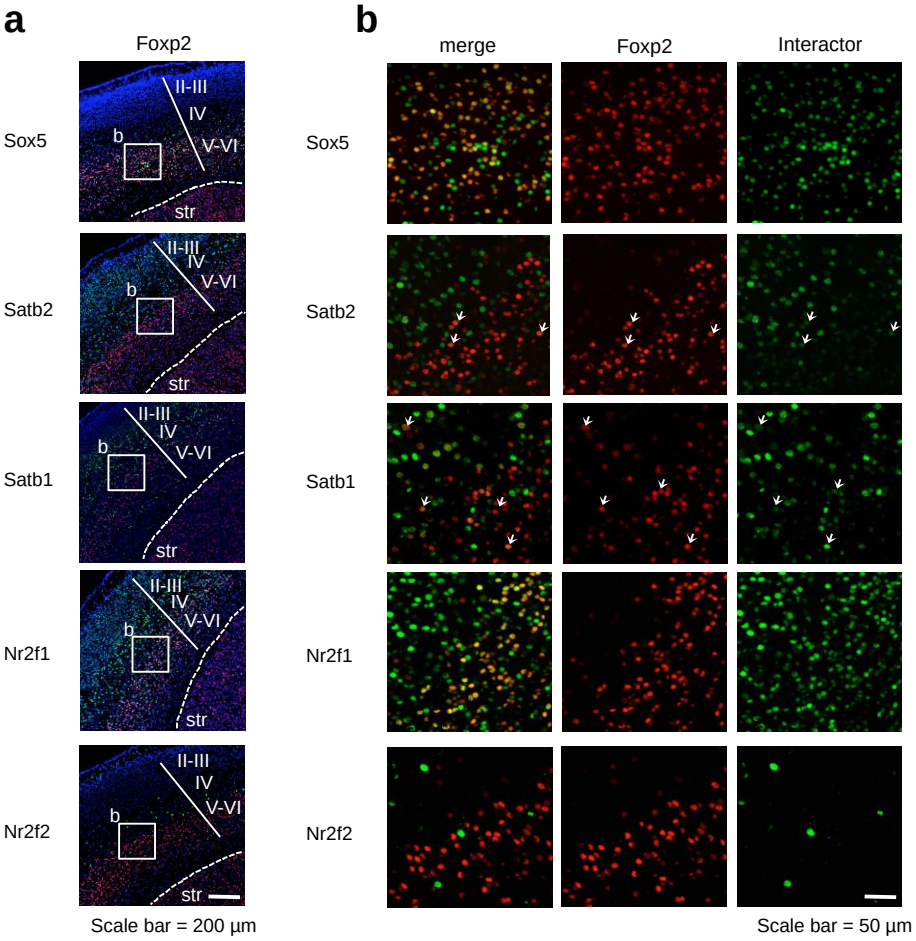


Figure 4.6: FOXP2 is co-expressed with SOX5, SATB1, SATB2 and NR2F2 in the cerebral cortex. (a) Immunofluorescence experiments to assess endogenous co-expression of Sox5, Satb1, Satb2, Nr2f1 and Nr2f2 with FoxP2 in the cerebral cortex of P3 mice brains. A white line indicates the cerebral cortex, layer numbers are indicated. A dashed white line demarks the striatum (str) (scale bar = 200 μm). **(b)** Magnified views of the region indicated with a white square in (a). White arrows indicate cells that co-express both FoxP2 and the FoxP2-interacting transcription factor (Scale bar = 50 μm). For **(a)** and **(b)**, Foxp2 protein expression is shown in red; FOXP2-interacting proteins, in green; and nuclei stained with DAPI, in blue

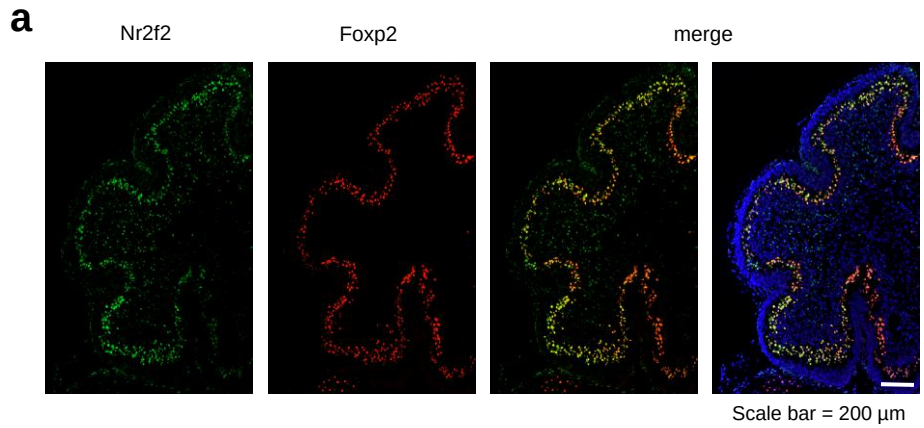


Figure 4.7: FOXP2 is co-expressed with NR2F2 in the Purkinje cells in the cerebellum Immunofluorescence experiments to assess endogenous co-expression of Nr2f2 with FoxP2 in the cerebellum of P3 mice brains. Foxp2 protein expression is shown in red; Nr2f2, in green; and nuclei stained with DAPI, in blue. (Scale bar = 200 μm).

modulating dendritic spine density (Balamotis *et al.* , 2012; Denaxa *et al.* , 2012). Nr2f1 regulates the differentiation of motor neurons, axonal projection and cortical arealization and, together with Nr2f2, plays a role in cell migration and regulation of neurogenesis (Armentano *et al.* , 2006, 2007; Faedo *et al.* , 2008; Tomassy *et al.* , 2010; Tripodi *et al.* , 2004). The interactions we demonstrated between these proteins and FOXP family members may therefore impact on a range of crucial processes in cortical development and maturation. Protein-protein interactions and their downstream consequences are naturally limited to the cell populations in which the proteins are co-expressed. Our analysis of co-expression of Foxp2 and interacting transcription factors in early post-natal mouse cortex shows that the subpopulations of co-expressing cells vary from a minor fraction of cells to the majority of deep layer projection neurons. Foxp1 and Foxp4 show different patterns of expression in the developing cortex compared to Foxp2, leading to further combinations of protein-protein interactions with differing functional outcomes.

In addition to roles in cortical development, our findings highlight potential interactions in the cerebellum. We found co-expression of Nr2f2 and Foxp2 in

Purkinje cells of the cerebellum; expression of the individual proteins in Purkinje cells has been reported previously in separate studies (Ferland *et al.* , 2003; Kim *et al.* , 2009a; Lai *et al.* , 2003). Nr2f2 regulates cerebellar growth and patterning (Kim *et al.* , 2009a), while Foxp2 has been implicated in dendritic outgrowth and arborization of Purkinje cells (Usui *et al.* , 2016) and mice lacking functional Foxp2 have small cerebellums (Groszer *et al.* , 2008). Foxp4, which is similarly expressed in Purkinje cells, may also be involved in the maintenance of dendritic arborization (Tam *et al.* , 2011).

FOXP1, FOXP2 and FOXP4 have all been implicated in human genetic disorders with neurodevelopmental phenotypes. Several of the FOXP-interacting transcription factors identified in this study are known to cause neurodevelopmental disorders through haploinsufficiency. SOX5 haploinsufficiency results in ID and language impairment (OMIM # 616803) (Lamb *et al.* , 2012; Nesbitt *et al.* , 2015). De novo mutations in SATB2 cause syndromic ID, usually with absent or near absent speech, and often cleft palate or dental anomalies (OMIM # 612313) (Bengani *et al.* , 2017; Zarate & Fish, 2016; Zarate *et al.* , 2017). Mutations in NR2F1 cause Bosch-Boonstra-Schaaf optic atrophy syndrome, a dominant disorder characterized by optic atrophy and ID, frequently accompanied by hypotonia, seizures, ASD and oromotor dysfunction (OMIM # 615722) (Bosch *et al.* , 2014; Chen *et al.* , 2016a). YY1 haploinsufficiency leads to a syndrome that includes ID often accompanied by motor problems and various congenital malformations (OMIM # 617557) (Gabriele *et al.* , 2017; Vissers *et al.* , 2010). ZMYM2 de novo mutations have been found in large-scale exome sequencing studies of cases of ASD (Neale *et al.* , 2012). However, since ZMYM2 is ubiquitously expressed, further cases of disruptions are needed in order to confidently establish its involvement in neurodevelopmental disorders. Of note, following completion of experiments for this study, recurrent *de novo* mutations in ZNF148, another transcription factor identified in our mass spectrometry screen, have been identified in individuals with developmental delay and ID (OMIM # 617260) (Stevens *et al.* , 2016). The differing levels of ID and speech/language impairment associated with disruption of FOXP proteins and their interacting transcription factors suggests that reduced levels of these transcription factors lead to differing, but potentially related, disturbances in downstream gene expression and cortical development.

Most of the identified FOXP-interacting transcription factors control cortical development in an interconnected fashion, cooperating with each other but also

forming positive and negative loops of regulation between them (Greig *et al.* , 2013). For instance, to maintain the specific neuronal identities in deeper layers of the cortex, SOX5 directly represses Fezf2, a key node of the cortical transcriptional network, in layer VI and the subplate (Shim *et al.* , 2012), which at the same time is repressed by TBR1 (Han *et al.* , 2011), an ASD-related transcription factor also known to interact with FOXP2 and FOXP1 (Deriziotis *et al.* , 2014a). In contrast, in layer V, SATB2 promotes the expression of SOX5 and Fezf2, while Fezf2 represses SATB2 (McKenna *et al.* , 2015). Also, NR2F1, SATB2 and SOX5 control the development of cortical projection neurons at least in part by repressing the same target gene: the transcription factor BCL11B (Alcamo *et al.* , 2008; Britanova *et al.* , 2008; Lai *et al.* , 2008; Tomassy *et al.* , 2010). Strikingly, NR2F1 and NR2F2 physically interact with transcription factors BCL11B and BCL11A (Chan *et al.* , 2013), which also plays roles in cortical development by repressing TBR1 (Cánovas *et al.* , 2015). BCL11A haploinsufficiency has recently been identified as a cause of sporadic ID (Dias *et al.* , 2016). Therefore, a number of genetically-distinct neurodevelopmental disorders may have shared roots in this complex regulatory network of transcription factors that orchestrates cortical development, and our results contribute to extend the characterization of this network. It will be interesting in the future to further determine the specific roles the FOXPs play in this network and, in general, better understand the transcriptional architecture underlying cortical development.

Clearly the proteins investigated here are only a subset of the biologically-important interactors of FOXP proteins. Our mass spectrometry screens provide the largest list of candidate FOXP-interacting proteins reported to date, and represent a resource for further investigation of protein networks surrounding these transcription factors. We elected to focus on transcription factors with potential roles in neurodevelopment, and a logical next step would be to consider proteins mediating transcription factor activity, such as chromatin modifying and remodelling complexes. For example, the candidate FOXP-interacting proteins include the chromatin remodeler SETD2, haploinsufficiency of which leads to Luscan-Lumish syndrome (OMIM # 616831), a neurodevelopmental disorder characterized by ID and macrocephaly accompanied by speech delay (Lumish *et al.* , 2015; Luscan *et al.* , 2014). In addition, FOXP proteins have roles in multiple tissues and candidate interactors in our screen may be of relevance to the function of FOXPs in non-neuronal contexts. For example, FOXP1, FOXP2 and FOXP4 are all expressed in the embryonic heart, and mice lacking

Foxp1 or Foxp4 die before birth due to heart defects (Ferland *et al.* , 2003; Li, 2004; Teufel *et al.* , 2003; Wang *et al.* , 2004). Heterozygous mutations in our confirmed FOXP-interaction transcription factor NR2F2 cause a variety of congenital heart defects (Al Turki *et al.* , 2014), and a critical role in heart development has also been described for our confirmed interactor YY1 (Beketaev *et al.* , 2015). Therefore the FOXPs may co-operate with NR2F2 and YY1 to regulate downstream target genes important for cardiac development.

Our strategy employed affinity purification coupled to mass spectrometry to screen for potential FOXP-interacting transcription factors, followed by validation of selected candidate interactors using a BRET assay. These two methods have different strengths in relation to identifying protein interaction partners. The affinity purification method gives a proteome-wide picture of protein interactions, but may miss weak or transient interactions because it is performed on cell lysates, and may suffer from false-positives due to non-specific and indirect interactions. In contrast, BRET is a targeted method which monitors protein interactions in living cells, can detect transient interactions, is less vulnerable to false positives, but may give false negatives due to the constraints on energy transfer (Deriziotis *et al.* , 2014b). We show here that this two-stage approach is effective in identifying high-confidence protein interaction partners with potential biological importance.

In conclusion, our findings situate the FOXPs within a broader molecular network that orchestrates cortical development, and thus provide novel clues as to how these proteins work to build a functioning brain, and why their deficiency leads to neurodevelopmental disorders.

4.5 SUPPLEMENTARY MATERIAL

Supplementary Table S4.1: Primers

Gene	Forward primer 5'-3'	Restriction enzyme	Reverse primer 5'-3'	Restriction enzyme
NFAT5	GGATCCTGGGCGGTGCTTGCAGC	BamHI	GCTAGCTTAAAAGGAGCCAGTCAAGTTGTTCC	NheI
NR2F1	GGATCCTGGCAATGGTAGTTAGCAGCTG	BamHI	TCTAGACTAGGAGCACTGGATGGACATG	XbaI
NR2F2	AGATCTTGGCAATGGTAGTCAGCACG	BglII	TCTAGATTATTGAATTGCCATATACGGCCAGTT	XbaI
SATB1	GGATCCTGGATCATTTGAACGAGGCAACT	BamHI	TCTAGATCAGTCTTTCAAATCAGTATTAATGTC	XbaI
SATB2	TGATCATGGAGCGGCGGAGCGAGA	BclI	TCTAGATTATCTCTGGTCAATTTCGGCAGGTG	XbaI
SOX5	GGATCCTGTCTTCCAAGCGACCAGCC	BamHI	TCTAGATCAGTTGGCTTGTCCTGCAAT	XbaI
TFDP1	GGATCCTGGCAAAAGATGCCGGTCTA	BamHI	TCTAGATCAGTCGTCCTCGTCATTCTCGT	XbaI
TP53	GGATCCTGGAGGAGCCGCAGTCAGAT	BamHI	TCTAGATCAGTCTGAGTCAGGCCCTTCTG	XbaI
YY1	GGATCCTGGCCTCGGGCGACACC	BamHI	TCTAGATCACTGGTTGTTTTTGGCCTTAGCATG	XbaI
ZBTB2	GAATCCTGGATTTGGCCAACCATGGA	BamHI	TCTAGATCAGTCTAGTAAGACGGTTTCTTGTTC	XbaI
ZMYM2	GGATCCTGGACACAAGTTCAGTGGGAGG	BamHI	GCTAGCTTAGTCTGTGTCTTCATCCAGTTC	NheI
ZNF687	GGATCCTGGGGGATATGAAGACCCCT	BamHI	TCTAGACTAGTTGTCCCCAACAGCCC	XbaI

Supplementary Table S4.2: List of identified candidate FOXP-interacting proteins

The compelete supplementary Table 1 can be accessed via this link:
https://docs.google.com/spreadsheets/d/1XAa2H6aKAkNeB8fBDhvMAToUHWGng2BJK2OE2P1GG5Q/edit?usp=sharing

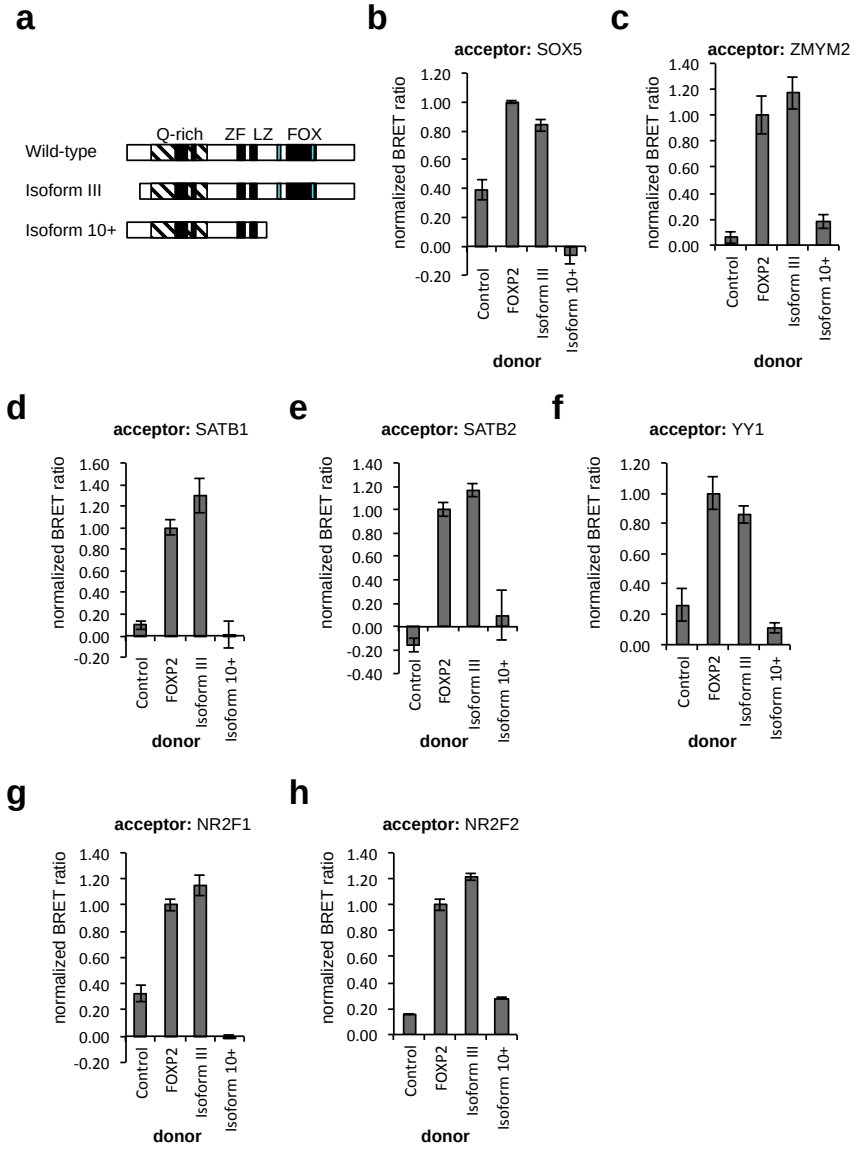
	Gene Symbol	Uniprot accession	description	overlap FOXP1	overlap FOXP2	overlap FOXP4	overlap FOXPs	Mascot score							
								IP1_FOXP1	IP2_FOXP1	IP3_FOXP1	IP1_FOXP2	IP2_FOXP2	IP3_FOXP2	IP2_FOXP4	IP3_FOXP4
1	FOXP4	Q8IVH2	Forkhead box	3	3	2	8	1330	1111	806	456	976	1363	2704	1376
2	FOXP2	O15409	Forkhead box	3	3	2	8	769	478	722	1040	1909	1769	1909	640
3	FOXP1	Q9H334	Forkhead box	3	3	2	8	2489	2140	1263	456	931	1363	1129	518
4	NFAT5	O94916	Nuclear factc	1	0	2	3	-	-	200	-	-	-	85	59
5	SMARCA5	O60264	SWI/SNF-rela	2	1	1	4	261	-	646	-	-	68	1094	-
6	SATB2	Q9UPW6	DNA-binding	2	1	1	4	401	-	350	329	-	-	45	-
7	BAP18	Q8IXM2	Chromatin cc	2	0	1	3	86	-	45	-	-	-	159	-
8	RNF2	Q99496	E3 ubiquitin-l	2	2	0	4	109	-	72	79	-	128	-	-
9	TRRAP	Q9Y4A5	Transformati	2	0	1	3	65	-	122	-	-	-	165	-
10	EMSY	Q7Z589	Protein EMSY	2	0	1	3	147	-	81	-	-	-	164	-
11	BPTF	Q12830	Nucleosome-	2	0	1	3	236	-	978	-	-	-	954	-
12	ZBTB2	Q8N680	Zinc finger an	2	1	0	3	62	-	93	58	-	-	-	-
13	CDC23	Q9UJX2	Cell division c	2	0	0	2	71	-	66	-	-	-	-	-
14	APOD	C9JF17	Apolipoprote	2	0	0	2	-	86	50	-	-	-	-	-
15	LIG3	P49916	DNA ligase 3	2	0	0	2	68	-	61	-	-	-	-	-
16	FRYL	O94915	Protein furry	2	0	0	2	82	-	1117	-	-	-	-	-
17	ZMYM2	Q9UBW7	Zinc finger M	1	1	2	4	-	-	409	-	-	45	306	127
18	UBAP2L	F8W726	Ubiquitin-ass	1	1	2	4	-	-	119	-	-	88	242	74
19	ZNF687	Q8N1G0	Zinc finger pr	1	1	1	3	-	-	275	-	-	83	-	53
20	ZMYND8	H7C4X9	Protein kinas	1	1	1	3	-	-	245	-	-	141	-	55
21	YWHAG	P61981	14-3-3 protei	1	1	1	3	-	-	185	-	-	74	-	68
22	TP53	A0A087WT	Cellular tumc	1	1	1	3	-	-	94	-	-	188	-	194
23	SUPT16H	Q9Y5B9	FACT comple	1	1	1	3	-	-	541	-	-	143	-	66
24	SMC3	Q9UQE7	Structural ma	1	1	1	3	-	-	473	-	-	71	-	191

Supplementary Table S4.3: Putative FOXP-interacting transcription factors

* Tentative association.

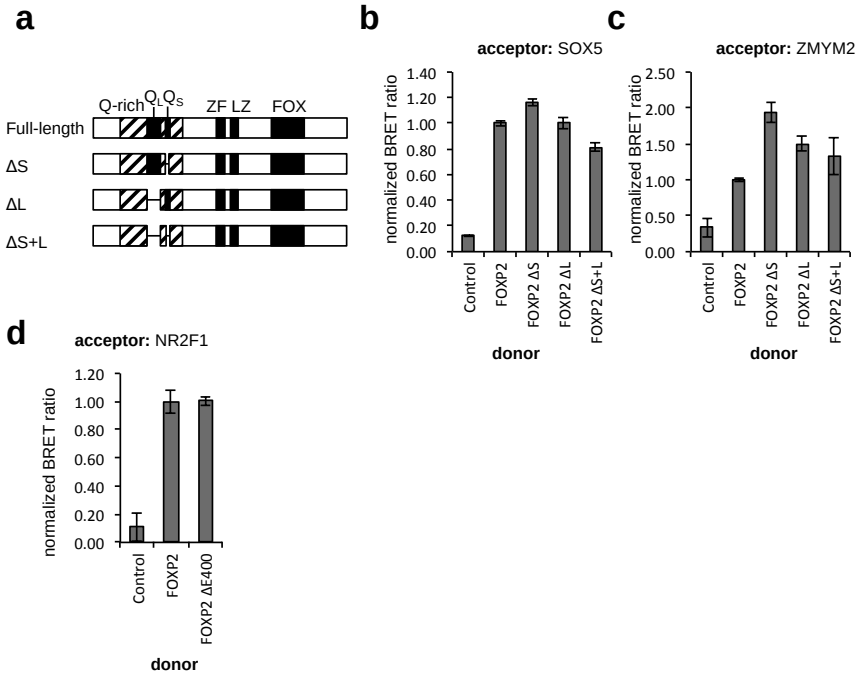
	Gene Symbol	Uniprot accession	Function	Associated Phenotype	Phenotype OMIM number	Inheritance
1	BNC2	Q6ZN30	Regulation of skin pigmentation			
2	CDC5L	Q99459	Regulation of cell cycle			
3	EMSY	Q7Z589	DNA repair			
4	FOXC1	Q12948	Embryonic and ocular development	Iris hypoplasia and glaucoma	601631	Autosomal dominant
5	FOXP1	Q9H334	Brain, heart and lung development	ID with language impairment	613670	Autosomal dominant
6	FOXP2	O15409	Brain, heart and lung development	Severe language impairment	602081	Autosomal dominant
7	FOXP4	Q8IVH2	Brain, heart and lung development	Developmental delay*		Autosomal recessive
8	FUBP1	E9PEB5	Regulation of cell cycle			
9	GATAD2B	Q8WXI9	Brain development	ID	615074	Autosomal dominant
10	HIC2	Q96JB3	Unknown function			
11	NFAT5	O94916	Regulation of inflammatory response			
12	NR2F1 (COUP-TFI)	P10589	Brain and heart development	ID with optic atrophy	615722	Autosomal dominant
13	NR2F2 (COUP-TFII)	P24468	Brain and heart development	Congenital heart defects	615779	Autosomal dominant
14	RFX1	P22670	Unknown function			
15	SATB1	Q01826	Regulation of immune system and brain development			
16	SATB2	Q9UPW6	Brain and bone development	ID with language impairment	612313	Autosomal dominant
17	SOX13	E9PPW0	T-cell development			
18	SOX5	F5H0I3	Brain and bone development	ID with language impairment	616803	Autosomal dominant
19	TFDP1	F5H452	Regulation of cell cycle	ID with language impairment*		Autosomal dominant
20	TOX	O94900	T-cell development			
21	TP53	P04637	Regulation of cell cycle	Li-Fraumeni syndrome	151623	Autosomal dominant
22	YY1	P25490	Embryonic development	ID	617557	Autosomal dominant
23	ZBTB2	Q8N680	Regulation of cell cycle			
24	ZBTB39	O15060	Unknown function			
25	ZFHX4	E7EVZ1	Brain and muscle development	Ptosis	178300	Autosomal dominant
26	ZMYM2	Q9UBW7	Immue system regulation			
27	ZNF148	Q9UQR1	Unknown function	ID with language impairment	617260	Autosomal dominant
28	ZNF687	Q8N1G0	Bone development	Paget disease of bone	616833	Autosomal dominant

	Gene Symbol	FOXP1 experiments (total =3)	FOXP2 experiments (total =3)	overlap FOXP4	Total experiments (total = 8)	Max Mascot Score			Previously investigated	Selected for validation	Fetal or adult brain protein expression (human
						FOXP1	FOXP2	FOXP4			
1	BNC2	1	1	1	3	268,00	61,00	69,00			No
2	CDC5L	1	0	0	1	174,00	-	-			Yes
3	EMSY	2	0	1	3	147,00	-	164,00			Yes
4	FOXC1	1	0	0	1	45,00	-	-			No
5	FOXP1	3	3	2	8	2489,00	1363,00	1129,00	Li et al.		Yes
6	FOXP2	3	3	2	8	769,00	1909,00	1909,00	Li et al.		Yes
7	FOXP4	3	3	2	8	1330,00	1363,00	2704,00	Li et al.		Yes
8	FUBP1	1	0	0	1	155,00	-	-			Yes
9	GATAD2B	1	0	0	1	111,00	-	-	Chokas et al.		Yes
10	HIC2	1	0	0	1	90,00	-	-			No
11	NFAT5	2	0	2	4	200,00	-	85,00		Yes	Yes
12	NR2F1 (COUP-TFI)	0	0	1	1	-	-	137,00		Yes	Yes
13	NR2F2 (COUP-TFII)	0	0	1	1	-	-	137,00		Yes	Yes
14	RFX1	0	0	1	1	-	-	244,00			Yes
15	SATB1	0	0	1	1	-	-	45,00		Yes	Yes
16	SATB2	2	1	1	4	401,00	329,00	45,00		Yes	Yes
17	SOX13	1	0	0	1	113,00	-	-			Yes
18	SOX5	0	1	0	1	-	98,00	-		Yes	Yes
19	TFDP1	0	1	0	1	-	97,00	-		Yes	No
20	TOX	1	0	0	1	66,00	-	-			No
21	TP53	1	1	1	3	94,00	188,00	194,00		Yes	No
22	YY1	1	0	0	1	160,00	-	-		Yes	Yes
23	ZBTB2	2	1	0	3	93,00	58,00	-		Yes	Yes
24	ZBTB39	1	1	0	2	87,00	72,00	-			No
25	ZFHX4	1	0	0	1	113,00	-	-			Yes
26	ZMYM2	1	1	2	4	409,00	45,00	306,00		Yes	Yes
27	ZNF148	1	0	0	1	88,00	-	-			Yes
28	ZNF687	1	1	1	3	275,00	83,00	53,00		Yes	No



Supplementary Figure 4.1: FOXP2 isoforms interaction with FOXP-interacting proteins. (a) Schematic representation of synthetic FOXP2 variants. Known domains are labeled: glutamine-rich (Q-rich) region (hatched shading), long (QL) and short (QS) poly-

Q tracts, zinc finger (ZF), leucine zipper (LZ) and forkhead domain (FOX) **(b-h)** BRET assays for for interaction of FOXP2 isoforms with FOXP-interacing proteins. HEK293 cells were transfected with Renilla luciferase-FOXP2 wild type (WT) or variants (donor) and YFP (acceptor) fusions of SOX5 **(b)**, ZMYM2 **(c)**, SATB1 **(d)**, SATB2 **(e)**, YY1 **(f)**, NR2F1 **(g)** or NR2F2 **(h)**. The control acceptor protein is a nuclear-targeted YFP. Values are mean corrected BRET ratios normalized to wild-type FOXP2 ± S.E.M (n=3).



Supplementary Figure 4.2: Synthetic FOXP2 variants interaction with SOX5, ZMYM2 and YY1. (a) Schematic representation synthetic FOXP2 poly-Q deletion variants. Known domains are labeled: glutamine-rich (Q-rich) region (hatched shading), long (QL) and short (QS) poly-Q tracts, zinc finger (ZF), leucine zipper (LZ) and forkhead domain (FOX) **(b-c)** BRET assays for interaction of FOXP2 poly-Q synthetic variants with FOXP-interacting proteins. HEK293 cells were transfected with Renilla luciferase-FOXP2 wild

type (WT) or variants (donor) and YFP (acceptor) fusions of SOX5 **(b)** or ZMYM2 **(c)**. The control acceptor protein is a nuclear-targeted YFP. Values are mean corrected BRET ratios normalized to wild-type FOXP2 ± S.E.M. (n = 3). **(d)** BRET assay for the interaction of a FOXP2 synthetic variant lacking residue E400 (key residue for homo-dimerization through the leucine zipper) with NR2F1. HEK293 cells were transfected with Renilla luciferase-FOXP2 wild type (WT) or variant (donor) and YFP (donor) fusion of NR2F1. The control acceptor protein is a nuclear-targeted YFP. Values are mean corrected BRET ratios normalized to wild-type FOXP2 ± S.E.M. (n = 3).

5

FUNCTIONAL CHARACTERIZATION OF BCL11A MUTATIONS IN NEURODEVELOPMENTAL DISORDERS[1]

Abstract: *De novo missense variants in BCL11A were identified in three unrelated infants with developmental delay through trio-based exome sequencing performed as part of the Deciphering Developmental Disorders study. The variants in these patients are notably clustered within the N-terminal region of the protein, further suggesting that disruption of BCL11A function is the cause of the disorder in these patients. BCL11A encodes a zinc finger transcription factor, expressed in both the hematopoietic system and the developing brain. Whole gene deletions affecting BCL11A have previously been described in individuals with neurodevelopmental disorders. A microdeletion only affecting BCL11A was found in a child with childhood apraxia of speech together with mild autistic traits, developmental delay and hypotonia. BCL11A haploinsufficiency has also been suggested to underlie some of the clinical features of the rare chromosome 2p16.1-p15 deletion syndrome, which is characterized by intellectual disability, dysmorphic features, microcephaly, and skeletal and organ abnormalities, although not all individuals with this syndrome have deletions encompassing BCL11A. The role of BCL11A in brain development is poorly understood, but the gene is expressed in the developing cortex, hippocampus, basal ganglia and cerebellum, and has recently been reported to play a role in the specification of subcortical projection neurons by repressing expression of TBR1, a gene recurrently mutated in cases of autism. The missense variants found in the new patients lie outside the zinc finger DNA-binding domains, in a region of unclear function. We therefore sought to characterize the effects of these variants on protein function in order to confirm their etiological role, and to illuminate the molecular mechanism of disorder. We found that all three variants disrupt the localization of BCL11A within the nucleus and abolish protein dimerization and transcriptional regulatory activity.*

1 Results in this chapter have been published as part of: Dias C.*, Estruch S.B.*, Graham S.A., McRae J., Sawiak S.J., Hurst J.A. , Joss S.K., Holder S.E., Morton J.E.V., Turner C., Thevenon J., Mellul K., Sanxhez-Andrade G., Ibarra-Soria X., Deriziotis P., Santos R.F., Lee S., Faivre L., Kleefstra T., Liu P., Hurles M.E., DDD Study, Fisher S.E., Logan D.W. (2016). BCL11A haploinsufficiency causes an intellectual disability syndrome and dysregulates transcription. The American Journal of Human Genetics, 99(2), 253-274. doi:10.1016 /j.ajhg.2016.05.030. *joint first authors

Our results therefore strongly support a causal role for BCL11A missense variants in the developmental delay of these patients, and contribute to the delineation of a novel monogenic intellectual disability syndrome. Furthermore the characterization of these variants reveals a key role for the N-terminal region of BCL11A in mediating protein dimerization and regulation of transcription, and suggest that this region may be a mutational hotspot.

5.1 INTRODUCTION

The Deciphering Developmental Disorders (DDD) study has contributed significantly to the understanding of neurodevelopmental disorders through the identification of novel causative genes on a large scale (Deciphering Developmental Disorders Study, 2015). Large studies of this kind could prove to be a powerful tool for discovery of language-related genes, as a large proportion of the patient cohort exhibit speech and language deficits. The strategy of identifying *de novo* variants in patients with severe sporadic disorders by exome sequencing can enable disorder genes to be identified based on the recurrence of mutations in the same gene in patients with similar phenotypes. For some genes there are insufficient patients, or the phenotype is not distinctive enough, to allow pathogenicity to be conclusively assigned to the variants based on statistical analysis of the sequencing data alone. In such cases, functional characterization of the variants can provide compelling evidence of pathogenicity.

The DDD identified *de novo* missense mutations in the *BCL11A* gene (also known as *CTIP1* or *EVI9*) in three patients with intellectual disability (ID). This gene is a promising candidate for involvement in neurodevelopmental disorder. Haploinsufficiency of BCL11A has been suggested to be a key contributor to the clinical features of the 2p15p16.1 microdeletion syndrome (OMIM # 612513), a very rare genetic disorder caused by a heterozygous deletions in the short arm of chromosome 2 that encompasses several different genes. The first 2p15p16.1 deletion was described in 2007 in two individuals who presented with phenotypes that included developmental delay, structural brain abnormalities, dysmorphic features, gross motor impairments, ID and autistic behaviors (Rajcan-Separovic *et al.* , 2007). Since then, several cases presenting similar phenotypic profiles have been reported with partially overlapping deletions of this chromosomal region that range from 0.3 to 5.7 Mb (Peter *et al.* , 2014). The 2p15p16.1 microdeletion syndrome presents a variable neurodevelopmental phenotype but some features are common in all the patients, such as ID, abnormal muscle tone, gross motor delays, growth retardation, non verbal or low verbal status and craniofacial and skeletal dysmorphic features (Peter *et al.* , 2014). In 2014, the smallest deletion, which involved only the *BCL11A* gene, was reported in an individual who presented with mild ID, motor delay, hypotonia and CAS (Peter *et al.* , 2014). The proband differed from other microdeletion cases in that he did not have any craniofacial or skeletal anomalies and he presented a

milder form of ID. Moreover, while other microdeletion cases exhibited mild to severe language delays, this individual presented a distinctive speech phenotype, specifically characterized by childhood apraxia of speech (CAS), as well as dysarthria, which was likely caused by the general hypotonia. Interestingly, the *BCL11A* gene is located within a dyslexia susceptibility candidate region (DYX3), previously identified in several dyslexia studies using linkage analysis, although evidence of association of common variants in the gene has not been reported (Carrion-Castillo *et al.* , 2013).

BCL11A is a zinc finger transcription factor that is expressed at high levels in the hematopoietic system and also in the fetal brain (Satterwhite *et al.* , 2001). BCL11A (CTIP-1, COUP-TF interacting partner 1) was first identified, together with its close homologue BCL11B (CTIP-2, COUP-TF interacting partner 2), as a binding partner of NR2F2 (COUP-TFII) in a yeast two-hybrid screen (Avram *et al.* , 2000). BCL11A acts as a co-repressor of NR2F1 and NR2F2 but it can also function as a transcription factor and bind DNA to regulate gene expression independently (Avram *et al.* , 2002). Alternative splicing of BCL11A produces multiple isoforms: the most abundant ones are known as the extra long (XL, accession AJ404611), long (L, accession AJ404612), short (S, accession AJ404613), and extra short (XS, accession AY692278) isoforms (Figure 5.1). BCL11A-XL is the most abundant isoform in blood cells while BCL11A-L is most abundant in the brain (Kuo & Hsueh, 2007; Liu *et al.* , 2006; Satterwhite *et al.* , 2001). Isoforms S and XS can act as natural antagonists of the L and XL isoforms by forming hetero-dimers with them and preventing their binding to DNA (Kuo *et al.* , 2009). All four isoforms share exons 1 and 2, which encode a non-canonical C2CH zinc finger domain (shaded green box in Figure 5.1). This type of zinc finger usually mediates protein-protein interactions and is a putative hetero- and homo-dimerization domain (Avram *et al.* , 2000; Liu *et al.* , 2006). All isoforms except for BCL11A-XS also contain a non DNA-binding C2H2 zinc finger domain (shaded grey). Only BCL11A-XL and –L have the DNA-binding C2H2 zinc fingers (shaded blue boxes in Figure 5.1), which allow these isoforms to regulate gene expression by direct binding to DNA. The proline-rich domain (shaded yellow in Figure 5.1), which usually mediates multiple transcriptional-related protein (Hsieh *et al.* , 1994); and a glutamate-rich domain (shaded pink), which is involved in regulating gene expression (Chou & Wang, 2015) are also specific to the L and XL isoforms. Although BCL11A is expressed in both the hematopoietic system and the developing brain (Kuo and Hsueh, 2007; Leid et

al., 2004; Saiki et al., 2000), its functions have primarily been studied in the hematopoietic system, where BCL11A is essential for B-cell development; in the absence of this gene no B-cells are produced (Liu *et al.* , 2003). BCL11A is also functionally important in the development of erythrocytes where it mediates the switching between fetal and adult forms of hemoglobin (Sankaran *et al.* , 2010). Fetal hemoglobin, which has an α 2γ 2 subunit composition, is able to bind oxygen with greater affinity than the adult form, which has an α 2β 2 composition, giving the developing fetus better access to oxygen from the mother bloodstream (Sankaran *et al.* , 2010). The switch from expression of γ subunits to expression of β subunits occurs over the first few months of postnatal life and crucially involves transcriptional repression of the γ-globin gene by BCL11A (Xu *et al.* , 2010). A single-nucleotide polymorphism (SNP) identified at the BCL11A locus decreases the expression of BCL11A in erythroblasts 3.5-fold, resulting in the continued expression of fetal γ-globin in adults (Sankaran *et al.* , 2010). Remarkably, reducing the levels of BCL11A in a mouse model of sickle cell disease (which is caused by mutation of the hemoglobin β subunit) reactivates transcription of fetal hemoglobin, which is sufficient to ameliorate the symptoms of the disease (Xu *et al.* , 2011). Currently, post-natal hematopoietic-specific suppression of BCL11A is being developed as a potential therapy for patients with sickle cell disease (Bjurström *et al.* , 2016; Hossain & Bungert, 2017).

The functions of BCL11A in the brain have been far less well studied than in the hematopoietic system. In the developing brain, BCL11A is expressed in several regions including the cortex, hippocampus, striatum, cerebellum, and spinal cord (Kuo & Hsueh, 2007; Leid *et al.* , 2004) and it has been shown to negatively regulate axon branching and dendrite outgrowth (Kuo *et al.* , 2009). Knocking down isoform L of Bcl11a in cultured neurons resulted in increased dendrite number and total dendrite length, as well as a higher number of axonal branches, as compared to control neurons. Down-regulation of Bcl11a-L also increased the arbor complexity of both axons and dendrites (Kuo *et al.* , 2009). Strikingly, overexpression of a construct that encoded a truncated version of Bcl11a-L (resembling Bcl11a-S) promoted neurite arborization, mimicking the effect of Bcl11a-L knockdown. This suggests that Bcl11a-S may function as an endogenous negative regulator for Bcl11a-L by interacting with it and preventing it from binding to DNA (Kuo *et al.* , 2009). The negative effects of Bcl11a-L on axon morphogenesis may be controlled in part by its physical interaction

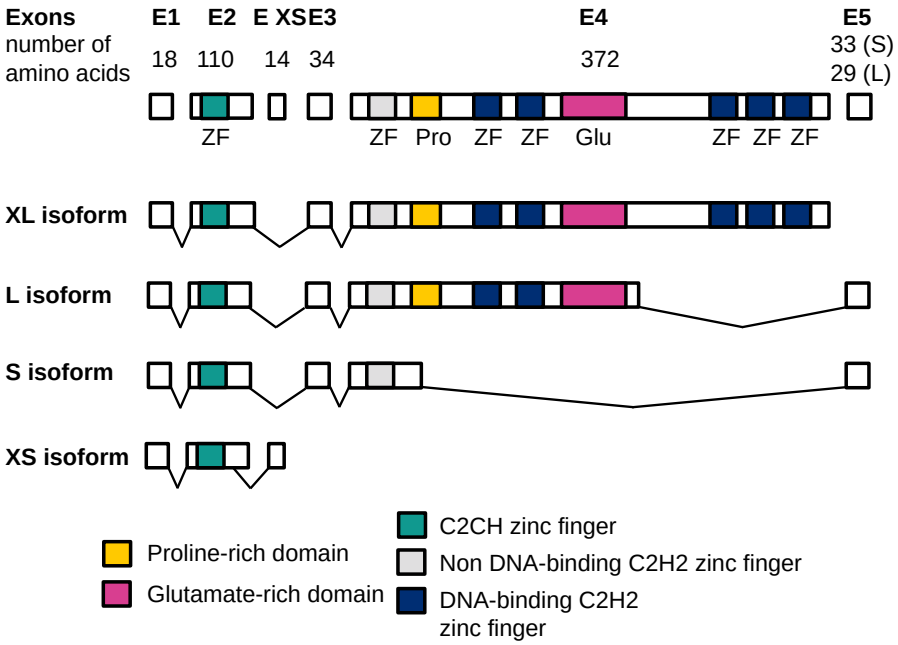


Figure 5.1: Schematic representation of BCL11A isoforms. BCL11A most abundant isoforms are the extra long (XL), long (L), short (S), and extra short (XS) isoforms. Known domains are shown: the glutamate-rich domain (Glu) is shaded in red; the proline-rich domain (Pro), in green; the C2CH zinc finger (ZF), in yellow; the non-DNA binding C2H2 zinc finger, in grey; and the DNA-binding C2H2 zinc finger domain, in dark blue. The different exons (E) and the numbers of amino acids that compose each exon are indicated at the top.

with the ID-related transcription factor CASK, since disruption of the interaction between CASK and Bcl11a increased axon outgrowth and branching (Kuo *et al.* , 2010).

More recently, Bcl11a has been reported to control the migration of cortical projection neurons by repressing the expression of Sema3c (Wiegreffe *et al.* , 2015) and to be involved in determining the identity of the different types of cortical projection neurons (Woodworth *et al.* , 2016). In the absence of Bcl11a, corticothalamic and deep-layer callosal projection neurons show impaired differentiation, and deep-layer neurons adopt characteristics of subcerebral pro-

jection neurons. In contrast, overexpression of Bcl11a in the brain results in suppressed expression of subcerebral projection neuron specific genes, resulting in an abnormal projection of this subtype of projection neuron; instead of projecting to subcerebral brain regions, they abnormally project to the thalamus and across the corpus callosum, resembling corticothalamic and callosal projection neurons (Woodworth *et al.* , 2016). This suggests that Bcl11a is crucial for determining and maintaining the identity of these specific subtypes of projection neurons.

Many known BCL11A interaction partners play important roles during neurodevelopment, such as NR2F1 and NR2F2 (Avram *et al.* , 2000, 2002; Chan *et al.* , 2013), which are critical determinants of cortical development; NR2E1 (Chan *et al.* , 2013; Estruch *et al.* , 2012), which is a neurogenesis regulator; or CASK (Kuo *et al.* , 2010), which carries out multiple functions in neurons. Notably, most of these are transcription factors, which suggests that they could interact with Bcl11a to cooperatively regulate downstream target genes important for brain development. Bcl11a target genes include Dcc and Map1b, which mediate the effects of Bcl11A on neurite branching and outgrowth (Kuo *et al.* , 2009), and the autism-related transcription factor Tbr1, the repression of which by Bcl11A is necessary for specification of cortical subcerebral projection neurons (Cánovas *et al.* , 2015). Interestingly, haploinsufficiency of TBR1 results in autism spectrum disorder (ASD) together with language impairment and etiological TBR1 variants disrupt the interaction with FOXP2, mutations of which lead to CAS (Deriziotis *et al.* , 2014a). Also, Sema3c, which is required to control migration of cortical projection neurons, is a major downstream target gene of Bcl11a (Wiegreffe *et al.* , 2015).

There is therefore growing evidence that BCL11A is an important regulator of neurodevelopment and that it is potentially implicated in neurodevelopmental disorders. In the work described in this chapter, I sought to confirm the role of *BCL11A* in neurodevelopmental disorder by performing functional characterization of BCL11A variants arising from point mutations in patients with such disorders. Assays in cell-based models support the hypothesis that the missense mutations result in loss of function of the mutated proteins *in vivo*, highlighting BCL11A haploinsufficiency as a cause of the disorder in the patients. These assays, combined with detailed clinical characterization of patients with BCL11A variants by our collaborators, including identification of additional cases with loss-of-function variants, provide the evidence to define a novel ID syndrome.

5.2 MATERIAL AND METHODS

DNA constructs

The coding sequences of BCL11A-S (NM_138559), BCL11A-L (NM_018014), NONO (NM_001145408) and NR2F1 (NM_005654) were amplified from human fetal brain cDNA using the primers in Table S1 and cloned into pCR2.1-TOPO (Invitrogen). The patient missense mutations were introduced using the Quik-Change Lightning SDM kit (Agilent). For expression of fusion proteins with *Renilla* luciferase, YFP and mCherry, cDNAs were subcloned into the pLuc, pYFP and pmCherry expression vectors, respectively, which have been described previously, using the BamHI and XbaI sites (Deriziotis, Graham, Estruch, & Fisher, 2014; Deriziotis, O'Roak, et al., 2014). For the mammalian one-hybrid assay, a vector for expression of BCL11A fused in frame with the yeast GAL4 DNA-binding domain was created by cutting and re-ligating pBIND (Promega) at the ClaI sites to remove the *Renilla* luciferase expression cassette. Wild-type and mutant forms of BCL11A-L were subcloned into the BamHI and XbaI sites of this vector. A reporter plasmid was generated by inserting a KpnI-NcoI fragment of pG5luc (Promega) containing five GAL4 binding sites and a minimal adenovirus major late promoter into the vector pGL4.23 (Promega), which contains a codon-optimized firefly luciferase gene. A plasmid containing *Renilla* luciferase downstream of the herpes simplex virus thymidine kinase promoter (pGL4.74, Promega) was used for normalization.

Cell culture

HEK293 cells (ECACC, cat. no. 85120602) were cultured in DMEM supplemented with 10% FBS. Transfections were performed using GeneJuice (Merck-Millipore) according to the manufacturer's instructions.

Western blotting

Cells were transfected in 6-well plates and cultured for 48 h. Cells were lysed for 10 min at 4° C with 100 mM Tris pH 7.5, 150 mM NaCl, 10 mM EDTA, 0.2% Triton X-100, 1% PMSF, and protease inhibitor cocktail. Cell lysates were

cleared by centrifugation at 10,000 x g for 3 min at 4° C. Proteins were resolved on 10% SDS-polyacrylamide gels and transferred to PVDF membranes using a TransBlot Turbo blotting apparatus (Bio-Rad). Membranes were blocked in phosphate buffered saline containing 5% non-fat milk powder and 0.1% Tween-20, and incubated overnight at 4° C with primary antibody. The following antibodies were used: anti-GFP (Clontech cat. no. 632380, 1:8000, for YFP constructs); anti-β -actin (Sigma cat. no. A5441, 1:10,000). After washing, membranes were incubated with horseradish peroxidase-conjugated goat anti-mouse or anti-rabbit IgG for 45 min at room temperature. Proteins were visualized using Novex ECL Chemiluminescent Substrate Reagent Kit (Invitrogen) and a ChemiDoc XRS+ imaging system (Bio-Rad).

Cellular assay fluorescence microscopy

Cells were seeded on coverslips coated with poly-L-lysine. Cells were cultured for 30 h post-transfection, and then fixed with methanol. Fluorescence images were acquired using Zeiss Axiovert A-1 or Axio Imager 2 fluorescence microscopes with ZEN Image software.

BRET assay

The BRET assay has been described in detail elsewhere (Deriziotis et al., 2014a). Briefly, cells were transfected in white, clear-bottomed 96-well plates, in triplicate, with 6 fmol *Renilla* luciferase fusion expression plasmid and 6 fmol YFP fusion expression plasmid (total mass of DNA was adjusted to 60 ng with filler plasmid). *Renilla* luciferase or YFP with a nuclear localization signal were used as controls. Cells were cultured for 48 h post-transfection. Enduren luciferase substrate (Promega) was added at a final concentration of 60 μM, and cells were cultured for a further 4 h. Luminescence was measured in a TECAN Infinite F200PRO microplate reader using the Blue1 and Green1 filters.

Mammalian one-hybrid assay

Cells were transfected in white, clear-bottomed 96-well plates, in triplicate, with 8.5 fmol pBIND-BCL11A, 5 fmol firefly luciferase reporter plasmid, and 2

fmol *Renilla* luciferase normalization plasmid (total mass of DNA was adjusted to 60 ng with filler plasmid). Cells were cultured for 48 h post-transfection. Firefly and *Renilla* luciferase activities were measured using the Dual Luciferase Reporter Assay (Promega).

5.3 RESULTS

5.3.1 *Identification of mutations in BCL11A*

In the DDD cohort six individuals were identified with *de novo* heterozygous variants in the *BCL11A* gene, from a total of 4,295 affected individuals studied using whole exome sequencing (Figure 5.2; Table 5.1). The first three variants to be discovered in the DDD cohort were all missense variants, which clustered together in exon 2 of *BCL11A* (Patients 1, 2 and 3) (Deciphering Developmental Disorders Study, 2015). Three additional variants discovered later in an expanded DDD cohort were all loss-of-function (LoF) variants (one nonsense, Patient 4, and two frameshift, Patients 5 and 6) (Figure 5.2b). Searching for additional BCL11A variants in other unpublished developmental disorder cohorts uncovered three additional patients with *de novo* LoF variants (Patients 7 to 9). Finally, two patients with LoF variants were identified among cohorts of individuals with ASD in published exome sequencing studies; the relevance of these variants to the patients' phenotypes had been unclear at the time of original publication (Patients 10 and 11). Thus, this study included a total of 11 patients, with 3 missense variants and 8 LoF variants.

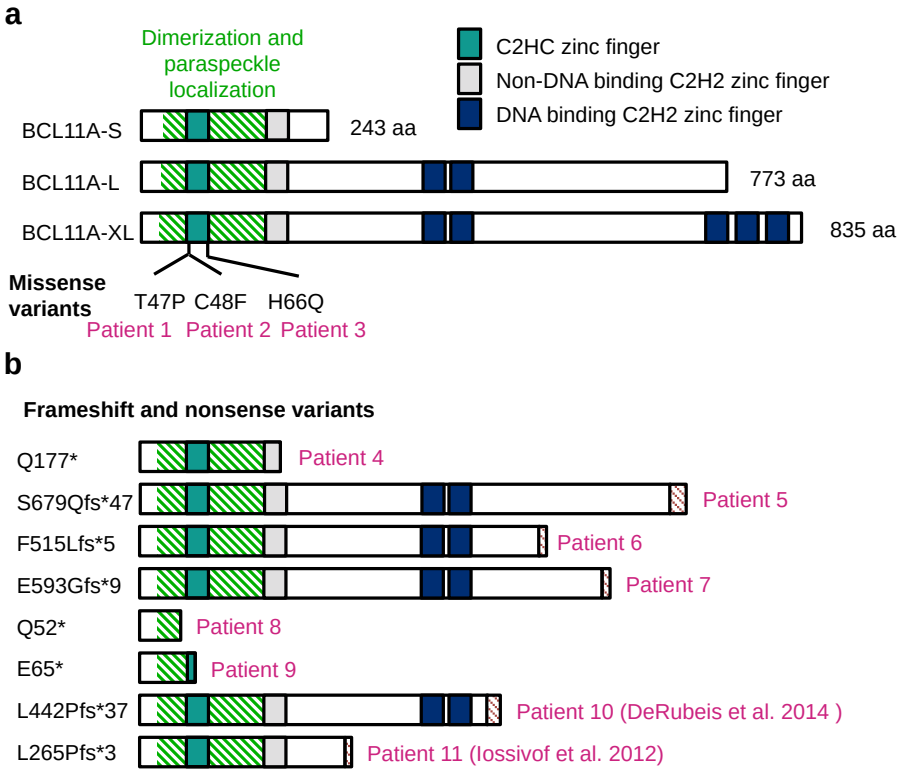


Figure 5.2: Variants in BCL11A found in patients with neurodevelopmental disorders. **(a)** Schematic representation of BCL11A isoforms S, L and XL, and BCL11A missense mutations. Known domains are shown: the C2CH zinc finger (ZF) is shaded in green; the non-DNA binding C2H2 zinc finger, in grey; and the DNA-binding C2H2 zinc finger domain, in dark blue. **(b)** Schematic representation of frameshift and nonsense BCL11A variants.

Patient #	1	2	3	4	5	6[a]	7	8	9	10[c]	11[d]	Summary of features
***BCL11A* Mutation**[b]	c.139A>C, p.**T47P**	c.143G>T, p.**C48F**	c.198C>A, p.**H66Q**	c.529C>T, p.**Q177***	c.2035_2037delTCCinsC, p.**S679Qfs*47**	c.1545delCinsGGCTTC, p.**F515Lfs*5**	c.1775_1776insTGGCTCAGCGG, p.**E593Gfs*9**	c.154C>T p.**Q52***	c.193C>T p.**E65***	c.1325_1325delT, p.**L442Pfs*37**	c.792_793insC, p.**L265Pfs*3**	3 missense 8 nonsense / frameshift
Decipher ID	262471	262952	261658	268026	275695	280953	n.a.	n.a.	n.a.	n.a	n.a	
Mutation class	Missense	Missense	Missense	Loss of function	Loss of function	Loss of function	Loss of function	Loss of function	Loss of function	Loss of function	Loss of function	
Sex	F	M	F	F	F	F	F	F	M	F	M	3M : 8 F
Microcephaly	–[f]	–[f]	+	+	+	+	+	–[g]	–[g]	n.a.	n.a.	5/9
Intellectual disability	Mild - moderate	Moderate - severe	Moderate	Moderate	Moderate	+[e]	Moderate	Moderate	Moderate	Severe	n.a.	10/10 (average moderate)
Developmental milestones (age of achievement, months)												
Sat independently	7	10	12	n.a.	11	n.a.	n.a.	14	12	n.a.	n.a.	~11 mo.
Walked independently	20	36	24	22	45	n.a.	n.a.	23	36	30	n.a.	~29.5 mo.
First words	22	27	36	16	n.a.[c] (80 words at 100 mo.)	24-30	n.a. (few words at 33 mo.)	60	36	n.a. (2 words at 6 years)	n.a.	~32 mo.
Craniofacial features												
Downslanting palpebral fissures	+	–	–	+	–	n.a.	+	+	–	n.a.	n.a.	4/8
Epicanthus	–	–	–	–	+	n.a.	+	–	+	n.a.	n.a.	
Strabismus	+	+	+	+	+	n.a.	+	+	+	n.a.	n.a.	8/8
Blue sclera in infancy	–	–	–	+	+	+	–	–	–	n.a.	n.a.	3/9
Flat midface	+	+	+	+	+	n.a.	–	–	+	n.a.	n.a.	6/8
Thin upper lip	+	+	+	+	+	n.a.	–	+	+	n.a.	n.a.	7/8
Everted lower lip	+	+	–	+	–	n.a.	+	+	+	n.a.	n.a.	6/8
Nose	Anteverted	Small nares	Anteverted; full tip	–	–	n.a.	Full tip	small nares; full tip	small nares; full tip	n.a.	n.a.	4/8
Micro/retrognathia	–	–	–	–	–	+[e]	–	+[g]	+[g]	n.a.	n.a.	3/9
Additional craniofacial features	Frontal upsweep; Cleft uvula	Coarse hair		Small mouth; High palate; Pointed chin	Small mouth; Plagiocephaly; Synophrys		Large tip of the nose, broad bridge; Flared eyebrows	Large tip of the nose; High palate	High nasal bridge	n.a.	n.a.	

							Telecanthus;					
External ear anomalies	–	–	+[A]	+[A]	+[A]	n.a.	+[A]	–	+[A]	n.a.	n.a.	5/8
Additional physical and neurologic features												
Joint hypermobility	+	–	+	+	+	n.a.	+	+	+	n.a.	n.a.	7/8
Short stature	–	–	–	–	–	+	+	–	–	n.a.	n.a.	2/9
Gait abnormalities	Broad based	Broad based; truncal ataxia	–	–	–	n.a.	–	–	Ataxia	n.a.	n.a.	3/8
Other	Anteriorly placed anus; dyspraxia		Fetal pads; Bilateral coxa valga; valgus foot deformity		Hernia repair	Congenital hip dislocation; Delayed bone age;	GE reflux;	Large 2[nd] metacarpals; Scoliosis;	Pectus excavatum,	n.a.	n.a.	
Behavioral features												
ASD	–	+	–	–	–	n.a.	–	–	–	+	+	3/10
Repetitive behavior	+	+	+	+	–	n.a.	–	–	–	–	n.a.	4/9
Other behavior problems	Emotional lability.	Recurrent hand flapping	Sensory abn.; self-injurious behavior.	Recurrent hand flapping/biting.	None reported	n.a.	None reported	Anxiety; Eating disorder.	None reported	attention deficit.	n.a.	6/9
Sleep disturbance	–	+	+	+	–	n.a.	–	–	–	+	n.a.	4/9
Additional investigations												
MRI	Slightly reduced WM volume	Small CV	n.a.	Normal	VM	n.a.	n.a.	Atrophy of the superior CV	Mild hypoplasia of the CC	n.a.	n.a.	
Hemoglobin F	20.8%	8%	8.7%	n.a.	26.3%	n.a.	n.a.	3.1%	8.6%	n.a.	n.a.	6/6

Table 5.1: Genetic and clinical characteristics of individuals with *de novo BCL11A* mutations. *BCL11A* mutations are annotated to transcript ENST00000335712, GRCh37.[B]Patient 6 has an additional probable pathogenic copy number variation (4.3Mb duplication: dup15q15.3q21.1). [C]Mutation reported by de Rubeis et. al., 2014. [D]Mutation reported by Iossifov et al, 2012. n.a.: not available. [E]OFC measured between percentiles 9 and 25. [F]Severity unknown. [G]no words at 28 mo. (months). ASD: autism spectrum disorder. Abn.: abnormalities. WM: white matter. VM: Ventriculomegaly. CV: cerebellar vermis. CC: corpus callosum. Patient ascertainment: Patients 1 to 6, the DDD study. Patients 7 and 8, clinical exome sequencing. Patient 9, PARI 2011 study.

Interestingly, given the exonic structure of *BCL11A*, LoF variants might affect each isoform differently (Table 5.2). Isoforms S and XS are unaffected by 5/8 patient LoF variants, suggesting that a functional S/XS isoform is insufficient for normal neurodevelopment, consistent with the lack of DNA-binding capacity for these isoforms (Table 5.2). All LoF variants are predicted to affect isoform L through nonsense-mediated RNA decay, whereas isoform XL might be affected by either premature truncation or nonsense-mediated decay, depending on the type of mutation. Given that isoform L is known to be the major isoform in the brain, all LoF variants might be expected to have comparable effects on neurodevelopment. However, it is possible that the differing effects on isoform XL may contribute to the observed variability in phenotypic profiles of the patients (Table 5.1).

Table 5.2: BCL11A patient variants effects on each isoform.
NMD, nonsense-mediated mRNA decay.

Patient #	BCL11A Variant	Predicted effect on each isoform		
		XL	L	S
1	p.T47P	Substitution	Substitution	Substitution
2	p.C48F	Substitution	Substitution	Substitution
3	p.H66Q	Substitution	Substitution	Substitution
4	p.Q177*	NMD	NMD	NMD
5	p.S679Qfs*47	Premature truncation	NMD	Unaffected
6	p.S515Lfs*5	Premature truncation	NMD	Unaffected
7	p.E593Gfs*9	Premature truncation	NMD	Unaffected
8	p.Q52*	NMD	NMD	NMD
9	p.E65*	NMD	NMD	NMD
10	p.L442Pfs*37	Premature truncation	NMD	Unaffected
11	p.L265Pfs*3	Premature truncation	NMD	Unaffected

5.3.2 *Clinical features of patients with BCL11A mutations*

Clinical characterization of the patients was carried out by C. Dias. All individuals identified with *BCL11A* mutations presented with global delay in developmental milestones, including speech and language delay. Most patients exhibit moderate ID, though the degree of cognitive dysfunction varies (Table 5.1). One of the two patients previously ascertained through an ASD study (De Rubeis *et al.* , 2014; Iossifov *et al.* , 2014) is also known to have severe ID (Patient 10); intellectual capacity was not reported in the other. Patient 2 has received a diagnosis of ASD, and four other patients present a spectrum of behaviour abnormalities including repetitive behaviour and sensory problems. Shared physical features among patients include joint laxity (87%), strabismus (100%), microcephaly (55%), thin upper lip and flat midface. 62% of patients have external ear abnormalities (Table 5.1), which are more severe in patients with nonsense and frameshift mutations. Patients 4, 5 and 6, all of whom have nonsense or frameshift mutations, had blue sclerae in infancy. As well as the *BCL11A* variant, Patient 6 also carries a probably pathogenic 4.3 Mb duplication of 15q15.3q21.1, and a contribution of this duplication to the phenotype, specifically to delayed skeletal maturation and short stature, cannot be excluded.

Fetal hemoglobin was significantly elevated in all patients in whom it was assessed, including all those with missense mutations in *BCL11A* (Table 5.1). The phenotypes of patients with missense and LoF variants are therefore similar, suggesting that the effect of the missense variants is to impair normal protein function. To determine if and how the missense variants affect protein function, I performed in silico analyses of the variants, followed by functional characterization using cell-based assays.

5.3.3 *Patient missense mutations cluster in a non-canonical zinc finger domain in BCL11A*

All three missense mutations are located in the N-terminal region of BCL11A, which is reported to be required for homo- and hetero-dimerization of BCL11A isoforms (Liu *et al.* , 2006), as well as for interaction with repressive nucleosome remodeling complexes (Cismasiu *et al.* , 2005). Specifically, these mutations cluster in the non-canonical C2HC zinc finger domain of BCL11A (Figure 5.3). A zinc finger is a small protein structure domain that mediates protein-

protein and protein-nucleic acid interactions. The C2H2 is the canonical zinc finger domain and forms a stable secondary structure by the coordination of a zinc atom by two cysteine residues and two histidine residues (Krishna *et al.* , 2003).

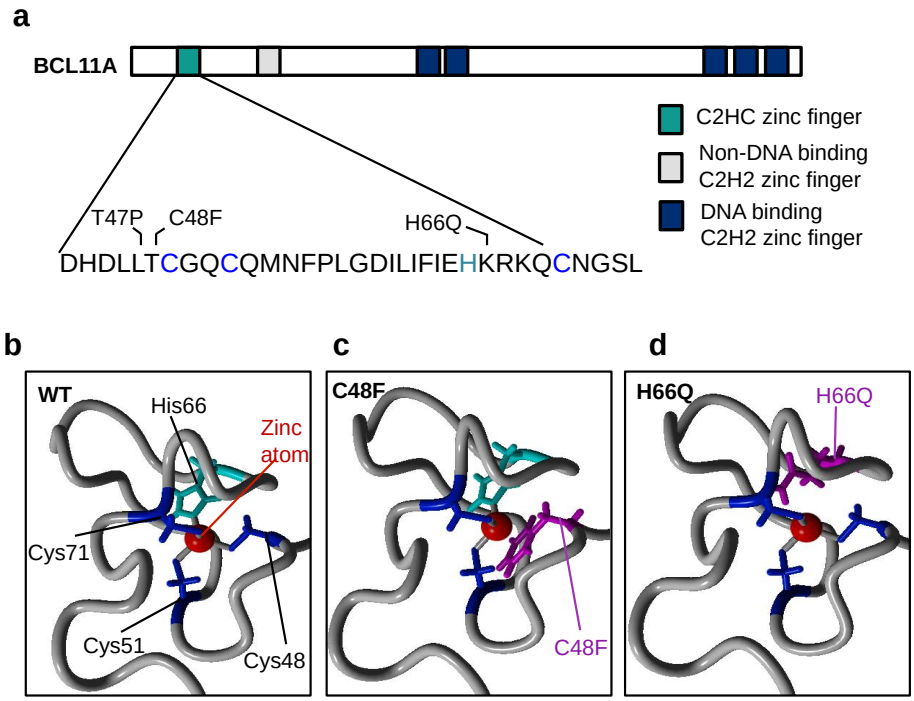


Figure 5.3: BCL11A missense substitutions cluster in the C2HC zinc finger domain of BCL11A. (a) Schematic representation of BCL11A isoform XL showing the three missense variants clustering in the C2HC zinc finger domain. The C2CH zinc finger (ZF) is shaded in yellow; the non-DNA binding C2H2 zinc finger, in grey; and the DNA-binding C2H2 zinc finger domain, in dark blue. Core cysteine and histidine residues of the domain are shown in dark and light blue, respectively. **(b-d)** 3D models of C2HC zinc finger (Protein Bata Bank [PDB] ID 2JYD) visualized with YASARA. The wildtype (WT) C2HC zinc finger **(b)** consists of three cysteine residues, in dark blue, and one histidine residue, in light blue, that hold in place the zinc atom, in red, to stabilize the secondary structure. BCL11A variants p.C48F **(c)** and p.H66Q **(d)**, in purple, directly disrupt core cysteine and histidine residues of the domain.

All of the zinc finger domains in BCL11A are of the C2H2 type, with the exception of the N-terminal C2HC domain (Avram *et al.* , 2000).The C2HC is a non-canonical type of zinc finger domain that utilizes 3 cysteine residues and one histidine residue to hold the zinc atom in place and stabilize the secondary structure (Krishna *et al.* , 2003). Molecular modelling suggests that the missense mutation p.T47P is adjacent to a central cysteine residue of this domain while p.C48F and p.H66Q affect core cysteine and histidine residues of the domain (Figure 5.3b-d). This suggests that all three missense mutations are likely to abolish zinc binding and disturb the folding of the C2HC zinc finger domain of BCL11A, resulting in the loss of molecular interactions involving this domain.

5.3.4 *Patient missense mutations disrupt cellular localization of BCL11A-L.*

To study the impact of the missense mutations on BCL11A protein function, I performed a battery of cellular assays focusing on BCL11A-L and BCL11A-S, the two isoforms of BCL11A reported to be more abundant in the brain (Kuo & Hsueh, 2007; Liu *et al.* , 2006; Satterwhite *et al.* , 2001). We generated mCherry and YFP-tagged versions of these isoforms containing each of the three patient missense mutations and expressed them in HEK293 cells (Figure 5.4).

The expression levels of wild-type and mutant BCL11A proteins were assessed by western blotting of lysates from transfected cells (Figure 5.4a). Wild-type BCL11A-L was detected at ~ 120 KDa and BCL11A-S at ~ 50 KDa. Proteins with the three missense mutations were of identical molecular weight and expressed in similar amounts to each corresponding wild-type BCL11A isoform.

Subcellular localization of BCL11A protein variants was examined in HEK293 cells transfected with mCherry-tagged BCL11A-L and BCL11A-S. Consistent with what was previously reported (Liu *et al.* , 2006), BCL11A-L was localized to the nucleus and BCL11A-S was observed only in the cytoplasm. The subcellular localization of the mutant forms of BCL11A-S was unaffected and indistinguishable from the wildtype form (Figure 5.4b). Wildtype BCL11A-L was found in nuclear paraspeckles in agreement with previous studies (Fox & Lamond, 2010). Nuclear paraspeckles are subnuclear structures that may be involved in the regulation of transcription and RNA processing, as inferred by the presence of paraspeckle specific proteins (i.e. NONO, PSPC1 and SFPQ) that are known to contribute to such processes (Fox & Lamond, 2010; Morimoto & Boerkoel, 2013). Strikingly, all three patient missense mutations disrupted the

paraspeckle distribution of BCL11A-L (Figure 5.4c) and showed a diffuse pattern of localization in the nucleus.

The nuclear paraspeckle localization of wild-type BCL11A-L was demonstrated by its co-localization with the paraspeckle-specific protein NONO (Figure 5.5a). Co-localization of the mutant forms of BCL11A-L with NONO was severely reduced but did not completely disappear. To further investigate this phenomenon, I used a bioluminescence resonance energy transfer (BRET) assay, which allows the monitoring of protein-protein interactions in living cells. In the BRET assay, a protein of interest is expressed as a fusion with *Renilla* luciferase (Luc) and a candidate interaction partner is expressed as a fusion

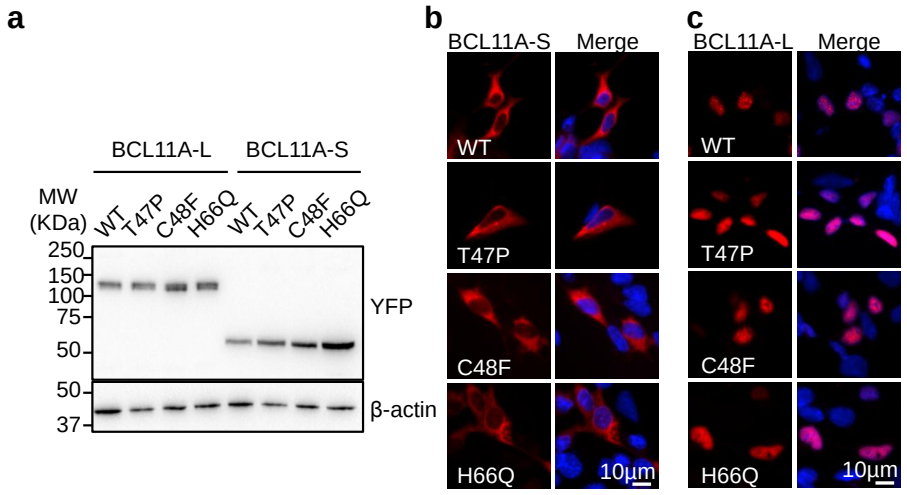


Figure 5.4: Expression of wildtype and mutant BCL11A-L and –S isoforms in HEK293 cells. (a) Western blot of mutant and wildtype BCL11A-L and BCL11A-S isoforms. HEK293 cells were transfected with mutant or wildtype (WT) BCL11A fused to YFP. Blots of whole cell lysates were probed with anti-YFP to detect BCL11A and with anti-β-actin to confirm equal loading. **(b)** Fluorescence micrographs of HEK293 cells transfected with wildtype (WT) or mutant short isoform of BCL11A fused to mCherry (red). Nuclei were stained with Hoechst 33342 (blue). **(c)** Fluorescence micrographs of HEK293 cells transfected with wildtype or mutant long isoform of BCL11A fused to mCherry (red). Nuclei were stained with Hoechst 33342 (blue). Scale bars represent 10 μm.

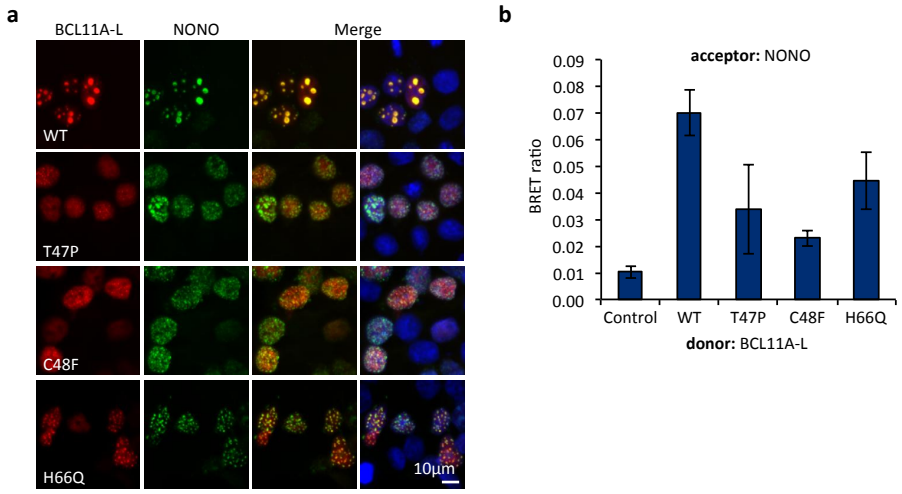


Figure 5.5: BCL11A missense substitutions alter paraspeckle localization of isoform L. (a) Fluorescence micrographs of cells transfected with wild-type or mutant BCL11A-L fused to mCherry (red) and NONO fused to YFP (green). Nuclei were stained with Hoechst 33342 (blue). Scale bars represent 10 μm. **(b)** Bioluminescence resonance energy transfer (BRET) assay for interaction of wild-type or mutant BCL11A-L with NONO. HEK293 cells were transfected with BCL11A-L fused to Renilla luciferase, and NONO fused to YFP. Values are mean corrected BRET ratios ± SEM (n = 3).

with yellow fluorescent protein (YFP). An interaction between the two proteins brings the Luc and YFP molecules into sufficient proximity to allow resonance energy transfer to occur upon addition of a luciferase substrate, shifting the wavelength of the emitted light from 480 nm to 530 nm. Using Luc-BCL11A-L and YFP-NONO fusion proteins, I confirmed that BCL11A-L interacted with NONO (Figure 5.5b), consistent with the co-localization observed in the nuclear paraspeckles. The patient mutations reduced, but did not completely abolish, the interaction of BCL11A-L with NONO, which together with the reduced co-localization suggested that a fraction of the expressed BCL11A mutant protein was still found in nuclear paraspeckles. Interestingly, NONO has recently been implicated in ID, therefore the interaction between BCL11A and NONO may represent a functional link between different ID syndromes (Mircsof *et al.*, 2015).

5.3.5 *Patient missense mutations disrupt homo- and hetero-dimerization*

The three patient missense mutations all lie within a non-canonical C2CH zinc finger domain and two of the mutations disrupt a core cysteine and a core histidine of this domain (Figure 5.3). This zinc finger domain is present in all BCL11A isoforms (Figure 5.1), and is within the region of the protein reported to mediate BCL11A dimerization, suggesting that it could potentially have a central role in mediating homo- and hetero-dimerization of BCL11A isoforms. I used the BRET assay coupled with fluorescence imaging of protein localization to study the effects of these variants on homo- and hetero-dimerization of BCL11A isoforms. The hetero-dimerization of BCL11A-L and BCL11A-XL with BCL11A-S mediates the translocation of BCL11A-S from the cytoplasm into the nucleus (Liu *et al.* , 2006). The results confirmed that wild-type BCL11A-L and BCL11A-S form homo- and heterodimers (Figure 5.6a,b; Figure 5.7a,b) (Kuo & Hsueh, 2007; Liu *et al.* , 2006) and that co-expression of BCL11A-L causes BCL11A-S to translocate from the cytoplasm into nuclear paraspeckles (Figures 5.6c, 5.7c) (Liu *et al.* , 2006). In contrast, BCL11A-L isoforms carrying mutations have a substantially reduced interaction with both L and S wild-type BCL11A isoforms (Figure 5.6a,b) and an abolished capacity to translocate wild-type BCL11A-S into the nucleus (Figure 5.6c). Analogous effects were observed when the patient mutations were introduced into BCL11A-S (Figure 5.7). Taken together, these results suggest that the variants could be blocking the correct folding of the C2CH zinc finger domain, thereby disrupting the dimerization ability of BCL11A.

5.3.6 *Patient missense mutations do not disrupt interaction with NR2F1*

BCL11A-L interacts with members of the nuclear receptor family 2F, NR2F1 and NR2F2, as well as with members of the nuclear receptor family 2E, NR2E1 and NR2E3, which are transcription factors expressed in neural tissues and with important roles in brain development (Avram *et al.* , 2000; Chan *et al.* , 2013; Estruch *et al.* , 2012). Moreover, mutations in NR2F1 have been reported in patients with ID (Bosch *et al.* , 2014). I used the BRET assay to confirm BCL11A interaction with NR2F1 and observed that none of the missense mutations in BCL11A-L disrupted the interaction. This finding is in line with the fact that the domains in BCL11A-L that mediate this interaction, which encom-

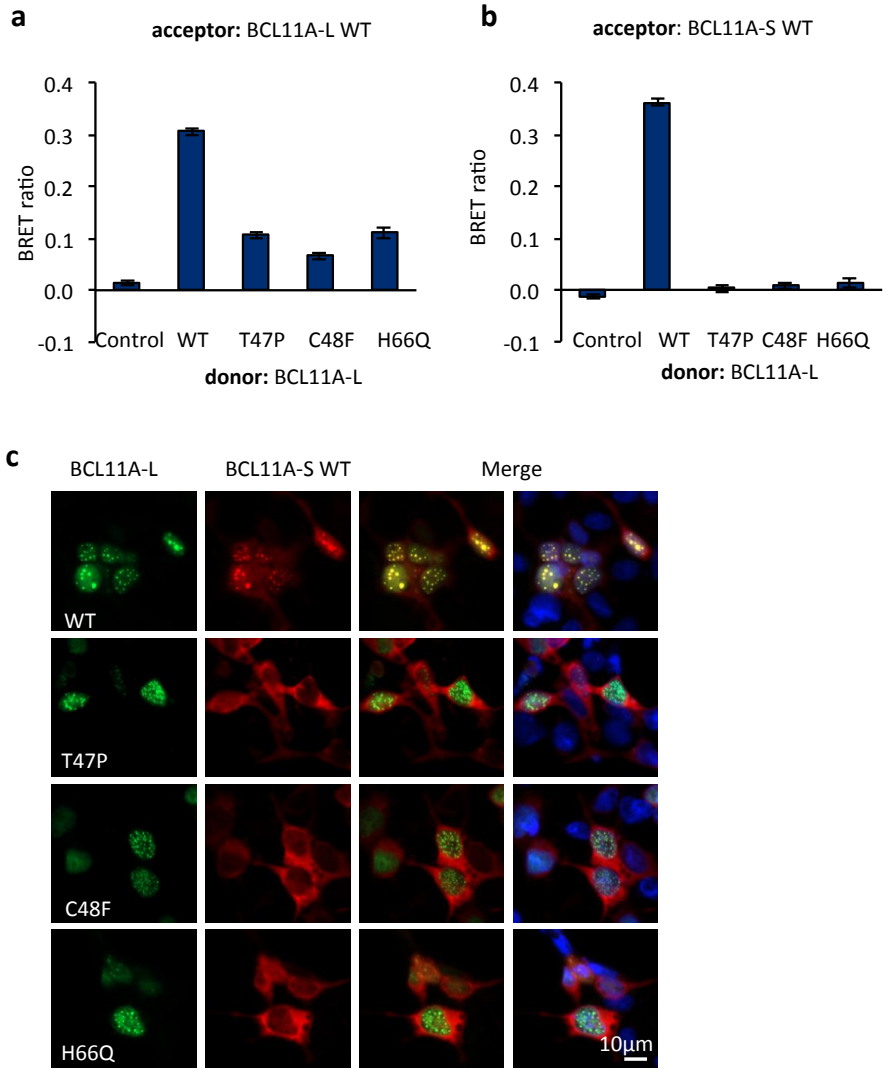


Figure 5.6: Missense substitutions in BCL11A-L disrupt homo and hetero-dimerization. (a) BRET assay for interaction of mutant BCL11A-L with wildtype (WT) BCL11A-L. HEK293 cells were transfected with wildtype or mutant BCL11A-L fused to Renilla luciferase (donor) and wildtype BCL11A-L fused to YFP (acceptor). The control donor protein is a nuclear-targeted luciferase. Values are mean corrected BRET ratios ± SEM (n = 3).

Figure 5.6: (Cont.) **(b)** BRET assay for interaction of mutant BCL11A-L with wild-type BCL11A-S. HEK293 cells were transfected with wildtype or mutant BCL11A-L fused to Renilla luciferase (donor) and wildtype BCL11A-S fused to YFP (acceptor). Values are mean corrected BRET ratios ± SEM (n = 3). **c** Fluorescence micrographs of HEK293 cells transfected with wildtype or mutant BCL11A-L fused to YFP (green), together with wildtype BCL11A-S fused to mCherry (red). Nuclei were stained with Hoechst 33342 (blue). Scale bar represents 10 μm.

pass residues 310-325 and 651-670, are located some distance away from the locations of the mutations (Chan *et al.* , 2013; Estruch *et al.* , 2012) (Figure 5.8a). When transfected alone, NR2F1 normally exhibits a diffuse nuclear distribution (Figure 5.8b) but when co-expressed with wild-type BCL11A-L, NR2F1 showed a nuclear speckled pattern that coincided with BCL11A-L (Figure 5.8c). In agreement with the BRET results, NR2F1 protein was diffuse in the nucleus when co-expressed with the patient BCL11A patient missense variants, which also localized in the nucleus in a diffuse manner (Figures 5.4c,5.8c).

5.3.7 *Patient missense mutations disrupt transcriptional activity*

BCL11A-L acts as a transcription factor by binding to DNA in a sequence-specific manner via two C2H2 zinc finger domains, while BCL11A-S lacks the DNA-binding zinc finger domains and is unable to bind to DNA (Figure 5.1) (Avram *et al.* , 2002; Liu *et al.* , 2006). We used a mammalian one-hybrid (M1H) assay to examine the effects of the patient missense mutations on the capacity of BCL11A to regulate transcription. BCL11A-L was fused to the DNA-binding domain of yeast GAL4 and co-transfected with a reporter plasmid containing five sequential GAL4 binding sites upstream of a luciferase gene (Figure 5.9a). Wild-type BCL11A-L produced a ∼ 2.5-fold activation of reporter transcription, showing that BCL11A can stimulate transcription in this system. The activation of reporter transcription was significantly reduced for all three mutant variants (p<0.05, one-way ANOVA followed by Bonferroni post-hoc test; Figure 5.9b), suggesting that these variants would have a reduced ability to regulate transcription of BCL11A target genes in vivo.

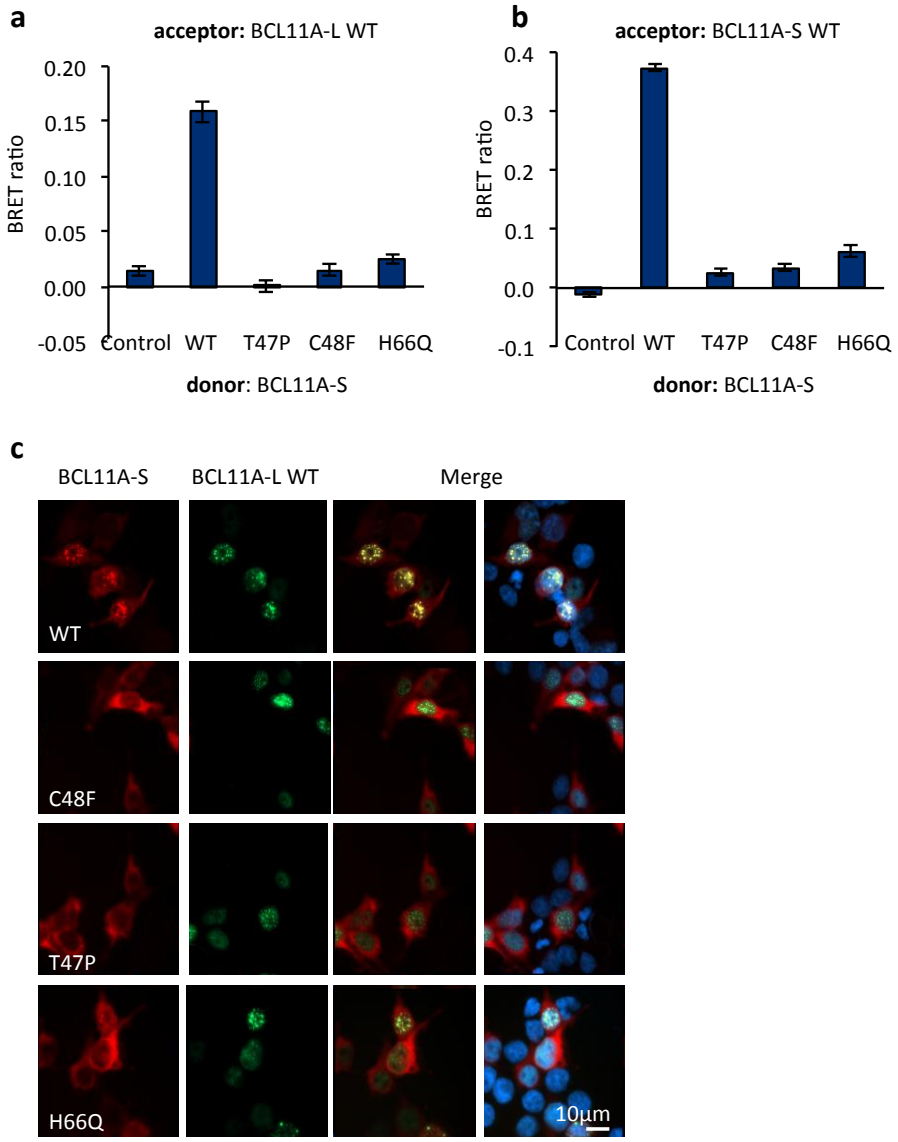


Figure 5.7: Missense substitutions in BCL11A-S disrupt homo and hetero-dimerization. (a) BRET assay for interaction of mutant BCL11A-S with wildtype (WT) BCL11A-S. HEK293 cells were transfected with wildtype or mutant BCL11A-S fused to Renilla luciferase (donor) and wildtype BCL11A-S fused to YFP (acceptor). The control donor protein is a nuclear-targeted luciferase. Values are mean corrected BRET ratios ± SEM (n = 3).

Figure 5.7: (Cont.) **(b)** BRET assay for interaction of mutant BCL11A-S with wild-type BCL11A-L. HEK293 cells were transfected with wildtype or mutant BCL11A-S fused to Renilla luciferase (donor) and wildtype BCL11A-L fused to YFP (acceptor). Values are mean corrected BRET ratios ± SEM (n = 3). **(c)** Fluorescence micrographs of HEK293 cells transfected with wildtype or mutant BCL11A-S fused to YFP (green), together with wildtype BCL11A-L fused to mCherry (red). Nuclei were stained with Hoechst 33342 (blue). Scale bars represent 10 μm.

5.4 DISCUSSION

The results of this study demonstrate that missense variants in BCL11A found in children with neurodevelopmental disorders have consistent deleterious effects on BCL11A localization, dimerization and transcriptional regulatory activity. Together with the clinical observation that patients with such variants have elevated fetal hemoglobin and a profile of other clinical features similar to loss-of-function mutations, the data support the hypothesis that these N-terminal missense mutations result in a loss of BCL11A function that results in neurodevelopmental impairments. The cellular assays suggest that *BCL11A* missense mutations are hypomorphic alleles, where the loss of function may not be complete. The missense mutations severely impair BCL11A transcriptional activity and dimerization but they do not completely abolish the localization of BCL11A in nuclear paraspeckles and do not alter its interaction with NR2F1. This could explain the similar but milder phenotypes seen in the patients with missense mutations, as compared to what is observed in patients with truncations, where loss of function of the affected allele is expected to be complete.

The work described in this chapter identifies a novel syndrome in individuals with missense and truncating *BCL11A* mutations that is most notably characterized by ID and persistence of fetal hemoglobin. While the patients do not present consistent recognizable dysmorphic features, the presence of mild dysmorphism with ID and persistence of fetal hemoglobin define a clinical syndrome, with the latter providing a valuable diagnostic tool. Persistence of fetal hemoglobin has also been noted in individuals with chromosomal microdeletions encompassing *BCL11A* (Basak *et al.* , 2015; Funnell *et al.* , 2015). In individuals with 2p15p16.1 deletions that do not disrupt the BCL11A coding sequence, testing for persistence of fetal hemoglobin could be a useful method for ascer-

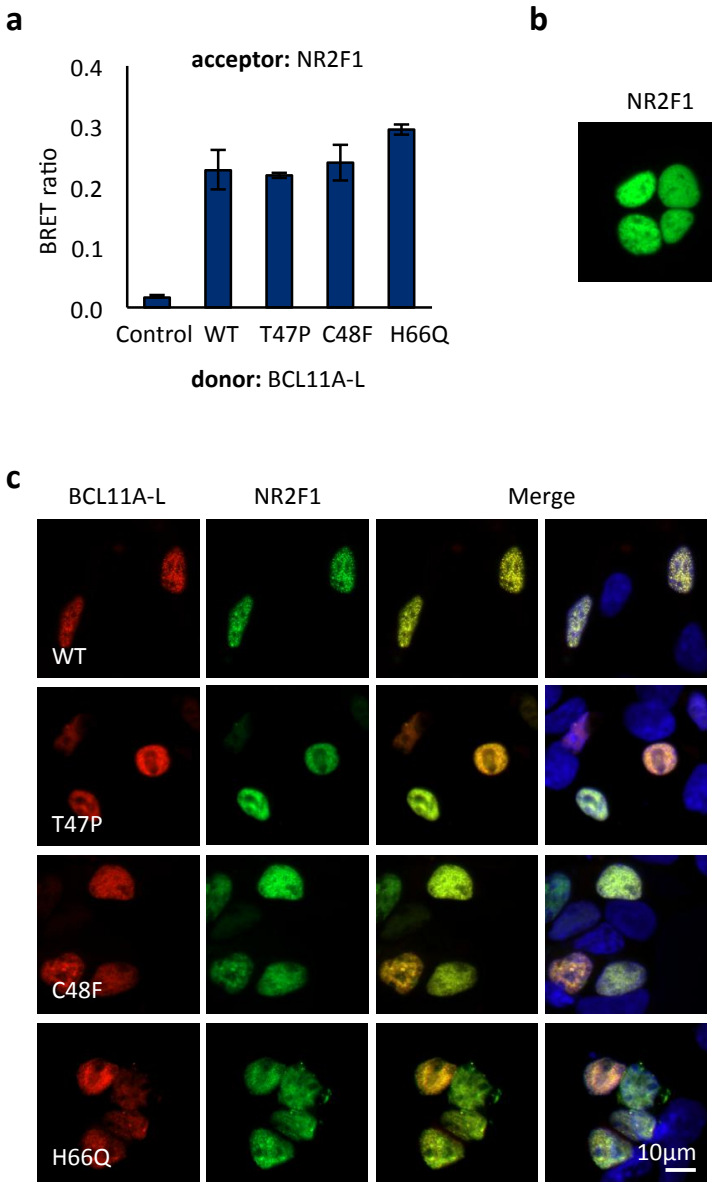


Figure 5.8: Missense substitutions in BCL11A-L do not affect interaction with NR2F1. **(a)** BRET assay for interaction of mutant BCL11A-L with wildtype (WT) NR2F1. HEK293 cells were transfected with wildtype or mutant BCL11A-L fused to Renilla luciferase (donor) and wildtype NR2F1 fused to YFP (acceptor). The control donor protein is a nuclear-targeted luciferase. Values are mean corrected BRET ratios ± SEM (n = 3).

Figure 5.8: (Cont.) **(b)** Fluorescence micrographs of HEK293 cells transfected with NR2F1 fused to YFP. **(c)** Fluorescence micrographs of HEK293 cells transfected with wildtype or mutant BCL11A-L fused to mCherry (red), together with NR2F1 fused to YFP (green). For **(b)** and **(c)**, Nuclei were stained with Hoechst 33342 (blue). Scale bar represents 10 μm.

taining whether critical regulatory sequences have been deleted, leading to loss of BCL11A expression. Fetal hemoglobin persistence is asymptomatic and does not affect the normal performance of the hematopoietic system in individuals with BCL11A disruptions.

BCL11A has previously been proposed as a susceptibility gene for ASD (De Rubeis *et al.* , 2014). However, despite a high frequency of behaviour abnormalities among patients with BCL11A disruptions, only 30% fulfill the diagnostic criteria for ASD, whereas all have ID. Thus in contrast to ID, ASD is a variable feature of the BCL11A syndrome. The microdeletion spanning only the *BCL11A* gene in a child with CAS suggested a strong involvement of this gene in speech and language development (Peter *et al.* , 2014). The findings here support a role of BCL11A in communication skills, but in this set of cases the language phenotype is predominantly characterized by a considerable delay of language development. Moreover, the clinical profiles also underscore the importance of this gene in other features, such as developmental delay, ID and hypotonia. A recent review of 36 cases of 2p15p16.1 microdeletion syndrome further confirmed that the phenotype caused by BCL11A haploinsufficiency includes disruptions of speech or language functions, but the manifestation of these signs is variable and the affected patients also present other mild cognitive deficits and weak muscular tone (Lévy *et al.* , 2017). Recently, a new frameshift mutation was identified in a child with CAS, dyspraxia, mild ID and hypotonia across motor systems including the oral and speech motor systems (Soblet *et al.* , 2017). Taken together, these findings point towards a broad role of BCL11A in the development of social cognitive abilities.

Results in this chapter have been published as part of a larger study that included a mouse model of BCL11A haploinsufficiency (Dias *et al.* , 2016). Notably, mice carrying only one copy of the *BCL11A* gene exhibit microcephaly together with impaired social behaviour, a phenotypic profile that is consistent

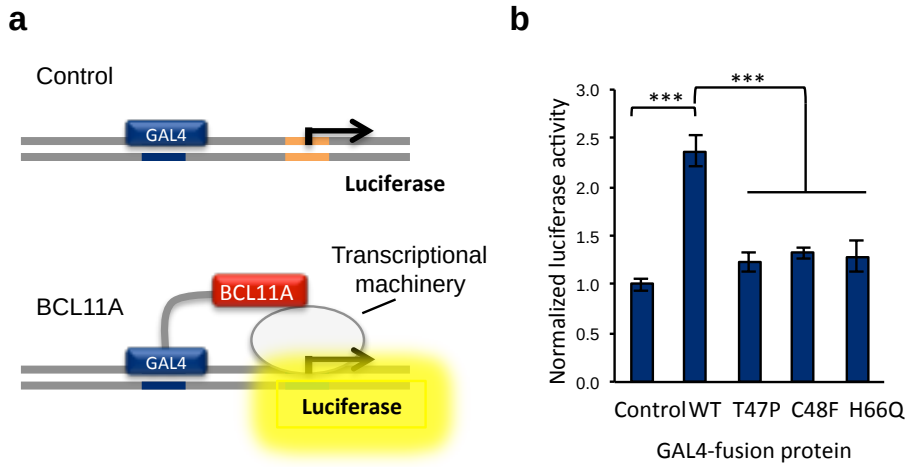


Figure 5.9: BCL11A missense substitutions disrupt transcriptional regulation. (a) Schematic representation of the mammalian one-hybrid (M1H) assay to asses transcriptional regulatory activity. **(b)** M1H assay for BCL11A-L transcriptional regulatory activity. HEK293 cells were transfected with wildtype or mutant BCL11A-L fused to the DNA-binding domain of GAL4, together with a firefly luciferase reporter plasmid containing GAL4 binding sites, and a Renilla luciferase normalization plasmid. Values are mean firefly luciferase activity normalized to Renilla luciferase activity ± SEM (n = 3), expressed relative to the control (***$p < 0.001$, one-way ANOVA followed by Bonferroni post hoc correction).

with that observed in patients with loss-of-function mutations of the human orthologue. This supports the hypothesis that haploinsufficiency of BCL11A is sufficient to cause neurodevelopmental disorder. Additionally, large-scale transcriptomic analyses of this mouse model revealed differences in gene expression in the cortex and the hippocampus between heterozygous and wildtype neonatal mice. Together with the impaired transcriptional activity that I demonstrated for the BCL11A missense mutations, these findings suggest that transcriptional dysregulation is an underlying etiological mechanism in the associated syndrome.

The functional experiments described in this chapter also add to the knowledge of BCL11A function, supporting the hypothesis that the zinc finger do-

main in the N-terminus of the protein is involved in dimerization and is necessary for transcriptional regulatory activity. Notably, the DNA-binding domains of BCL11A remain intact in patients with missense BCL11A variants, but regulation of transcription is impaired. Thus, cell-based assays and patient features (clinical and hematological) indicate the importance of protein-protein interactions for the role of BCL11A in regulating gene expression. This implies that the N-terminus region of BCL11A has an underappreciated significance for transcriptional repression of fetal hemoglobin and other target genes. This region may prove to be a hotspot for disorder-related missense mutations in BCL11A as more cases come to light. The suite of assays employed here to investigate the functional consequences of missense mutations in BCL11A proved to be highly effective in detecting the deleterious effects of these variants and therefore provides a framework for functional assessment of novel BCL11A variants in the future.

6

GENERAL DISCUSSION

6.1 SUMMARY

The aim of this dissertation was to shed light on the biological foundations of human speech and language by investigating the molecular function of transcription factors related to monogenic language-related disorders. As described in **Chapter 1**, the complex foundations of speech and language entail a challenging search for the underlying genes. Large-scale genetic association studies have revealed a few common genetic variants with small contributions to speech and language, but the most fruitful entry points into the biological bases have come from investigating monogenic language-related disorders, such as cases of childhood apraxia of speech (CAS) caused by rare *FOXP2* mutations. *FOXP2* was the first gene to be clearly linked with speech and language, and the vast majority of current knowledge on relevant molecular mechanisms comes from investigating its functions in cellular and animal models. *FOXP2* encodes a transcription factor that belongs to the FOXP subgroup of forkhead-box proteins. In this subgroup, the related proteins FOXP2, FOXP1 and FOXP4 are expressed in the central nervous system and play overlapping yet distinct roles in brain development. FOXP1 is the closest homologue to FOXP2 and has been implicated in an intellectual disability (ID) syndrome that includes language impairment.

The common aim of **Chapters 2** to **4** was to identify and investigate novel FOXP interaction partners. Deciphering the physical interaction networks of the FOXP proteins can reveal valuable insights into the ways these transcription factors contribute to brain development and how are they regulated, as well as pinpointing novel language-related candidate genes. Prior to the work described here, this particular aspect of the molecular biology of the FOXPs had been scarcely studied. The two most commonly used protein-protein interaction screening methods are yeast two-hybrid (Y2H) assays and affinity purification followed by mass spectrometry (AP/MS). Previous studies by our group had used both these techniques to search for FOXP2 interacting proteins pro-

viding a list of potential candidate partners that awaited confirmation with different methods.

In **Chapter 2**, I validated a putative interaction between FOXP2 and PIAS1 found in prior Y2H screens and demonstrated that it mediates the post-translational modification (PTM) of FOXP2 with SUMO molecules. I identified a major single sumoylation site at K674 conserved in FOXP2 orthologues and paralogues, which suggests that sumoylation of this protein has an evolutionarily ancient role. Abolishing the major sumoylation site disrupted FOXP2 sumoylation but did not produce any changes in subcellular localization, stability, transcriptional regulation, dimerization with wild-type FOXP2, or interaction with the co-repressor CtBP1. Sumoylation is a fast and dynamic process, which implies that only a minor fraction of protein is sumoylated at a given time. Therefore, the functional consequences of FOXP2 sumoylation may be subtle and only detectable in the context of a developing organism. Additionally, I analysed sumoylation of the FOXP2 variant R553H, which disrupts the forkhead-box domain and is known to be the cause of CAS in the large multigenerational KE family. I found that the mutated protein is sumoylated to a lesser extent than wild-type FOXP2 and shows a reduced interaction with PIAS proteins. These results suggest that the forkhead-box domain may also be required for proper sumoylation of FOXP2.

In **Chapter 3**, I sought to confirm the putative interactions of FOXP2 with NONO and SFPQ, which were found in AP/MS screens previously performed in our group. As neurodevelopmental transcription factors that interact with each other, NONO and SFPQ were promising candidates for interaction with FOXP2. I used a range of different complementary assays to test for FOXP2 interactions with SFPQ and NONO, but such interactions could not be detected in any of the experiments. In contrast, FOXP2 homo-dimerization and NONO interaction with SFPQ were successfully observed with the techniques employed. Based on the findings of this chapter, together with particular experimental caveats of the previously performed AP/MS screens, it seems likely that NONO and SFPQ were false positives in the original FOXP2 protein-protein interaction screens. Thus, this work suggested that new mass spectrometry screens should be performed in order to robustly identify novel FOXP2 interaction partners.

Consequently, in **Chapter 4**, I performed new AP/MS screens to search for FOXP2 and also for FOXP1 and FOXP4 protein-protein interactions, providing the most extensive list of putative FOXP interacting-proteins reported to date.

I characterized the interaction between the FOXPs and 7 neurally-expressed transcription factors– SOX5, NR2F1, NR2F2, SATB1, SATB2, ZMYM2 and YY1. I demonstrated overlapping expression of these transcription factors with the FOXPs in regions of the developing brain, which suggests that they can cooperate *in vivo* to regulate downstream genes in neural sites of co-expression. Several of the interacting transcription factors are known to be crucial for brain development and are disrupted in individuals with neurodevelopmental and/or psychiatric disorders. I showed that these interactions are disrupted by FOXP1 and FOXP2 mutations that cause language-related neurodevelopmental disorders, revealing novel functional links between transcription factors underlying such conditions. Finally, I identified specific regions in FOXP2 mediating the interactions. The findings in this chapter highlight the power of proteomic-based tools in expanding our understanding of the role of FOXP proteins and other neural transcription factors in brain development, and in the biology of speech and language.

In **Chapter 5**, I set the FOXPs aside to focus on a novel language-related candidate gene, *BCL11A*. Recent advances in DNA sequencing have revolutionized the field of genetics making it easier, faster and cheaper to find molecular causes of neurodevelopmental disorders. This is fueling the discovery of novel genes with putative roles in neurodevelopment, and in speech and language, one of which is *BCL11A*. This gene encodes a transcription factor that has mostly been studied in relation to its functions in the hematopoietic system; few investigations have been carried out in a neuronal context. A microdeletion affecting only the *BCL11A* gene found in a patient with CAS and mild ID highlighted a possible role of this gene in speech and language. Deletions involving *BCL11A* had been previously found in individuals with a neurodevelopmental phenotype that included some common features such as ID and variable degrees of language impairment. In this chapter, I characterized the functional impacts of three newly identified BCL11A missense variants found in several individuals with ID and delayed developmental milestones, including speech and language delay. The substitutions in BCL11A cluster in the zinc-finger domain in the N-terminal region. My functional assays revealed that BCL11A missense mutations negatively affect several aspects of protein function including subcellular localization, dimerization and transcriptional activity. Overall, these experiments confirmed the pathogenic role of the missense mutations, and uncovered roles of the N-terminal zinc-finger domain of BCL11A in protein dimer-

ization, subcellular localization and transcriptional activity. By combining the functional data with detailed clinical descriptions of the patient phenotypes obtained by our collaborators, we were able to define a novel monogenic neurodevelopmental disorder characterized by ID, persistence of fetal hemoglobin, and other variable features (OMIM 617101).

6.2 FINE-TUNING FOXP2 WITH POST-TRANSLATIONAL MODIFICATIONS

Transcription factors are often regulated by post-translational modifications (PTMs) (Filtz *et al.* , 2014; Skelly *et al.* , 2016). PTMs are reversible protein chemical modifications that can transiently modulate protein function in a myriad of ways (Seo & Lee, 2004). Typically, transcription factors are subject to a considerable number of PTMs, which are often interconnected with each other forming complex regulatory systems (Filtz *et al.* , 2014; Skelly *et al.* , 2016). For instance, it is known that FOXP3 can be acetylated, ubiquitinylated and phosphorylated and that these PTMs affect many core aspects of FOXP3 function, such as transcriptional activity and protein stability (van Loosdregt & Coffer, 2014). In contrast, the regulation of the other FOXP proteins (FOXP1/2/4) by PTMs has been clearly underexplored.

In chapter 2, I demonstrated that FOXP2 is subject to sumoylation, the first PTM reported for the major isoform of this protein. This finding was also reported by two independent studies within the same period of time when we published the results of chapter 2 (Meredith *et al.* , 2015; Usui *et al.* , 2016). The three investigations agree on the basic molecular mechanisms of FOXP2 sumoylation: it occurs at lysine K674 and is promoted by members of the PIAS family. However, there are some discrepancies regarding the effects that this PTM has on FOXP2 function (Table 6.1). The consequences of sumoylation are varied but the underlying principle is that it alters inter- and/or intramolecular interactions to change protein localization, stability, and/or activity. My work in HEK293 cells showed that the K674R substitution in FOXP2, which abolishes the sumoylation site, behaves in a similar way to the wild-type protein in all the assays that I performed (Table 6.1).

Conflicting with the work in this dissertation and with the investigations by Meredith et al., Usui and colleagues reported that inhibition of sumoylation resulted in an increased cytoplasmic localization of FOXP2, and that the K674R mutant was found both in the nucleus and in the cytoplasm in human HEK293

Table 6.1: Comparison of the effects of FOXP2 sumoylation on protein function observed in three independent studies. The effects of sumoylation on the functions of FOXP2 have been investigated in three independent studies (Chapter 2, (Meredith *et al.* , 2015; Usui *et al.* , 2016)) The different investigations found that some FOXP2 protein functions were altered by sumoylation (Affected) and others were not (Unaffected).

	Protein stability	**Protein-Protein interactions**	**Subcellular localization**	**Transcriptional activity**	
				HEK293	**hNP**
Meredith et al.	Unaffected	Not tested	Unaffected	Affected	Not tested
Chapter 2	Unaffected	Unaffected	Unaffected	Unaffected	Not tested
Usui et al.	Unaffected	Unaffected	Affected	Unaffected	Affected

cells, and in the mouse brain (cortical layer 6 neurons and Purkinje cells in the cerebellum) (Usui *et al.* , 2016). Thus, Usui et al. suggested that sumoylation is required for nuclear localization of FOXP2 and that this PTM affects its transcriptional activity, as cytoplasmic protein is not able to bind to DNA. However, they did not find any substantial differences between wild-type and K674R FOXP2 in the transcriptional repression of a canonical FOXP2 motif, in line with the findings in Chapter 2. Conversely, Meredith et al. found that K674R FOXP2 exhibited a reduced repression of *DISC1*, *SRPX2* and *MIR200c* promoters compared to wild-type protein (Meredith *et al.* , 2015). However, though they were statistically significant, the differences between K674R and wild-type FOXP2 were relatively small. The modulation of FOXP2 transcriptional activity by sumoylation might be affecting an aspect of protein function that has not been assessed in any of the studies, such as a specific DNA or protein interaction. Alternatively, it could be the case that the functional consequences of sumoylation are subtle, and/or cell-type and promoter dependent. Usui et al. found that CNTNAP2 mRNA levels were higher in human neural progenitor cells expressing FOXP2 K674R than in those expressing the wild-type protein, possibly suggesting that in human neuronal progenitor cells, FOXP2 sumoylation is necessary to repress CNTNAP2 expression (Usui *et al.* , 2016). However,

none of the putative functional consequences of FOXP2 sumoylation has been replicated in (at least) two independent studies and additional confirmatory studies are needed.

In any case, the effects of sumoylation may precisely fine-tune FOXP2 activity in such a way that they may only be evident in the context of a developing brain. It has been shown that in the brain, sumoylation has an impact on diverse aspects of neuronal morphology and function, and that dysregulation of the SUMO pathway often leads to neurological and neurodegenerative disorders (Henley *et al.* , 2014). Strikingly, the work by Usui and colleagues suggests that FOXP2 sumoylation in mice may be important for motor control and ultrasonic vocalizations through the regulation of Purkinje cell growth in the developing cerebellum (Usui *et al.* , 2016). In the study, they detected a sumoylated form of FOXP2 protein in the neonatal cerebellum. According to their data, almost all FOXP2 is sumoylated at specific neonatal stages (P7, P10, P14), which are critical time points of cerebellar development (Usui *et al.* , 2016). This is an unusual finding since other studies suggest that the fast dynamics of sumoylation means that only a minor fraction of a protein is sumoylated at a certain time point (Wilson, 2017). FOXP2 knockdown in the neonatal cerebellum resulted in a reduction of neuron outgrowth, a phenotype that could be rescued by overexpressing the wild-type protein but not with a K674R mutant (Usui *et al.* , 2016). Moreover, FOXP2 knockdown also altered motor function, as well as the number and length of ultrasonic vocalizations, phenotypes that again could only be rescued with wild-type FOXP2 but not with the K674R variant (Usui *et al.* , 2016). In the future it will be crucial to confirm these findings with, for instance, the generation of transgenic mice that express FOXP2 K674R only in the cerebellum.

The sumoylation site in FOXP2 is conserved among its paralogues and orthologues, suggesting an ancestral role for this modification in the FOXP family. In Chapter 2, I showed that FOXP1, like FOXP2, interacts with members of the PIAS family and with SUMO molecules, which indicates that FOXP1 is probably sumoylated. Indeed, a subsequent study confirmed that FOXP1 is sumoylated at the same conserved site as FOXP2 (K670 in FOXP1, K674 in FOXP2) (Rocca *et al.* , 2017). That study demonstrated that FOXP1 sumoylation regulates transcription cooperating with the CtBP1 co-repressor complex to promote dendritic morphogenesis in rat neurons (Rocca *et al.* , 2017). In HEK293 cells, FOXP1 mutant K670R did not repress transcription in a luciferase assay

using the SV40 promoter, a common binding promoter of FOXP1 and FOXP2. Conversely, in chapter 2 I show, in the same cell line, that FOXP2 K674R repressed the SV40 promoter with the same strength as wild-type FOXP2. In neuronal cell lines, Rocca and colleagues reported that FOXP1 sumoylation is required for the regulation of CNTNAP2 expression, similarly to the findings that Usui and colleagues reported for FOXP2 (Usui *et al.* , 2016; Rocca *et al.* , 2017). This suggests that sumoylation affects the specific transcriptional activities of FOXP1 and FOXP2 in different but overlapping ways to regulate neuronal development.

The rapid and dynamic nature of sumoylation makes it a well-suited mechanism to modify the activity of proteins such as FOXP2 and FOXP1 in response to activity within neural circuits. Remarkably, activity-dependent calcium signaling regulates the ability of FOXP1 sumoylation to recruit the CtBP1 co-repressor complex, as neuronal activity results in the loss of HDAC2 from this complex (Rocca *et al.* , 2017). In the future, it will be interesting to determine how FOXP2 sumoylation may be regulated in the brain in response to environmental cues, and/or whether there are other mechanisms that regulate neural FOXP1 sumoylation.

FOXP1/2 sumoylation is probably not the only PTM regulating transcription factor function; hence, additional FOXP1/2 PTMs remain to be identified. Various PTMs are known to fine-tune FOXP3 function, some of which lie on residues that are conserved in the other FOXPs (van Loosdregt & Coffer, 2014). For instance, FOXP3 lysine residues 268 and 393 are modified by ubiquitin to regulate FOXP3 degradation (van Loosdregt *et al.* , 2013). These sites are conserved in the other FOXPs and lie in the leucine zipper and forkhead domain, respectively. Additionally, lysine 268 can be acetylated, which has been shown to regulate both FOXP3 protein stability and transcriptional activity (Kwon *et al.* , 2012). These findings pinpoint sites in FOXP1 and FOXP2 that may be post-translationally modified, uncovering additional PTMs that potentially regulate FOXP1 and FOXP2 other than sumoylation.

PTMs such as acetylation, ubiquitinylation or phosphorylation in FOXP1 and FOXP2 could be identified utilizing proteomic mass-spectrometry approaches. Additionally, it would be interesting to analyze missense variants in patients with neurodevelopmental disorders to see if these affect PTM sites.

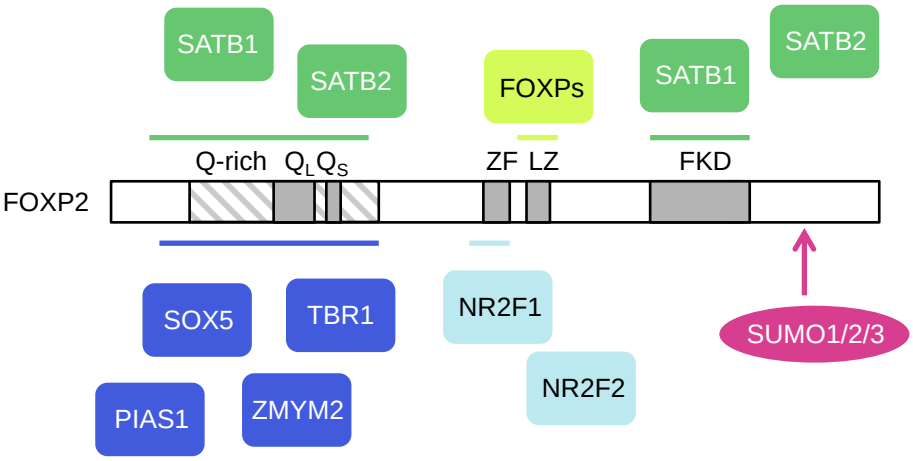


Figure 6.1: Protein interaction regions of FOXP2. Schematic representation of FOXP2 protein. Known domains are shown in grey: the glutamine-rich region (Q-rich), the polyglutamine tracts shaded in a dashed pattern, the zinc finger (ZF), the leucine zipper (LZ), the forkhead domain (FOX) and the nuclear localization signals (NLS). The regions in FOXP2 that mediate specific protein-protein interactions are indicated with horizontal lines, which are colour-coded for each interaction partner. The sumoylation site is indicated with an arrow.

6.3 NOVEL FUNCTIONAL REGIONS IN FOXP PROTEINS

When I started this thesis, the roles of the FOXP forkhead-box domain in DNA binding and of the leucine zipper in homo- and hetero-dimerization were well known, but the functional significance of other protein regions remained to be discovered. The protein-protein interaction work of this dissertation has uncovered new functions for uncharacterized regions of FOXP2, such as the N- and C-terminal parts, and for known domains, such as the zinc finger and the forkhead-box domains (Figure 6.1).

The N-terminal region of FOXP2 seems to coordinate several different protein-protein interactions. This part of FOXP2 is essential for the interaction with the transcription factors TBR1, SOX5, ZMYM2 and also the PIAS enzymes (Chapters 2 and 4) (Deriziotis *et al.* , 2014a),and may additionally be important for the interaction with SATB1 and SATB2 (Chapter 4). The N-terminal part of FOXP2

is an intrinsically disordered region (IDR) (Viscardi *et al.* , 2017). This kind of region is often found in transcription factors, in which it can also mediate different protein interactions (Shammas, 2017; Wang *et al.* , 2015). IDRs lack stable three-dimensional structures, which allows for a wide range of conformations (Dyson & Wright, 2005). Their versatility in adopting different conformations enables interactions with multiple proteins (Hsu *et al.* , 2012).

The conformational plasticity of IDRs also confers accessibility to PTM enzymes, such as PIAS proteins, and many PTMs occur preferentially in IDRs (Mittag *et al.* , 2010). In line with this, FOXP2 sumoylation takes place at residue K674 in the C-terminal region, which is also intrinsically disordered. Many other PTMs may be occurring in the N-terminal and C-terminal regions of FOXP2. As discussed in the previous section, it would be very interesting to identify more PTMs regulating FOXP2 and investigate how they crosstalk with each other to precisely control FOXP2 function.

Results in chapter 4 suggest that both the N-terminal and the forkhead-box domain in FOXP2 are required for the interaction with SATB1 and SATB2 (Figure 6.1). This means that either the SATB proteins physically interact with both domains or that FOXP2 has to be bound to DNA in order for the interaction to occur.

Notably, the FOXP2 substitution p.R533H, which is located in the forkhead-box domain and prevents the mutant protein from binding to DNA, disrupts most of the FOXP2 protein-protein interactions and also its ability to be sumoylated (Chapters 2 and 4), despite the fact that most of the mapped FOXP2 binding-regions for these interactions do not include this domain. A hypothesis to explain this paradox would be that some FOXP2 protein-protein interactions are DNA-dependent and can only occur if FOXP2 is bound to DNA. Alternatively, it could be the case that the substitution p.R553H destabilizes the forkhead-box domain, causing other protein regions to adopt an aberrant conformation that renders them inaccessible to other proteins.

Finally, the BRET results in chapter 4 suggest the zinc-finger domain of FOXP2 probably mediates the interaction with the transcription factors NR2F1 and NR2F2. Accordingly, this domain contains a site that partially overlaps with a known NR2F binding motif (Chan *et al.* , 2013). This would be the first indication of the functional role of the zinc-finger domain in FOXP2. However, further experiments involving the mutation of the core residues of the NR2F binding-

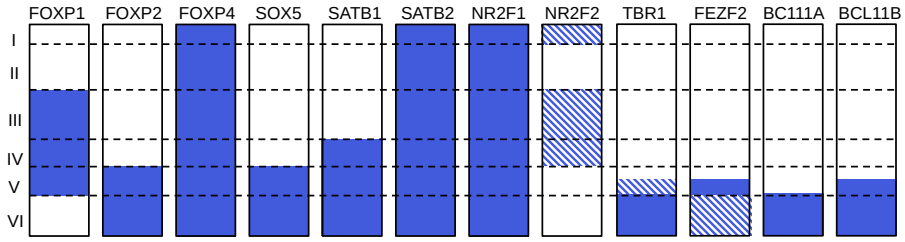


Figure 6.2: Cortical expression patterns in the developing brain of transcription factors studied or mentioned in this thesis. Schematic vertical representation of the different layers of the cortex. Dashed lines separate the different cortical layers and shaded are the layers in which each transcription factor is expressed. Dashed shading indicates low levels of expression (Alcamo *et al.* , 2008; Cánovas *et al.* , 2015; Denaxa *et al.* , 2012; Ferland *et al.* , 2003; Lai *et al.* , 2008; McKenna *et al.* , 2015; Szemes *et al.* , 2006; Takahashi *et al.* , 2008; Tripodi *et al.* , 2004; Woodworth *et al.* , 2016)

motif in FOXP2 would need to be done in order to confirm that this region does indeed mediate these interactions.

6.4 A TRANSCRIPTION FACTOR NETWORK UNDERLYING CORTICAL DEVELOPMENT

The cerebral cortex processes higher order cognitive functions, such as learning, planning and social behaviours, and also motor skills, which are critical neuronal mechanisms underpinning speech and language (Donovan & Basson, 2016; Krishnan *et al.* , 2016). Cortical development is tightly controlled by transcription factors that cooperate with each other to regulate gene expression forming complex molecular networks. Intriguingly, several of the FOXP-interacting transcription factors characterized in chapter 4, namely SOX5, SATB1, SATB2, NR2F1 and NR2F2, as well as the FOXPs, are expressed in the developing cortex. TBR1, which is able to interact with FOXP1 and FOXP2, is also expressed in this part of the brain (Deriziotis *et al.* , 2014a). These transcription factors display distinct yet overlapping patterns of expression throughout the different layers of the cortex (Figure 6.2), indicating that interactions could be occurring *in vivo* in specific layers of the cortex.

Most of these transcription factors are known to play important roles in the developing cortex. Among other functions, FOXP2 regulates neurite outgrowth and synaptic plasticity in cortical neurons, as well as in other brain areas (Enard *et al.* , 2009; Groszer *et al.* , 2008; Reimers-Kipping *et al.* , 2011; Vernes *et al.* , 2011). A role for FOXP2 and also FOXP1 in controlling migration of cortical and striatal neurons has also been reported in a few studies (Li *et al.* , 2015b).It is well known that SOX5, TBR1 and SATB2 are essential transcription factors for the development of a normal cortex. SOX5 controls the timing of the generation of the distinct corticofugal neuron subtypes and TBR1 is crucial for the differentiation and identity determination of deep-layer projection neurons(Lai *et al.* , 2008; Cánovas *et al.* , 2015). SATB2 is an important regulator of the differentiation and axonal path finding of cortical projection neurons (Britanova *et al.* , 2008).In contrast, its close homologue SATB1 is mostly known for its role in the immune system, although there is a study indicating that in the brain SATB1 helps regulate cortical interneuron development (Denaxa *et al.* , 2012). The nuclear receptors NR2F1 and NR2F2 are also involved in cortical development. NR2F1 regulates the differentiation of cortical motor neurons, axonal projection and cortical arealization and, together with NR2F2, plays roles in cell migration and regulation of neurogenesis (Tomassy *et al.* , 2010; Tripodi *et al.* , 2004).

The functions of the protein-protein interactions identified here remains unknown but there are different possible mechanisms (Figure 6.3). One plausible scenario is that two transcription factors bind overlapping or adjacent binding sites and cooperate with each other to regulate common target genes (Figure 6.3a). For example, a transcription factor may require an interacting partner to first contact DNA so it can be brought in close proximity to bind a specific low-affinity binding site (Deplancke *et al.* , 2016). Two interacting transcription factors can also be mutually dependent if the binding affinity of the pair together is greater than the sum of the individual affinities(Deplancke *et al.* , 2016). Another possibility is that two interacting transcription factors contact distal DNA sites creating a DNA loop that can alter the chromatin state and conformation (Deplancke *et al.* , 2016) (Figure 6.3b). Alternatively, one transcription factor could prevent the other from binding DNA by interacting with its binding domain or blocking its accessibility (Figure 6.3c), as has been suggested for the interaction of Foxp2 and Nxk2.1 in the lung (Zhou *et al.* , 2008).

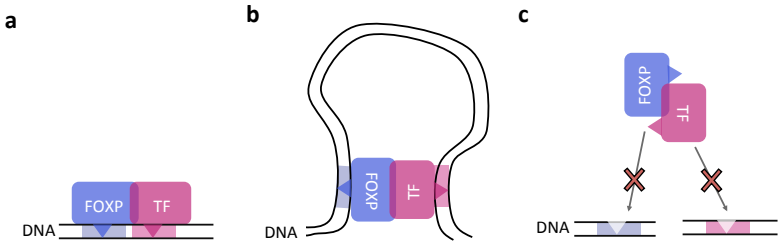


Figure 6.3: Hypothetical different types of interplay between the FOXPs and their interacting-transcription factors. (a) Cooperative binding to overlapping or adjacent binding sites. **(b)** Binding to distant cis-elements forming a DNA loop. **(c)** Interaction that blocks the binding of one ore both transcription factors to their DNA-binding sites. TF, Transcription factor.

Interestingly, many of the cortical FOXP-interacting transcription factors regulate the expression of each other and are connected through a network of both transcriptional regulation and protein-protein interactions, which also involves other language- and neurodevelopment-related transcription factors (Figure 6.4). In layer VI and in the developing subplate, SOX5 and TBR1 directly repress the transcription of FEZF2, which is essential for maintaining the specific neuronal identities of deep cortical neurons (Shim *et al.* , 2012; Han *et al.* , 2011).

Besides interacting with FOXP2, SOX5 and TBR1 have also been hypothesized to regulate FOXP2 expression in the developing cortex (Becker, 2016). In layer V, FEZF2 represses SATB2, while SATB2 promotes the expression of FEZF2 and also SOX5 (McKenna *et al.* , 2015). The repression of the transcription factor BCL11B by NR2F1, SATB2 and SOX5 is important to control the development of cortical projection neurons (Alcamo *et al.* , 2008; Britanova *et al.* , 2008; Lai *et al.* , 2008; Tomassy *et al.* , 2010). BCL11B physically interacts with BCL11A, and also NR2F1 and NR2F2 interact with both BCL11B and BCL11A (Chan *et al.* , 2013; Estruch *et al.* , 2012). Moreover, in the developing cortex, BCL11A represses TBR1 and also regulates SEMA3C, a signaling molecule that

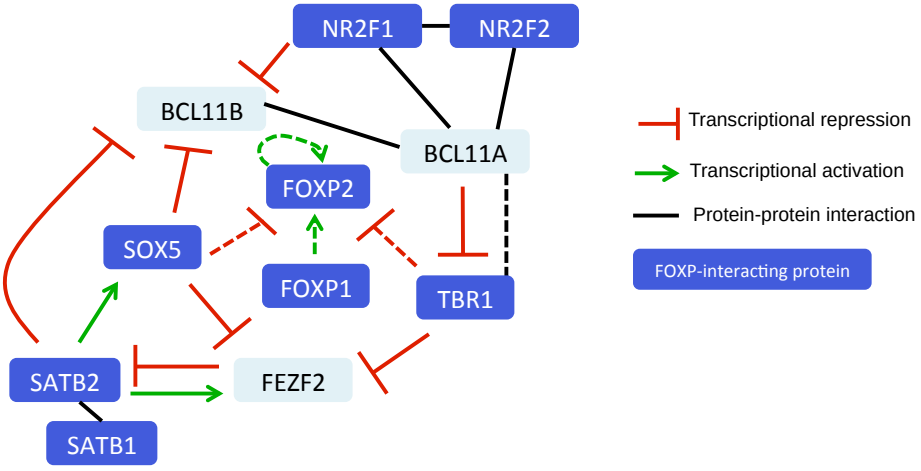


Figure 6.4: Potential molecular network of cortical transcription factors. Schematic representation of the molecular interactions between FOXP- or language-related cortical transcription factors. Red lines indicate transcriptional repression; green, transcriptional activation and black, protein-protein interactions. Dashed lines indicate unpublished findings from our group. FOXP-interacting transcription factors are dark-shaded.

provides guidance cues to migrating neurons (Cánovas *et al.* , 2015). It is important, though, to bear in mind that all of these regulatory and physical interactions do not occur in the same neuronal subtypes and at the same time during development, which adds an extra level of complexity to this intricate transcription factor network.

Given their important roles in cortical development, it is perhaps not surprising that the disruption of the genes encoding some of these transcription factors result in neurodevelopmental disorders. Haploinsufficiency of *SOX5*, *SATB2*, *NR2F1* or *YY1* results in ID syndromes, which partially overlap with the phenotypes caused by mutations of *FOXP1* and *FOXP2* (Bacon & Rappold, 2012; Bosch *et al.* , 2014; Gabriele *et al.* , 2017; Nesbitt *et al.* , 2015; Zarate & Fish, 2016). *SOX5*- and *SATB2*-related syndromes are particularly interesting because subjects with heterozygous mutations in these genes present severely altered language functions (Lamb *et al.* , 2012; Nesbitt *et al.* , 2015; Zarate & Fish, 2016). *BCL11A* haploinsufficiency leads to a neurodevelopmental syndrome that includes motor and language delays (Chapter 5) and, intriguingly,

work in our group recently identified two rare intronic *SEMA3C* variants in multiple affected members of a family with developmental dyslexia Carrion-Castillo (2016).

De novo mutations in TBR1 found in ASD patients have been shown to disrupt its interaction with FOXP2 or the ID-linked protein CASK mutations disrupting FOXP2 also negatively altered the FOXP2-TBR1 interaction (Deriziotis *et al.* , 2014a). In chapter 4, I have shown that FOXP1 and FOXP2 pathogenic variants disrupt many of the interactions with these transcription factors, which points towards a convergent molecular etiology for distinct yet overlapping neurodevelopmental disorders. In the future, it would be interesting to investigate the other side of the coin, looking at how pathogenic mutations in each of the FOXP-interacting transcription factors affect their association with the FOXPs, and how this network of cortical transcription factors may be disrupted.
Overall, this shows that neurodevelopmental disorders linked to these transcription factors are connected at the molecular level. The data presented here support the hypothesis that shared molecular networks underlying cortical development are involved in distinct neurodevelopmental disorders such as ASD, ID and speech and language disorders (Kwan, 2013; Voineagu *et al.* , 2011). Because these are transcription factors, the connections between them are not only physical interactions between proteins but also negative or positive regulatory links. This implies that disrupting one gene can affect both the expression and function of many other genes. Depending on which gene is affected, the network can be destabilized in diverse ways leading to distinct neurodevelopmental phenotypes with variable degrees of severity. The disruption of one gene of this network can also potentially lead to normal cortical development since the network might be readjusted to compensate for the dysfunction of the gene during development.

FOXP-interacting transcription factors are also co-expressed with the FOXPs in further brain regions other than the cortex, and hence, some of these interactions may also have other functions in the brain beyond cortical development. In chapter 4, I reported that FOXP2 and NR2F2 physically interact and showed that they are specifically co-expressed in the Purkinje cells of the cerebellum. Other FOXP2 interacting partners, namely FOXP4, CTBP1, and CTBP2, are also expressed in Purkinje cells. The cerebellum is classically known for its role in motor functions but it is also involved in non-motor cognitive skills such as speech and language (Mariën *et al.* , 2014). Among other structural alterations

of the brain, individuals with CAS carrying *FOXP2* mutations exhibit reduced gray matter density in parts of the cerebellum, which could be relevant for their speech and language impairment (Vargha-Khadem *et al.* , 2005). Disruptions of Foxp2 in mice result in a reduction of cerebellar size (Groszer *et al.* , 2008). Similarly, NR2F2 conditional knockout in the cerebellum results in decreased cerebellar growth and, reduced differentiation of cerebellar neurons (Kim *et al.* , 2009a)

Because disruptions of NR2F2 and FOXP2 lead to similar cerebellar phenotypes it is reasonable to hypothesize that the physical interaction between FOXP2 and NR2F2 may be involved in cerebellar growth. To date, the molecular behaviour of FOXP2 and NR2F2 in cerebellar development has been scarcely investigated. One study suggests that the decreased cerebellar growth upon knocking out NR2F2 may be in part mediated by a reduced expression of IGF-1 (Kim *et al.* , 2009a). Also, another study reported that CNTNAP2 mRNA levels are increased in the cerebellum of mice carrying a Foxp2 p.R552H mutation (corresponding to p.R553H in its human orthologue), but did not demonstrate that these increased CNTNAP2 mRNA levels were causing a reduced cerebellar size (Fujita-Jimbo & Momoi, 2014). In any case, additional work encompassing all possible downstream target genes will be needed to unravel how FOXP2 and NR2F2 take part in the development of the cerebellum and what behavioural consequences that may have.

This work highlights the importance of investigating the molecular links between genes identified in complex neurodevelopmental disorders. The complex heterogeneity of language-related disorders and the substantial comorbidity between neurodevelopmental disorders suggests that complex shared molecular networks underlie their etiologies. Therefore, in order to comprehend the molecular neurobiological foundations of language-related disorders, we not only have to keep on searching for associated genes in patient cohorts but also we need to decipher whether these genes are functionally linked to each other. In this dissertation I have only considered rare language-related gene variants, but common variation should also be taken into account when investigating language-related networks, as it is likely that common and rare gene variants converge on the same molecular pathways.

The transcription factors studied in this dissertation might represent just the tip of the iceberg of a broader molecular network. A logical next step would be to further characterize and expand this network, first, by assessing whether

there are other protein-protein interactions involving these transcription factors, and, second, by identifying more interaction partners. This could be done in a relatively efficient way with methods like the BRET assay and AP/MS. What would be more challenging but particularly valuable is to investigate in which specific neuron subtypes and developmental time points these interactions occur.

Additionally, identifying the downstream target genes of these transcription factors would also be extremely informative. Research on FOXP2 neural target genes revealed that these participate in key pathways involved in brain development (Spiteri *et al.* , 2007; Vernes *et al.* , 2007, 2011). Likewise, ChIP-seq experiments for TBR1 and FOXP1 highlighted that their downstream pathways are enriched for genes involved in ASD (Araujo *et al.* , 2015; Notwell *et al.* , 2016). Finally, it would also be interesting to search for shared target genes of different transcription factors and see whether they are also involved in brain development and/or mutated in neurodevelopmental disorders.

Systems biology methodologies combined with the systematic characterization of neural gene function make it possible to fully and deeply characterize language-related molecular networks. This type of approach combines experimental data with statistical methods to model large sets of gene networks.

6.5 THE EXPANDING FOXP INTERACTOME

A major part of this thesis has involved the identification and characterization of FOXP-interacting proteins. In Chapter 2 I characterized the interaction between FOXP2 and PIAS proteins and discovered that this mediated the sumoylation of FOXP2, and in Chapter 4 I identified and characterized seven FOXP-interacting transcription factors, thereby expanding the FOXP interactome (Figure 6.5).

In chapter 4, I searched for FOXP-interacting proteins using affinity purification coupled by mass spectrometry (AP/MS). The fact that my screens identified several known FOXP interactors (i.e. FOXP1, FOXP2, FOXP4, GATAD2B, CTBP1, CTBP2) indicated that the technical approach was effective (Chapter 4, Supplementary Table S4.2). The AP/MS screens identified more than 100 putative FOXP-interacting proteins (Chapter 4, Supplementary Table S4.2), of which I chose to focus on 12 transcription factors for validation. Therefore, within the identified proteins there could be more interesting candidate FOXP-interactors

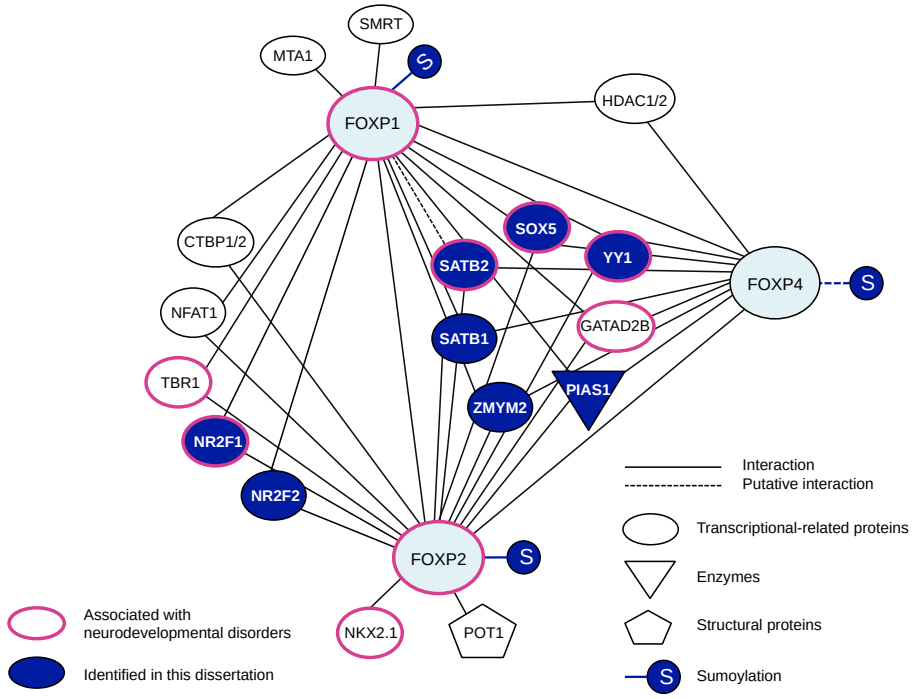


Figure 6.5: The FOXP interactome. Network diagram of FOXP protein-protein interactions and post-translational modifications at least confirmed with two different experimental approaches. Lines indicate protein-protein interactions. Protein-protein interactions identified in this dissertation are shaded in dark blue purple font. A pink border indicates that the genes encoding these proteins have been associated with neurodevelopmental disorders. Shaded in light blue are the members of the FOXP family. A dashed line indicates an unclear interaction. Different shapes represent different types of proteins: ovals are transcription factors; triangles, enzymes; and pentagons, structural proteins. A small circle represents sumoylation.

to confirm and further characterize in the future. Examples include the transcription factor ZNF148, recurrent *de novo* mutations of which have been identified in individuals with developmental delay and ID (Stevens *et al.* , 2016), and the chromatin remodeler SETD2, haploinsufficiency of which leads to a neurodevelopmental syndrome characterized by ID and macrocephaly accompanied by speech delay (Lumish *et al.* , 2015; Luscan *et al.* , 2014).

One of the major limitations of AP/MS lies in the difficulty to detect weak or transient protein-protein interactions (such as interactions with enzymes), which, presumably, might have not been captured in our screens. Protein-protein interaction screening methods that allow the detection of weak and transient interactions, such as Y2H, represent a reasonable complementary approach to AP/MS screens. In chapter 2, I characterized the interaction of FOXP2 with PIAS1, which was first identified in Y2H screens previously performed in our group. These screens also identified additional putative FOXP-interacting proteins (Chapter 2, Supplementary Table S2.3), but these need to be further confirmed. Additionally, it could be worthwhile to perform new Y2H screens using different batches of brain cDNA prey libraries. Because the composition and relative abundances of cDNAs in each batch is variable (Sainz, 1993), the cDNAs encoding some FOXP-interacting proteins may have not been present or could have been too little expressed in the cDNA batch used to perform the Y2H screens.

Similarly, in the AP/MS screens, FOXP-interacting proteins (such as TBR1) might have not been identified because they are not expressed in the cell line employed (i.e. HEK293 cells). Also, this cell line may lack specific biological cues necessary for the FOXPs to bind some protein partners. Therefore, a logical next step would be to carry out the experiments using different cell lines, such as neural cell lines or tissue that endogenously expresses FOXP proteins. However, this approach hits upon a challenge because FOXP2-expressing cell types are post-mitotic neurons that no longer divide and are therefore challenging to grow and manipulate. Therefore, obtaining a sufficient amount of FOXP2-expressing neurons to perform AP/MS screens is not clearly feasible.

Alternatively, recently emerged genome-editing techniques such as CRISPR/Cas9 could be used to genetically engineer a mouse line that expresses endogenous FOXPs fused to an epitope tag, therefore enabling a more efficient and specific affinity purification of FOXP complexes. Future advances in the proteomics field, such as continued improvements in the sensitivity of mass spectrometry

methods and increasingly sophisticated data analysis to discriminate between spurious results and biologically relevant interactions will lead to more effective ways of detecting protein-protein interactions, which will help expand the neural FOXP interactome.

6.6 THE VALUE OF FUNCTIONAL CHARACTERIZATION STUDIES IN LANGUAGE-RELATED DISORDERS

Advanced DNA sequencing techniques are accelerating the discovery of genes involved in complex neurodevelopmental conditions such as speech and language impairment. Therefore, it is becoming more and more important to discern which genetic variants are causal and which ones are incidental to the phenotype of interest. Predictive algorithms are commonly used to rank sequence variants according to their predicted degree of pathogenicity. *In silico* modeling of protein-coding mutations in the three-dimensional structure of the protein is also a valuable tool that can help predict the impact of a mutation on the structure of the protein. However, data from computational methods need to be interpreted with caution, as they are predictions and can seldom provide sufficient proof of causality. Functional characterization in living model systems is therefore crucial to clarify a pathogenic role for a suspected causal variant. Moreover, this kind of functional work also sheds light on the molecular functions of the protein being disrupted and therefore on the biology underlying the disorder.

The relevance of functional characterization of variants in speech and language impairment has been principally proven by studies of FOXP2 variants in cases of CAS. p.R553H, which was found to co-segregate with CAS in a large multigenerational pedigree, was the first FOXP2 variant to be functionally characterized (Vernes *et al.* , 2006). This substitution is located in the forkhead-box domain of the encoded protein and functional assays in cellular models revealed that it severely disrupts FOXP2's binding to DNA and transcriptional activity as well as its subcellular localization, confirming a causal role in the disorder. Variants p.R328* and p.Q390Vfs* , which result in truncated FOXP2 proteins that lack the forkhead-box domain, also impair the ability of FOXP2 to bind to DNA (Estruch *et al.* , 2016a).

In Chapter 4, I found that disorder-related FOXP1 and FOXP2 variants showed altered interactions with the newly discovered FOXP-interacting transcription factors. The protein interaction assays revealed that FOXP2 variants p.Q17L,

p.M406T, p.P416T and p.N597H exhibited similar behaviour to the wild-type protein, further supporting the view that they are incidental to the phenotype, as observed in a previous study (Estruch *et al.* , 2016a). In contrast, the etiological FOXP2 variants p.R553H, p.R328* and p.Q390Vfs* 7, as well as all of the FOXP1 variants that I assessed, had an abolished or reduced ability to interact with most of the FOXP-interacting transcription factors. These findings further support the pathogenic roles of these variants, and add novel insights into the molecular mechanisms by which they can lead to neurodevelopmental disorders (Estruch *et al.* , 2016a; Sollis *et al.* , 2016).

Intriguingly, the interaction between FOXP2 p.R553H variant and SATB2 was slightly less reduced than for the other transcription factors, and co-expression of both proteins led to the mislocalization of SATB2 in the cytoplasm. In a similar way, FOXP1 variants p.R465G and p.R514C retained the ability to interact with SOX5, which was abnormally redistributed into nuclear aggregates upon co-expression with these variants. The same effects were observed for the interaction between FOXP2 p.R465G variant and ZMYM2. These findings suggest that the relevant variants might additionally interfere with the normal function of a subset of interacting transcription factors, which may also contribute to pathology. This is in line with previous studies that propose a partial dominant-negative effect for these variants by interfering with the function of wild-type FOXPs (Estruch *et al.* , 2016a; Sollis *et al.* , 2016).

In Chapter 5, I functionally characterized three *de novo* missense BCL11A variants identified in patients with neurodevelopmental disorders that included language impairment. The three substitutions lie within a non-canonical zinc finger domain with previously unknown function and affected many aspects of protein function such as subcellular localization, homo-dimerization and transcriptional activity, suggesting that the loss of function of BCL11A caused the disorder. Moreover, this implies that the non-canonical zinc finger domain and, more concretely, the specific residues mutated are involved in the aforementioned BCL11A protein functions. This work highlights the power of functional characterization not only in ascertaining the contribution of particular variants to disorder, but also in the discovery of novel molecular aspects of the function of the protein. Identifying new functional protein domains can help interpret further variants in that region identified in patients in the future.

Results in chapter 5 were published as a part of a study that included the characterization of a mouse model of BCL11A haploinsufficiency (Dias *et al.*

, 2016). Mice carrying only one copy of the gene showed abnormal social behaviour and microcephaly, in line with the phenotype observed in human patients. Furthermore, comparative transcriptomic analyses between heterozygous and wild-type neonatal brains revealed transcriptional dysregulation in the cortex and the hippocampus of the haploinsufficient mice, consistent with the results from my luciferase assays. Interestingly, in the differentially expressed genes there is a considerable enrichment for genes belonging to pathways involved in neurodevelopment, such as semaphorin-plexin and slit-robo signaling pathways. Notably, members of the Semaphorin and the Robo families of genes belonging to these pathways have been associated with dyslexia and SLI, respectively (Carrion-Castillo, 2016; Chen *et al.* , 2017; Hannula-Jouppi *et al.* , 2005; St Pourcain *et al.* , 2014a). This work demonstrates how coupling functional experiments in cellular and animal models allows for a thorough delineation of the molecular etiology underlying neurodevelopmental disorders that include speech and language impairment.

The rapid pace of discovery of language-related genes, pathways and networks will require the development of efficient, systematic methods to interrogate gene function. Functional interpretation of genetic variants is nowadays one of the major bottlenecks in the field, as experiments using cellular and animal models are laborious, expensive and time-consuming and it is not yet feasible to perform them in a systematic way to filter or rank variants from sequencing studies. Existing methods that perturb the expression or inhibit the function of a gene product involve systematic mutagenesis using RNA interference or overexpression approaches using cDNA, ORF and miRNA expression libraries. Although these are useful methods, they do not allow the modeling of single nucleotide changes.

The recently emerged genome editing technique CRISPR/Cas9 is beginning to be used for the systematic functional analysis of whole genomes (Rauscher *et al.* , 2017). Similarly, massively parallel single-amino acid mutagenesis techniques, which enable missense mutational scans, can also be extremely useful to interpret single nucleotide variants identified by clinical sequencing (Kitzman *et al.* , 2015; Shendure & Fields, 2016). Approaches like this allow for the characterization of both coding and non-coding DNA variation, but they have been more extensively used to assess regulatory DNA regions, such as enhancers (Patwardhan *et al.* , 2012). This is perhaps not surprising, since the impact of mutations in regulatory regions can be clearly assessed by measuring gene ex-

pression. For instance, Cas9-mediated in situ saturating mutagenesis has been used to characterize an erythroid enhancer of human BCL11A, which is subject to common genetic variation associated with the fetal hemoglobin levels (Canver *et al.* , 2015). Conversely, the effects of protein-coding substitutions can be multiple and unpredictable, therefore it is more challenging to find simple scalable tests that can be coupled to high-throughput methodologies. However, this has been successfully done to functionally analyze the impact of nearly 2000 missense substitutions of uncertain significance in the tumor suppressor gene *BRCA1* on its core functions (Starita *et al.* , 2015).

High-throughput methods like the above-mentioned represent highly promising approaches to systematically characterize the impact of missense variants in language-related disorders.

6.7 THE BIG PICTURE OF FUNCTIONAL GENOMICS IN LANGUAGE-RELATED DISORDERS.

Developmental disorders of speech and language affect approximately 10% of children at school entry and have major impacts on educational success, social relationships and job opportunities (Norbury *et al.* , 2016). Speech therapy can significantly ameliorate certain speech and language deficits in children with these disorders (Ullrich *et al.* , 2009). According to the American speech-language-hearing association, it is important to start these therapies as early as possible to guarantee an optimal outcome, since the earlier the intervention, the more effective it is in restoring speech and language abilities for a given subject.

However, probands with language-related disorders may present with complex phenotypes in which often some cognitive deficits are overlooked. Functional genomics studies in language-related disorders can offer the possibility of a genotype first-approach in the diagnosis of neurodevelopmental disorders (Stessman *et al.* , 2014). A comprehensive characterization of the genetic causes of monogenic neurodevelopmental syndromes will allow for earlier and clearer diagnoses by genotyping the proband prior to fully characterizing the phenotype in sporadic cases.

The molecular characterization of monogenic language-related disorders can also provide valuable biomarkers to facilitate reliable low-cost diagnosis of a specific syndrome. This is the case for the BCL11A-related syndrome described in chapter 5. Because of the non-neural role of BCL11A in repressing fetal

hemoglobin, haploinsufficiency leads to abnormally elevated fetal hemoglobin levels. This hematopoietic phenotype does not represent a health risk for the patient and can be successfully used as a physiological biomarker, enabling detection of this syndrome with a simple blood test coupled to genetic analysis. It is therefore important to also understand the non-neuronal functions of language-related genes and use this knowledge to comprehend the complete phenotype.

In this dissertation, I take monogenic language-related disorders as entry points to the molecular basis of speech and language abilities. Ideally, the best cases to do so would be the ones that solely present language impairment. However, patients with language impairment in isolation are not normally referred for genetic analysis because the phenotype is not usually considered in a clinical context. Genetic data on this type of subjects usually come from research studies and are not as abundant as those for other neurodevelopmental conditions (i.e. ASD, schizophrenia). Hence, it has been challenging to recruit sufficiently large cohorts to guarantee the necessary power to carry out genome-wide association studies.

So far, the most convincing monogenic cases of clearly disproportionate effects on the speech and language domain are those arising from FOXP2 mutations. Undoubtedly, the discovery of FOXP2 disruption in the KE family was a groundbreaking finding and this has provided an extremely valuable molecular window into the neurobiological basis of speech and language. However, speech and language skills are certainly influenced by numerous other genetic factors, most of which remain to be discovered.

Exploiting FOXP2 molecular networks has successfully led to the discovery of novel language-related genes that are implicated in speech and language-related disorders, such as FOXP2-interacting transcription factors FOXP1, TBR1, SOX5 and SATB2 or target genes like *CNTNAP2*. The phenotypes caused by mutations in these genes are broader than the one typically caused by FOXP2 haploinsufficiency, including additional cognitive impairments such as ID or ASD. However, these genes also offer molecular gateways into speech and language disorders. Language is deeply intermingled with other cognitive abilities at genetic, molecular, anatomical, and phenotypic levels. Therefore it is necessary to contemplate other behavioural traits when investigating aspects of speech and language impairment.

Intriguingly, BCL11A is indirectly connected with FOXP1 and FOXP2, since they share NR2F1 and NR2F2 as interaction partners (Figure 6.4). But, most likely, not all the genetic factors influencing speech and language will necessary revolve around FOXP2 and it will be necessary to study different genetic pathways in order to fully comprehend the genetic foundations of speech and language. For instance, probands with GRIN2A mutations present a wide range of epilepsy syndromes together with severe speech impairment that can include dysarthria, speech dyspraxia, and both receptive and expressive language delay (Turner *et al.* , 2015). GRIN2A encodes a subunit of NMDA receptors, ionotropic glutamate receptors that regulate mediate calcium transport through the membranes of excitatory synapses. Another NMDA receptor subunit is encoded by GRIN2B, mutations of which lead to a wide range of neurodevelopmental disorders including motor and language disorders, ASD, attention deficit hyperactivity disorder (ADHD), developmental delay, epilepsy and schizophrenia (Hu *et al.* , 2016). The study of the molecular dynamics of GRIN2A and GRIN2B gene function in the excitatory synapses, or other language-related molecular pathways yet to be discovered, can also tell us something about speech and language.

However, as discussed earlier, the monogenic cases that may be of most direct relevance for understanding the neurobiology of speech and language are those in which language impairments occur against a background of normal non-verbal cognitive abilities. Chen and colleagues recently reported whole-exome sequencing in a cohort of 43 probands with specific language impairment (SLI) and found several potentially pathogenic genetic variants. Some of these were in genes that were already related to language, such as *GRIN2A*, *ROBO1* or *ERC1*, but a few others may point to new candidate language-related genes, such as *OXR1*, *SCN9A* and *KMT2D* (Chen *et al.* , 2017).

In the future, intensive analyses of genotype-phenotype correlations of a range of neurodevelopmental disorders will be crucial to decipher how specific genetic factors contribute to speech and language. Advances in computational methodologies will eventually allow generation of comprehensive networks to model the complicated molecular links underlying speech and language. Incorporating data from experimental studies into large *in silico* molecular networks will be of vital importance to enhance the veracity of such models and support their biological relevance.

Computational systems biology methods allow the effective integration of multiple large data sets that come from different sources. These will be essential to keep up with the rapid pace at which genetic data is being generated using next-generation sequencing techniques and the potentially large amount of data that will arise from systematic functional characterizations. Moreover, to fully understand the biology of speech and language, findings from other disciplines such as neuroscience, evolutionary anthropology and linguistics will also need to be taken into account and integrated with molecular genetic data.

Adegbola, Abidemi A, Cox, Gerald F, Bradshaw, Elizabeth M, Hafler, David A, Gimelbrant, Alexander, & Chess, Andrew. 2015. Monoallelic expression of the human FOXP2 speech gene. *Proceedings of the National Academy of Sciences of the United States of America*, **112**(22), 6848–6854.

Al Turki, Saeed, Manickaraj, Ashok K, Mercer, Catherine L, Gerety, Sebastian S, Hitz, Marc-Phillip, Lindsay, Sarah, D'Alessandro, Lisa C A, Swaminathan, G Jawahar, Bentham, Jamie, Arndt, Anne-Karin, Louw, Jacoba, Low, Jacoba, Breckpot, Jeroen, Gewillig, Marc, Thienpont, Bernard, Abdul-Khaliq, Hashim, Harnack, Christine, Hoff, Kirstin, Kramer, Hans-Heiner, Schubert, Stephan, Siebert, Reiner, Toka, Okan, Cosgrove, Catherine, Watkins, Hugh, Lucassen, Anneke M, O'Kelly, Ita M, Salmon, Anthony P, Bu'lock, Frances A, Granados-Riveron, Javier, Setchfield, Kerry, Thornborough, Chris, Brook, J David, Mulder, Barbara, Klaassen, Sabine, Bhattacharya, Shoumo, Devriendt, Koen, Fitzpatrick, David F, UK10K Consortium, Wilson, David I, Mital, Seema, & Hurles, Matthew E. 2014. Rare variants in NR2F2 cause congenital heart defects in humans. *American journal of human genetics*, **94**(4), 574–585.

Alcamo, Elizabeth A, Chirivella, Laura, Dautzenberg, Marcel, Dobreva, Gergana, Fariñas, Isabel, Grosschedl, Rudolf, & McConnell, Susan K. 2008. Satb2 regulates callosal projection neuron identity in the developing cerebral cortex. *Neuron*, **57**(3), 364–377.

Alcock, K J, Passingham, R E, Watkins, K, & Vargha-Khadem, F. 2000. Pitch and timing abilities in inherited speech and language impairment. *Brain and Language*, **75**(1), 34–46.

Alvarez, J D, Yasui, D H, Niida, H, Joh, T, Loh, D Y, & Kohwi-Shigematsu, T. 2000. The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. *Genes & development*, **14**(5), 521–535.

Amelio, Antonio L, Miraglia, Loren J, Conkright, Juliana J, Mercer, Becky A, Batalov, Serge, Cavett, Valerie, Orth, Anthony P, Busby, Jennifer, Hogenesch, John B, & Conkright, Michael D. 2007. A coactivator trap identifies NONO (p54nrb) as a component of the cAMP-signaling pathway. *Proceedings of the National Academy of Sciences of the United States of America*, **104**(51), 20314–20319.

Anderson, Garret R, Galfin, Timothy, Xu, Wei, Aoto, Jason, Malenka, Robert C, & Südhof, Thomas C. 2012. Candidate autism gene screen identifies critical role for cell-

adhesion molecule CASPR2 in dendritic arborization and spine development. *Proceedings of the National Academy of Sciences of the United States of America*, **109**(44), 18120–18125.

Anthoni, Heidi, Zucchelli, Marco, Matsson, Hans, Müller-Myhsok, Bertram, Fransson, Ingegerd, Schumacher, Johannes, Massinen, Satu, Onkamo, Päivi, Warnke, Andreas, Griesemann, Heide, Hoffmann, Per, Nopola-Hemmi, Jaana, Lyytinen, Heikki, Schulte-Körne, Gerd, Kere, Juha, Nöthen, Markus M, & Peyrard-Janvid, Myriam. 2007. A locus on 2p12 containing the co-regulated MRPL19 and C2ORF3 genes is associated to dyslexia. *Human molecular genetics*, **16**(6), 667–677.

Araujo, Daniel J, Anderson, Ashley G, Berto, Stefano, Runnels, Wesley, Harper, Matthew, Ammanuel, Simon, Rieger, Michael A, Huang, Hung-Chung, Rajkovich, Kacey, Loerwald, Kristofer W, Dekker, Joseph D, Tucker, Haley O, Dougherty, Joseph D, Gibson, Jay R, & Konopka, Genevieve. 2015. FoxP1 orchestration of ASD-relevant signaling pathways in the striatum. *Genes & development*, **29**(20), 2081–2096.

Arber, Silvia. 2008. FoxP1: conducting the Hox symphony in spinal motor neurons. *Nature neuroscience*, **11**(10), 1122–1124.

Armentano, Maria, Filosa, Alessandro, Andolfi, Gennaro, & Studer, Michèle. 2006. COUP-TFI is required for the formation of commissural projections in the forebrain by regulating axonal growth. *Development (Cambridge, England)*, **133**(21), 4151–4162.

Armentano, Maria, Chou, Shen-Ju, Tomassy, Giulio Srubek, Leingärtner, Axel, O'Leary, Dennis D M, & Studer, Michèle. 2007. COUP-TFI regulates the balance of cortical patterning between frontal/motor and sensory areas. *Nature neuroscience*, **10**(10), 1277–1286.

Avram, D, Fields, A, Pretty On Top, K, Nevrivy, D J, Ishmael, J E, & Leid, M. 2000. Isolation of a novel family of C(2)H(2) zinc finger proteins implicated in transcriptional repression mediated by chicken ovalbumin upstream promoter transcription factor (COUP-TF) orphan nuclear receptors. *The Journal of biological chemistry*, **275**(14), 10315–10322.

Avram, Dorina, Fields, Andrew, Senawong, Thanaset, Topark-Ngarm, Acharawan, & Leid, Mark. 2002. COUP-TF (chicken ovalbumin upstream promoter transcription factor)-interacting protein 1 (CTIP1) is a sequence-specific DNA binding protein. *The Biochemical journal*, **368**(Pt 2), 555–563.

Bacon, C, Schneider, M, Le Magueresse, C, Froehlich, H, Sticht, C, Gluch, C, Monyer, H, & Rappold, G A. 2015. Brain-specific Foxp1 deletion impairs neuronal development and causes autistic-like behaviour. *Molecular psychiatry*, **20**(5), 632–639.

Bacon, Claire, & Rappold, Gudrun A. 2012. The distinct and overlapping phenotypic spectra of FOXP1 and FOXP2 in cognitive disorders. *Human genetics*, **131**(11), 1687–1698.

Balamotis, Michael A, Tamberg, Nele, Woo, Young Jae, Li, Jingchuan, Davy, Brian, Kohwi-Shigematsu, Terumi, & Kohwi, Yoshinori. 2012. Satb1 ablation alters temporal expression of immediate early genes and reduces dendritic spine density during postnatal brain development. *Molecular and cellular biology*, **32**(2), 333–347.

Balci, Tugce B, Sawyer, Sarah L, Davila, Jorge, Humphreys, Peter, & Dyment, David A. 2015. Brain malformations in a patient with deletion 2p16.1: A refinement of the phenotype to BCL11A. *European journal of medical genetics*, **58**(6-7), 351–354.

Barry, J G, Yasin, I, & Bishop, D V M. 2007. Heritable risk factors associated with language impairments. *Genes, brain, and behavior*, **6**(1), 66–76.

Basak, Anindita, Hancarova, Miroslava, Ulirsch, Jacob C, Balci, Tugce B, Trkova, Marie, Pelisek, Michal, Vlckova, Marketa, Muzikova, Katerina, Cermak, Jaroslav, Trka, Jan, Dyment, David A, Orkin, Stuart H, Daly, Mark J, Sedlacek, Zdenek, & Sankaran, Vijay G. 2015. BCL11A deletions result in fetal hemoglobin persistence and neurodevelopmental alterations. *The Journal of clinical investigation*, **125**(6), 2363–2368.

Beauclair, Guillaume, Bridier-Nahmias, Antoine, Zagury, Jean-François, Saïb, Ali, & Zamborlini, Alessia. 2015. JASSA: a comprehensive tool for prediction of SUMOylation sites and SIMs. *Bioinformatics (Oxford, England)*, **31**(21), 3483–3491.

Becker, Martin. 2016 (Nov.). *On the identification of FOXP2 gene enhancers and their role in brain development.* Ph.D. thesis, Radboud University of Nijmegen.

Beketaev, Ilimbek, Zhang, Yi, Kim, Eun Young, Yu, Wei, Qian, Ling, & Wang, Jun. 2015. Critical role of YY1 in cardiac morphogenesis. *Developmental Dynamics*, **244**(5), 669–680.

Belaguli, Narasimhaswamy S, Zhang, Mao, Brunicardi, F Charles, & Berger, David H. 2012. Forkhead box protein A2 (FOXA2) protein stability and activity are regulated by sumoylation. *PloS one*, **7**(10), e48019.

Belton, Emma, Salmond, Claire H, Watkins, Kate E, Vargha-Khadem, Faraneh, & Gadian, David G. 2003. Bilateral brain abnormalities associated with dominantly inherited verbal and orofacial dyspraxia. *Human brain mapping*, **18**(3), 194–200.

Benayoun, Bérénice A, Caburet, Sandrine, & Veitia, Reiner A. 2011. Forkhead transcription factors: key players in health and disease. *Trends in genetics : TIG*, **27**(6), 224–232.

Bengani, Hemant, Handley, Mark, Alvi, Mohsan, Ibitoye, Rita, Lees, Melissa, Lynch, Sally Ann, Lam, Wayne, Fannemel, Madeleine, Nordgren, Ann, Malmgren, H, Kvarnung, M, Mehta, Sarju, McKee, Shane, Whiteford, Margo, Stewart, Fiona, Connell, Fiona, Clayton-Smith, Jill, Mansour, Sahar, Mohammed, Shehla, Fryer, Alan, Morton, Jenny, UK10K Consortium, Grozeva, Detelina, Asam, Tara, Moore, David, Sifrim, Alejandro, McRae, Jeremy, Hurles, Matthew E, Firth, Helen V, Raymond, F Lucy, Kini, Usha, Nellåker, Christoffer, DDD Study, & FitzPatrick, David R. 2017. Clinical and molecular consequences of disease-associated de novo mutations in SATB2. *Genetics in medicine : official journal of the American College of Medical Genetics*, Feb.

Bishop, D V, North, T, & Donlan, C. 1995. Genetic basis of specific language impairment: evidence from a twin study. *Developmental medicine and child neurology*, **37**(1), 56–71.

Bishop, D V M. 2010. Overlaps between autism and language impairment: phenomimicry or shared etiology? *Behavior genetics*, **40**(5), 618–629.

Bishop, D V M, & Hayiou-Thomas, M E. 2008. Heritability of specific language impairment depends on diagnostic criteria. *Genes, brain, and behavior*, **7**(3), 365–372.

Bjurström, Carmen F, Mojadidi, Michelle, Phillips, John, Kuo, Caroline, Lai, Stephen, Lill, Georgia R, Cooper, Aaron, Kaufman, Michael, Urbinati, Fabrizia, Wang, Xiaoyan, Hollis, Roger P, & Kohn, Donald B. 2016. Reactivating Fetal Hemoglobin Expression in Human Adult Erythroblasts Through BCL11A Knockdown Using Targeted Endonucleases. *Molecular therapy. Nucleic acids*, **5**, e351.

Bonkowsky, Joshua L, Wang, Xu, Fujimoto, Esther, Lee, Ji Eun, Chien, Chi-Bin, & Dorsky, Richard I. 2008. Domain-specific regulation of foxP2 CNS expression by lef1. *BMC developmental biology*, **8**(1), 103.

Bosch, Daniëlle G M, Boonstra, F Nienke, Gonzaga-Jauregui, Claudia, Xu, Mafei, de Ligt, Joep, Jhangiani, Shalini, Wiszniewski, Wojciech, Muzny, Donna M, Yntema, Helger G, Pfundt, Rolph, Vissers, Lisenka E L M, Spruijt, Liesbeth, Blokland, Ellen A W, Chen, Chun-An, Baylor-Hopkins Center for Mendelian Genomics, Lewis, Richard A, Tsai, Sophia Y, Gibbs, Richard A, Tsai, Ming-Jer, Lupski, James R, Zoghbi, Huda Y, Cremers, Frans P M, de Vries, Bert B A, & Schaaf, Christian P. 2014. NR2F1 mutations cause optic atrophy with intellectual disability. *American journal of human genetics*, **94**(2), 303–309.

Bowers, J Michael, & Konopka, Genevieve. 2012a. ASD-relevant Animal Models of the Foxp Family of Transcription Factors. *Autism-open access*, **Suppl 1**(10).

Bowers, J Michael, & Konopka, Genevieve. 2012b. The role of the FOXP family of transcription factors in ASD. *Disease markers*, **33**(5), 251–260.

Britanova, Olga, de Juan Romero, Camino, Cheung, Amanda, Kwan, Kenneth Y, Schwark, Manuela, Gyorgy, Andrea, Vogel, Tanja, Akopov, Sergey, Mitkovski, Miso, Agoston, Denes, Sestan, Nenad, Molnár, Zoltán, & Tarabykin, Victor. 2008. Satb2 is a postmitotic determinant for upper-layer neuron specification in the neocortex. *Neuron*, **57**(3), 378–392.

Bruce, Heather A, & Margolis, Russell L. 2002. FOXP2: novel exons, splice variants, and CAG repeat length stability. *Human genetics*, **111**(2), 136–144.

Budayeva, Hanna G, & Cristea, Ileana M. 2014. A mass spectrometry view of stable and transient protein interactions. *Advances in experimental medicine and biology*, **806**(Chapter 11), 263–282.

Butland, Stefanie L, Devon, Rebecca S, Huang, Yong, Mead, Carri-Lyn, Meynert, Alison M, Neal, Scott J, Lee, Soo Sen, Wilkinson, Anna, Yang, George S, Yuen, Macaire M S, Hayden, Michael R, Holt, Robert A, Leavitt, Blair R, & Ouellette, B F Francis. 2007. CAG-encoded polyglutamine length polymorphism in the human genome. *BMC genomics*, **8**(1), 126.

Calvete, Oriol, Martinez, Paula, Garcia-Pavia, Pablo, Benitez-Buelga, Carlos, Paumard-Hernández, Beatriz, Fernandez, Victoria, Dominguez, Fernando, Salas, Clara, Romero-Laorden, Nuria, Garcia-Donas, Jesus, Carrillo, Jaime, Perona, Rosario, Triviño, Juan Carlos, Andrés, Raquel, Cano, Juana María, Rivera, Bárbara, Alonso-Pulpon, Luis, Setien, Fernando, Esteller, Manel, Rodriguez-Perales, Sandra, Bougeard, Gaelle, Frebourg, Tierry, Urioste, Miguel, Blasco, Maria A, & Benítez, Javier. 2015. A mutation in the POT1 gene is responsible for cardiac angiosarcoma in TP53-negative Li-Fraumeni-like families. *Nature communications*, **6**(Sept.), 8383.

Cánovas, José, Berndt, F Andrés, Sepúlveda, Hugo, Aguilar, Rodrigo, Veloso, Felipe A, Montecino, Martín, Oliva, Carlos, Maass, Juan C, Sierralta, Jimena, & Kukuljan, Manuel. 2015. The Specification of Cortical Subcerebral Projection Neurons Depends on the Direct Repression of TBR1 by CTIP1/BCL11a. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, **35**(19), 7552–7564.

Canver, Matthew C, Smith, Elenoe C, Sher, Falak, Pinello, Luca, Sanjana, Neville E, Shalem, Ophir, Chen, Diane D, Schupp, Patrick G, Vinjamur, Divya S, Garcia, Sara P, Luc, Sidinh, Kurita, Ryo, Nakamura, Yukio, Fujiwara, Yuko, Maeda, Takahiro, Yuan, Guo-Cheng, Zhang, Feng, Orkin, Stuart H, & Bauer, Daniel E. 2015. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature*, **527**(7577), 192–197.

Carr, Christopher W, Moreno-De-Luca, Daniel, Parker, Colette, Zimmerman, Holly H, Ledbetter, Nikki, Martin, Christa Lese, Dobyns, William B, & Abdul-Rahman, Omar A. 2010. Chiari I malformation, delayed gross motor skills, severe speech delay, and epileptiform discharges in a child with FOXP1 haploinsufficiency. *European journal of human genetics : EJHG*, **18**(11), 1216–1220.

Carrion-Castillo, Amaia. 2016 (Nov.). *Deciphering common and rare genetic effects on reading ability*. Ph.D. thesis, Radboud University of Nijmegen.

Carrion-Castillo, Amaia, Franke, Barbara, & Fisher, Simon E. 2013. Molecular Genetics of Dyslexia: An Overview. *Dyslexia (Chichester, England)*, **19**(4), 214–240.

Chabchoub, E, Vermeesch, J R, de Ravel, T, de Cock, P, & Fryns, J-P. 2008. The facial dysmorphy in the newly recognised microdeletion 2p15-p16.1 refined to a 570 kb region in 2p15. *Journal of medical genetics*, **45**(3), 189–192.

Chae, Wook-Jin, Henegariu, Octavian, Lee, Sang-Kyou, & Bothwell, Alfred L M. 2006. The mutant leucine-zipper domain impairs both dimerization and suppressive function of Foxp3 in T cells. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(25), 9631–9636.

Chan, Chun Ming, Fulton, Joel, Montiel-Duarte, Cristina, Collins, Hilary M, Bharti, Neetu, Wadelin, Frances R, Moran, Paula M, Mongan, Nigel P, & Heery, David M. 2013. A signature motif mediating selective interactions of BCL11A with the NR2E/F subfamily of orphan nuclear receptors. *Nucleic acids research*, **41**(21), 9663–9679.

Charng, Wu-Lin, Karaca, Ender, Coban Akdemir, Zeynep, Gambin, Tomasz, Atik, Mehmed M, Gu, Shen, Posey, Jennifer E, Jhangiani, Shalini N, Muzny, Donna M, Doddapaneni, Harsha, Hu, Jianhong, Boerwinkle, Eric, Gibbs, Richard A, Rosenfeld, Jill A, Cui, Hong, Xia, Fan, Manickam, Kandamurugu, Yang, Yaping, Faqeih, Eissa A, Al Asmari, Ali, Saleh, Mohammed A M, El-Hattab, Ayman W, & Lupski, James R. 2016. Exome sequencing in mostly consanguineous Arab families with neurologic disease provides a high potential molecular diagnosis rate. *BMC medical genomics*, **9**(1), 42.

Chatila, T A, Blaeser, F, Ho, N, Lederman, H M, Voulgaropoulos, C, Helms, C, & Bowcock, A M. 2000. JM2, encoding a fork head-related protein, is mutated in X-linked autoimmunity-allergic disregulation syndrome. *The Journal of clinical investigation*, **106**(12), R75–81.

Chen, Chun-An, Bosch, Daniëlle G M, Cho, Megan T, Rosenfeld, Jill A, Shinawi, Marwan, Lewis, Richard Alan, Mann, John, Jayakar, Parul, Payne, Katelyn, Walsh, Laurence, Moss, Timothy, Schreiber, Allison, Schoonveld, Cheri, Monaghan, Kristin G, Elmslie, Frances, Douglas, Ganka, Boonstra, F Nienke, Millan, Francisca, Cremers, Frans

P M, McKnight, Dianalee, Richard, Gabriele, Juusola, Jane, Kendall, Fran, Ramsey, Keri, Anyane-Yeboa, Kwame, Malkin, Elfrida, Chung, Wendy K, Niyazov, Dmitriy, Pascual, Juan M, Walkiewicz, Magdalena, Veluchamy, Vivekanand, Li, Chumei, Hisama, Fuki M, de Vries, Bert B A, & Schaaf, Christian. 2016a. The expanding clinical phenotype of Bosch-Boonstra-Schaaf optic atrophy syndrome: 20 new cases and possible genotype-phenotype correlations. *Genetics in medicine : official journal of the American College of Medical Genetics*, **18**(11), 1143–1150.

Chen, Xiaowei Sylvia, Reader, Rose H, Hoischen, Alexander, Veltman, Joris A, Simpson, Nuala H, Francks, Clyde, Newbury, Dianne F, & Fisher, Simon E. 2017. Next-generation DNA sequencing identifies novel gene variants and pathways involved in specific language impairment. *Scientific reports*, **7**(Apr.), 46105.

Chen, Yi-Chuan, Kuo, Hsiao-Ying, Bornschein, Ulrich, Takahashi, Hiroshi, Chen, Shih-Yun, Lu, Kuan-Ming, Yang, Hao-Yu, Chen, Gui-May, Lin, Jing-Ruei, Lee, Yi-Hsin, Chou, Yun-Chia, Cheng, Sin-Jhong, Chien, Cheng-Ting, Enard, Wolfgang, Hevers, Wulf, Pääbo, Svante, Graybiel, Ann M, & Liu, Fu-Chin. 2016b. Foxp2 controls synaptic wiring of corticostriatal circuits and vocal communication by opposing Mef2c. *Nature neuroscience*, **19**(11), 1513–1522.

Chiu, Yi-Chi, Li, Ming-Yang, Liu, Yuan-Hsuan, Ding, Jing-Ya, Yu, Jenn-Yah, & Wang, Tsu-Wei. 2014. Foxp2 regulates neuronal differentiation and neuronal subtype specification. *Developmental neurobiology*, **74**(7), 723–738.

Chokas, Ann L, Trivedi, Chinmay M, Lu, Min Min, Tucker, Philip W, Li, Shanru, Epstein, Jonathan A, & Morrisey, Edward E. 2010. Foxp1/2/4-NuRD interactions regulate gene expression and epithelial injury response in the lung via regulation of interleukin-6. *The Journal of biological chemistry*, **285**(17), 13304–13313.

Chou, Chia-Cheng, & Wang, Andrew H-J. 2015. Structural D/E-rich repeats play multiple roles especially in gene regulation through DNA/RNA mimicry. *Molecular bioSystems*, **11**(8), 2144–2151.

Cismasiu, Valeriu B, Adamo, Karen, Gecewicz, Jennifer, Duque, Javier, Lin, Qishan, & Avram, Dorina. 2005. BCL11B functionally associates with the NuRD complex in T lymphocytes to repress targeted promoter. *Oncogene*, **24**(45), 6753–6764.

Clovis, Yoanne M, Enard, Wolfgang, Marinaro, Federica, Huttner, Wieland B, & De Pietri Tonelli, Davide. 2012. Convergent repression of Foxp2 3'UTR by miR-9 and miR-132 in embryonic mouse neocortex: implications for radial migration of neurons. *Development (Cambridge, England)*, **139**(18), 3332–3342.

Corominas, Roser, Yang, Xinping, Lin, Guan Ning, Kang, Shuli, Shen, Yun, Ghamsari, Lila, Broly, Martin, Rodriguez, Maria, Tam, Stanley, Trigg, Shelly A, Fan, Changyu, Yi, Song, Tasan, Murat, Lemmens, Irma, Kuang, Xingyan, Zhao, Nan, Malhotra, Dheeraj, Michaelson, Jacob J, Vacic, Vladimir, Calderwood, Michael A, Roth, Frederick P, Tavernier, Jan, Horvath, Steve, Salehi-Ashtiani, Kourosh, Korkin, Dmitry, Sebat, Jonathan, Hill, David E, Hao, Tong, Vidal, Marc, & Iakoucheva, Lilia M. 2014. Protein interaction network of alternatively spliced isoforms from brain links genetic risk factors for autism. *Nature communications*, **5**(Apr.), 3650.

Coutinho, Pedro, Pavlou, Sofia, Bhatia, Shipra, Chalmers, Kevin J, Kleinjan, Dirk A, & van Heyningen, Veronica. 2011. Discovery and assessment of conserved Pax6 target genes and enhancers. *Genome research*, **21**(8), 1349–1359.

Dasen, Jeremy S, De Camilli, Alessandro, Wang, Bin, Tucker, Philip W, & Jessell, Thomas M. 2008. Hox Repertoires for Motor Neuron Diversity and Connectivity Gated by a Single Accessory Factor, FoxP1. *Cell*, **134**(2), 304–316.

de Leeuw, N, Pfundt, R, Koolen, D A, Neefs, I, Scheltinga, I, Mieloo, H, Sistermans, E A, Nillesen, W, Smeets, D F, de Vries, B B A, & Knoers, N V A M. 2008. A newly recognised microdeletion syndrome involving 2p15p16.1: narrowing down the critical region by adding another patient detected by genome wide tiling path array comparative genomic hybridisation analysis. *Journal of medical genetics*, **45**(2), 122–124.

De Rubeis, Silvia, He, Xin, Goldberg, Arthur P, Poultney, Christopher S, Samocha, Kaitlin, Cicek, A Erucment, Kou, Yan, Liu, Li, Fromer, Menachem, Walker, Susan, Singh, Tarinder, Klei, Lambertus, Kosmicki, Jack, Shih-Chen, Fu, Aleksic, Branko, Biscaldi, Monica, Bolton, Patrick F, Brownfeld, Jessica M, Cai, Jinlu, Campbell, Nicholas G, Carracedo, Angel, Chahrour, Maria H, Chiocchetti, Andreas G, Coon, Hilary, Crawford, Emily L, Curran, Sarah R, Dawson, Geraldine, Duketis, Eftichia, Fernandez, Bridget A, Gallagher, Louise, Geller, Evan, Guter, Stephen J, Hill, R Sean, Ionita-Laza, Juliana, Jimenz Gonzalez, Patricia, Kilpinen, Helena, Klauck, Sabine M, Kolevzon, Alexander, Lee, Irene, Lei, Irene, Lei, Jing, Lehtimäki, Terho, Lin, Chiao-Feng, Ma'ayan, Avi, Marshall, Christian R, McInnes, Alison L, Neale, Benjamin, Owen, Michael J, Ozaki, Noriio, Parellada, Mara, Parr, Jeremy R, Purcell, Shaun, Puura, Kaija, Rajagopalan, Deepthi, Rehnström, Karola, Reichenberg, Abraham, Sabo, Aniko, Sachse, Michael, Sanders, Stephan J, Schafer, Chad, Schulte-Rüther, Martin, Skuse, David, Stevens, Christine, Szatmari, Peter, Tammimies, Kristiina, Valladares, Otto, Voran, Annette, Li-San, Wang, Weiss, Lauren A, Willsey, A Jeremy, Yu, Timothy W, Yuen, Ryan K C, DDD Study, Homozygosity Mapping Collaborative for Autism, UK10K Consortium, Cook, Edwin H, Freitag, Christine M, Gill, Michael, Hultman, Christina M, Lehner, Thomas, Palotie, Aaarno, Schellenberg, Gerard D, Sklar, Pamela,

State, Matthew W, Sutcliffe, James S, Walsh, Christiopher A, Scherer, Stephen W, Zwick, Michael E, Barett, Jeffrey C, Cutler, David J, Roeder, Kathryn, Devlin, Bernie, Daly, Mark J, & Buxbaum, Joseph D. 2014. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*, **515**(7526), 209–215.

Deciphering Developmental Disorders Study. 2015. Large-scale discovery of novel genetic causes of developmental disorders. *Nature*, **519**(7542), 223–228.

Denaxa, Myrto, Kalaitzidou, Melanie, Garefalaki, Anna, Achimastou, Angeliki, Lasrado, Reena, Maes, Tamara, & Pachnis, Vassilis. 2012. Maturation-promoting activity of SATB1 in MGE-derived cortical interneurons. *Cell reports*, **2**(5), 1351–1362.

Deplancke, Bart, Alpern, Daniel, & Gardeux, Vincent. 2016. The Genetics of Transcription Factor DNA Binding Variation. *Cell*, **166**(3), 538–554.

Deriziotis, Pelagia, O'Roak, Brian J, Graham, Sarah A, Estruch, Sara B, Dimitropoulou, Danai, Bernier, Raphael A, Gerdts, Jennifer, Shendure, Jay, Eichler, Evan E, & Fisher, Simon E. 2014a. De novo TBR1 mutations in sporadic autism disrupt protein functions. *Nature communications*, **5**(Sept.), 4954.

Deriziotis, Pelagia, Graham, Sarah A, Estruch, Sara B, & Fisher, Simon E. 2014b. Investigating protein-protein interactions in live cells using bioluminescence resonance energy transfer. *Journal of visualized experiments : JoVE*, May.

Devanna, Paolo, Middelbeek, Jeroen, & Vernes, Sonja C. 2014. FOXP2 drives neuronal differentiation by interacting with retinoic acid signaling pathways. *Frontiers in cellular neuroscience*, **8**, 305.

Dias, Cristina, Estruch, Sara B, Graham, Sarah A, McRae, Jeremy, Sawiak, Stephen J, Hurst, Jane A, Joss, Shelagh K, Holder, Susan E, Morton, Jenny E V, Turner, Claire, Thevenon, Julien, Mellul, Kelly, Sánchez-Andrade, Gabriela, Ibarra-Soria, Ximena, Deriziotis, Pelagia, Santos, Rui F, Lee, Song-Choon, Faivre, Laurence, Kleefstra, Tjitske, Liu, Pentao, Hurles, Mathew E, DDD Study, Fisher, Simon E, & Logan, Darren W. 2016. BCL11A Haploinsufficiency Causes an Intellectual Disability Syndrome and Dysregulates Transcription. *American journal of human genetics*, **99**(2), 253–274.

Diella, Francesca, Haslam, Niall, Chica, Claudia, Budd, Aidan, Michael, Sushama, Brown, Nigel P, Trave, Gilles, & Gibson, Toby J. 2008. Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Frontiers in bioscience : a journal and virtual library*, **13**(May), 6580–6603.

Diella, Francesca, Chabanis, Sophie, Luck, Katja, Chica, Claudia, Ramu, Chenna, Nerlov, Claus, & Gibson, Toby J. 2009. KEPE–a motif frequently superimposed on sumoylation sites in metazoan chromatin proteins and transcription factors. *Bioinformatics (Oxford, England)*, **25**(1), 1–5.

Dong, Xuesen, Sweet, Joan, Challis, John R G, Brown, Theodore, & Lye, Stephen J. 2007. Transcriptional activity of androgen receptor is modulated by two RNA splicing factors, PSF and p54nrb. *Molecular and cellular biology*, **27**(13), 4863–4875.

Donovan, Alex P A, & Basson, M Albert. 2016. The neuroanatomy of autism - a developmental perspective. *Journal of anatomy*, Sept.

Dyson, H Jane, & Wright, Peter E. 2005. Intrinsically unstructured proteins and their functions. *Nature reviews. Molecular cell biology*, **6**(3), 197–208.

Elahi, Fanny M. 2008 (Aug.). *Functional studies of FOXP2— a gene involved in brain development and language acquisition* . Ph.D. thesis, Oxford University, Oxford University.

Enard, Wolfgang, Gehre, Sabine, Hammerschmidt, Kurt, Hölter, Sabine M, Blass, Torsten, Somel, Mehmet, Brückner, Martina K, Schreiweis, Christiane, Winter, Christine, Sohr, Reinhard, Becker, Lore, Wiebe, Victor, Nickel, Birgit, Giger, Thomas, Müller, Uwe, Groszer, Matthias, Adler, Thure, Aguilar, Antonio, Bolle, Ines, Calzada-Wack, Julia, Dalke, Claudia, Ehrhardt, Nicole, Favor, Jack, Fuchs, Helmut, Gailus-Durner, Valérie, Hans, Wolfgang, Hölzlwimmer, Gabriele, Javaheri, Anahita, Kalaydjiev, Svetoslav, Kallnik, Magdalena, Kling, Eva, Kunder, Sandra, Mossbrugger, Ilona, Naton, Beatrix, Racz, Ildikó, Rathkolb, Birgit, Rozman, Jan, Schrewe, Anja, Busch, Dirk H, Graw, Jochen, Ivandic, Boris, Klingenspor, Martin, Klopstock, Thomas, Ollert, Markus, Quintanilla-Martinez, Leticia, Schulz, Holger, Wolf, Eckhard, Wurst, Wolfgang, Zimmer, Andreas, Fisher, Simon E, Morgenstern, Rudolf, Arendt, Thomas, de Angelis, Martin Hrabé, Fischer, Julia, Schwarz, Johannes, & Pääbo, Svante. 2009. A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell*, **137**(5), 961–971.

Engelen, Erik, Akinci, Umut, Bryne, Jan Christian, Hou, Jun, Gontan, Cristina, Moen, Maaike, Szumska, Dorota, Kockx, Christel, van Ijcken, Wilfred, Dekkers, Dick H W, Demmers, Jeroen, Rijkers, Erik-Jan, Bhattacharya, Shoumo, Philipsen, Sjaak, Pevny, Larysa H, Grosveld, Frank G, Rottier, Robbert J, Lenhard, Boris, & Poot, Raymond A. 2011. Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. *Nature genetics*, **43**(6), 607–611.

Estruch, Sara B, Buzón, Víctor, Carbó, Laia R, Schorova, Lenka, Lüders, Jens, & Estébanez-Perpiñá, Eva. 2012. The oncoprotein BCL11A binds to orphan nuclear receptor TLX and potentiates its transrepressive function. *PloS one*, **7**(6), e37963.

Estruch, Sara B, Graham, Sarah A, Chinnappa, Swathi M, Deriziotis, Pelagia, & Fisher, Simon E. 2016a. Functional characterization of rare FOXP2 variants in neurodevelopmental disorder. *Journal of Neurodevelopmental Disorders*, **8**(1), 354.

Estruch, Sara B, Graham, Sarah A, Deriziotis, Pelagia, & Fisher, Simon E. 2016b. The language-related transcription factor FOXP2 is post-translationally modified with small ubiquitin-like modifiers. *Scientific reports*, **6**(Feb.), 20911.

Faedo, Andrea, Tomassy, Giulio Srubek, Ruan, Youlin, Teichmann, Hannah, Krauss, Stefan, Pleasure, Samuel J, Tsai, Sophia Y, Tsai, Ming-Jer, Studer, Michèle, & Rubenstein, John L R. 2008. COUP-TFI coordinates cortical patterning, neurogenesis, and laminar fate and modulates MAPK/ERK, AKT, and beta-catenin signaling. *Cerebral cortex (New York, N.Y. : 1991)*, **18**(9), 2117–2131.

Fannemel, Madeleine, Barøy, Tuva, Holmgren, Asbjørn, Rødningen, Olaug K, Haugsand, Trine M, Hansen, Børre, Frengen, Eirik, & Misceo, Doriana. 2014. Haploinsufficiency of XPO1 and USP34 by a de novo 230 kb deletion in 2p15, in a patient with mild intellectual disability and cranio-facial dysmorphisms. *European journal of medical genetics*, **57**(9), 513–519.

Félix, Têmis Maria, Petrin, Aline Lourenço, Sanseverino, Maria Teresa Vieira, & Murray, Jeffrey C. 2010. Further characterization of microdeletion syndrome involving 2p15-p16.1. *American journal of medical genetics. Part A*, **152A**(10), 2604–2608.

Ferland, Russell J, Cherry, Timothy J, Preware, Patricia O, Morrisey, Edward E, & Walsh, Christopher A. 2003. Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain. *The Journal of comparative neurology*, **460**(2), 266–279.

Feuk, Lars, Kalervo, Aino, Lipsanen-Nyman, Marita, Skaug, Jennifer, Nakabayashi, Kazuhiko, Finucane, Brenda, Hartung, Danielle, Innes, Micheil, Kerem, Batsheva, Nowaczyk, Malgorzata J, Rivlin, Joseph, Roberts, Wendy, Senman, Lili, Summers, Anne, Szatmari, Peter, Wong, Virginia, Vincent, John B, Zeesman, Susan, Osborne, Lucy R, Cardy, Janis Oram, Kere, Juha, Scherer, Stephen W, & Hannula-Jouppi, Katariina. 2006. Absence of a paternally inherited FOXP2 gene in developmental verbal dyspraxia. *American journal of human genetics*, **79**(5), 965–972.

Filtz, Theresa M, Vogel, Walter K, & Leid, Mark. 2014. Regulation of transcription factor activity by interconnected post-translational modifications. *Trends in pharmacological sciences*, **35**(2), 76–85.

Finkel, T, Duc, J, Fearon, E R, Dang, C V, & Tomaselli, G F. 1993. Detection and modulation in vivo of helix-loop-helix protein-protein interactions. *The Journal of biological chemistry*, **268**(1), 5–8.

Fisher, S E, Vargha-Khadem, F, Watkins, K E, Monaco, A P, & Pembrey, M E. 1998. Localisation of a gene implicated in a severe speech and language disorder. *Nature genetics*, **18**(2), 168–170.

Fisher, Simon E, & Scharff, Constance. 2009. FOXP2 as a molecular window into speech and language. *Trends in genetics : TIG*, **25**(4), 166–177.

Fisher, Simon E, & Vernes, Sonja C. 2015. Genetics and the Language Sciences. *Annual Review of Linguistics*, **1**(1), 289–310.

Fisher, Simon E, Lai, Cecilia S L, & Monaco, Anthony P. 2003. Deciphering the genetic basis of speech and language disorders. *Annual review of neuroscience*, **26**(1), 57–80.

Florisson, Joyce M G, Mathijssen, Irene M J, Dumee, Belinda, Hoogeboom, Jeannette A M, Poddighe, Pino J, Oostra, Ben A, Frijns, Jean Pierre, Koster, Linda, de Klein, Annelies, Eussen, Bert, de Vries, Bert B A, Swagemakers, Sigrid, van der Spek, Peter J, & Verkerk, Annemieke J M H. 2013. Complex craniosynostosis is associated with the 2p15p16.1 microdeletion syndrome. *American journal of medical genetics. Part A*, **161A**(2), 244–253.

Fontenot, Jason D, Gavin, Marc A, & Rudensky, Alexander Y. 2003. Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. *Nature immunology*, **4**(4), 330–336.

Fox, Archa H, & Lamond, Angus I. 2010. Paraspeckles. *Cold Spring Harbor perspectives in biology*, **2**(7), a000687–a000687.

Fox, Archa H, Bond, Charles S, & Lamond, Angus I. 2005. P54nrb forms a heterodimer with PSP1 that localizes to paraspeckles in an RNA-dependent manner. *Molecular biology of the cell*, **16**(11), 5304–5315.

Francks, Clyde, Paracchini, Silvia, Smith, Shelley D, Richardson, Alex J, Scerri, Tom S, Cardon, Lon R, Marlow, Angela J, MacPhie, I Laurence, Walter, Janet, Pennington, Bruce F, Fisher, Simon E, Olson, Richard K, DeFries, John C, Stein, John F, & Monaco, Anthony P. 2004. A 77-kilobase region of chromosome 6p22.2 is associated with dyslexia in families from the United Kingdom and from the United States. *American journal of human genetics*, **75**(6), 1046–1058.

French, C A, Jin, X, Campbell, T G, Gerfen, E, Groszer, M, Fisher, S E, & Costa, R M. 2012. An aetiological Foxp2 mutation causes aberrant striatal activity and alters plasticity during skill learning. *Molecular psychiatry*, **17**(11), 1077–1085.

French, Catherine A, & Fisher, Simon E. 2014. What can mice tell us about Foxp2 function? *Current opinion in neurobiology*, **28**(Oct.), 72–79.

Fu, Lijuan, Shi, Zhimin, Luo, Guanzheng, Tu, Weihong, Wang, XiuJie, Fang, Zhide, & Li, XiaoChing. 2014. Multiple microRNAs regulate human FOXP2 gene expression by targeting sequences in its 3' untranslated region. *Molecular brain*, **7**(1), 71.

Fujita-Jimbo, Eriko, & Momoi, Takashi. 2014. Specific expression of FOXP2 in cerebellum improves ultrasonic vocalization in heterozygous but not in homozygous Foxp2 (R552H) knock-in pups. *Neuroscience letters*, **566**(Apr.), 162–166.

Funnell, Alister P W, Prontera, Paolo, Ottaviani, Valentina, Piccione, Maria, Giambona, Antonino, Maggio, Aurelio, Ciaffoni, Fiorella, Stehling-Sun, Sandra, Marra, Manuela, Masiello, Francesca, Varricchio, Lilian, Stamatoyannopoulos, John A, Migliaccio, Anna R, & Papayannopoulou, Thalia. 2015. 2p15-p16.1 microdeletions encompassing and proximal to BCL11A are associated with elevated HbF in addition to neurologic impairment. *Blood*, **126**(1), 89–93.

Gabriele, Michele, Vulto-van Silfhout, Anneke T, Germain, Pierre-Luc, Vitriolo, Alessandro, Kumar, Raman, Douglas, Evelyn, Haan, Eric, Kosaki, Kenjiro, Takenouchi, Toshiki, Rauch, Anita, Steindl, Katharina, Frengen, Eirik, Misceo, Doriana, Pedurupillay, Christeen Ramane J, Stromme, Petter, Rosenfeld, Jill A, Shao, Yunru, Craigen, William J, Schaaf, Christian P, Rodriguez-Buritica, David, Farach, Laura, Friedman, Jennifer, Thulin, Perla, McLean, Scott D, Nugent, Kimberly M, Morton, Jenny, Nicholl, Jillian, Andrieux, Joris, Stray-Pedersen, Asbjørg, Chambon, Pascal, Patrier, Sophie, Lynch, Sally A, Kjaergaard, Susanne, Tørring, Pernille M, Brasch-Andersen, Charlotte, Ronan, Anne, van Haeringen, Arie, Anderson, Peter J, Powis, Zöe, Brunner, Han G, Pfundt, Rolph, Schuurs-Hoeijmakers, Janneke H M, van Bon, Bregje W M, Lelieveld, Stefan, Gilissen, Christian, Nillesen, Willy M, Vissers, Lisenka E L M, Gecz, Jozef, Koolen, David A, Testa, Giuseppe, & de Vries, Bert B A. 2017. YY1 Haploinsufficiency Causes an Intellectual Disability Syndrome Featuring Transcriptional and Chromatin Dysfunction. *American journal of human genetics*, **100**(6), 907–925.

Gagliardi, Alessia, Mullin, Nicholas P, Ying Tan, Zi, Colby, Douglas, Kousa, Anastasia I, Halbritter, Florian, Weiss, Jason T, Felker, Anastasia, Bezstarosti, Karel, Favaro, Rebecca, Demmers, Jeroen, Nicolis, Silvia K, Tomlinson, Simon R, Poot, Raymond A, & Chambers, Ian. 2013. A direct physical interaction between Nanog and Sox2 regulates embryonic stem cell self-renewal. *The EMBO journal*, **32**(16), 2231–2247.

Garcia-Calero, Elena, Botella-Lopez, Arancha, Bahamonde, Olga, Perez-Balaguer, Ariadna, & Martinez, Salvador. 2016. FoxP2 protein levels regulate cell morphology changes and migration patterns in the vertebrate developing telencephalon. *Brain Structure and Function*, **221**(6), 2905–2917.

Geiss-Friedlander, Ruth, & Melchior, Frauke. 2007. Concepts in sumoylation: a decade on. *Nature reviews. Molecular cell biology*, **8**(12), 947–956.

Golebiowski, Filip, Matic, Ivan, Tatham, Michael H, Cole, Christian, Yin, Yili, Nakamura, Akihiro, Cox, Jürgen, Barton, Geoffrey J, Mann, Matthias, & Hay, Ronald T. 2009. System-wide changes to SUMO modifications in response to heat shock. *Science signaling*, **2**(72), ra24–ra24.

Golson, Maria L, & Kaestner, Klaus H. 2016. Fox transcription factors: from development to disease. *Development (Cambridge, England)*, **143**(24), 4558–4570.

Graham, Sarah A, & Fisher, Simon E. 2015. Understanding Language from a Genomic Perspective. *Annual review of genetics*, **49**(1), 131–160.

Graham, Sarah A, Deriziotis, Pelagia, & Fisher, Simon E. 2015. Insights into the genetic foundations of human communication. *Neuropsychology review*, **25**(1), 3–26.

Greig, Luciano C, Woodworth, Mollie B, Greppi, Chloé, & Macklis, Jeffrey D. 2016. Ctip1 Controls Acquisition of Sensory Area Identity and Establishment of Sensory Input Fields in the Developing Neocortex. *Neuron*, **90**(2), 261–277.

Greig, Luciano Custo, Woodworth, Mollie B, Galazo, Maria J, Padmanabhan, Hari, & Macklis, Jeffrey D. 2013. Molecular logic of neocortical projection neuron specification, development and diversity. *Nature reviews. Neuroscience*, **14**(11), 755–769.

Groszer, Matthias, Keays, David A, Deacon, Robert M J, de Bono, Joseph P, Prasad-Mulcare, Shweta, Gaub, Simone, Baum, Muriel G, French, Catherine A, Nicod, Jérôme, Coventry, Julie A, Enard, Wolfgang, Fray, Martin, Brown, Steve D M, Nolan, Patrick M, Pääbo, Svante, Channon, Keith M, Costa, Rui M, Eilers, Jens, Ehret, Günter, Rawlins, J Nicholas P, & Fisher, Simon E. 2008. Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits. *Current biology : CB*, **18**(5), 354–362.

Gwizdek, Carole, Cassé, Frédéric, & Martin, Stéphane. 2013. Protein sumoylation in brain development, neuronal morphology and spinogenesis. *Neuromolecular medicine*, **15**(4), 677–691.

Haesler, Sebastian, Rochefort, Christelle, Georgi, Benjamin, Licznerski, Pawel, Osten, Pavel, & Scharff, Constance. 2007. Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. *PLoS biology*, **5**(12), e321.

Hamdan, Fadi F, Daoud, Hussein, Rochefort, Daniel, Piton, Amélie, Gauthier, Julie, Langlois, Mathieu, Foomani, Gila, Dobrzeniecka, Sylvia, Krebs, Marie-Odile, Joober, Ridha, Lafrenière, Ronald G, Lacaille, Jean-Claude, Mottron, Laurent, Drapeau, Pierre, Beauchamp, Miriam H, Phillips, Michael S, Fombonne, Eric, Rouleau, Guy A, & Michaud, Jacques L. 2010. De novo mutations in FOXP1 in cases with intellectual disability, autism, and language impairment. *American journal of human genetics*, **87**(5), 671–678.

Hammarström, Harald. 2016. Linguistic diversity and language evolution. *Journal of Language Evolution*, **1**(1), 19–29.

Han, Wenqi, Kwan, Kenneth Y, Shim, Sungbo, Lam, Mandy M S, Shin, Yurae, Xu, Xuming, Zhu, Ying, Li, Mingfeng, & Sestan, Nenad. 2011. TBR1 directly represses Fezf2 to control the laminar origin and development of the corticospinal tract. *Proceedings of the National Academy of Sciences of the United States of America*, **108**(7), 3041–3046.

Hancarova, Miroslava, Simandlova, Martina, Drabova, Jana, Mannik, Katrin, Kurg, Ants, & Sedlacek, Zdenek. 2013. A patient with de novo 0.45 Mb deletion of 2p16.1: the role of BCL11A, PAPOLG, REL, and FLJ16341 in the 2p15-p16.1 microdeletion syndrome. *American journal of medical genetics. Part A*, **161A**(4), 865–870.

Hannenhalli, Sridhar, & Kaestner, Klaus H. 2009. The evolution of Fox genes and their role in development and disease. *Nature reviews. Genetics*, **10**(4), 233–240.

Hannula-Jouppi, Katariina, Kaminen-Ahola, Nina, Taipale, Mikko, Eklund, Ranja, Nopola-Hemmi, Jaana, Kääriäinen, Helena, & Kere, Juha. 2005. The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. *PLoS genetics*, **1**(4), e50.

Hasegawa, Yuta, Yoshida, Daisuke, Nakamura, Yuki, & Sakakibara, Shin-ichi. 2014. Spatiotemporal distribution of SUMOylation components during mouse brain development. *The Journal of comparative neurology*, **522**(13), 3020–3036.

Hayes, Sheri, Malacrida, Beatrice, Kiely, Maeve, & Kiely, Patrick A. 2016. Studying protein-protein interactions: progress, pitfalls and solutions. *Biochemical Society transactions*, **44**(4), 994–1004.

Henley, Jeremy M, Craig, Tim J, & Wilkinson, Kevin A. 2014. Neuronal SUMOylation: mechanisms, physiology, and roles in neuronal dysfunction. *Physiological reviews*, **94**(4), 1249–1285.

Heston, Jonathan B, & White, Stephanie A. 2015. Behavior-linked FoxP2 regulation enables zebra finch vocal learning. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, **35**(7), 2885–2894.

Hoischen, Alexander, Krumm, Niklas, & Eichler, Evan E. 2014. Prioritization of neurodevelopmental disease genes by discovery of new mutations. *Nature neuroscience*, **17**(6), 764–772.

Horn, Denise, Kapeller, Johannes, Rivera-Brugués, Núria, Moog, Ute, Lorenz-Depiereux, Bettina, Eck, Sebastian, Hempel, Maja, Wagenstaller, Janine, Gawthrope, Alex, Monaco, Anthony P, Bonin, Michael, Riess, Olaf, Wohlleber, Eva, Illig, Thomas, Bezzina, Connie R, Franke, Andre, Spranger, Stephanie, Villavicencio-Lorini, Pablo, Seifert, Wenke, Rosenfeld, Jochen, Klopocki, Eva, Rappold, Gudrun A, & Strom, Tim M. 2010. Identification of FOXP1 deletions in three unrelated patients with mental retardation and significant speech and language deficits. *Human mutation*, **31**(11), E1851–60.

Hossain, Mir A, & Bungert, Jörg. 2017. Genome Editing for Sickle Cell Disease: A Little BCL11A Goes a Long Way. *Molecular therapy : the journal of the American Society of Gene Therapy*, **25**(3), 561–562.

Hsieh, M, Tintut, Y, & Gralla, J D. 1994. Functional roles for the glutamines within the glutamine-rich region of the transcription factor sigma 54. *The Journal of biological chemistry*, **269**(1), 373–378.

Hsu, Wei-Lun, Oldfield, Christopher, Meng, Jingwei, Huang, Fei, Xue, Bin, Uversky, Vladimir N, Romero, Pedro, & Dunker, A Keith. 2012. Intrinsic protein disorder and protein-protein interactions. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 116–127.

Hu, Chun, Chen, Wenjuan, Myers, Scott J, Yuan, Hongjie, & Traynelis, Stephen F. 2016. Human GRIN2B variants in neurodevelopmental disorders. *Journal of pharmacological sciences*, **132**(2), 115–121.

Hu, Hui, Wang, Bin, Borde, Madhuri, Nardone, Julie, Maika, Shan, Allred, Laura, Tucker, Philip W, & Rao, Anjana. 2006. Foxp1 is an essential transcriptional regulator of B cell development. *Nature immunology*, **7**(8), 819–826.

Huang, Ying, Zhang, Lei, Song, Ning-Ning, Hu, Ze-Lan, Chen, Jia-Yin, & Ding, Yu-Qiang. 2011. Distribution of Satb1 in the central nervous system of adult mice. *Neuroscience research*, **71**(1), 12–21.

Hucthagowder, Vishwanathan, Liu, Ta-Chiang, Paciorkowski, Alex R, Thio, Liu Lin, Keller, Martin S, Anderson, Christopher D, Herman, Thomas, Dehner, Louis P, Grange, Dorothy K, & Kulkarni, Shashikant. 2012. Chromosome 2p15p16.1 microdeletion syndrome: 2.5 Mb deletion in a patient with renal anomalies, intractable seizures and a choledochal cyst. *European journal of medical genetics*, **55**(8-9), 485–489.

Iossifov, Ivan, O'Roak, Brian J, Sanders, Stephan J, Ronemus, Michael, Krumm, Niklas, Levy, Dan, Stessman, Holly A, Witherspoon, Kali T, Vives, Laura, Patterson, Karynne E, Smith, Joshua D, Paeper, Bryan, Nickerson, Deborah A, Dea, Jeanselle, Dong, Shan, Gonzalez, Luis E, Mandell, Jeffrey D, Mane, Shrikant M, Murtha, Michael T, Sullivan, Catherine A, Walker, Michael F, Waqar, Zainulabedin, Wei, Liping, Willsey, A Jeremy, Yamrom, Boris, Lee, Yoon-ha, Grabowska, Ewa, Dalkic, Ertugrul, Wang, Zihua, Marks, Steven, Andrews, Peter, Leotta, Anthony, Kendall, Jude, Hakker, Inessa, Rosenbaum, Julie, Ma, Beicong, Rodgers, Linda, Troge, Jennifer, Narzisi, Giuseppe, Yoon, Seungtai, Schatz, Michael C, Ye, Kenny, McCombie, W Richard, Shendure, Jay, Eichler, Evan E, State, Matthew W, & Wigler, Michael. 2014. The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, **515**(7526), 216–221.

Ishida, Hironori, Masuhiro, Yoshikazu, Fukushima, Akie, Argueta, Jose Guillermo Martinez, Yamaguchi, Noboru, Shiota, Susumu, & Hanazawa, Shigemasa. 2005. Identification and characterization of novel isoforms of human DP-1: DP-1alpha regulates the transcriptional activity of E2F1 as well as cell cycle progression in a dominant-negative manner. *The Journal of biological chemistry*, **280**(26), 24642–24648.

Izumi, Hiroto, McCloskey, Asako, Shinmyozu, Kaori, & Ohno, Mutsuhito. 2014. p54nrb/NonO and PSF promote U snRNA nuclear export by accelerating its export complex assembly. *Nucleic acids research*, **42**(6), 3998–4007.

Jackson, Brian C, Carpenter, Christopher, Nebert, Daniel W, & Vasiliou, Vasilis. 2010. Update of human and mouse forkhead box (FOX) gene families. *Human genomics*, **4**(5), 345–352.

Jakobs, Astrid, Koehnke, Jesko, Himstedt, Fabian, Funk, Martin, Korn, Bernhard, Gaestel, Matthias, & Niedenthal, Rainer. 2007. Ubc9 fusion-directed SUMOylation (UFDS): a method to analyze function of protein SUMOylation. *Nature methods*, **4**(3), 245–250.

Jeon, Bu-Nam, Choi, Won-Il, Yu, Mi-Young, Yoon, A-Rum, Kim, Myung-Hwa, Yun, Chae-Ok, & Hur, Man-Wook. 2009. ZBTB2, a novel master regulator of the p53 pathway. *The Journal of biological chemistry*, **284**(27), 17935–17946.

Jepsen, Kristen, Gleiberman, Anatoli S, Shi, Can, Simon, Daniel I, & Rosenfeld, Michael G. 2008. Cooperative regulation in development by SMRT and FOXP1. *Genes & development*, **22**(6), 740–745.

Kahyo, T, Nishida, T, & Yasuda, H. 2001. Involvement of PIAS1 in the sumoylation of tumor suppressor p53. *Molecular cell*, **8**(3), 713–718.

Kanai, Yoshimitsu, Dohmae, Naoshi, & Hirokawa, Nobutaka. 2004. Kinesin transports RNA: isolation and characterization of an RNA-transporting granule. *Neuron*, **43**(4), 513–525.

Kaneko, Syuzo, Rozenblatt-Rosen, Orit, Meyerson, Matthew, & Manley, James L. 2007. The multifunctional protein p54nrb/PSF recruits the exonuclease XRN2 to facilitate pre-mRNA 3' processing and transcription termination. *Genes & development*, **21**(14), 1779–1789.

Kang, Changsoo, & Drayna, Dennis. 2011. Genetics of speech and language disorders. *Annual review of genomics and human genetics*, **12**(1), 145–164.

Kang, Changsoo, Riazuddin, Sheikh, Mundorff, Jennifer, Krasnewich, Donna, Friedman, Penelope, Mullikin, James C, & Drayna, Dennis. 2010. Mutations in the lysosomal enzyme-targeting pathway and persistent stuttering. *The New England journal of medicine*, **362**(8), 677–685.

Kim, Bum Jun, Takamoto, Norio, Yan, Jun, Tsai, Sophia Y, & Tsai, Ming-Jer. 2009a. Chicken Ovalbumin Upstream Promoter-Transcription Factor II (COUP-TFII) regulates growth and patterning of the postnatal mouse cerebellum. *Developmental biology*, **326**(2), 378–391.

Kim, Jonghwan, Cantor, Alan B, Orkin, Stuart H, & Wang, Jianlong. 2009b. Use of in vivo biotinylation to study protein-protein and protein-DNA interactions in mouse embryonic stem cells. *Nature protocols*, **4**(4), 506–517.

Kim, Kee K, Kim, Yong C, Adelstein, Robert S, & Kawamoto, Sachiyo. 2011. Fox-3 and PSF interact to activate neural cell-specific alternative splicing. *Nucleic acids research*, **39**(8), 3064–3078.

Kimura, Shioko. 2003. [Homeodomain transcription factor T/EBP/NKX2.1 in development and differentiation of the thyroid and lung]. *Seikagaku. The Journal of Japanese Biochemical Society*, **75**(12), 1493–1504.

Kitzman, Jacob O, Starita, Lea M, Lo, Russell S, Fields, Stanley, & Shendure, Jay. 2015. Massively parallel single-amino-acid mutagenesis. *Nature methods*, **12**(3), 203–6– 4 p following 206.

Knott, Gavin J, Bond, Charles S, & Fox, Archa H. 2016. The DBHS proteins SFPQ, NONO and PSPC1: a multipurpose molecular scaffold. *Nucleic acids research*, **44**(9), 3989–4004.

Konstantoulas, Constantine James, Parmar, Malin, & Li, Meng. 2010. FoxP1 promotes midbrain identity in embryonic stem cell-derived dopamine neurons by regulating Pitx3. *Journal of neurochemistry*, **113**(4), 836–847.

Kotaja, Noora, Karvonen, Ulla, Janne, Olli A, & Palvimo, Jorma J. 2002. PIAS proteins modulate transcription factors by functioning as SUMO-1 ligases. *Molecular and cellular biology*, **22**(14), 5222–5234.

Kowalska, Elzbieta, Ripperger, Jürgen A, Muheim, Christine, Maier, Bert, Kurihara, Yasuyuki, Fox, Archa H, Kramer, Achim, & Brown, Steven A. 2012. Distinct roles of DBHS family members in the circadian transcriptional feedback loop. *Molecular and cellular biology*, **32**(22), 4585–4594.

Kowalska, Elzbieta, Ripperger, Juergen A, Hoegger, Dominik C, Bruegger, Pascal, Buch, Thorsten, Birchler, Thomas, Mueller, Anke, Albrecht, Urs, Contaldo, Claudio, & Brown, Steven A. 2013. NONO couples the circadian clock to the cell cycle. *Proceedings of the National Academy of Sciences of the United States of America*, **110**(5), 1592–1599.

Krishna, S Sri, Majumdar, Indraneel, & Grishin, Nick V. 2003. Structural classification of zinc fingers: survey and summary. *Nucleic acids research*, **31**(2), 532–550.

Krishnan, Saloni, Watkins, Kate E, & Bishop, Dorothy V M. 2016. Neurobiological Basis of Language Learning Difficulties. *Trends in cognitive sciences*, **20**(9), 701–714.

Kuhl, Patricia K. 2004. Early language acquisition: cracking the speech code. *Nature reviews. Neuroscience*, **5**(11), 831–843.

Kuo, Ting-Yu, & Hsueh, Yi-Ping. 2007. Expression of zinc finger transcription factor Bcl11A/Evi9/CTIP1 in rat brain. *Journal of Neuroscience Research*, **85**(8), 1628–1636.

Kuo, Ting-Yu, Hong, Chen-Jei, & Hsueh, Yi-Ping. 2009. Bcl11A/CTIP1 regulates expression of DCC and MAP1b in control of axon branching and dendrite outgrowth. *Molecular and cellular neurosciences*, **42**(3), 195–207.

Kuo, Ting-Yu, Hong, Chen-Jei, Chien, Hsu-Ling, & Hsueh, Yi-Ping. 2010. X-linked mental retardation gene CASK interacts with Bcl11A/CTIP1 and regulates axon branching and outgrowth. *Journal of Neuroscience Research*, **88**(11), 2364–2373.

Kwan, Kenneth Y. 2013. Transcriptional dysregulation of neocortical circuit assembly in ASD. *International review of neurobiology*, **113**, 167–205.

Kwon, Hye-Sook, Lim, Hyung W, Wu, Jessica, Schnölzer, Martina, Verdin, Eric, & Ott, Melanie. 2012. Three novel acetylation sites in the Foxp3 transcription factor regulate the suppressive activity of regulatory T cells. *Journal of immunology (Baltimore, Md. : 1950)*, **188**(6), 2712–2721.

Laffin, Jennifer J S, Raca, Gordana, Jackson, Craig A, Strand, Edythe A, Jakielski, Kathy J, & Shriberg, Lawrence D. 2012. Novel candidate genes and regions for childhood apraxia of speech identified by array comparative genomic hybridization. *Genetics in medicine : official journal of the American College of Medical Genetics*, **14**(11), 928–936.

Lai, C S, Fisher, S E, Hurst, J A, Levy, E R, Hodgson, S, Fox, M, Jeremiah, S, Povey, S, Jamison, D C, Green, E D, Vargha-Khadem, F, & Monaco, A P. 2000. The SPCH1 region on human 7q31: genomic characterization of the critical interval and localization of translocations associated with speech and language disorder. *American journal of human genetics*, **67**(2), 357–368.

Lai, C S, Fisher, S E, Hurst, J A, Vargha-Khadem, F, & Monaco, A P. 2001. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, **413**(6855), 519–523.

Lai, Cecilia S L, Gerrelli, Dianne, Monaco, Anthony P, Fisher, Simon E, & Copp, Andrew J. 2003. FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. *Brain : a journal of neurology*, **126**(Pt 11), 2455–2462.

Lai, Tina, Jabaudon, Denis, Molyneaux, Bradley J, Azim, Eiman, Arlotta, Paola, Menezes, Joao R L, & Macklis, Jeffrey D. 2008. SOX5 controls the sequential generation of distinct corticofugal neuron subtypes. *Neuron*, **57**(2), 232–247.

Lamb, Allen N, Rosenfeld, Jill A, Neill, Nicholas J, Talkowski, Michael E, Blumenthal, Ian, Girirajan, Santhosh, Keelean-Fuller, Debra, Fan, Zheng, Pouncey, Jill, Stevens, Cathy, Mackay-Loder, Loren, Terespolsky, Deborah, Bader, Patricia I, Rosenbaum, Kenneth, Vallee, Stephanie E, Moeschler, John B, Ladda, Roger, Sell, Susan, Martin, Judith, Ryan, Shawnia, Jones, Marilyn C, Moran, Rocio, Shealy, Amy, Madan-Khetarpal, Suneeta, McConnell, Juliann, Surti, Urvashi, Delahaye, Andrée, Heron-Longe, Bénédicte, Pipiras, Eva, Benzacken, Brigitte, Passemard, Sandrine, Verloes, Alain, Isidor, Bertrand, Le Caignec, Cedric, Glew, Gwen M, Opheim, Kent E, Descartes, Maria, Eichler, Evan E, Morton, Cynthia C, Gusella, James F, Schultz, Roger A, Ballif, Blake C, & Shaffer, Lisa G. 2012. Haploinsufficiency of SOX5 at 12p12.1 is associated with developmental delays with prominent language delay, behavior problems, and mild dysmorphic features. *Human mutation*, **33**(4), 728–740.

Lambert, Nelle, Wermenbol, Vanessa, Pichon, Bruno, Acosta, Sandra, van den Ameele, Jelle, Perazzolo, Camille, Messina, Diana, Musumeci, Maria-Franca, Dessars, Barbara, De Leener, Anne, Abramowicz, Marc, & Vilain, Catheline. 2014. A familial heterozygous null mutation of MET in autism spectrum disorder. *Autism research : official journal of the International Society for Autism Research*, **7**(5), 617–622.

Le Fevre, Anna K, Taylor, Sharelle, Malek, Neva H, Horn, Denise, Carr, Christopher W, Abdul-Rahman, Omar A, O'Donnell, Sherindan, Burgess, Trent, Shaw, Marie, Gecz, Jozef, Bain, Nicole, Fagan, Kerry, & Hunter, Matthew F. 2013. FOXP1 mutations cause intellectual disability and a recognizable phenotype. *American journal of medical genetics. Part A*, **161A**(12), 3166–3175.

Lee, Florence Y, Faivre, Emily J, Suzawa, Miyuki, Lontok, Erik, Ebert, Daniel, Cai, Fang, Belsham, Denise D, & Ingraham, Holly A. 2011. Eliminating SF-1 (NR5A1) sumoylation in vivo results in ectopic hedgehog signaling and disruption of endocrine development. *Developmental cell*, **21**(2), 315–327.

Leid, Mark, Ishmael, Jane E, Avram, Dorina, Shepherd, David, Fraulob, Valérie, & Dollé, Pascal. 2004. CTIP1 and CTIP2 are differentially expressed during mouse embryogenesis. *Gene expression patterns : GEP*, **4**(6), 733–739.

Lennertz, Leonhard, Rujescu, Dan, Wagner, Michael, Frommann, Ingo, Schulze-Rauschenbach, Svenja, Schuhmacher, Anna, Landsberg, Martin W, Franke, Petra, Möller, Hans-Jürgen, Wölwer, Wolfgang, Gaebel, Wolfgang, Häfner, Heinz, Maier, Wolfgang, & Mössner, Rainald. 2011. Novel schizophrenia risk gene TCF4 influences verbal learning and memory functioning in schizophrenia patients. *Neuropsychobiology*, **63**(3), 131–136.

Lennon, P A, Cooper, M L, Peiffer, D A, Gunderson, K L, Patel, A, Peters, Sarika, Cheung, S W, & Bacino, C A. 2007. Deletion of 7q31.1 supports involvement of FOXP2 in language impairment: clinical report and review. *American journal of medical genetics. Part A*, **143A**(8), 791–798.

Levinson, Stephen C, & Holler, Judith. 2014. The origin of human multi-modal communication. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **369**(1651), 20130302.

Lévy, Jonathan, Coussement, Aurélie, Dupont, Céline, Guimiot, Fabien, Baumann, Clarisse, Viot, Géraldine, Passemard, Sandrine, Capri, Yline, Drunat, Séverine, Verloes, Alain, Pipiras, Eva, Benzacken, Brigitte, Dupont, Jean-Michel, & Tabet, Anne-Claude. 2017. Molecular and clinical delineation of 2p15p16.1 microdeletion syndrome. *American journal of medical genetics. Part A*, **173**(8), 2081–2087.

Lewis, B A, & Thompson, L A. 1992. A study of developmental speech and language disorders in twins. *Journal of speech and hearing research*, **35**(5), 1086–1094.

Lewis, Barbara A, Freebairn, Lisa A, Hansen, Amy J, Miscimarra, Lara, Iyengar, Sudha K, & Taylor, H Gerry. 2007. Speech and language skills of parents of children with speech sound disorders. *American journal of speech-language pathology*, **16**(2), 108–118.

Li, S. 2004. Advanced Cardiac Morphogenesis Does Not Require Heart Tube Fusion. *Science*, **305**(5690), 1619–1622.

Li, Shanru, Weidenfeld, Joel, & Morrisey, Edward E. 2004. Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions. *Molecular and cellular biology*, **24**(2), 809–822.

Li, X, Wang, W, Wang, J, Malovannaya, A, Xi, Y, Li, W, Guerra, R, Hawke, D H, Qin, J, & Chen, J. 2015a. Proteomic analyses reveal distinct chromatin-associated and soluble transcription factor complexes. *Molecular Systems Biology*, **11**(1), 775–775.

Li, Xue, Xiao, Jian, Fröhlich, Henning, Tu, Xiaomeng, Li, Lianlian, Xu, Yue, Cao, Huateng, Qu, Jia, Rappold, Gudrun A, & Chen, Jie-Guang. 2015b. Foxp1 regulates cortical radial migration and neuronal morphogenesis in developing cerebral cortex. *PloS one*, **10**(5), e0127671.

Liang, J-S, Shimojima, K, Sugiura, C, Une, Y, Ohno, K, & Yamamoto, T. 2009. A newly recognised microdeletion syndrome of 2p15-16.1 manifesting moderate developmental delay, autistic behaviour, short stature, microcephaly, and dysmorphic features: a new patient with 3.2 Mb deletion. *Journal of medical genetics*, **46**(9), 645–647.

Liégeois, F, Morgan, A T, Connelly, A, & Vargha-Khadem, F. 2011. Endophenotypes of FOXP2: dysfunction within the human articulatory network. *European journal of paediatric neurology : EJPN : official journal of the European Paediatric Neurology Society*, **15**(4), 283–288.

Liégeois, Frédérique, Baldeweg, Torsten, Connelly, Alan, Gadian, David G, Mishkin, Mortimer, & Vargha-Khadem, Faraneh. 2003. Language fMRI abnormalities associated with FOXP2 gene mutation. *Nature neuroscience*, **6**(11), 1230–1237.

Liu, Bin, Mink, Sheldon, Wong, Kelly A, Stein, Natalie, Getman, Crescent, Dempsey, Paul W, Wu, Hong, & Shuai, Ke. 2004. PIAS1 selectively inhibits interferon-inducible genes and is important in innate immunity. *Nature immunology*, **5**(9), 891–898.

Liu, Hui, Ippolito, Gregory C, Wall, Jason K, Niu, Teresa, Probst, Loren, Lee, Baeck-Seung, Pulford, Karen, Banham, Alison H, Stockwin, Luke, Shaffer, Arthur L, Staudt,

Louis M, Das, Chhaya, Dyer, Martin J S, & Tucker, Philip W. 2006. Functional studies of BCL11A: characterization of the conserved BCL11A-XL splice variant and its interaction with BCL6 in nuclear paraspeckles of germinal center B cells. *Molecular cancer*, **5**(1), 18.

Liu, Pentao, Keller, Jonathan R, Ortiz, Mariaestela, Tessarollo, Lino, Rachel, Rivka A, Nakamura, Takuro, Jenkins, Nancy A, & Copeland, Neal G. 2003. Bcl11a is essential for normal lymphoid development. *Nature immunology*, **4**(6), 525–532.

Loriol, Céline, Parisot, Joséphine, Poupon, Gwénola, Gwizdek, Carole, & Martin, Stéphane. 2012. Developmental regulation and spatiotemporal redistribution of the sumoylation machinery in the rat central nervous system. *PloS one*, **7**(3), e33757.

Lozano, Reymundo, Vino, Arianna, Lozano, Cristina, Fisher, Simon E, & Deriziotis, Pelagia. 2015. A de novo FOXP1 variant in a patient with autism, intellectual disability and severe speech and language impairment. *European journal of human genetics : EJHG*, **23**(12), 1702–1707.

Lu, Min Min, Li, Shanru, Yang, Honghua, & Morrisey, Edward E. 2002. Foxp4: a novel member of the Foxp subfamily of winged-helix genes co-expressed with Foxp1 and Foxp2 in pulmonary and gut tissues. *Mechanisms of development*, **119 Suppl 1**(Dec.), S197–202.

Lumish, Heidi S, Wynn, Julia, Devinsky, Orrin, & Chung, Wendy K. 2015. Brief Report: SETD2 Mutation in a Child with Autism, Intellectual Disabilities and Epilepsy. *Journal of autism and developmental disorders*, **45**(11), 3764–3770.

Luscan, Armelle, Laurendeau, Ingrid, Malan, Valérie, Francannet, Christine, Odent, Sylvie, Giuliano, Fabienne, Lacombe, Didier, Touraine, Renaud, Vidaud, Michel, Pasmant, Eric, & Cormier-Daire, Valérie. 2014. Mutations in SETD2 cause a novel overgrowth condition. *Journal of medical genetics*, **51**(8), 512–517.

Ma, Mandy, Adams, Heather R, Seltzer, Laurie E, Dobyns, William B, & Paciorkowski, Alex R. 2016. Phenotype Differentiation of FOXG1 and MECP2 Disorders: A New Method for Characterization of Developmental Encephalopathies. *The Journal of pediatrics*, **178**(Nov.), 233–240.e10.

MacDermot, Kay D, Bonora, Elena, Sykes, Nuala, Coupe, Anne-Marie, Lai, Cecilia S L, Vernes, Sonja C, Vargha-Khadem, Faraneh, McKenzie, Fiona, Smith, Robert L, Monaco, Anthony P, & Fisher, Simon E. 2005. Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. *American journal of human genetics*, **76**(6), 1074–1080.

Maricic, Tomislav, Günther, Viola, Georgiev, Oleg, Gehre, Sabine, Curlin, Marija, Schreiweis, Christiane, Naumann, Ronald, Burbano, Hernán A, Meyer, Matthias, Lalueza-Fox, Carles, de la Rasilla, Marco, Rosas, Antonio, Gajovic, Srecko, Kelso, Janet, Enard, Wolfgang, Schaffner, Walter, & Pääbo, Svante. 2013. A recent evolutionary change affects a regulatory element in the human FOXP2 gene. *Molecular biology and evolution*, **30**(4), 844–852.

Mariën, Peter, Ackermann, Herman, Adamaszek, Michael, Barwood, Caroline H S, Beaton, Alan, Desmond, John, De Witte, Elke, Fawcett, Angela J, Hertrich, Ingo, Küper, Michael, Leggio, Maria, Marvel, Cherie, Molinari, Marco, Murdoch, Bruce E, Nicolson, Roderick I, Schmahmann, Jeremy D, Stoodley, Catherine J, Thürling, Markus, Timmann, Dagmar, Wouters, Ellen, & Ziegler, Wolfram. 2014 (June). Consensus paper: Language and the cerebellum: an ongoing enigma. *Pages 386–410 of: Cerebellum (London, England)*. Department of Clinical and Experimental Neurolinguistics, CLIN, Vrije Universiteit Brussel, Brussels, Belgium, peter.marien5@telenet.be.

Marongiu, Mara, Deiana, Manila, Meloni, Alessandra, Marcia, Loredana, Puddu, Alessandro, Cao, Antonio, Schlessinger, David, & Crisponi, Laura. 2010. The forkhead transcription factor Foxl2 is sumoylated in both human and mouse: sumoylation affects its stability, localization, and activity. *PloS one*, **5**(3), e9477.

McKenna, William L, Ortiz-Londono, Christian F, Mathew, Thomas K, Hoang, Kendy, Katzman, Sol, & Chen, Bin. 2015. Mutual regulation between Satb2 and Fezf2 promotes subcerebral projection neuron identity in the developing cerebral cortex. *Proceedings of the National Academy of Sciences of the United States of America*, **112**(37), 11702–11707.

Meerschaut, Ilse, Rochefort, Daniel, Revencu, Nicole, Petre, Justine, Corsello, Christina, Rouleau, Guy A, Hamdan, Fadi F, Michaud, Jacques L, Morton, Jenny, Radley, Jessica, Ragge, Nicola, Garcia-Minaur, Sixto, Lapunzina, Pablo, Bralo, Maria Palomares, Mori, Maria Angeles, Moortgat, Stephanie, Benoit, Valerie, Mary, Sandrine, Bockaert, Nele, Oostra, Ann, Vanakker, Olivier, Velinov, Milen, de Ravel, Thomy Jl, Mekahli, Djalila, Sebat, Jonathan, Vaux, Keith K, DiDonato, Nataliya, Hanson-Kahn, Andrea K, Hudgins, Louanne, Dallapiccola, Bruno, Novelli, Antonio, Tarani, Luigi, Andrieux, Joris, Parker, Michael J, Neas, Katherine, Ceulemans, Berten, Schoonjans, An-Sofie, Prchalova, Darina, Havlovicova, Marketa, Hancarova, Miroslava, Budişteanu, Magdalena, Dheedene, Annelies, Menten, Bjorn, Dion, Patrick A, Lederer, Damien, & Callewaert, Bert. 2017. FOXP1-related intellectual disability syndrome: a recognisable entity. *Journal of medical genetics*, **54**(9), 613–623.

Mellacheruvu, Dattatreya, Wright, Zachary, Couzens, Amber L, Lambert, Jean-Philippe, St-Denis, Nicole A, Li, Tuo, Miteva, Yana V, Hauri, Simon, Sardiu, Mihaela E, Low,

Teck Yew, Halim, Vincentius A, Bagshaw, Richard D, Hubner, Nina C, Al-Hakim, Abdallah, Bouchard, Annie, Faubert, Denis, Fermin, Damian, Dunham, Wade H, Goudreault, Marilyn, Lin, Zhen-Yuan, Badillo, Beatriz Gonzalez, Pawson, Tony, Durocher, Daniel, Coulombe, Benoit, Aebersold, Ruedi, Superti-Furga, Giulio, Colinge, Jacques, Heck, Albert J R, Choi, Hyungwon, Gstaiger, Matthias, Mohammed, Shabaz, Cristea, Ileana M, Bennett, Keiryn L, Washburn, Mike P, Raught, Brian, Ewing, Rob M, Gingras, Anne-Claude, & Nesvizhskii, Alexey I. 2013. The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nature methods*, **10**(8), 730–736.

Mendoza, Ezequiel, Tokarev, Kirill, Düring, Daniel N, Retamosa, Eva Camarillo, Weiss, Michael, Arpenik, Nshdejan, & Scharff, Constance. 2015. Differential coexpression of FoxP1, FoxP2, and FoxP4 in the Zebra Finch (Taeniopygia guttata) song system. *The Journal of comparative neurology*, **523**(9), 1318–1340.

Meng, Haiying, Smith, Shelley D, Hager, Karl, Held, Matthew, Liu, Jonathan, Olson, Richard K, Pennington, Bruce F, DeFries, John C, Gelernter, Joel, O'Reilly-Pol, Thomas, Somlo, Stefan, Skudlarski, Pawel, Shaywitz, Sally E, Shaywitz, Bennett A, Marchione, Karen, Wang, Yu, Paramasivam, Murugan, LoTurco, Joseph J, Page, Grier P, & Gruen, Jeffrey R. 2005. DCDC2 is associated with reading disability and modulates neuronal development in the brain. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(47), 17053–17058.

Meredith, Leslie J, Wang, Chiung-Min, Nascimento, Leticia, Liu, Runhua, Wang, Lizhong, & Yang, Wei-Hsiung. 2015. The Key Regulator for Language and Speech Development, FOXP2, is a Novel Substrate for SUMOylation. *Journal of Cellular Biochemistry*, **117**(2), 426–438.

Meulmeester, Erik, & Melchior, Frauke. 2008. Cell biology: SUMO. *Nature*, **452**(7188), 709–711.

Miller, Julie E, Spiteri, Elizabeth, Condro, Michael C, Dosumu-Johnson, Ryan T, Geschwind, Daniel H, & White, Stephanie A. 2008. Birdsong decreases protein levels of FoxP2, a molecule required for human speech. *Journal of neurophysiology*, **100**(4), 2015–2025.

Mircsof, Dennis, Langouët, Maéva, Rio, Marlène, Moutton, Sébastien, Siquier-Pernet, Karine, Bole-Feysot, Christine, Cagnard, Nicolas, Nitschke, Patrick, Gaspar, Ludmila, Žnidarič, Matej, Alibeu, Olivier, Fritz, Ann-Kristina, Wolfer, David P, Schröter, Aileen, Bosshard, Giovanna, Rudin, Markus, Koester, Christina, Crestani, Florence, Seebeck, Petra, Boddaert, Nathalie, Prescott, Katrina, DDD Study, Hines, Rochelle, Moss, Steven J, Fritschy, Jean-Marc, Munnich, Arnold, Amiel, Jeanne, Brown, Steven A,

Tyagarajan, Shiva K, & Colleaux, Laurence. 2015. Mutations in NONO lead to syndromic intellectual disability and inhibitory synaptic defects. *Nature neuroscience*, **18**(12), 1731–1736.

Mittag, Tanja, Kay, Lewis E, & Forman-Kay, Julie D. 2010. Protein dynamics and conformational disorder in molecular recognition. *Journal of molecular recognition : JMR*, **23**(2), 105–116.

Mizutani, Akifumi, Matsuzaki, Ayumi, Momoi, Mariko Y, Fujita, Eriko, Tanabe, Yuko, & Momoi, Takashi. 2007. Intracellular distribution of a speech/language disorder associated FOXP2 mutant. *Biochemical and biophysical research communications*, **353**(4), 869–874.

Morgan, Angela, Fisher, Simon E, Scheffer, Ingrid, & Hildebrand, Michael. 2016. FOXP2-Related Speech and Language Disorders. *Gene Reviews*.

Morikawa, Yoshihiro, Komori, Tadasuke, Hisaoka, Tomoko, & Senba, Emiko. 2009. Detailed Expression Pattern of Foxp1 and Its Possible Roles in Neurons of the Spinal Cord during Embryogenesis. *Developmental Neuroscience*, **31**(6), 511–522.

Morimoto, Marie, & Boerkoel, Cornelius F. 2013. The role of nuclear bodies in gene expression and disease. *Biology*, **2**(3), 976–1033.

Mukamel, Zohar, Konopka, Genevieve, Wexler, Eric, Osborn, Gregory E, Dong, Hongmei, Bergman, Mica Y, Levitt, Pat, & Geschwind, Daniel H. 2011. Regulation of MET by FOXP2, genes implicated in higher cognitive dysfunction and autism risk. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, **31**(32), 11437–11442.

Murugan, Malavika, Harward, Stephen, Scharff, Constance, & Mooney, Richard. 2013. Diminished FoxP2 levels affect dopaminergic modulation of corticostriatal signaling important to song variability. *Neuron*, **80**(6), 1464–1476.

Nacerddine, Karim, Lehembre, François, Bhaumik, Mantu, Artus, Jérôme, Cohen-Tannoudji, Michel, Babinet, Charles, Pandolfi, Pier Paolo, & Dejean, Anne. 2005. The SUMO pathway is essential for nuclear integrity and chromosome segregation in mice. *Developmental cell*, **9**(6), 769–779.

Nazaryan, Lusine, Stefanou, Eunice G, Hansen, Claus, Kosyakova, Nadezda, Bak, Mads, Sharkey, Freddie H, Mantziou, Theodora, Papanastasiou, Anastasios D, Velissariou, Voula, Liehr, Thomas, Syrrou, Maria, & Tommerup, Niels. 2014. The strength of combined cytogenetic and mate-pair sequencing techniques illustrated by a germline chromothripsis rearrangement involving FOXP2. *European journal of human genetics : EJHG*, **22**(3), 338–343.

Neale, Benjamin M, Kou, Yan, Liu, Li, Ma'ayan, Avi, Samocha, Kaitlin E, Sabo, Aniko, Lin, Chiao-Feng, Stevens, Christine, Wang, Li-San, Makarov, Vladimir, Polak, Paz, Yoon, Seungtai, Maguire, Jared, Crawford, Emily L, Campbell, Nicholas G, Geller, Evan T, Valladares, Otto, Schafer, Chad, Liu, Han, Zhao, Tuo, Cai, Guiqing, Lihm, Jayon, Dannenfelser, Ruth, Jabado, Omar, Peralta, Zuleyma, Nagaswamy, Uma, Muzny, Donna, Reid, Jeffrey G, Newsham, Irene, Wu, Yuanqing, Lewis, Lora, Han, Yi, Voight, Benjamin F, Lim, Elaine, Rossin, Elizabeth, Kirby, Andrew, Flannick, Jason, Fromer, Menachem, Shakir, Khalid, Fennell, Tim, Garimella, Kiran, Banks, Eric, Poplin, Ryan, Gabriel, Stacey, DePristo, Mark, Wimbish, Jack R, Boone, Braden E, Levy, Shawn E, Betancur, Catalina, Sunyaev, Shamil, Boerwinkle, Eric, Buxbaum, Joseph D, Cook, Edwin H, Devlin, Bernie, Gibbs, Richard A, Roeder, Kathryn, Schellenberg, Gerard D, Sutcliffe, James S, & Daly, Mark J. 2012. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*, **485**(7397), 242–245.

Neils, J, & Aram, D M. 1986. Family history of children with developmental language disorders. *Perceptual and motor skills*, **63**(2 Pt 1), 655–658.

Nelson, Christopher S, Fuller, Chris K, Fordyce, Polly M, Greninger, Alexander L, Li, Hao, & DeRisi, Joseph L. 2013. Microfluidic affinity and ChIP-seq analyses converge on a conserved FOXP2-binding motif in chimp and human, which enables the detection of evolutionarily novel targets. *Nucleic acids research*, **41**(12), 5991–6004.

Nesbitt, Addie, Bhoj, Elizabeth J, McDonald Gibson, Kristin, Yu, Zhenming, Denenberg, Elizabeth, Sarmady, Mahdi, Tischler, Tanya, Cao, Kajia, Dubbs, Holly, Zackai, Elaine H, & Santani, Avni. 2015. Exome sequencing expands the mechanism of SOX5-associated intellectual disability: A case presentation with review of sox-related disorders. *American journal of medical genetics. Part A*, **167A**(11), 2548–2554.

Newbury, D F, & Monaco, A P. 2010. Genetic advances in the study of speech and language disorders. *Neuron*, **68**(2), 309–320.

Newbury, Dianne F, & Monaco, Anthony P. 2002. Molecular genetics of speech and language disorders. *Current opinion in pediatrics*, **14**(6), 696–701.

Newbury, Dianne F, Winchester, Laura, Addis, Laura, Paracchini, Silvia, Buckingham, Lyn-Louise, Clark, Ann, Cohen, Wendy, Cowie, Hilary, Dworzynski, Katharina, Everitt, Andrea, Goodyer, Ian M, Hennessy, Elizabeth, Kindley, A David, Miller, Laura L, Nasir, Jamal, O'Hare, Anne, Shaw, Duncan, Simkin, Zoe, Simonoff, Emily, Slonims, Vicky, Watson, Jocelynne, Ragoussis, Jiannis, Fisher, Simon E, Seckl, Jonathon R, Helms, Peter J, Bolton, Patrick F, Pickles, Andrew, Conti-Ramsden, Gina, Baird, Gillian, Bishop, Dorothy V M, & Monaco, Anthony P. 2009. CMIP and ATP2C2

modulate phonological short-term memory in language impairment. *American journal of human genetics*, **85**(2), 264–272.

Ninkovic, Jovica, Steiner-Mezzadri, Andrea, Jawerka, Melanie, Akinci, Umut, Masserdotti, Giacomo, Petricca, Stefania, Fischer, Judith, von Holst, Alexander, Beckers, Johanes, Lie, Chichung D, Petrik, David, Miller, Erik, Tang, Jiong, Wu, Jiang, Lefebvre, Véronique, Demmers, Jeroen, Eisch, Amelia, Metzger, Daniel, Crabtree, Gerald, Irmler, Martin, Poot, Raymond, & Götz, Magdalena. 2013. The BAF complex interacts with Pax6 in adult neural progenitors to establish a neurogenic cross-regulatory transcriptional network. *Cell stem cell*, **13**(4), 403–418.

Norbury, Courtenay Frazier, Gooch, Debbie, Wray, Charlotte, Baird, Gillian, Charman, Tony, Simonoff, Emily, Vamvakas, George, & Pickles, Andrew. 2016. The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study. *Journal of child psychology and psychiatry, and allied disciplines*, **57**(11), 1247–1257.

Notwell, James H, Heavner, Whitney E, Darbandi, Siavash Fazel, Katzman, Sol, McKenna, William L, Ortiz-Londono, Christian F, Tastad, David, Eckler, Matthew J, Rubenstein, John L R, McConnell, Susan K, Chen, Bin, & Bejerano, Gill. 2016. TBR1 regulates autism risk genes in the developing neocortex. *Genome research*, **26**(8), 1013–1022.

O'Roak, Brian J, Deriziotis, Pelagia, Lee, Choli, Vives, Laura, Schwartz, Jerrod J, Girirajan, Santhosh, Karakoc, Emre, Mackenzie, Alexandra P, Ng, Sarah B, Baker, Carl, Rieder, Mark J, Nickerson, Deborah A, Bernier, Raphael, Fisher, Simon E, Shendure, Jay, & Eichler, Evan E. 2011. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nature genetics*, **43**(6), 585–589.

Pal, Deb K. 2011. Epilepsy and neurodevelopmental disorders of language. *Current opinion in neurology*, **24**(2), 126–131.

Palka, Chiara, Alfonsi, Melissa, Mohn, Angelika, Cerbo, Renato, Guanciali Franchi, Paolo, Fantasia, Donatella, Morizio, Elisena, Stuppia, Liborio, Calabrese, Giuseppe, Zori, Roberto, Chiarelli, Francesco, & Palka, Giandomenico. 2012. Mosaic 7q31 deletion involving FOXP2 gene associated with language impairment. *Pediatrics*, **129**(1), e183–8.

Palmesino, Elena, Rousso, David L, Kao, Tzu-Jen, Klar, Avihu, Laufer, Ed, Uemura, Osamu, Okamoto, Hitoshi, Novitch, Bennett G, & Kania, Artur. 2010. Foxp1 and lhx1 coordinate motor neuron migration with axon trajectory choice by gating Reelin signalling. *PLoS biology*, **8**(8), e1000446.

Pariani, Mitchel J, Spencer, Andrew, Graham, John M, & Rimoin, David L. 2009. A 785kb deletion of 3p14.1p13, including the FOXP1 gene, associated with speech delay, contractures, hypertonia and blepharophimosis. *European journal of medical genetics*, **52**(2-3), 123–127.

Park, Yoojin, Lee, Ja-Myong, Hwang, Min-Young, Son, Gi-hoon, & Geum, Dongho. 2013. NonO binds to the CpG island of oct4 promoter and functions as a transcriptional activator of oct4 gene expression. *Molecules and cells*, **35**(1), 61–69.

Patwardhan, Rupali P, Hiatt, Joseph B, Witten, Daniela M, Kim, Mee J, Smith, Robin P, May, Dalit, Lee, Choli, Andrie, Jennifer M, Lee, Su-In, Cooper, Gregory M, Ahituv, Nadav, Pennacchio, Len A, & Shendure, Jay. 2012. Massively parallel functional dissection of mammalian enhancers in vivo. *Nature biotechnology*, **30**(3), 265–270.

Peñagarikano, Olga, Abrahams, Brett S, Herman, Edward I, Winden, Kellen D, Gdalyahu, Amos, Dong, Hongmei, Sonnenblick, Lisa I, Gruver, Robin, Almajano, Joel, Bragin, Anatol, Golshani, Peyman, Trachtenberg, Joshua T, Peles, Elior, & Geschwind, Daniel H. 2011. Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell*, **147**(1), 235–246.

Peng, Yun, Huentelman, Matthew, Smith, Christopher, & Qiu, Shenfeng. 2013. MET receptor tyrosine kinase as an autism genetic risk factor. *International review of neurobiology*, **113**, 135–165.

Petek, Erwin, Windpassinger, Christian, Simma, Burkhard, Mueller, Thomas, Wagner, Klaus, & Kroisel, Peter M. 2003. Molecular characterisation of a 15 Mb constitutional de novo interstitial deletion of chromosome 3p in a boy with developmental delay and congenital anomalies. *Journal of human genetics*, **48**(6), 283–287.

Peter, Beate, Matsushita, Mark, Oda, Kaori, & Raskind, Wendy. 2014. De novo microdeletion of BCL11A is associated with severe speech sound disorder. *American journal of medical genetics. Part A*, **164A**(8), 2091–2096.

Piccione, Maria, Piro, Ettore, Serraino, Francesca, Cavani, Simona, Ciccone, Roberto, Malacarne, Michela, Pierluigi, Mauro, Vitaloni, Marianna, Zuffardi, Orsetta, & Corsello, Giovanni. 2012. Interstitial deletion of chromosome 2p15-16.1: report of two patients and critical review of current genotype-phenotype correlation. *European journal of medical genetics*, **55**(4), 238–244.

Poot, Martin. 2015. Connecting the CNTNAP2 Networks with Neurodevelopmental Disorders. *Molecular syndromology*, **6**(1), 7–22.

Precious, S V, Kelly, C M, Reddington, A E, Vinh, N N, Stickland, R C, Pekarik, V, Scherf, C, Jeyasingham, R, Glasbey, J, Holeiter, M, Jones, L, Taylor, M V, & Rosser, A E. 2016. FoxP1 marks medium spiny neurons from precursors to maturity and is required for their differentiation. *Experimental neurology*, **282**(Aug.), 9–18.

Prontera, Paolo, Bernardini, Laura, Stangoni, Gabriela, Capalbo, Anna, Rogaia, Daniela, Romani, Rita, Ardisia, Carmela, Dallapiccola, Bruno, & Donti, Emilio. 2011. Deletion 2p15-16.1 syndrome: case report and review. *American journal of medical genetics. Part A*, **155A**(10), 2473–2478.

Rajcan-Separovic, E, Harvard, C, Liu, X, McGillivray, B, Hall, J G, Qiao, Y, Hurlburt, J, Hildebrand, J, Mickelson, E C R, Holden, J J A, & Lewis, M E S. 2007. Clinical and molecular cytogenetic characterisation of a newly recognised microdeletion syndrome involving 2p15-16.1. *Journal of medical genetics*, **44**(4), 269–276.

Ramsay, Andrew J, Quesada, Víctor, Foronda, Miguel, Conde, Laura, Martínez-Trillos, Alejandra, Villamor, Neus, Rodríguez, David, Kwarciak, Agnieszka, Garabaya, Cecilia, Gallardo, Mercedes, López-Guerra, Mónica, López-Guillermo, Armando, Puente, Xose S, Blasco, Maria A, Campo, Elías, & López-Otín, Carlos. 2013. POT1 mutations cause telomere dysfunction in chronic lymphocytic leukemia. *Nature genetics*, **45**(5), 526–530.

Rauscher, Benedikt, Heigwer, Florian, Breinig, Marco, Winter, Jan, & Boutros, Michael. 2017. GenomeCRISPR - a database for high-throughput CRISPR/Cas9 screens. *Nucleic acids research*, **45**(D1), D679–D686.

Ravasi, Timothy, Suzuki, Harukazu, Cannistraci, Carlo Vittorio, Katayama, Shintaro, Bajic, Vladimir B, Tan, Kai, Akalin, Altuna, Schmeier, Sebastian, Kanamori-Katayama, Mutsumi, Bertin, Nicolas, Carninci, Piero, Daub, Carsten O, Forrest, Alistair R R, Gough, Julian, Grimmond, Sean, Han, Jung-Hoon, Hashimoto, Takehiro, Hide, Winston, Hofmann, Oliver, Kamburov, Atanas, Kaur, Mandeep, Kawaji, Hideya, Kubosaki, Atsutaka, Lassmann, Timo, van Nimwegen, Erik, MacPherson, Cameron Ross, Ogawa, Chihiro, Radovanovic, Aleksandar, Schwartz, Ariel, Teasdale, Rohan D, Tegnér, Jesper, Lenhard, Boris, Teichmann, Sarah A, Arakawa, Takahiro, Ninomiya, Noriko, Murakami, Kayoko, Tagami, Michihira, Fukuda, Shiro, Imamura, Kengo, Kai, Chikatoshi, Ishihara, Ryoko, Kitazume, Yayoi, Kawai, Jun, Hume, David A, Ideker, Trey, & Hayashizaki, Yoshihide. 2010. An atlas of combinatorial transcriptional regulation in mouse and man. *Cell*, **140**(5), 744–752.

Reader, Rose H, Covill, Laura E, Nudel, Ron, & Newbury, Dianne F. 2014. Genome-Wide Studies of Specific Language Impairment. *Current behavioral neuroscience reports*, **1**(4), 242–250.

Reimers-Kipping, S, Hevers, W, Pääbo, S, & Enard, W. 2011. Humanized Foxp2 specifically affects cortico-basal ganglia circuits. *Neuroscience*, **175**(Feb.), 75–84.

Reinstein, Eyal, Tzur, Shay, Cohen, Rony, Bormans, Concetta, & Behar, Doron M. 2016. Intellectual disability and non-compaction cardiomyopathy with a de novo NONO mutation identified by exome sequencing. *European journal of human genetics : EJHG*, **24**(11), 1635–1638.

Reuter, Miriam S, Riess, Angelika, Moog, Ute, Briggs, Tracy A, Chandler, Kate E, Rauch, Anita, Stampfer, Miriam, Steindl, Katharina, Gläser, Dieter, Joset, Pascal, DDD Study, Krumbiegel, Mandy, Rabe, Harald, Schulte-Mattler, Uta, Bauer, Peter, Beck-Wödl, Stefanie, Kohlhase, Jürgen, Reis, André, & Zweier, Christiane. 2016. FOXP2 variants in 14 individuals with developmental speech and language disorders broaden the mutational and clinical spectrum. *Journal of medical genetics*, Aug., jmedgenet–2016–104094.

Rice, Gregory M, Raca, Gordana, Jakielski, Kathy J, Laffin, Jennifer J, Iyama-Kurtycz, Christina M, Hartley, Sigan L, Sprague, Rae E, Heintzelman, Anne T, & Shriberg, Lawrence D. 2012. Phenotype of FOXP2 haploinsufficiency in a mother and son. *American journal of medical genetics. Part A*, **158A**(1), 174–181.

Rocca, Daniel L, Wilkinson, Kevin A, & Henley, Jeremy M. 2017. SUMOylation of FOXP1 regulates transcriptional repression via CtBP1 to drive dendritic morphogenesis. *Scientific reports*, **7**(1), 877.

Rochefort, Christelle, He, Xiaolu, Scotto-Lomassese, Sophie, & Scharff, Constance. 2007. Recruitment of FoxP2-expressing neurons to area X varies during song development. *Developmental neurobiology*, **67**(6), 809–817.

Rodenas-Cuadrado, Pedro, Ho, Joses, & Vernes, Sonja C. 2014. Shining a light on CNTNAP2: complex functions to complex disorders. *European journal of human genetics : EJHG*, **22**(2), 171–178.

Rodenas-Cuadrado, Pedro, Pietrafusa, Nicola, Francavilla, Teresa, La Neve, Angela, Striano, Pasquale, & Vernes, Sonja C. 2016. Characterisation of CASPR2 deficiency disorder–a syndrome involving autism, epilepsy and language impairment. *BMC medical genetics*, **17**(1), 8.

Roll, Patrice, Rudolf, Gabrielle, Pereira, Sandrine, Royer, Barbara, Scheffer, Ingrid E, Massacrier, Annick, Valenti, Maria-Paola, Roeckel-Trevisiol, Nathalie, Jamali, Sarah, Beclin, Christophe, Seegmuller, Caroline, Metz-Lutz, Marie-Noëlle, Lemainque, Arnaud, Delepine, Marc, Caloustian, Christophe, de Saint Martin, Anne, Bruneau, Nadine,

Depétris, Danièle, Mattéi, Marie-Geneviève, Flori, Elisabeth, Robaglia-Schlupp, Andrée, Lévy, Nicolas, Neubauer, Bernd A, Ravid, Rivka, Marescaux, Christian, Berkovic, Samuel F, Hirsch, Edouard, Lathrop, Mark, Cau, Pierre, & Szepetowski, Pierre. 2006. SRPX2 mutations in disorders of language cortex and cognition. *Human molecular genetics*, **15**(7), 1195–1207.

Roll, Patrice, Vernes, Sonja C, Bruneau, Nadine, Cillario, Jennifer, Ponsole-Lenfant, Magali, Massacrier, Annick, Rudolf, Gabrielle, Khalife, Manal, Hirsch, Edouard, Fisher, Simon E, & Szepetowski, Pierre. 2010. Molecular networks implicated in speech-related disorders: FOXP2 regulates the SRPX2/uPAR complex. *Human molecular genetics*, **19**(24), 4848–4860.

Rolland, Thomas, Tasan, Murat, Charloteaux, Benoit, Pevzner, Samuel J, Zhong, Quan, Sahni, Nidhi, Yi, Song, Lemmens, Irma, Fontanillo, Celia, Mosca, Roberto, Kamburov, Atanas, Ghiassian, Susan D, Yang, Xinping, Ghamsari, Lila, Balcha, Dawit, Begg, Bridget E, Braun, Pascal, Brehme, Marc, Broly, Martin P, Carvunis, Anne-Ruxandra, Convery-Zupan, Dan, Corominas, Roser, Coulombe-Huntington, Jasmin, Dann, Elizabeth, Dreze, Matija, Dricot, Amélie, Fan, Changyu, Franzosa, Eric, Gebreab, Fana, Gutierrez, Bryan J, Hardy, Madeleine F, Jin, Mike, Kang, Shuli, Kiros, Ruth, Lin, Guan Ning, Luck, Katja, MacWilliams, Andrew, Menche, Jörg, Murray, Ryan R, Palagi, Alexandre, Poulin, Matthew M, Rambout, Xavier, Rasla, John, Reichert, Patrick, Romero, Viviana, Ruyssinck, Elien, Sahalie, Julie M, Scholz, Annemarie, Shah, Akash A, Sharma, Amitabh, Shen, Yun, Spirohn, Kerstin, Tam, Stanley, Tejeda, Alexander O, Trigg, Shelly A, Twizere, Jean-Claude, Vega, Kerwin, Walsh, Jennifer, Cusick, Michael E, Xia, Yu, Barabási, Albert-László, Iakoucheva, Lilia M, Aloy, Patrick, De Las Rivas, Javier, Tavernier, Jan, Calderwood, Michael A, Hill, David E, Hao, Tong, Roth, Frederick P, & Vidal, Marc. 2014. A proteome-scale map of the human interactome network. *Cell*, **159**(5), 1212–1226.

Rousso, David L, Pearson, Caroline Alayne, Gaber, Zachary B, Miquelajauregui, Amaya, Li, Shanru, Portera-Cailliau, Carlos, Morrisey, Edward E, & Novitch, Bennett G. 2012. Foxp-mediated suppression of N-cadherin regulates neuroepithelial character and progenitor maintenance in the CNS. *Neuron*, **74**(2), 314–330.

Rytinki, Miia M, Kaikkonen, Sanna, Pehkonen, Petri, Jaaskelainen, Tiina, & Palvimo, Jorma J. 2009. PIAS proteins: pleiotropic interactors associated with SUMO. *Cellular and molecular life sciences : CMLS*, **66**(18), 3029–3041.

Sainz, J. 1993. Human cDNA libraries and brain banks. *Journal of neural transmission. Supplementum*, **39**, 135–141.

Sakai, Yasunari, Shaw, Chad A, Dawson, Brian C, Dugas, Diana V, Al-Mohtaseb, Zaina, Hill, David E, & Zoghbi, Huda Y. 2011. Protein interactome reveals converging molecular pathways among autism disorders. *Science translational medicine*, **3**(86), 86ra49–86ra49.

Sankaran, Vijay G, Xu, Jian, & Orkin, Stuart H. 2010. Transcriptional silencing of fetal hemoglobin by BCL11A. *Annals of the New York Academy of Sciences*, **1202**(1), 64–68.

Santti, H, Mikkonen, L, Anand, A, Hirvonen-Santti, S, Toppari, J, Panhuysen, M, Vauti, F, Perera, M, Corte, G, Wurst, W, Jänne, O A, & Palvimo, J J. 2005. Disruption of the murine PIASx gene results in reduced testis weight. *Journal of molecular endocrinology*, **34**(3), 645–654.

Satterwhite, E, Sonoki, T, Willis, T G, Harder, L, Nowak, R, Arriola, E L, Liu, H, Price, H P, Gesk, S, Steinemann, D, Schlegelberger, B, Oscier, D G, Siebert, R, Tucker, P W, & Dyer, M J. 2001. The BCL11 gene family: involvement of BCL11A in lymphoid malignancies. *Blood*, **98**(12), 3413–3420.

Seo, Jawon, & Lee, Kong-Joo. 2004. Post-translational modifications and their biological functions: proteomic analysis and systematic approaches. *Journal of biochemistry and molecular biology*, **37**(1), 35–44.

Shammas, Sarah L. 2017. Mechanistic roles of protein disorder within transcription. *Current opinion in structural biology*, **42**(Feb.), 155–161.

Shav-Tal, Yaron, & Zipori, Dov. 2002. PSF and p54(nrb)/NonO–multi-functional nuclear proteins. *FEBS letters*, **531**(2), 109–114.

Shaw, Gerry, Morse, Silas, Ararat, Miguel, & Graham, Frank L. 2002. Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, **16**(8), 869–871.

Shendure, Jay, & Fields, Stanley. 2016. Massively Parallel Genetics. *Genetics*, **203**(2), 617–619.

Shi, Can, Zhang, Xiaobin, Chen, Zhiping, Sulaiman, Karina, Feinberg, Mark W, Ballantyne, Christie M, Jain, Mukesh K, & Simon, Daniel I. 2004. Integrin engagement regulates monocyte differentiation through the forkhead transcription factor Foxp1. *The Journal of clinical investigation*, **114**(3), 408–418.

Shi, Can, Sakuma, Masashi, Mooroka, Toshifumi, Liscoe, Alison, Gao, Huiyun, Croce, Kevin J, Sharma, Arjun, Kaplan, David, Greaves, David R, Wang, Yunmei, & Simon,

Daniel I. 2008. Down-regulation of the forkhead transcription factor Foxp1 is required for monocyte differentiation and macrophage function. *Blood*, **112**(12), 4699–4711.

Shim, Sungbo, Kwan, Kenneth Y, Li, Mingfeng, Lefebvre, Véronique, & Sestan, Nenad. 2012. Cis-regulatory control of corticospinal system development and evolution. *Nature*, **486**(7401), 74–79.

Shriberg, Lawrence D, Ballard, Kirrie J, Tomblin, J Bruce, Duffy, Joseph R, Odell, Katharine H, & Williams, Charles A. 2006. Speech, prosody, and voice characteristics of a mother and daughter with a 7;13 translocation affecting FOXP2. *Journal of speech, language, and hearing research : JSLHR*, **49**(3), 500–525.

Shu, W, Yang, H, Zhang, L, Lu, M M, & Morrisey, E E. 2001. Characterization of a new subfamily of winged-helix/forkhead (Fox) genes that are expressed in the lung and act as transcriptional repressors. *The Journal of biological chemistry*, **276**(29), 27488–27497.

Shu, Weiguo, Cho, Julie Y, Jiang, Yuhui, Zhang, Minhua, Weisz, Donald, Elder, Gregory A, Schmeidler, James, De Gasperi, Rita, Sosa, Miguel A Gama, Rabidou, Donald, Santucci, Anthony C, Perl, Daniel, Morrisey, Edward, & Buxbaum, Joseph D. 2005. Altered ultrasonic vocalization in mice with a disruption in the Foxp2 gene. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(27), 9643–9648.

Shu, Weiguo, Lu, Min Min, Zhang, Yuzhen, Tucker, Philip W, Zhou, Deying, & Morrisey, Edward E. 2007. Foxp2 and Foxp1 cooperatively regulate lung and esophagus development. *Development (Cambridge, England)*, **134**(10), 1991–2000.

Silbereis, John C, Pochareddy, Sirisha, Zhu, Ying, Li, Mingfeng, & Sestan, Nenad. 2016. The Cellular and Molecular Landscapes of the Developing Human Central Nervous System. *Neuron*, **89**(2), 248–268.

Sin, Cora, Li, Hongyan, & Crawford, Dorota A. 2015. Transcriptional regulation by FOXP1, FOXP2, and FOXP4 dimerization. *Journal of molecular neuroscience : MN*, **55**(2), 437–448.

Siper, Paige M, De Rubeis, Silvia, Trelles, Maria Del Pilar, Durkin, Allison, Di Marino, Daniele, Muratet, Francois, Frank, Yitzchak, Lozano, Reymundo, Eichler, Evan E, Kelly, Morgan, Beighley, Jennifer, Gerdts, Jennifer, Wallace, Arianne S, Mefford, Heather C, Bernier, Raphael A, Kolevzon, Alexander, & Buxbaum, Joseph D. 2017. Prospective investigation of FOXP1 syndrome. *Molecular autism*, **8**, 57.

Skelly, Michael J, Frungillo, Lucas, & Spoel, Steven H. 2016. Transcriptional regulation by complex interplay between post-translational modifications. *Current opinion in plant biology*, **33**(Oct.), 126–132.

Smith, Amena W, Holden, Kenton R, Dwivedi, Alka, Dupont, Barbara R, & Lyons, Michael J. 2015. Deletion of 16q24.1 supports a role for the ATP2C2 gene in specific language impairment. *Journal of child neurology*, **30**(4), 517–521.

Smith, Kenny, & Kirby, Simon. 2008. Cultural evolution: implications for understanding the human language faculty and its evolution. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **363**(1509), 3591–3603.

Smith, Ngaio C, & Matthews, Jacqueline M. 2016. Mechanisms of DNA-binding specificity and functional gene regulation by transcription factors. *Current opinion in structural biology*, **38**(June), 68–74.

Soblet, Julie, Dimov, Ivan, Graf von Kalckreuth, Clemens, Cano-Chervel, Julie, Baijot, Simon, Pelc, Karin, Sottiaux, Martine, Vilain, Catheline, Smits, Guillaume, & Deconinck, Nicolas. 2017. BCL11A frameshift mutation associated with dyspraxia and hypotonia affecting the fine, gross, oral, and speech motor systems. *American journal of medical genetics. Part A*, Sept.

Sollis, Elliot, Graham, Sarah A, Vino, Arianna, Froehlich, Henning, Vreeburg, Maaike, Dimitropoulou, Danai, Gilissen, Christian, Pfundt, Rolph, Rappold, Gudrun A, Brunner, Han G, Deriziotis, Pelagia, & Fisher, Simon E. 2016. Identification and functional characterization of de novo FOXP1 variants provides novel insights into the etiology of neurodevelopmental disorder. *Human molecular genetics*, **25**(3), 546–557.

Song, Wenjia, Li, Wenyan, Feng, Jinong, Heston, Leonard L, Scaringe, William A, & Sommer, Steve S. 2008. Identification of high risk DISC1 structural variants with a 2attributable risk for schizophrenia. *Biochemical and biophysical research communications*, **367**(3), 700–706.

Spiteri, Elizabeth, Konopka, Genevieve, Coppola, Giovanni, Bomar, Jamee, Oldham, Michael, Ou, Jing, Vernes, Sonja C, Fisher, Simon E, Ren, Bing, & Geschwind, Daniel H. 2007. Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. *American journal of human genetics*, **81**(6), 1144–1157.

Srivastava, Siddharth, Cohen, Julie S, Vernon, Hilary, Barañano, Kristin, McClellan, Rebecca, Jamal, Leila, Naidu, SakkuBai, & Fatemi, Ali. 2014. Clinical whole exome sequencing in child neurology practice. *Annals of neurology*, **76**(4), 473–483.

St Pourcain, Beate, Cents, Rolieke A M, Whitehouse, Andrew J O, Haworth, Claire M A, Davis, Oliver S P, O'Reilly, Paul F, Roulstone, Susan, Wren, Yvonne, Ang, Qi W, Velders, Fleur P, Evans, David M, Kemp, John P, Warrington, Nicole M, Miller, Laura, Timpson, Nicholas J, Ring, Susan M, Verhulst, Frank C, Hofman, Albert, Rivadeneira, Fernando, Meaburn, Emma L, Price, Thomas S, Dale, Philip S, Pillas, Demetris, Ylisherva, Anneli, Rodriguez, Alina, Golding, Jean, Jaddoe, Vincent W V, Jarvelin, Marjo-Riitta, Plomin, Robert, Pennell, Craig E, Tiemeier, Henning, & Davey Smith, George. 2014a. Common variation near ROBO2 is associated with expressive vocabulary in infancy. *Nature communications*, **5**(Sept.), 4831.

St Pourcain, Beate, Skuse, David H, Mandy, William P, Wang, Kai, Hakonarson, Hakon, Timpson, Nicholas J, Evans, David M, Kemp, John P, Ring, Susan M, McArdle, Wendy L, Golding, Jean, & Smith, George Davey. 2014b. Variability in the common genetic architecture of social-communication spectrum phenotypes during childhood and adolescence. *Molecular autism*, **5**(1), 18.

Starita, Lea M, Young, David L, Islam, Muhtadi, Kitzman, Jacob O, Gullingsrud, Justin, Hause, Ronald J, Fowler, Douglas M, Parvin, Jeffrey D, Shendure, Jay, & Fields, Stanley. 2015. Massively Parallel Functional Analysis of BRCA1 RING Domain Variants. *Genetics*, **200**(2), 413–422.

Stepanenko, A A, & Dmitrenko, V V. 2015. HEK293 in cell biology and cancer research: phenotype, karyotype, tumorigenicity, and stress-induced genome-phenotype evolution. *Gene*, **569**(2), 182–190.

Stessman, Holly A, Bernier, Raphael, & Eichler, Evan E. 2014. A genotype-first approach to defining the subtypes of a complex disease. *Cell*, **156**(5), 872–877.

Stevens, Servi J C, van Essen, Anthonie J, van Ravenswaaij, Conny M A, Elias, Abdallah F, Haven, Jaclyn A, Lelieveld, Stefan H, Pfundt, Rolph, Nillesen, Willy M, Yntema, Helger G, van Roozendaal, Kees, Stegmann, Alexander P, Gilissen, Christian, & Brunner, Han G. 2016. Truncating de novo mutations in the Krüppel-type zinc-finger gene ZNF148 in patients with corpus callosum defects, developmental delay, short stature, and dysmorphisms. *Genome medicine*, **8**(1), 131.

Stranneheim, H, & Wedell, A. 2016. Exome and genome sequencing: a revolution for the discovery and diagnosis of monogenic disorders. *Journal of internal medicine*, **279**(1), 3–15.

Stroud, James C, Wu, Yongqing, Bates, Darren L, Han, Aidong, Nowick, Katja, Pääbo, Svante, Tong, Harry, & Chen, Lin. 2006. Structure of the forkhead domain of FOXP2 bound to DNA. *Structure (London, England : 1993)*, **14**(1), 159–166.

Sun, Sihuai, Yang, Xiaobing, Wang, Yao, & Shen, Xihui. 2016. In Vivo Analysis of Protein-Protein Interactions with Bioluminescence Resonance Energy Transfer (BRET): Progress and Prospects. *International journal of molecular sciences*, **17**(10), 1704.

Sussel, L, Marin, O, Kimura, S, & Rubenstein, J L. 1999. Loss of Nkx2.1 homeobox gene function results in a ventral to dorsal molecular respecification within the basal telencephalon: evidence for a transformation of the pallidum into the striatum. *Development (Cambridge, England)*, **126**(15), 3359–3370.

Szemes, Marianna, Gyorgy, Andrea, Paweletz, Cloud, Dobi, Albert, & Agoston, Denes V. 2006. Isolation and characterization of SATB2, a novel AT-rich DNA binding protein expressed in development- and cell-specific manner in the rat brain. *Neurochemical research*, **31**(2), 237–246.

Taipale, Mikko, Kaminen, Nina, Nopola-Hemmi, Jaana, Haltia, Tuomas, Myllyluoma, Birgitta, Lyytinen, Heikki, Muller, Kurt, Kaaranen, Minna, Lindsberg, Perttu J, Hannula-Jouppi, Katariina, & Kere, Juha. 2003. A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. *Proceedings of the National Academy of Sciences of the United States of America*, **100**(20), 11553–11558.

Takahashi, Kaoru, Liu, Fu-Chin, Hirokawa, Katsuiku, & Takahashi, Hiroshi. 2003. Expression of Foxp2, a gene involved in speech and language, in the developing and adult striatum. *Journal of Neuroscience Research*, **73**(1), 61–72.

Takahashi, Kaoru, Liu, Fu-Chin, Hirokawa, Katsuiku, & Takahashi, Hiroshi. 2008. Expression of Foxp4 in the developing and adult rat forebrain. *Journal of Neuroscience Research*, **86**(14), 3106–3116.

Tam, W Y, Leung, C K Y, Tong, K K, & Kwan, K M. 2011. Foxp4 is essential in maintenance of Purkinje cell dendritic arborization in the mouse cerebellum. *Neuroscience*, **172**(Jan.), 562–571.

Tamura, Shinobu, Morikawa, Yoshihiro, Iwanishi, Hiroki, Hisaoka, Tomoko, & Senba, Emiko. 2003. Expression pattern of the winged-helix/forkhead transcription factor Foxp1 in the developing central nervous system. *Gene Expression Patterns*, **3**(2), 193–197.

Tanabe, Yuko, Fujita, Eriko, & Momoi, Takashi. 2011. FOXP2 promotes the nuclear translocation of POT1, but FOXP2(R553H), mutation related to speech-language disorder, partially prevents it. *Biochemical and biophysical research communications*, **410**(3), 593–596.

Tatham, Michael H, Rodriguez, Manuel S, Xirodimas, Dimitris P, & Hay, Ronald T. 2009. Detection of protein SUMOylation in vivo. *Nature protocols*, **4**(9), 1363–1371.

Tatham, Michael H, Matic, Ivan, Mann, Matthias, & Hay, Ronald T. 2011. Comparative proteomic analysis identifies a role for SUMO in protein quality control. *Science signaling*, **4**(178), rs4–rs4.

Teramitsu, Ikuko, & White, Stephanie A. 2006. FoxP2 regulation during undirected singing in adult songbirds. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, **26**(28), 7390–7394.

Teramitsu, Ikuko, Kudo, Lili C, London, Sarah E, Geschwind, Daniel H, & White, Stephanie A. 2004. Parallel FoxP1 and FoxP2 expression in songbird and human brain predicts functional interaction. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, **24**(13), 3152–3163.

Teufel, Andreas, Wong, Eric A, Mukhopadhyay, Mahua, Malik, Nasir, & Westphal, Heiner. 2003. FoxP4, a novel forkhead transcription factor. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, **1627**(2-3), 147–152.

Thevenon, Julien, Callier, Patrick, Andrieux, Joris, Delobel, Bruno, David, Albert, Sukno, Sylvie, Minot, Delphine, Mosca Anne, Laure, Marle, Nathalie, Sanlaville, Damien, Bonnet, Marlène, Masurel-Paulet, Alice, Levy, Fabienne, Gaunt, Lorraine, Farrell, Sandra, Le Caignec, Cedric, Toutain, Annick, Carmignac, Virginie, Mugneret, Francine, Clayton-Smith, Jill, Thauvin-Robinet, Christel, & Faivre, Laurence. 2013. 12p13.33 microdeletion including ELKS/ERC1, a new locus associated with childhood apraxia of speech. *European journal of human genetics : EJHG*, **21**(1), 82–88.

Thompson, Christopher K, Schwabe, Fabian, Schoof, Alexander, Mendoza, Ezequiel, Gampe, Jutta, Rochefort, Christelle, & Scharff, Constance. 2013. Young and intense: FoxP2 immunoreactivity in Area X varies with age, song stereotypy, and singing in male zebra finches. *Frontiers in neural circuits*, **7**, 24.

Tomassy, Giulio Srubek, De Leonibus, Elvira, Jabaudon, Denis, Lodato, Simona, Alfano, Christian, Mele, Andrea, Macklis, Jeffrey D, & Studer, Michèle. 2010. Area-specific temporal control of corticospinal motor neuron differentiation by COUP-TFI. *Proceedings of the National Academy of Sciences of the United States of America*, **107**(8), 3576–3581.

Tomblin, J B. 1989. Familial concentration of developmental language impairment. *The Journal of speech and hearing disorders*, **54**(2), 287–295.

Tomblin, J Bruce, O'Brien, Marlea, Shriberg, Lawrence D, Williams, Charles, Murray, Jeff, Patil, Shivanand, Bjork, Jonathan, Anderson, Steve, & Ballard, Kirrie. 2009. Language features in a mother and daughter of a chromosome 7;13 translocation involving FOXP2. *Journal of speech, language, and hearing research : JSLHR*, **52**(5), 1157–1174.

Trimarchi, Jeffrey M, Cho, Seo-Hee, & Cepko, Constance L. 2009. Identification of genes expressed preferentially in the developing peripheral margin of the optic cup. *Developmental Dynamics*, **238**(9), 2327–2329.

Tripodi, Marco, Filosa, Alessandro, Armentano, Maria, & Studer, Michèle. 2004. The COUP-TF nuclear receptors regulate cell migration in the mammalian basal forebrain. *Development (Cambridge, England)*, **131**(24), 6119–6129.

Tsui, David, Vessey, John P, Tomita, Hideaki, Kaplan, David R, & Miller, Freda D. 2013. FoxP2 regulates neurogenesis during embryonic cortical development. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, **33**(1), 244–258.

Turner, Samantha J, Hildebrand, Michael S, Block, Susan, Damiano, John, Fahey, Michael, Reilly, Sheena, Bahlo, Melanie, Scheffer, Ingrid E, & Morgan, Angela T. 2013. Small intragenic deletion in FOXP2 associated with childhood apraxia of speech and dysarthria. *American journal of medical genetics. Part A*, **161A**(9), 2321–2326.

Turner, Samantha J, Mayes, Angela K, Verhoeven, Andrea, Mandelstam, Simone A, Morgan, Angela T, & Scheffer, Ingrid E. 2015. GRIN2A: an aptly named gene for speech dysfunction. *Neurology*, **84**(6), 586–593.

Ullrich, D, Ullrich, K, & Marten, M. 2009. [Long-term school-development of children with impaired language skills: 15-years follow up after dismissal from child care centres with speech therapy]. *Laryngo- rhino- otologie*, **88**(10), 647–652.

Usui, Noriyoshi, Co, Marissa, Harper, Matthew, Rieger, Michael A, Dougherty, Joseph D, & Konopka, Genevieve. 2016. Sumoylation of FOXP2 Regulates Motor Function and Vocal Communication Through Purkinje Cell Development. *Biological psychiatry*, **0**(0).

Utine, G Eda, Haliloglu, Goknur, Volkan-Salanci, Bilge, Cetinkaya, Arda, Kiper, Pelin O, Alanay, Yasemin, Aktas, Dilek, Anlar, Banu, Topcu, Meral, Boduroglu, Koray, & Alikasifoglu, Mehmet. 2014. Etiological yield of SNP microarrays in idiopathic intellectual disability. *European journal of paediatric neurology : EJPN : official journal of the European Paediatric Neurology Society*, **18**(3), 327–337.

van den Berg, Debbie L C, Snoek, Tim, Mullin, Nick P, Yates, Adam, Bezstarosti, Karel, Demmers, Jeroen, Chambers, Ian, & Poot, Raymond A. 2010. An Oct4-centered protein interaction network in embryonic stem cells. *Cell stem cell*, **6**(4), 369–381.

van Loosdregt, Jorg, & Coffer, Paul J. 2014. Post-translational modification networks regulating FOXP3 function. *Trends in immunology*, **35**(8), 368–378.

van Loosdregt, Jorg, Fleskens, Veerle, Fu, Juan, Brenkman, Arjan B, Bekker, Cornelis P J, Pals, Cornelieke E G M, Meerding, Jenny, Berkers, Celia R, Barbi, Joseph, Gröne, Andrea, Sijts, Alice J A M, Maurice, Madelon M, Kalkhoven, Eric, Prakken, Berent J, Ovaa, Huib, Pan, Fan, Zaiss, Dietmar M W, & Coffer, Paul J. 2013. Stabilization of the transcription factor Foxp3 by the deubiquitinase USP7 increases Treg-cell-suppressive capacity. *Immunity*, **39**(2), 259–271.

Vargha-Khadem, F, Watkins, K E, Price, C J, Ashburner, J, Alcock, K J, Connelly, A, Frackowiak, R S, Friston, K J, Pembrey, M E, Mishkin, M, Gadian, D G, & Passingham, R E. 1998. Neural basis of an inherited speech and language disorder. *Proceedings of the National Academy of Sciences of the United States of America*, **95**(21), 12695–12700.

Vargha-Khadem, Faraneh, Gadian, David G, Copp, Andrew, & Mishkin, Mortimer. 2005. FOXP2 and the neuroanatomy of speech and language. *Nature reviews. Neuroscience*, **6**(2), 131–138.

Veltman, Joris A, & Brunner, Han G. 2012. De novo mutations in human genetic disease. *Nature reviews. Genetics*, **13**(8), 565–575.

Vernes, Sonja C, Nicod, Jérôme, Elahi, Fanny M, Coventry, Julie A, Kenny, Niamh, Coupe, Anne-Marie, Bird, Louise E, Davies, Kay E, & Fisher, Simon E. 2006. Functional genetic analysis of mutations implicated in a human speech and language disorder. *Human molecular genetics*, **15**(21), 3154–3167.

Vernes, Sonja C, Spiteri, Elizabeth, Nicod, Jérôme, Groszer, Matthias, Taylor, Jennifer M, Davies, Kay E, Geschwind, Daniel H, & Fisher, Simon E. 2007. High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders. *American journal of human genetics*, **81**(6), 1232–1250.

Vernes, Sonja C, MacDermot, Kay D, Monaco, Anthony P, & Fisher, Simon E. 2009. Assessing the impact of FOXP1 mutations on developmental verbal dyspraxia. *European journal of human genetics : EJHG*, **17**(10), 1354–1358.

Vernes, Sonja C, Oliver, Peter L, Spiteri, Elizabeth, Lockstone, Helen E, Puliyadi, Rathi, Taylor, Jennifer M, Ho, Joses, Mombereau, Cedric, Brewer, Ariel, Lowy, Ernesto, Nicod, Jérôme, Groszer, Matthias, Baban, Dilair, Sahgal, Natasha, Cazier, Jean-Baptiste, Ragoussis, Jiannis, Davies, Kay E, Geschwind, Daniel H, & Fisher, Simon E. 2011. Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. *PLoS genetics*, **7**(7), e1002145.

Vihma, Hanna, Luhakooder, Mirjam, Pruunsild, Priit, & Timmusk, Tonis. 2016. Regulation of different human NFAT isoforms by neuronal activity. *Journal of neurochemistry*, **137**(3), 394–408.

Villanueva, Pía, Nudel, Ron, Hoischen, Alexander, Fernández, María Angélica, Simpson, Nuala H, Gilissen, Christian, Reader, Rose H, Jara, Lillian, Echeverry, María Magdalena, Francks, Clyde, Baird, Gillian, Conti-Ramsden, Gina, O'Hare, Anne, Bolton, Patrick F, Hennessy, Elizabeth R, SLI Consortium, Palomino, Hernán, Carvajal-Carmona, Luis, Veltman, Joris A, Cazier, Jean-Baptiste, De Barbieri, Zulema, Fisher, Simon E, & Newbury, Dianne F. 2015. Exome sequencing in an admixed isolated population indicates NFXL1 variants confer a risk for specific language impairment. *PLoS genetics*, **11**(3), e1004925.

Viscardi, Lucas Henriques, Tovo-Rodrigues, Luciana, Pare, Pamela, Fagundes, Nelson Jurandi Rosa, Salzano, Francisco Mauro, Paixao-Cortes, Vanessa Rodrigues, Bau, Claiton Henrique Dotto, & Bortolini, Maria Catira. 2017. FOXP in Tetrapoda: Intrinsically Disordered Regions, Short Linear Motifs and their evolutionary significance. *Genetics and molecular biology*, **40**(1), 181–190.

Vissers, Lisenka E L M, de Ligt, Joep, Gilissen, Christian, Janssen, Irene, Steehouwer, Marloes, de Vries, Petra, van Lier, Bart, Arts, Peer, Wieskamp, Nienke, del Rosario, Marisol, van Bon, Bregje W M, Hoischen, Alexander, de Vries, Bert B A, Brunner, Han G, & Veltman, Joris A. 2010. A de novo paradigm for mental retardation. *Nature genetics*, **42**(12), 1109–1112.

Voineagu, Irina, Wang, Xinchen, Johnston, Patrick, Lowe, Jennifer K, Tian, Yuan, Horvath, Steve, Mill, Jonathan, Cantor, Rita M, Blencowe, Benjamin J, & Geschwind, Daniel H. 2011. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*, **474**(7351), 380–384.

Vorstman, Jacob A S, Parr, Jeremy R, Moreno-De-Luca, Daniel, Anney, Richard J L, Nurnberger, John I Jr, & Hallmayer, Joachim F. 2017. Autism genetics: opportunities and challenges for clinical translation. *Nature reviews. Genetics*, **18**(6), 362–376.

Walker, Rosie M, Hill, Alison E, Newman, Alice C, Hamilton, Gillian, Torrance, Helen S, Anderson, Susan M, Ogawa, Fumiaki, Derizioti, Pelagia, Nicod, Jérôme, Vernes, Sonja C, Fisher, Simon E, Thomson, Pippa A, Porteous, David J, & Evans, Kathryn L. 2012. The DISC1 promoter: characterization and regulation by FOXP2. *Human molecular genetics*, **21**(13), 2862–2872.

Wang, Bin, Lin, Danjuan, Li, Chuan, & Tucker, Philip. 2003. Multiple domains define the expression and regulatory properties of Foxp1 forkhead transcriptional repressors. *The Journal of biological chemistry*, **278**(27), 24259–24268.

Wang, Bin, Weidenfeld, Joel, Lu, Min Min, Maika, Shanna, Kuziel, William A, Morrisey, Edward E, & Tucker, Philip W. 2004. Foxp1 regulates cardiac outflow tract, endocardial cushion morphogenesis and myocyte proliferation and maturation. *Development (Cambridge, England)*, **131**(18), 4477–4487.

Wang, Feng, Marshall, Christopher B, & Ikura, Mitsuhiko. 2015. Forkhead followed by disordered tail: The intrinsically disordered regions of FOXO3a. *Intrinsically disordered proteins*, **3**(1), e1056906.

Watkins, K E, Vargha-Khadem, F, Ashburner, J, Passingham, R E, Connelly, A, Friston, K J, Frackowiak, R S J, Mishkin, M, & Gadian, D G. 2002. MRI analysis of an inherited speech and language disorder: structural brain abnormalities. *Brain : a journal of neurology*, **125**(Pt 3), 465–478.

Wen, Donghua, Xu, Zhijian, Xia, Li, Liu, Xinyi, Tu, Yaoyao, Lei, Hu, Wang, Weiwei, Wang, Tongdan, Song, Lili, Ma, Chunmin, Xu, Hanzhang, Zhu, Weiliang, Chen, Guoqiang, & Wu, Yingli. 2014. Important role of SUMOylation of Spliceosome factors in prostate cancer cells. *Journal of proteome research*, **13**(8), 3571–3582.

Wiegreffe, Christoph, Simon, Ruth, Peschkes, Katharina, Kling, Carolin, Strehle, Michael, Cheng, Jin, Srivatsa, Swathi, Liu, Pentao, Jenkins, Nancy A, Copeland, Neal G, Tarabykin, Victor, & Britsch, Stefan. 2015. Bcl11a (Ctip1) Controls Migration of Cortical Projection Neurons through Regulation of Sema3c. *Neuron*, **87**(2), 311–325.

Willemsen, Marjolein H, Nijhof, Bonnie, Fenckova, Michaela, Nillesen, Willy M, Bongers, Ernie M H F, Castells-Nobau, Anna, Asztalos, Lenke, Viragh, Erika, van Bon, Bregje W M, Tezel, Emre, Veltman, Joris A, Brunner, Han G, de Vries, Bert B A, de Ligt, Joep, Yntema, Helger G, van Bokhoven, Hans, Isidor, Bertrand, Le Caignec, Cedric, Lorino, Elsa, Asztalos, Zoltan, Koolen, David A, Vissers, Lisenka E L M, Schenck, Annette, & Kleefstra, Tjitske. 2013. GATAD2B loss-of-function mutations cause a recognisable syndrome with intellectual disability and are associated with learning deficits and synaptic undergrowth in Drosophila. *Journal of medical genetics*, **50**(8), 507–514.

Wilson, Van G. 2017. Introduction to Sumoylation. *Advances in experimental medicine and biology*, **963**(Chapter 1), 1–12.

Wolfe, Scot A, Grant, Robert A, & Pabo, Carl O. 2003. Structure of a designed dimeric zinc finger protein bound to DNA. *Biochemistry*, **42**(46), 13401–13409.

Wong, Kelly A, Kim, Rachel, Christofk, Heather, Gao, Jing, Lawson, Gregory, & Wu, Hong. 2004. Protein inhibitor of activated STAT Y (PIASy) and a splice variant lack-

ing exon 6 enhance sumoylation but are not essential for embryogenesis and adult life. *Molecular and cellular biology*, **24**(12), 5577–5586.

Woodworth, Mollie B, Greig, Luciano C, Liu, Kevin X, Ippolito, Gregory C, Tucker, Haley O, & Macklis, Jeffrey D. 2016. Ctip1 Regulates the Balance between Specification of Distinct Projection Neuron Subtypes in Deep Cortical Layers. *Cell reports*, **15**(5), 999–1012.

Wu, Yongqing, Borde, Madhuri, Heissmeyer, Vigo, Feuerer, Markus, Lapan, Ariya D, Stroud, James C, Bates, Darren L, Guo, Liang, Han, Aidong, Ziegler, Steven F, Mathis, Diane, Benoist, Christophe, Chen, Lin, & Rao, Anjana. 2006. FOXP3 controls regulatory T cell function through cooperation with NFAT. *Cell*, **126**(2), 375–387.

Xu, Jian, Sankaran, Vijay G, Ni, Min, Menne, Tobias F, Puram, Rishi V, Kim, Woojin, & Orkin, Stuart H. 2010. Transcriptional silencing of gamma-globin by BCL11A involves long-range interactions and cooperation with SOX6. *Genes & development*, **24**(8), 783–798.

Xu, Jian, Peng, Cong, Sankaran, Vijay G, Shao, Zhen, Esrick, Erica B, Chong, Bryan G, Ippolito, Gregory C, Fujiwara, Yuko, Ebert, Benjamin L, Tucker, Philip W, & Orkin, Stuart H. 2011. Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. *Science*, **334**(6058), 993–996.

Zarate, Yuri A, & Fish, Jennifer L. 2016. SATB2-associated syndrome: Mechanisms, phenotype, and practical recommendations. *American journal of medical genetics. Part A*, Oct.

Zarate, Yuri A, Kalsner, Louisa, Basinger, Alice, Jones, Julie R, Li, Chumei, Szybowska, Marta, Xu, Zhuo Luan, Vergano, Samantha, Caffrey, Aisling R, Gonzalez, Cruz Velasco, Dubbs, Holly, Zackai, Elaine, Millan, Francisca, Telegrafi, Aida, Baskin, Berivan, Person, Richard, Fish, Jennifer L, & Everman, David B. 2017. Genotype and Phenotype in 12 additional individuals with SATB2-Associated Syndrome. *Clinical genetics*, Jan.

Zeesman, Susan, Nowaczyk, Małgorzata J M, Teshima, Ikuko, Roberts, Wendy, Cardy, Janis Oram, Brian, Jessica, Senman, Lili, Feuk, Lars, Osborne, Lucy R, & Scherer, Stephen W. 2006. Speech and language impairment and oromotor dyspraxia due to deletion of 7q31 that involves FOXP2. *American journal of medical genetics. Part A*, **140**(5), 509–514.

Zhang, Yuzhen, Li, Shanru, Yuan, Lijun, Tian, Ying, Weidenfeld, Joel, Yang, Jifu, Liu, Feiyan, Chokas, Ann L, & Morrisey, Edward E. 2010. Foxp1 coordinates cardiomyocyte proliferation through both cell-autonomous and nonautonomous mechanisms. *Genes & development*, **24**(16), 1746–1757.

Zhao, Qi, Xie, Yubin, Zheng, Yueyuan, Jiang, Shuai, Liu, Wenzhong, Mu, Weiping, Liu, Zexian, Zhao, Yong, Xue, Yu, & Ren, Jian. 2014. GPS-SUMO: a tool for the prediction of sumoylation sites and SUMO-interaction motifs. *Nucleic acids research*, **42**(Web Server issue), W325–30.

Zhou, Beiyun, Zhong, Qian, Minoo, Parviz, Li, Changgong, Ann, David K, Frenkel, Baruch, Morrisey, Edward E, Crandall, Edward D, & Borok, Zea. 2008. Foxp2 inhibits Nkx2.1-mediated transcription of SP-C via interactions with the Nkx2.1 homeodomain. *American journal of respiratory cell and molecular biology*, **38**(6), 750–758.

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. *American journal of medical genetics. Part A*, **158A**(1), 254–256.

APPENDIX I: FUNCTIONAL CHARACTERIZATION OF RARE FOXP2 VARIANTS IN NEURODEVELOPMENTAL DISORDER

Estruch et al. Journal of Neurodevelopmental Disorders (2016) 8:44
DOI 10.1186/s11689-016-9177-2

Journal of Neurodevelopmental Disorders

RESEARCH Open Access

Functional characterization of rare FOXP2 variants in neurodevelopmental disorder

Sara B. Estruch[1†], Sarah A. Graham[1†], Swathi M. Chinnappa[1], Pelagia Deriziotis[1] and Simon E. Fisher[1,2*]

Abstract

Background: Heterozygous disruption of *FOXP2* causes a rare form of speech and language impairment. Screens of the *FOXP2* sequence in individuals with speech/language-related disorders have identified several rare protein-altering variants, but their phenotypic relevance is often unclear. *FOXP2* encodes a transcription factor with a forkhead box DNA-binding domain, but little is known about the functions of protein regions outside this domain.

Methods: We performed detailed functional analyses of seven rare FOXP2 variants found in affected cases, including three which have not been previously characterized, testing intracellular localization, transcriptional regulation, dimerization, and interaction with other proteins. To shed further light on molecular functions of FOXP2, we characterized the interaction between this transcription factor and co-repressor proteins of the C-terminal binding protein (CTBP) family. Finally, we analysed the functional significance of the polyglutamine tracts in FOXP2, since tract length variations have been reported in cases of neurodevelopmental disorder.

Results: We confirmed etiological roles of multiple FOXP2 variants. Of three variants that have been suggested to cause speech/language disorder, but never before been characterized, only one showed functional effects. For the other two, we found no effects on protein function in any assays, suggesting that they are incidental to the phenotype. We identified a CTBP-binding region within the N-terminal portion of FOXP2. This region includes two amino acid substitutions that occurred on the human lineage following the split from chimpanzees. However, we did not observe any effects of these amino acid changes on CTBP binding or other core aspects of FOXP2 function. Finally, we found that FOXP2 variants with reduced polyglutamine tracts did not exhibit altered behaviour in cellular assays, indicating that such tracts are non-essential for core aspects of FOXP2 function, and that tract variation is unlikely to be a highly penetrant cause of speech/language disorder.

Conclusions: Our findings highlight the importance of functional characterization of novel rare variants in FOXP2 in assessing the contribution of such variants to speech/language disorder and provide further insights into the molecular function of the FOXP2 protein.

Keywords: Transcription factor, Speech, Language, Functional genetics, Neuroscience

Background

FOXP2 is a member of the forkhead box (FOX) family of transcription factors and has crucial roles in the development of the brain and other organs [1, 2]. Heterozygous disruptions of the *FOXP2* gene cause a rare and severe speech and language disorder (OMIM 602081) [3]. This disorder was first reported in a three-generation pedigree (the KE family), in which approximately half of the individuals have difficulties with learning to make the co-ordinated orofacial movements required for speech (childhood apraxia of speech, CAS), together with wide-ranging impairments in comprehension and production of spoken and written language, but without major deficits in other aspects of cognitive functioning [4]. All affected members of the family were found to carry a missense variant in *FOXP2* that alters a critical residue within the DNA-recognition helix of the FOX domain and thus prevents DNA binding and regulation of transcription [4–6]. A number of individuals have since been reported to

* Correspondence: Simon.Fisher@mpi.nl
†Equal contributors
[1]Language and Genetics Department, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, the Netherlands
[2]Donders Institute for Brain, Cognition and Behaviour, Radboud University, 6525 EN Nijmegen, the Netherlands

present with severe speech/language difficulties together with heterozygous whole gene deletions or chromosomal translocations disrupting *FOXP2*, confirming the necessity of two functional copies of this gene for typical speech and language development [3].

Screening of the *FOXP2* coding region for protein-altering variants has been performed in a few small cohorts of children with speech articulation disorders similar to those reported in the KE family [7–9]. In addition, to address the possibility that *FOXP2* disruption might also be a factor in other disorders characterized by speech/language problems, similar screens have been performed in individuals with specific language impairment, speech sound disorder, autism, schizophrenia, and epilepsy of the speech cortex [10–18]. There are no common non-synonymous variants in FOXP2 in the general population, and relatively little coding sequence variation has been observed in individuals with speech/language-related disorders, indicating that *FOXP2* disruptions are a rare cause of such disorders, which likely have a highly heterogeneous genetic basis. Nonetheless, screening for *FOXP2* variants in individuals with neurodevelopmental phenotypes has identified a small number of rare protein-altering variants, including five missense variants, one stop-gain variant, one 2-bp deletion resulting in a frame-shift, and several in-frame insertions or deletions of glutamine residues within polyglutamine tracts. However, the contribution of individual rare variants to disorder often remains unclear because the genetic evidence in isolation is insufficient to confirm a causal or contributory role, and the effect of the variant on protein function is unknown.

To clarify the etiological contribution of the rare FOXP2 variants reported to date in individuals with neurodevelopmental disorders, we performed functional characterization of these variants by assaying their effects on a range of molecular properties. In addition, we characterize the interaction between FOXP2 and the co-repressors of the C-terminal binding protein (CTBP) family, which may have a central role in FOXP2-mediated transcriptional repression. Finally, we provide the first detailed examination of the role of the polyglutamine tract in FOXP2 function, in order to shed light on the contribution of tract length variation to disorder.

Methods

DNA constructs

The cloning of human FOXP2 (NM_014491), FOXP1 (NM_032682), and CTBP1 (NM_001328) and mouse Foxp2 (NM_053242) has been described previously [6, 19]. The coding sequence of CTBP2 (NM_001329) was amplified from human foetal brain cDNA using the primers listed in Additional file 1. The FOXP2 p.Q17L, p.M406T, p.P416T, p.R553H, p.N597H, p.N303T, p.S325N, and p.N303T/p.S325N variants were generated by site-directed mutagenesis using the Quick-Change Site-Directed Mutagenesis kit (Stratagene) following the manufacturer's protocol. (Note that the numbering of FOXP2 variants throughout this manuscript is given with respect to the coding sequence of the predominant isoform, NM_014491, which normally encodes a 715 amino-acid protein.) Primers used in site-directed mutagenesis are listed in Additional file 2. The FOXP2 p.R328* variant and synthetic truncated forms of FOXP2 were generated using the primers listed in Additional file 1. Synthetic FOXP2 variants with reduced polyglutamine tracts were generated using a PCR-based strategy [20]. All FOXP2 variants were initially generated in an intermediary plasmid, pCR2.1-TOPO (Life Technologies), and the entire coding sequence of FOXP2 was verified by Sanger sequencing before being subcloned into the final expression vector. For expression of fusion proteins with *Renilla* luciferase, yellow fluorescent protein (YFP), mCherry, and three tandem N-terminal Myc tags, cDNAs were subcloned into the pLuc, pYFP, pmCherry, and pMyc expression vectors, respectively, which have been described previously [19–21]. The SRPX2 luciferase reporter plasmid was generated by subcloning a 1146-bp region of the SRPX2 promoter into the promoterless firefly luciferase vector pGL4.23 (Promega), as described previously [21]. All constructs were verified by Sanger sequencing. Plasmid sequences are available upon request.

Cell culture and transfection

HEK293 were obtained from ECACC (cat. no. 85120602) and cultured in DMEM supplemented with 10% foetal bovine serum. Transfections were performed using GeneJuice (Merck-Millipore) according to the manufacturer's instructions.

Western blotting

HEK293 cells were transfected in 6-well plates and cultured for 48 h. Cells were lysed for 10 min at 4 °C with 100 mM Tris pH 7.5, 150 mM NaCl, 10 mM EDTA, 0.2% Triton X-100, 1% PMSF, and protease inhibitor cocktail. Cell lysates were cleared by centrifugation at 10,000×*g* for 3 min at 4 °C. Proteins were resolved on 10% SDS-polyacrylamide gels and transferred to PVDF membranes using a TransBlot Turbo Blotting apparatus (Bio-Rad). Membranes were blocked in PBS containing 5% milk and 0.1% Tween-20 and incubated overnight at 4 °C with primary antibody. The following antibodies were used: anti-GFP (Clontech cat. no. 632380, 1:8000, for YFP constructs); anti-Myc tag (Abcam cat. no. ab9106, 1:1000); and anti-β-actin (Sigma cat. no. A5441, 1:10,000). After washing, membranes were incubated with horseradish peroxidase-conjugated goat anti-mouse or anti-rabbit IgG for 45 min at room temperature. Proteins were visualized

using Novex ECL Chemiluminescent Substrate Reagent Kit (Invitrogen) and a ChemiDoc XRS+ imaging system (Bio-Rad).

BRET assay

Bioluminescence resonance energy transfer (BRET) assays were performed as described [19]. Briefly, HEK293 cells were transfected in 96-well plates with DNA plasmids encoding YFP- and luciferase-fusion proteins. After 36–48 h, Enduren Live Cell Luciferase Substrate (Promega) was added at a final concentration of 60 μM. Cells were cultured for a further 4 h, and emission readings (integrated over 10 s) were taken using a TECAN F200PRO microplate reader using the Blue1 and Green1 filter sets. Expression levels of the YFP-fusion proteins were measured by taking fluorescent readings using the filter set and dichroic mirror suitable for green fluorescent protein (excitation 480 nm, emission 535 nm). The corrected BRET ratio was calculated with the following formula: $[\text{Green1}_{(\text{experimental condition})}/\text{Blue1}_{(\text{experimental condition})}] - [\text{Green1}_{(\text{control condition})}/\text{Blue1}_{(\text{control condition})}]$. The control conditions used luciferase or YFP fused to a C-terminal nuclear localization signal.

Fluorescence microscopy

HEK293 cells were seeded on coverslips coated with poly-L-lysine. Cells were cultured for 30 h post-transfection and then fixed with methanol. Nuclei were stained with Hoechst 33342. Fluorescence images were acquired using an Axiovert A-1 fluorescent microscope with ZEN Image software (Zeiss).

Fluorescence-based quantitation of protein expression levels

HEK293 cells were transfected in triplicate with YFP-tagged FOXP2 variants and mCherry, in clear-bottomed black 96-well plates. Cells were cultured at 37 °C with 5% CO_2 in a TECAN M200PRO microplate reader equipped with a Gas Control Module for live-cell kinetic assays. Fluorescence intensity was measured 48 h post-transfection. For each well and time point, the background-subtracted YFP intensity was divided by the background-subtracted mCherry intensity. Triplicate conditions were averaged.

Luciferase reporter assays

HEK293 cells were seeded in clear-bottomed white 96-well plates and transfected in triplicate. For the SV40 assay, cells were transfected with 12 ng of pGL3-promoter firefly luciferase reporter construct containing the SV40 promoter (Promega), 5 ng of pRL-TK *Renilla* luciferase normalization control (Promega), and 16 ng of YFP-FOXP2 (wild-type or variant) or YFP control construct. For the SRPX2 assay, cells were transfected with 4.3 ng of SRPX2 luciferase reporter construct, 5 ng of pGL4.74 *Renilla* luciferase normalization control (Promega), and 45 ng of YFP-FOXP2 (wild-type or variant) or YFP control construct. After 48 h, luciferase activity was measured in a TECAN F200PRO microplate reader using the Dual-Luciferase Reporter Assay system (Promega).

Statistical analysis

The statistical significance of the luciferase reporter assays and BRET assays was analysed using a one-way analysis of variance (ANOVA) followed by Bonferroni's post hoc correction.

Results

Rare FOXP2 variants implicated in neurodevelopmental disorder

We examined seven rare FOXP2 variants that have been observed in individuals with neurodevelopmental disorders, including five missense variants, one stop-gain variant, and one frameshift variant (Table 1, Fig. 1a). For three of the variants examined (p.N597H, p.P416T, p.Q390Vfs*7), this is the first report of any functional characterization. The remaining four variants have been studied previously to varying extents and are included here for comparison with the uncharacterized variants. All of the variants were additionally characterized using novel assays that have not been used to study FOXP2 in prior literature. Two of the variants are of particular interest because of their uncertain significance with regard to the phenotype of the affected cases. The p.N597H variant was found by targeted sequencing in a proband with CAS and was described as a likely pathogenic variant, but it was not ascertained if this variant occurred de novo, and no functional characterization was performed [7]. The p.M406T variant was identified in a proband with an epilepsy-aphasia spectrum disorder (focal epilepsy with continuous spike-and-waves during sleep), cognitive and language deficits, and polymicrogyria of the left rolandic operculum [15]. The variant was also carried by two siblings who were not known to have any neurological abnormality and was inherited from the father, who did not display any neurological or MRI abnormalities. It was suggested that this variant plays an etiological role in the epilepsy-aphasia spectrum disorder observed in the proband, acting as a risk factor with incomplete penetrance [15]. However, the contribution of this variant to the phenotype is tentative because of the lack of segregation with the disorder, the atypical phenotypic presentation in comparison to other cases of *FOXP2* disruption, and the limited functional characterization performed to date.

The seven FOXP2 variants were expressed as fusion polypeptides with YFP in HEK293 cells. The expression of these YFP-tagged variants (and of all other YFP-tagged forms of FOXP2 used in this study) was verified by western blotting, as shown in Additional file 3. Direct measurement of fluorescence intensity in live cells indicated that the

Table 1 Rare FOXP2 variants in individuals with neurodevelopmental disorders

Variant description[a]	Frequency in ExAC[b]	Phenotype and inheritance pattern	Role in disorder[c]	Reference
p.Q17L (missense) c.50A > T chr7:114426561A > T rs201649896	56/120570 (1 homozygote)	Found in a proband with CAS but not in an affected sibling. Parental genotypes not determined	Probably incidental	MacDermot et al. [8]
p.M406T (missense) c.1217 T > C chr7:114653960 T > C no rs ID	Not observed	Found in a proband with rolandic epilepsy and polymicrogyria, in two unaffected siblings and in unaffected father	Uncertain significance	Roll et al. [15]
p.P416T (missense) c.1246C > A chr7:114653989C > A rs369313543	1/121328	Present in two siblings with severe stuttering. Absent in affected father. Inherited from mother, who does not stutter but has oral motor impairments	Probably incidental	Turner et al. [9]
p.R553H (missense) c.1658G > A chr7:114662075G > A rs121908377	Not observed	Segregates with CAS in three generations of the KE family	Causal	Lai et al. [4]
p.N597H (missense) c.1789A > C chr7:114663469A > C no rs ID	1/120986	Found in a proband with CAS. Parental genotypes not determined	Uncertain significance	Laffin et al. [7]
p.R328* (stop-gain) c.982C > T chr7:114642616C > T rs121908378	Not observed	Present in a proband with CAS and in affected sibling. Inherited from affected mother	Causal	MacDermot et al. [8]
p.Q390Vfs*7 (frameshift) c.1168_1169del chr7: 114652276_ 114652277del no rs ID	Not observed	De novo variant in a proband with sporadic CAS, dysarthria, and fine motor apraxia	Causal	Turner et al. [9]

[a]Variants are described in accordance with Human Genome Variation Society recommendations (www.hgvs.org/mutnomen, accessed June 2016) with reference to the major transcript NM_014491.3 (ENST00000350908). Genomic coordinates refer to the hg38 assembly. The rs ID number is provided for variants that are present in dbSNP

[b]Variant allele frequency in the Exome Aggregation Consortium (ExAC) dataset (http://exac.broadinstitute.org, accessed June 2016)

[c]Variants are described as causal if they segregate perfectly with childhood apraxia of speech (CAS) affection status, or if they occurred de novo in sporadic cases, and if they additionally have been demonstrated to cause loss of protein function, or are very likely to do so because of protein truncation. Variants are described as probably incidental if they do not segregate with CAS and are observed in the Exome Aggregation Consortium (ExAC) dataset. Other variants are described as of uncertain significance

Note: This table does not include newly described variants that were reported by Reuter et al. [58] after the completion of the present study

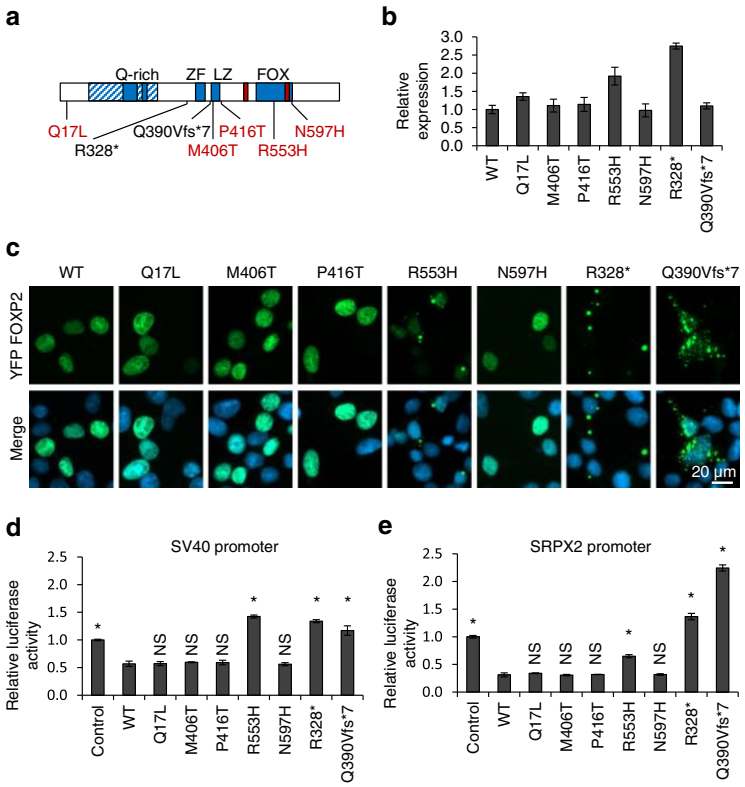


Fig. 1 Functional characterization of rare FOXP2 variants. **a** Schematic representation of the FOXP2 protein showing rare variants found in individuals with neurodevelopmental disorders. Stop-gain and frameshift variants are shown in black and missense variants in *red*. Known domains are labelled: glutamine-rich (*Q-rich*) region (hatched shading) including polyglutamine tracts (*solid shading*), zinc finger (*ZF*), leucine zipper (*LZ*) and forkhead domain (*FOX*). Nuclear localization signals are indicated with *red bars*. **b** Fluorescence-based measurement of FOXP2 expression level. HEK293 cells were transfected with YFP-FOXP2, together with mCherry for normalization. Fluorescence intensity was measured 48 h post-transfection. Values are mean YFP/mCherry fluorescence ratios ± S.D. ($n = 3$), expressed relative to the value for wild-type (WT) FOXP2. **c** Fluorescence micrographs of HEK293 cells transfected with YFP-FOXP2. Nuclei were stained with Hoechst 33342. **d**, **e** Luciferase reporter assays for transcriptional regulatory activity of FOXP2. HEK293 cells were transfected with a firefly luciferase reporter vector containing the SV40 promoter (**d**) or the human *SRPX2* promoter (**e**), together with a *Renilla* luciferase normalization plasmid and YFP-FOXP2, or YFP alone (control). Values are mean relative luciferase activity ± S.D. ($n = 3$), expressed relative to the control. *Asterisks* indicate significant differences compared to wild-type (WT) FOXP2 ($p < 0.05$, one-way ANOVA followed by Bonferroni post hoc correction). *NS* not significant. Exact *p* values for both the SV40 and the *SRPX2* assays are <0.0001 for the control and the R553H, R328*, and Q390Vfs*7 variants, and >0.9999 for all other variants

missense variants and the truncated p.Q390Vfs*7 variant are expressed at broadly similar levels to the wild-type protein, whereas the p.R553H and p.R328* variants may have slightly increased expression (Fig. 1b). In cells of people carrying the p.R328* and p.Q390Vfs*7 variants, the expression levels of these protein variants are likely to be very low due to nonsense-mediated decay of the aberrant transcripts, although this has not been formally tested due to the lack of *FOXP2* expression in accessible tissue [8, 9]. The inclusion of these truncated variants in functional assays provides a useful comparison with missense variants because the truncated variants lack the DNA-binding domain and hence cannot function normally in transcriptional regulation.

Effects of variants on nuclear localization and transcriptional regulation

The effect of the variants on FOXP2 localization was examined by direct imaging of the YFP-fusion proteins in transfected cells (Fig. 1c). Wild-type FOXP2 exhibited nuclear localization, with exclusion from nucleoli, as reported previously (Fig. 1c) [6, 22]. The two previously

characterized etiological variants, p.R553H and p.R328*, showed disrupted nuclear localization and formation of aggregates, also consistent with earlier reports (Fig. 1c) [6, 22]. The uncharacterized Q390Vfs*7 variant formed cytoplasmic aggregates similar to those observed for the p.R328* variant, consistent with the loss of endogenous nuclear localization signals in these truncated variants (Fig. 1a,c) [22]. The p.Q390Vfs*7 variant therefore has a similarly deleterious effect on protein localization as the two known etiological variants. None of the other missense variants showed any sign of abnormal localization, including the putatively pathogenic p.N597H and p.M406T variants (Fig. 1c). Altered localization has been reported previously for the p.M406T variant, but in multiple independent experiments, we did not observe any loss of nuclear localization for this variant [15]. Thus, with the exception of the p.R553H variant, all the missense variants retain the nuclear localization necessary for transcriptional regulatory activity.

To assess the ability of the FOXP2 variants to regulate transcription, we performed luciferase reporter assays using the SV40 promoter, a viral promoter which is repressed by FOXP2 [6, 23]. The etiological p.R553H and p.R328* variants exhibited the expected loss of repression activity in this assay [6] (Fig. 1d). The uncharacterized p.Q390Vfs*7 variant showed a comparable loss of repression to the other etiological variants (Fig. 1f). In contrast, the remaining four missense variants showed similar activity to the wild-type protein (Fig. 1d). To verify these results using an endogenous human FOXP2 target gene, we performed luciferase reporter assays using a region of the *SRPX2* promoter [15]. As observed for the SV40 promoter, the truncated variants and the p.R553H variant showed loss of repressive activity, whereas the other missense variants did not differ significantly in activity from the wild-type protein (Fig. 1e). The p.M406T variant has previously been reported to show reduced transcriptional repression activity in relation to the *SRPX2* promoter [15]. Small differences between the promoter regions and vector backbones used in this and the previous study might account for these conflicting results. Nonetheless, our data indicate that the p.M406T variant does not exhibit a generalized reduction in transcriptional regulatory activity.

Effects of variants on protein dimerization

FOXP2 and other proteins of the FOXP subfamily form dimers via their leucine zipper domains, a property which appears to be essential for transcriptional regulatory activity [24] (Fig. 1a). In FOXP3, loss of dimerization capacity as a result of an in-frame single amino acid deletion in the leucine zipper domain results in an immunological disorder known as IPEX syndrome (MIM 304790), with disease severity comparable to that resulting from loss of DNA-binding activity in FOXP3 [25, 26]. Variants affecting the dimerization of FOXP2 might therefore be a cause of speech/language disorder. Additionally, FOXP2 variants that have lost DNA-binding capacity but retain dimerization ability might interfere with the function of wild-type protein in cells of affected individuals, and thus exert a dominant-negative effect. Notably, FOXP2 not only forms homodimers but can also heterodimerize with FOXP1 and FOXP4, proteins which have partially overlapping expression patterns in the developing brain [24, 27–29]. (Note that FOXP3 is expressed only in haematopoietic cells.) Heterodimerization among certain FOXP proteins may therefore be an important mechanism for differential regulation of target genes in different neuronal subtypes. Of note, heterozygous disruption of *FOXP1* results in a severe neurodevelopmental phenotype that includes language deficits, potentially reflecting dysregulation of some of the same target genes impacted by *FOXP2* disruption [30, 31].

Two of the five FOXP2 missense variants reported in individuals with neurodevelopmental disorders lie within or near the leucine zipper domain, which spans residues p.V388-L409 (Fig. 1a). These variants (p.M406T and p.P416T) could therefore affect FOXP2 dimerization. The interaction between FOXP2 variants and wild-type FOXP2 was assayed using a BRET assay, which enables protein-protein interactions to be monitored in live cells [19]. The interaction of the FOXP2 variants with wild-type FOXP1 and FOXP4 was also examined. As expected, wild-type FOXP2 was able to homodimerize and also to heterodimerize with FOXP1 and FOXP4 (Fig. 2a–c) [24]. The truncated p.R328* and p.Q390fs*7 variants were not able to dimerize with the wild-type proteins, consistent with the loss of the leucine zipper domain in these variants (Fig. 2a–c). Thus, even if these abnormal proteins are present in cells of affected individuals despite the activation of nonsense-mediated decay mechanisms, they could not interfere with the function of wild-type proteins via dimerization. In contrast, the p.R553H variant found in the KE family has a normal leucine zipper domain and showed only a modest reduction in interaction with wild-type FOXP2, FOXP1, and FOXP4 (Fig. 2a–c). Furthermore, co-transfection of wild-type FOXP2 with the p.R553H variant led to some wild-type protein being mislocalized to the cytoplasm (Fig. 2d). Mislocalization of wild-type protein was not observed upon co-transfection with the non-interacting variants p.R328* and p.Q390Vfs*7 (Fig. 2d). The p.R553H variant might therefore interfere with the function of wild-type FOXP2 in cells of people carrying this variant, and this effect could contribute to the phenotype observed in the affected members of the KE family. All four of the other missense variants generally demonstrated normal dimerization capacity (Fig. 2a–c). The p.P416T variant showed a statistically significant difference in interaction with FOXP1 (Fig. 2b),

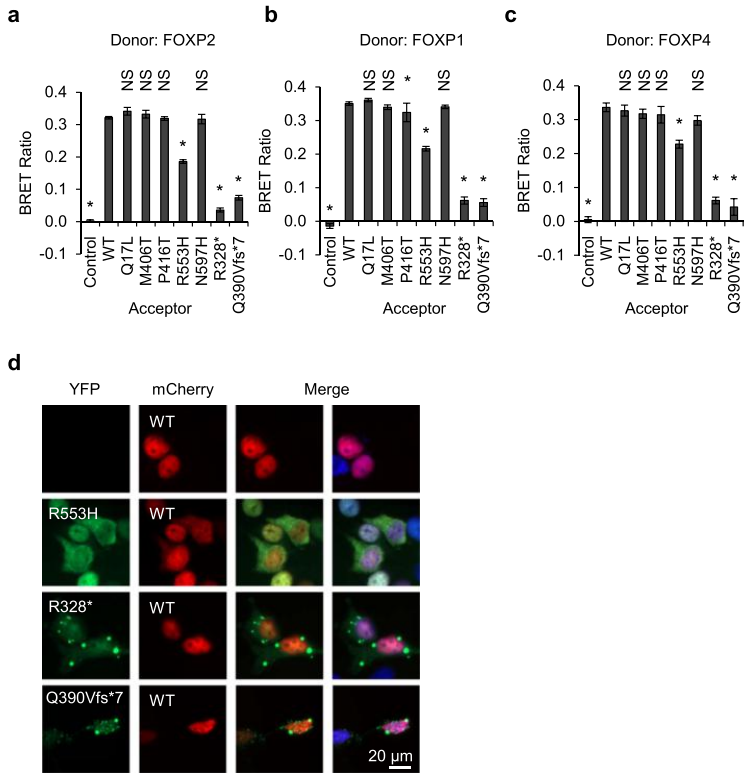


Fig. 2 Interaction of rare FOXP2 variants with wild-type FOXP proteins. **a**–**c** Bioluminescence resonance energy transfer (*BRET*) assays for interaction of FOXP2 variants with wild-type (*WT*) FOXP proteins. HEK293 cells were transfected with YFP-FOXP2 variants (acceptor) and *Renilla* luciferase (donor) fusions of FOXP2 (**a**), FOXP1 (**b**) or FOXP4 (**c**). The control acceptor protein is a nuclear-targeted YFP. Values are mean corrected BRET ratios ± S.D. ($n = 3$). *Asterisks* indicate significant differences compared to wild-type (*WT*) FOXP2 ($p < 0.05$, one-way ANOVA followed by Bonferroni post hoc correction). *NS* not significant. Exact *p* values for **a** are <0.0001 for the control and the R553H, R328* and Q390Vfs*7 variants and >0.9999 for all other variants. Exact *p* values for **b** are <0.0001 for the control and the R553H, R328* and Q390Vfs*7 variants, 0.024 for the P416T variant, and >0.9999 for all other variants. Exact *p* values for **c** are <0.0001 for the control and the R553H, R328* and Q390Vfs*7 variants, 0.088 for the N597H variant, and >0.9999 for all other variants. **d** Fluorescence micrographs of HEK293 cells transfected with YFP-fusions of the FOXP2 variants p.R553H, p.R328* and p.Q390Vfs*7, together with wild-type FOXP2 fused to mCherry. Nuclei were stained with Hoechst 33342

but this difference was not consistently observed across independent experiments. Therefore, despite being located in or near the leucine zipper domain, the p.M406T and p.P416T variants do not appear to affect FOXP2 dimerization and are unlikely to exert a pathogenic effect via this mechanism.

Effects of variants on interaction with the co-repressors CTBP1 and CTBP2

A further mechanism by which variants in FOXP2 might lead to speech/language disorder is through disruption of protein-protein interactions between FOXP2 and crucial mediators of transcriptional regulation. Relatively, little is known about the interaction partners of FOXP2; however, proteins of the CTBP family have been identified as candidate FOXP2 interactors in multiple independent yeast two-hybrid screens [24, 32–34]. The CTBP family consists of two proteins, CTBP1 and CTBP2, which function as co-repressors for multiple transcription factors [35, 36]. The interaction between FOXP2 and CTBP1 has previously been validated by us using the BRET assay, and by others via co-immunoprecipitation, but the interaction with CTBP2 has not yet been confirmed [19, 24]. Using the BRET assay, we validated the interaction between CTBP2 and FOXP2 and found that this interaction produces a notably higher BRET signal than the

interaction between CTBP1 and FOXP2 (Fig. 3a,b). The difference in the magnitude of the BRET signal may reflect subcellular localization differences between the CTBPs: CTBP2 is wholly nuclear and therefore shows a large degree of co-localization with wild-type FOXP2, whereas CTBP1 is found in both the cytoplasm and nucleus, and therefore has an overall lower degree of co-localization with FOXP2 (Fig. 3c) [37, 38]. Strikingly, we found that CTBP1 and CTBP2 are also able to interact with FOXP1, but not with FOXP4, in line with previously reported findings for CTBP1 (Fig. 3a,b) [24]. The lack of interaction between FOXP4 and CTBPs suggests that the FOXP proteins have partially divergent mechanisms for transcriptional regulation, which is particularly interesting given the ability of FOXP2 to dimerize with FOXP4 (Fig. 2c).

We next evaluated the effects of rare variants in FOXP2 on the interaction with CTBP1/2. The p.R553H variant found in the KE family showed slightly reduced interaction with CTBP2, but increased interaction with CTBP1, when compared with wild-type FOXP2 (Fig. 3d,e). Interestingly, the p.R553H variant displayed a mixed nuclear and cytoplasmic localization when transfected alone or with CTBP2, but became predominantly cytoplasmic in the presence of CTBP1 (Figs. 1e and 3f,g). In contrast, the nuclear localization of wild-type FOXP2 was not affected by co-transfection with CTBP1 (Fig. 3c). The cytoplasmic retention of the p.R553H variant, but not of wild-type FOXP2, upon co-transfection with CTBP1 may be a consequence of the loss of DNA-binding capacity in the p.R553H variant. These co-transfection experiments indicate that the p.R553H variant shows greater co-localization with CTBP1 but reduced co-localization with CTBP2, compared with wild-type FOXP2, consistent with the BRET data on interactions between these proteins (Fig. 3d,e). Although altered interaction with CTBP1/2 probably does not play a substantive role in the speech/language pathology in the KE family, the retained interaction between the p.R553H variant and CTBP1/2 is notable because it indicates that DNA-binding activity is not a prerequisite for interaction of FOXP2 with co-repressors of the CTBP family. The four other missense variants did not differ substantially from the wild-type protein in their interaction with CTBP1/2 in the BRET assay (Fig. 3d,e). These four variants thus have comparable properties to wild-type FOXP2 in assays of subcellular localization, transcriptional repression, protein dimerization, and interaction with co-repressor proteins, suggesting that all of these variants are benign rare polymorphisms (Table 2). In particular, we do not find any support for the etiological roles suggested previously for the p.M406T and p.N597H variants [7, 15].

Interestingly, the p.R328* and p.Q390Vfs*7 variants partially retained the ability to interact with CTBP1/2, despite lacking a large proportion of the normal FOXP2 polypeptide (Fig. 3d,e). The interaction with CTBP1 in particular was only mildly affected (Fig. 3d). Furthermore, co-transfection of CTBP1 with the p.R328* and p.Q390Vfs*7 variants resulted in clear co-localization of CTBP1 with the truncated variants within cytoplasmic aggregates (Fig. 3d,f). The interaction with CTBP2 was more severely affected, but co-localization of CTBP2 and the p.R328* variant within aggregates was still readily apparent (Fig. 3e,g). The interaction between these severely truncated FOXP2 variants and CTBP1/2 suggests that some key determinants of CTBP binding lie near the N-terminus of FOXP2.

Characterization of the interaction between FOXP proteins and CTBPs

The retention of CTBP binding by the severely truncated p.R328* and p.Q390Vfs*7 variants (Fig. 3d,e) prompted us to investigate further the interaction between FOXP2 and CTBPs. The CTBP binding site in FOXP2 was previously suggested to be a PLNLV motif at residues 422–426, due to the similarity of this motif to the consensus CTBP-binding motif PXDLS (Fig. 4a) [24, 36]. The PLNLV motif is conserved in FOXP1 but not FOXP4, consistent with the lack of interaction between CTBP1 and FOXP4 [24] (Fig. 3a,b). However, the residual interaction of CTBPs with the p.R328* and p.Q390Vfs*7 variants indicates that the PLNLV motif is not essential for the interaction and that the CTBP binding site is at least partially localized to a more N-terminal region of FOXP2 (Fig. 3d,e). Furthermore, one of the yeast two-hybrid screens that identified CTBP1 and CTBP2 as FOXP2 interactors employed a fragment of FOXP2 encompassing residues 122–382, indicating that this region, which lacks the PLNLV motif, is sufficient for the interaction [34].

To map the region of FOXP2 involved in binding to CTBP2, we performed a BRET assay using a series of synthetic truncated versions of FOXP2 (Fig. 4a) [21]. Truncated protein variants lacking the endogenous nuclear localization signals were targeted to the nucleus by addition of an artificial localization signal to the protein C-terminus (Fig. 4a,b). The efficacy of this truncation series in mapping interaction domains was demonstrated using a BRET assay for FOXP2 homodimerization (Fig. 4c). Truncation of FOXP2 after residue 423 did not interfere with homodimerization, whereas truncation after residue 329 dramatically reduced the interaction, reflecting the loss of the critical leucine zipper domain (Fig. 4c) [24].

In a BRET assay for interaction with CTBP2, a FOXP2 variant truncated after residue 487 showed a similar level of interaction to the full-length protein (Fig. 4d), indicating that the C-terminal region of FOXP2 including the FOX domain is not involved in the interaction with CTBPs. More severely truncated forms of FOXP2 showed a progressive reduction in interaction with CTBP2. The

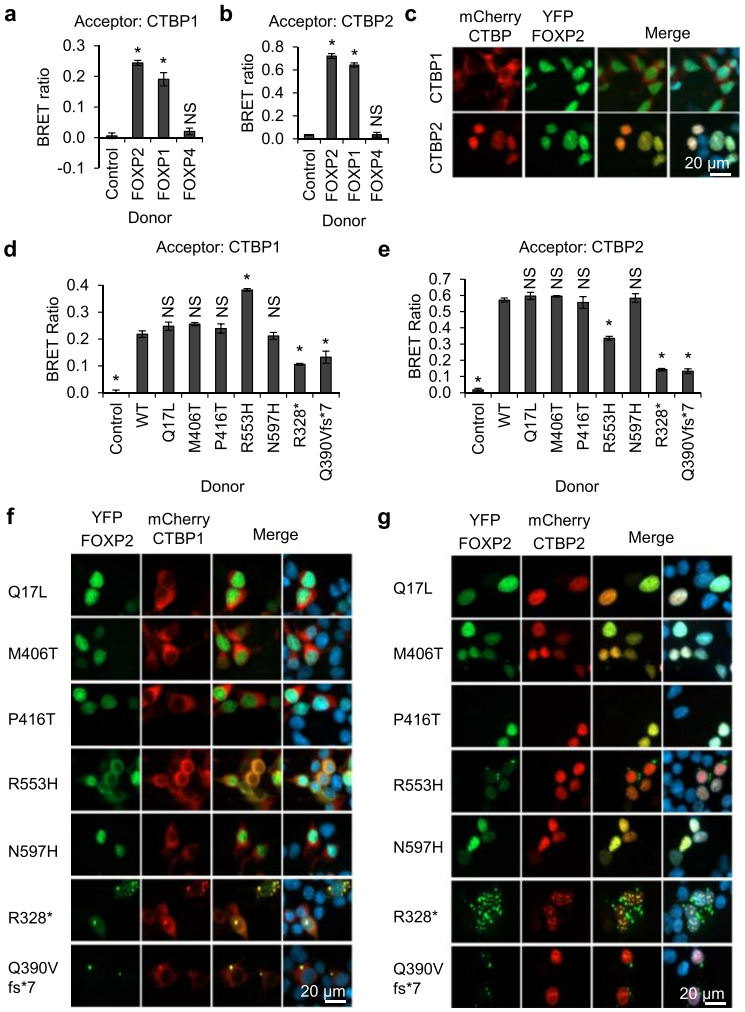


Fig. 3 Interaction of rare FOXP2 variants with CTBP1/2. **a**, **b** BRET assays for interaction of FOXP proteins with CTBP1 and CTBP2. HEK293 cells were transfected with FOXP1, FOXP2, or FOXP4 fused to *Renilla* luciferase (donor) and CTBP1 (**a**) or CTBP2 (**b**) fused to YFP (acceptor). The control donor protein is a nuclear-targeted luciferase. Values are mean corrected BRET ratios ± S.D. ($n = 3$). *Asterisks* indicate significant differences compared to control ($p < 0.05$, one-way ANOVA followed by Bonferroni post hoc correction). *NS* not significant. Exact p values for both **a** and **b** are <0.0001 for FOXP2 and FOXP1 and >0.9999 for FOXP4. **c** Fluorescence micrographs of HEK293 cells transfected with mCherry-CTBP1 or mCherry-CTBP2, together with YFP-FOXP2. Nuclei were stained with Hoechst 33342. **d**, **e** BRET assays for interaction of rare FOXP2 variants with CTBP1 and CTBP2. HEK293 cells were transfected with FOXP2 variants fused to *Renilla* luciferase (donor) and CTBP1 (**d**) or CTBP2 (**e**) fused to YFP (acceptor). The control donor protein is a nuclear-targeted luciferase. Values are mean corrected BRET ratios ± S.D. ($n = 3$). *Asterisks* indicate significant differences compared to wild-type (*WT*) FOXP2 ($p < 0.05$, one-way ANOVA followed by Bonferroni post hoc correction). *NS* not significant. Exact p values for **d** are <0.0001 for the control and the R553H, R328* and Q390Vfs*7 variants, 0.221 for the M406T variant, and >0.9999 for all other variants. Exact p values for **e** are <0.0001 for the control and the R553H, R328* and Q390Vfs*7 variants and >0.9999 for all other variants. **f**, **g** Fluorescence micrographs of HEK293 cells transfected with YFP-FOXP2 and mCherry-CTBP1 (**f**) or mCherry-CTBP2 (**g**). Nuclei were stained with Hoechst 33342

Table 2 Summary of functional characterization of rare FOXP2 variants in individuals with neurodevelopmental disorders

Variant[a]	Localization	Transcriptional repression		Dimerization with FOXPs			Interaction with CTBPs		Role in disorder[b]
		SV40	SRPX2	FOXP2	FOXP1	FOXP4	CTBP1	CTBP2	
p.Q17L	+	+	+	+	+	+	+	+	Incidental
p.M406T	+	+	+	+	+	+	+	+	Incidental
p.P416T	+	+	+	+	+	+	+	+	Incidental
p.R553H	–	–	–	–	–	–	–	–	Causal
p.N597H	+	+	+	+	+	+	+	+	Incidental
p.R328*	–	–	–	–	–	–	–	–	Causal
P.Q390Vfs*7	–	–	–	–	–	–	–	–	Causal

+ Behaviour of the variant is comparable to that of the wild-type protein in this assay, – Behaviour of the variant differs from that of the wild-type protein in this assay

[a]Variants are described in accordance with Human Genome Variation Society recommendations (www.hgvs.org/mutnomen, accessed June 2016) with reference to the major transcript NM_014491.3 (ENST00000350908)

[b]Probable role of the variant in the disorder in the affected individual, based on the results of functional characterization

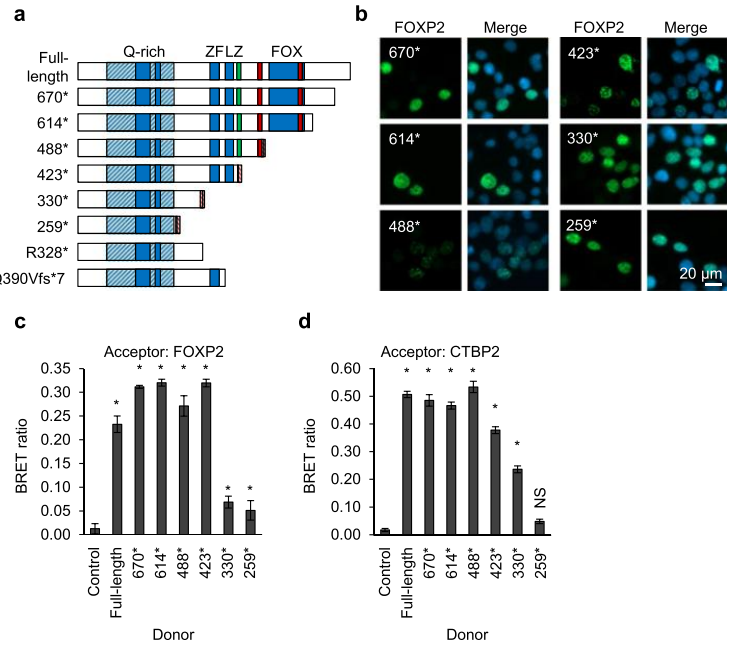


Fig. 4 Mapping of the CTBP binding site in FOXP2. **a** Schematic representation of synthetic truncated forms of FOXP2. The R328* and Q390Vfs*7 variants identified in patients are included for comparison. Known domains are labelled: glutamine-rich (*Q-rich*) region (*hatched shading*) including polyglutamine tracts (*solid shading*), zinc finger (*ZF*), leucine zipper (*LZ*), and forkhead domain (FOX). The PLNLV motif is indicated with a *green bar*. Nuclear localization signals are indicated with *red bars*. A synthetic nine-residue nuclear targeting sequence (*hatched red bars*) was appended to the C-terminus of variants which lack one or both of the endogenous nuclear localization signals. **b** Fluorescence micrographs of HEK293 cells transfected with synthetic truncated FOXP2 variants. Nuclei were stained with Hoechst 33342. **c**, **d** BRET assay for interaction of synthetic truncated FOXP2 variants with full-length FOXP2 and CTBP2. HEK293 cells were transfected with truncated FOXP2 variants fused to *Renilla* luciferase (donor) and FOXP2 (**c**) or CTBP2 (**d**) fused to YFP (acceptor). The control donor protein is a nuclear-targeted luciferase. Values are mean corrected BRET ratios ± S.D. ($n = 3$). *Asterisks* indicate significant differences compared to control ($p < 0.05$, one-way ANOVA followed by Bonferroni post hoc correction). *NS* not significant. Exact p values for **c** are 0.002 for FOXP2.330*, 0.025 for FOXP2.259* and <0.0001 for all other variants. Exact p values for **d** are 0.001 for FOXP2.423*, 0.012 for FOXP2.330*, 0.97 for FOXP2.259*, and <0.0001 for all other variants

substantial retention of interaction between CTBP2 and FOXP2 forms truncated after residue 422 or residue 329 indicates that the PLNLV motif, zinc finger, and leucine zipper domains are not essential for interaction, although they may enhance it, and furthermore, FOXP2 dimerization is not necessary for CTBP2 binding (Fig. 4d). Only the shortest FOXP2 variant, which is truncated after residue 258, showed a complete or near-complete loss of interaction (Fig. 4d). The protein region including residues 259–329 may therefore be sufficient for interaction with CTBP2.

The region of FOXP2 encompassing residues 259–329 corresponds to a single exon in which non-synonymous nucleotide substitutions between different species are relatively common in comparison to the exons encoding highly conserved elements such as the FOX and leucine zipper domains [39–41]. In particular, this exon includes two non-synonymous nucleotide substitutions that occurred after the split between the human and chimpanzee lineages, but before the split between modern humans and Neanderthals (Fig. 5a) [39, 42]. Mice carrying a 'humanized' version of Foxp2, containing these two recent amino acid changes, exhibit altered neuronal morphology and electrophysiology and subtle behavioural changes [43–45]. Strikingly, the morphological and electrophysiological changes have a direction of effect opposite to that observed in mice with only one functional copy of *Foxp2*, suggesting that the changes might have been selected for due to an enhancement of protein function [1, 46].

The molecular mechanisms underlying the neurobiological changes observed in mice carrying humanized *Foxp2* are unknown, but might entail an alteration in the strength of the interaction between Foxp2 and one or more proteins involved in Foxp2-mediated transcriptional regulation. Therefore, to test if the recent amino acid changes affect interaction with CTBPs, we introduced the ancestral amino acid residues, individually and in combination, into human FOXP2 (Fig. 5a). Versions of human FOXP2 carrying the ancestral amino acids were compared with the natural human and mouse proteins in functional assays. Introduction of the ancestral amino acids did not result in any change in protein localization or expression, and furthermore, the human and mouse versions of the protein were indistinguishable in these assays (Fig. 5b,c). Notably, the versions of FOXP2 carrying the ancestral amino acids also did not differ from the natural human protein in assays of transcriptional repression using the SV40 and *SRPX2* promoters, and there was no difference between the human and mouse orthologues in these assays (Fig. 5d,e). Homodimerization was also comparable for all variants tested (Fig. 5f). Finally, the versions of FOXP2 carrying the ancestral amino acids did not differ from the natural human protein in their interaction with CTBP1 and CTBP2 (Fig. 5g,h). Therefore, despite the CTBP binding site in FOXP2 encompassing the protein region that includes the recent amino acid changes, it is unlikely that an alteration in the strength of the interaction between FOXP2 and CTBP1/2 is responsible for the neurobiological differences observed in mice carrying a partially humanized version of Foxp2.

The FOXP2 polyglutamine tract and neurodevelopmental disorder

FOXP2 contains a large polyglutamine tract of 40 residues (p.Q152-Q191) and a small tract of 10 residues (p.Q200-Q209), separated by 8 residues of intervening polypeptide (Fig. 6a). The tracts are located in the N-terminal portion of the protein, within a larger glutamine-rich region (Fig. 6a). Of note, the large tract in FOXP2 is the longest polyglutamine tract in any human protein in healthy individuals [47]. Abnormal expansions of polyglutamine tracts in at least nine different proteins result in neurodegenerative diseases [48]. However, FOXP2 is not regarded as a strong candidate for involvement in such disorders because its large polyglutamine tract shows relatively little length variation in the human population [47]. The mixture of CAG and CAA codons without long pure CAG repeats probably makes the tract less prone to expansion, and no expansions were detected in a sample of 142 individuals with progressive movement disorders [47, 49]. However, some rare variations in tract length have been observed in individuals with neurodevelopmental disorders (Additional file 4) [8, 11, 13, 14, 17, 18]. Most of the observed variants are deletions of 1–6 residues from the large tract (Additional file 4). Such deletions have also been observed in the general population (Exome Aggregation Consortium (ExAC), Cambridge, MA (http://exac.broadinstitute.org)), suggesting that they are probably not high-penetrance causal variants in cases of severe disorder, but might still represent risk factors, or produce a milder phenotype.

The significance of the polyglutamine tracts to the functions of the FOXP2 protein remains unknown. Polyglutamine tracts may be critical to the function of some proteins [50], and the absence of equivalent tracts in human FOXP1 and FOXP4 (despite presence of glutamine-rich regions) raises the possibility that tract expansion in FOXP2 may have contributed to divergence in FOXP protein function. To investigate the roles of the polyglutamine tracts in FOXP2, we engineered versions of the protein with reduced tract lengths reflecting the presumed ancestral protein sequence. The long tract in FOXP2 underwent expansion early in tetrapod evolution and has subsequently remained relatively stable: ray-finned fish species have a cluster of only 3 glutamine residues in place of the long tract in human FOXP2 [51, 52], whereas the coelacanth, an evolutionary intermediate between ray-finned fish and tetrapods, has a 25-residue tract,

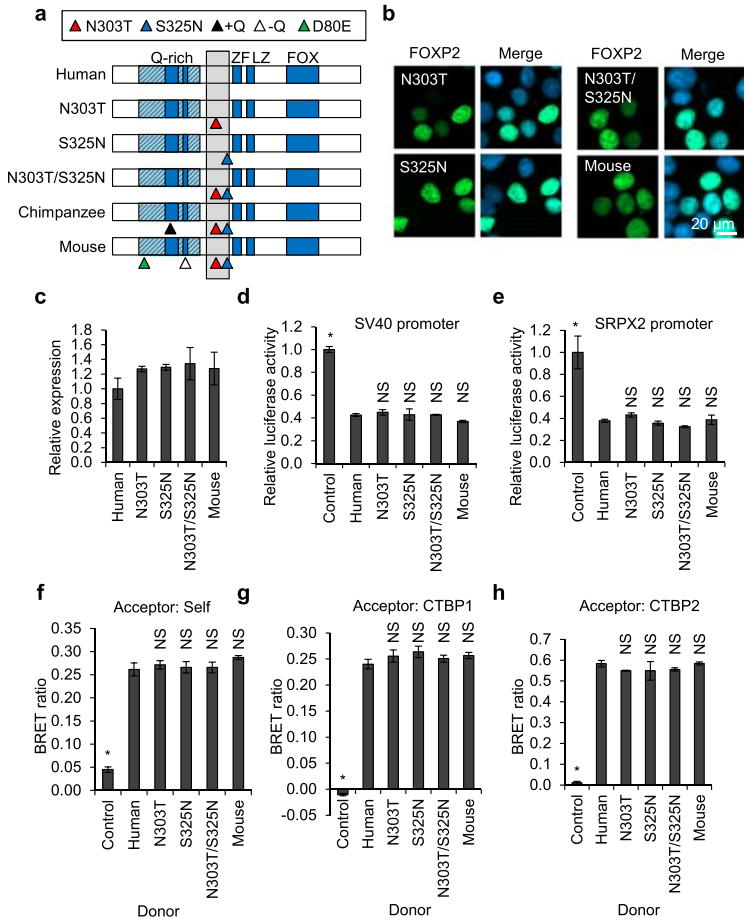


Fig. 5 (See legend on next page.)

(See figure on previous page.)
Fig. 5 Characterization of FOXP2 variants with ancestral amino acid substitutions. **a** Schematic representation of FOXP2 variants with ancestral amino acid substitutions. The N303T, S325N and N303T/S325N constructs are synthetic variants of human FOXP2 carrying ancestral amino acid substitutions. The chimpanzee and mouse orthologues of FOXP2 are included for comparison. Amino acid differences relative to human FOXP2 are indicated by arrowheads: N303T (*red*), S325N (*blue*), D80E (*green*), glutamine insertion (+Q, *black*), and glutamine deletion (–Q, *white*). Known domains are labelled: glutamine-rich (*Q-rich*) region (*hatched shading*) including polyglutamine tracts (*solid shading*), zinc finger (*ZF*), leucine zipper (*LZ*), and forkhead domain (*FOX*). The minimal CTBP interaction region determined using the BRET assay is indicated by the *grey-shaded box*. **b** Fluorescence micrographs of HEK293 cells transfected with FOXP2 variants with ancestral amino acid substitutions. Nuclei were stained with Hoechst 33342. **c** Fluorescence-based measurement of the expression level of FOXP2 variants with ancestral amino acid substitutions. HEK293 cells were transfected with YFP-FOXP2, together with mCherry for normalization. Fluorescence intensity was measured 48 h post-transfection. Values are mean YFP/mCherry fluorescence ratios ± S.D. ($n = 3$), relative to the value for human FOXP2. **d**, **e** Luciferase reporter assays for transcriptional regulatory activity of FOXP2 variants with ancestral amino acid substitutions. HEK293 cells were transfected with a luciferase reporter vector containing the SV40 promoter (**d**) or the human *SRPX2* promoter (**e**), together with a *Renilla* luciferase normalization plasmid, and YFP-FOXP2 or YFP alone (control). Values are mean relative luciferase activity ± S.D. ($n = 3$), expressed relative to the control. *Asterisks* indicate significant differences compared to human FOXP2 (*$p < 0.05$, one-way ANOVA followed by Bonferroni post hoc correction). *NS* not significant. Exact p values for **d** are <0.0001 for the control, >0.9999 for the N303T, S325N, and N303T/S325N variants and 0.2730 for the mouse protein. Exact p values for **e** are <0.0001 for the control and >0.9999 for the N303T, S325N, N303T/S325N and mouse variants. **f**–**h** BRET assays for protein-protein interactions of FOXP2 variants with ancestral amino acid substitutions. HEK293 cells were transfected with FOXP2 variants with ancestral amino acid substitutions fused to *Renilla* luciferase (donor), together with YFP (acceptor) fusions of the same FOXP2 variants (**f**), CTBP1 (**g**), or CTBP2 (**h**). The control donor protein is a nuclear-targeted luciferase and the control acceptor protein is a nuclear-targeted YFP. Values are mean corrected BRET ratios ± S.D ($n = 3$). *Asterisks* indicate significant differences compared to wild-type (WT) FOXP2 ($p < 0.05$, one-way ANOVA followed by Bonferroni post hoc correction). *NS* not significant. Exact p values for **f** are <0.0001 for the control, >0.9999 for the N303T, S325N, and N303T/S325N variants and 0.434 for the mouse protein. Exact p values for **g** are <0.0001 for the control and >0.9999 for the N303T, S325N, N303T/S325N and mouse variants. Exact p values for **h** are <0.0001 for the control, >0.9999 for the N303T, S325N, and mouse variants and 0.10 for the N303T/S325N variant

and tetrapod species typically have tracts of 35–41 residues (Additional file 5). We therefore reduced the length of the long tract in human FOXP2 to 3 residues. Similarly, the shorter tract was also reduced to 3 residues.

The two tract length reductions were engineered both individually and together (Fig. 6a). The resulting synthetic FOXP2 variants were assayed for a range of molecular properties including subcellular localization (Fig. 6b), expression level (Fig. 6c), transcriptional repression of the SV40 and *SRPX2* promoters (Fig. 6d,e), homodimerization (Fig. 6f), and interaction with CTBPs (Fig. 6g,h). Across these assays, there were no large differences between wild-type FOXP2 and versions of the protein with reduced polyglutamine tracts. Versions of FOXP2 lacking either the large tract or both tracts showed a small but statistically significant reduction in repressive ability in the SV40 luciferase reporter assay and also a small but statistically significant increase in homodimerization and CTBP1 interaction (Fig. 6e,f,g). However, these effects were not consistently observed across independent experiments and therefore may not represent genuine biological differences. The lack of any substantial differences between the wild-type protein and variants with reduced polyglutamine tracts in our assays suggests that the presence of the tracts may not have a marked impact on fundamental aspects of FOXP2 biology.

Discussion

Our detailed functional characterization of rare FOXP2 variants reported in individuals with neurodevelopmental disorders confirms the etiological role of the Q390Vfs*7 variant, which has similarly deleterious effects on protein function to the two previously characterized pathogenic variants, p.R553H and p.R328*, in our assays. The data presented here point to potential diversity in molecular mechanisms in *FOXP2*-related speech/language disorder. Variants which abolish both transcriptional regulatory activity and protein dimerization, such as the p.R328* and p.Q390Vfs*7 variants, may act as null alleles, resulting in haploinsufficiency of *FOXP2*. In contrast, variants which show a loss of transcriptional regulatory activity but retain the ability to dimerize, like the p.R553H variant, may additionally interfere with the functioning of wild-type protein, as has been suggested for comparable variants in FOXP1 [31]. Variants producing these dominant-negative effects might result in a more severe phenotype. However, it will be necessary to identify further cases of FOXP2-related speech/language disorder resulting from missense variants in order to assess differences in phenotypic outcome.

In our assays, the p.Q17L, p.M406T, p.P416T, or p.N597H variants do not display any substantial effects on the core functions of the FOXP2 protein, suggesting that they may be incidental variants. Notably, the p.N597H variant was previously described as a likely causal variant in a child with CAS [7]. While we cannot rule out that the p.N597H variant disrupts an aspect of FOXP2 function not tested here, subtle effects on protein function are unlikely to lead to a speech/language disorder of comparable severity to complete loss-of-function variants. This case illustrates that, given the high level of rare coding variation in the general population (The 1000 Genomes Project Consortium, 2015), novel rare variants in known disorder-related genes cannot automatically be regarded

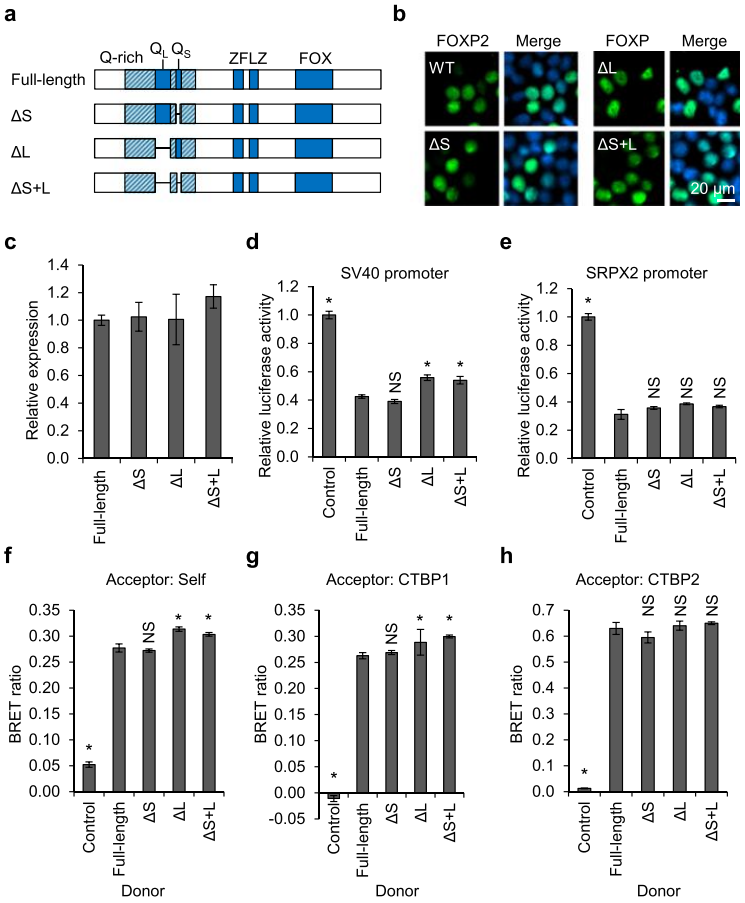


Fig. 6 Characterization of synthetic FOXP2 variants with reduced polyglutamine tracts. **a** Schematic representation of synthetic FOXP2 variants with reduced polyglutamine tracts. Known domains are labelled: glutamine-rich (*Q-rich*) region (*hatched shading*) including the long polyglutamine tract (Q_L) and short polyglutamine tract (Q_S), zinc finger (*ZF*), leucine zipper (*LZ*), and forkhead domain (*FOX*). The Δ*S* variant has a short polyglutamine tract reduced from 10 to 3 residues. The Δ*L* variant has a long polyglutamine tract reduced from 40 to 3 residues. The Δ*S* + *L* variant has shortened versions of both tracts. **b** Fluorescence micrographs of HEK293 cells transfected with FOXP2 variants with reduced polyglutamine tracts. Nuclei were stained with Hoechst 33342. **c** Fluorescence-based measurement of expression level for FOXP2 variants with reduced polyglutamine tracts. HEK293 cells were transfected with YFP-FOXP2 variants together with mCherry for normalization. Fluorescence intensity was measured 48 h post-transfection. Values are mean YFP/mCherry fluorescence ratios ± S.D. ($n = 3$), expressed relative to the value for full-length FOXP2. **d**, **e** Luciferase reporter assays for transcriptional regulatory activity of FOXP2 variants with reduced polyglutamine tracts. Cells were transfected with a luciferase reporter vector containing the SV40 promoter (**d**) or the human *SRPX2* promoter (**e**), together with a *Renilla* luciferase normalization plasmid, and YFP-FOXP2 or YFP alone (control). Values are mean relative luciferase activity ± S.D. ($n = 3$), expressed relative to the control. *Asterisks* indicate significant differences compared to full-length FOXP2 ($p < 0.05$, one-way ANOVA followed by Bonferroni post hoc correction). *NS* not significant. Exact *p* values for **d** are <0.0001 for the control, 0.4892 for Δ*S*, 0.0003 for Δ*L*, and 0.0010 for Δ*S* + *L*. Exact *p* values for **e** are <0.0001 for the control, 0.0768 for Δ*S*, 0.0633 for Δ*L*, and 0.1506 for Δ*S* + *L*. **f**–**h** BRET assays for protein-protein interactions of FOXP2 variants with reduced polyglutamine tracts. HEK293 cells were transfected with FOXP2 variants with reduced polyglutamine tracts fused to *Renilla* luciferase (donor) and YFP (acceptor) fusions of the same variants (**f**), CTBP1 (**g**) or CTBP2 (**h**). The control donor protein is a nuclear-targeted luciferase and the control acceptor protein is a nuclear-targeted YFP. Values are mean corrected BRET ratios ± S.D. ($n = 3$). *Asterisks* indicate significant differences compared to wild-type (WT) FOXP2 ($p < 0.05$, one-way ANOVA followed by Bonferroni post hoc correction). *NS* not significant. Exact *p* values for **f** are <0.0001 for the control, Δ*L* and Δ*S* + *L*, and >0.9999 for Δ*S*. Exact *p* values for **g** are <0.0001 for the control, >0.9999 for Δ*S*, 0.038 for Δ*L*, and 0.024 for Δ*S* + *L*. Exact *p* values for **h** are <0.0001 for the control, 0.4 for Δ*S*, and >0.9999 for Δ*L* and Δ*S* + *L*

as causal. This is particularly true if it is not possible to establish segregation of the variant with disorder (or de novo occurrence of a variant in sporadic disorder) or if the disorder is genetically heterogeneous and many genetic risk factors remain unknown, as is the situation for CAS. Functional characterization of novel rare FOXP2 variants is therefore essential to provide a concrete diagnosis of FOXP2-related disorder and to shed light on the aspects of protein function that are disrupted. Our experiments provide a framework for the characterization of novel FOXP2 variants uncovered through future genetic analyses of individuals with speech/language disorder.

Concurrent with the present study, two further cases of probable *FOXP2*-related speech/language disorder were added to the DECIPHER database (patient IDs 271859 and 271246) [53]. One DECIPHER case presented with delayed speech and language development, dysarthria, and pulmonic stenosis and carries a de novo p.R553H variant identical to that observed in the KE family. The probability of recurrence for this particular variant is elevated because it involves mutation of a CpG site. Indeed, pathological mutation events have been observed at the equivalent CpG site in the FOX proteins FOXC2, FOXE1, FOXF1, FOXL2, and FOXP1 [31, 54–57]. The second probable new case of *FOXP2*-related speech/language disorder in DECIPHER presented with delayed speech and language development, delayed fine motor development, strabismus and tall stature and carries a novel de novo stop-gain variant, p.R564*, truncating the FOXP2 protein within the FOX domain. These two new DECIPHER cases underline the characteristic manifestation of heterozygous FOXP2 disruption as a motor speech disorder. In addition, fine motor delays have now been reported in connection with two FOXP2 variants (p.R564* and p.Q390Vfs*7), suggesting that the effects of FOXP2 disruption on motor coordination may in some cases extend beyond the orofacial movements required for speech, consistent with the broader motor skill learning deficits observed in mice with heterozygous FOXP2 disruption [1, 9].

Very recently, while the current manuscript was under review, eight further families with potential *FOXP2*-related speech/language disorder have been newly described, based on sequencing data [58]. Although functional characterizations were not carried out in that study, the new cases suggest that there is some variation in the nature and severity of the speech and language deficits resulting from *FOXP2* disruption and that there may also be variable cognitive impairments and behavioural anomalies in affected individuals. Of particular note, the newly described cohort included two novel missense variants in the FOX domain of FOXP2, but unlike the p.R553H variant found in the KE family, these missense variants do not occur in the DNA-binding helix of the domain. Future functional analyses of these newly identified variants would therefore be valuable to reveal how they might interfere with protein function and to confirm that these variants play a causal role in the disorder. The description of these eight new families represents a step towards a 'genotype-first' approach to the characterization of the *FOXP2*-related phenotype [59]. However, sequencing of *FOXP2* in much larger cohorts of language-impaired individuals, preferably coupled to functional analyses, will be necessary to obtain a complete picture of the clinical spectrum associated with *FOXP2*-related disorder and to gain further insights into the role of FOXP2 in neurodevelopment.

In addition to assessing the effects of rare variants on FOXP2 protein function, we performed the first detailed characterization of the interactions between FOXP2 and the transcriptional co-repressors CTBP1 and CTBP2, finding that these interactions are independent of FOXP2 homodimerization and DNA binding. Our results point to an extended CTBP binding site in FOXP2 within residues 258–487, possibly encompassing multiple subsites. Previous yeast two-hybrid screens identified residues 122–382 in FOXP2 as sufficient for interaction with CTBPs [34]. We propose a narrower critical region of 71 residues, from residue 259 to 329. This minimal region contains both of the amino acid substitutions that have occurred since the divergence of the human and chimpanzee lineages. However, we did not find any effect of these substitutions on CTBP interaction, and the human and mouse versions of FOXP2 exhibit comparable levels of interaction with CTBPs. Changes in CTBP interaction do not therefore appear to underlie the neurobiological changes observed in mice carrying a partially humanized version of *Foxp2*. The amino acid changes on the human lineage may instead result in altered interaction with as yet unidentified protein interaction partners or in changes to post-translational modifications. In addition, we confirmed that the FOXP2 paralog FOXP1 also interacts with CTBPs, whereas FOXP4 does not, pointing to a divergence in mechanisms of transcriptional regulation within the FOXP subfamily.

The CTBPs are conserved vertebrate proteins that interact with several different transcription factors and mediate transcriptional repression primarily through recruitment of the histone deacetylases HDAC1/2 [36]. The CTBPs are widely expressed during embryonic development and have essential and partially overlapping roles in the development of multiple organs [35, 36]. They are expressed in the brain from a very early stage, although their specific roles in neurodevelopment have not been extensively investigated [35, 36]. CTBPs may therefore be key players in FOXP2- and FOXP1-mediated transcriptional repression across multiple organs and developmental stages, including in neurodevelopment. In addition, CTBPs may play a role in the post-translational modification of FOXP2. We and other groups have recently

demonstrated that FOXP2 is post-translationally modified by SUMOylation [21, 60, 61]. The essential SUMOylation enzyme UBC9 is a core component of the protein complex containing CTBP1, and CTBP1 has been suggested to function as a platform for protein SUMOylation [62]. Furthermore, the SUMO E3 ligase PIAS1, which promotes FOXP2 SUMOylation, is an auxilliary component of the CTBP1 complex [21, 62]. If CTBPs are among the key mediators of FOXP-mediated transcriptional repression, variants in FOXP2 or FOXP1 which disrupt the interaction with CTBPs might have a similar impact on neurodevelopment to variants which disrupt DNA binding, and an assessment of the effects on CTBP binding should therefore be included when characterizing novel, putatively pathogenic variants in FOXP proteins.

Finally, we reported the first detailed analysis of the role of the polyglutamine tracts in FOXP2 function. In our experiments, shortening the polyglutamine tracts in FOXP2 did not have a substantial impact on protein expression, nuclear localization, homodimerization, interaction with CTBPs, or transcriptional repression activity. The tracts do not therefore appear to be essential for core aspects of protein function in cellular assays. However, the tracts may have critical roles in vivo that are not apparent in cell models, and the generation of mice carrying a version of Foxp2 lacking the polyglutamine tracts could therefore be informative. Deletion of the polyglutamine tracts in a small number of other proteins has been found to affect behaviour in mouse models. For example, deletion of the polyglutamine tract in mouse huntingtin results in subtle behavioural changes relating to learning and memory and motor coordination [63]. Replacement of mouse POU3F2 with a *Xenopus* orthologue that lacks polyglutamine tracts and other homopolymer amino acid repeats resulted in neurochemical changes and a dramatic deterioration in pup retrieval behaviour [64]. The polyglutamine tract in mouse SRY has important roles in protein stabilization and transcriptional activation that are essential for sex determination [65]. It is also possible that the polyglutamine tracts in FOXP2 do not have critical roles in normal protein function. The tracts lie within a glutamine-rich protein region that is conserved in FOXP1 and FOXP4 and may be prone to developing expanded stretches of glutamine residues. In FOXP1, a different cluster of glutamine residues has undergone expansion in the rodent lineage to produce tracts of 37 residues in the mouse and 39 in rat, compared to 7 residues in primates. Expansions of up to ~40 glutamine residues within the glutamine-rich region of FOXP proteins may therefore have negligible effect on protein function.

The failure to observe any substantial effects of polyglutamine tract reduction on core aspects of FOXP2 function makes it less likely that tract length reductions contribute to risk for neurodevelopmental disorder. However, one recent study suggested that a FOXP2 polyglutamine tract length variant tentatively associated with speech sound disorder (p.Q172del) exhibits altered transcriptional repression activity in a cell model [18, 66]. The p.Q172del variant was reported to increase transcript levels of the FOXP2 target gene *CNTNAP2* in transfected cells, whereas transcript levels were reduced in cells transfected with wild-type protein [66]. Nonetheless, both the variant and wild-type forms of FOXP2 produced increases in protein levels of CNTNAP2, with the variant having a stronger effect [66]. It is therefore unclear what effect the p.Q172del variant might have on transcriptional regulation in vivo, but our experiments suggest that the effect of a single glutamine deletion is likely to be extremely small. The p.Q172del variant is observed in the general population with a minor allele frequency of 0.6 % in the ExAC database (714 of 115484 chromosomes), and larger tract length reductions are also present in the ExAC database at very low frequency. Studies of larger numbers of individuals with FOXP2 proteins with standard-length and reduced-length tracts are therefore necessary to determine if tract length reduction might confer an increased risk for neurodevelopmental disorder or a reduction in linguistic abilities.

Conclusions

By performing detailed functional characterization of rare variants in FOXP2 found in individuals with neurodevelopmental disorders, we confirm the causal role of one recently reported uncharacterized variant and provide further characterization of two further causal variants. In addition, we highlight two variants which had been suggested to contribute to the neurodevelopmental disorders in the affected individuals, but which do not have detrimental effects on protein function, suggesting that they are in fact incidental. These findings underline the importance of performing functional characterization of novel variants identified in individuals with neurodevelopmental disorders, even when the affected gene has previously been implicated in disorder. The provision of accurate molecular diagnoses in cases of neurodevelopmental disorder will require consideration of protein function data together with detailed clinical observations and genetic characterization of affected individuals and their parents on a genome-wide scale.

Our research also provides important new insights into the biology of the FOXP2 protein, expanding our knowledge of the molecular functions of this key neural transcription factor. Our characterization of the interaction of FOXP2 with co-repressor proteins of the CTBP family provides additional mechanistic insight into FOXP2-mediated repression of transcription. We also report the first detailed examination of the role of the polyglutamine tracts in FOXP2. We find that these tracts are not essential to the core molecular functions of the

FOXP2 protein, suggesting that variations in tract length are unlikely to be a highly penetrant cause of neurodevelopmental disorder.

Additional files

Additional file 1: DNA sequences of primers used for molecular cloning. (PDF 239 kb)

Additional file 2: DNA sequences of primers used for site-directed mutagenesis. (PDF 234 kb)

Additional file 3: Western blots of YFP-tagged FOXP2 variants. (PDF 521 kb)

Additional file 4: Summary of studies examining FOXP2 polyglutamine tract length variation in individuals with neurodevelopmental and neuropsychiatric disorders. (PDF 251 kb)

Additional file 5: FOXP2 polyglutamine tract lengths in selected vertebrates. (PDF 325 kb)

Abbreviations

BRET: Bioluminescence resonance energy transfer; CAS: Childhood apraxia of speech; CTBP: C-terminal binding protein; FOX: Forkhead Box; MRI: Magnetic resonance imaging; YFP: Yellow fluorescent protein

Acknowledgements

We wish to thank Elliot Sollis for help generating DNA constructs.

Funding

This work was supported by the Max Planck Society.

Availability of data and materials

The data supporting the conclusions of this article are included within the article and its additional files.

Authors' contributions

SEF and SAG designed and supervised the study. SBE, SAG, SMC and PD performed the experiments. SAG, SBE and SEF wrote the manuscript. All authors were involved in data analysis and interpretation, and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Received: 29 June 2016 Accepted: 8 November 2016
Published online: 28 November 2016

References

1. Groszer M, Keays DA, Deacon RM, de Bono JP, Prasad-Mulcare S, Gaub S, Baum MG, French CA, Nicod J, Coventry JA, Enard W, Fray M, Brown SD, Nolan PM, Paabo S, Channon KM, Costa RM, Eilers J, Ehret G, Rawlins JN, Fisher SE. Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits. Curr Biol. 2008;18:354–62. doi:10.1016/j.cub.2008.01.060.
2. Shu W, Cho JY, Jiang Y, Zhang M, Weisz D, Elder GA, Schmeidler J, De Gasperi R, Sosa MA, Rabidou D, Santucci AC, Perl D, Morrisey E, Buxbaum JD. Altered ultrasonic vocalization in mice with a disruption in the Foxp2 gene. Proc Natl Acad Sci U S A. 2005;102:9643–8. doi:10.1073/pnas.0503739102.
3. Graham SA, Fisher SE. Understanding language from a genomic perspective. Annu Rev Genet. 2015;49:131–60. doi:10.1146/annurev-genet-120213-092236.
4. Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. A forkhead-domain gene is mutated in a severe speech and language disorder. Nature. 2001;413:519–23. doi:10.1038/35097076.
5. Stroud JC, Wu Y, Bates DL, Han A, Nowick K, Paabo S, Tong H, Chen L. Structure of the forkhead domain of FOXP2 bound to DNA. Structure. 2006;14:159–66. doi:10.1016/j.str.2005.10.005.
6. Vernes SC, Nicod J, Elahi FM, Coventry JA, Kenny N, Coupe AM, Bird LE, Davies KE, Fisher SE. Functional genetic analysis of mutations implicated in a human speech and language disorder. Hum Mol Genet. 2006;15:3154–67. doi:10.1093/hmg/ddl392.
7. Laffin JJ, Raca G, Jackson CA, Strand EA, Jakielski KJ, Shriberg LD. Novel candidate genes and regions for childhood apraxia of speech identified by array comparative genomic hybridization. Genet Med. 2012;14:928–36. doi:10.1038/gim.2012.72.
8. MacDermot KD, Bonora E, Sykes N, Coupe AM, Lai CS, Vernes SC, Vargha-Khadem F, McKenzie F, Smith RL, Monaco AP, Fisher SE. Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. Am J Hum Genet. 2005;76:1074–80. doi:10.1086/430841.
9. Turner SJ, Hildebrand MS, Block S, Damiano J, Fahey M, Reilly S, Bahlo M, Scheffer IE, Morgan AT. Small intragenic deletion in FOXP2 associated with childhood apraxia of speech and dysarthria. Am J Med Genet A. 2013;161A:2321–6. doi:10.1002/ajmg.a.36055.
10. Gauthier J, Joober R, Mottron L, Laurent S, Fuchs M, De Kimpe V, Rouleau GA. Mutation screening of FOXP2 in individuals diagnosed with autistic disorder. Am J Med Genet A. 2003;118A:172–5. doi:10.1002/ajmg.a.10105.
11. Laroche F, Ramoz N, Leroy S, Fortin C, Rousselot-Paillet B, Philippe A, Colleaux L, Bresson JL, Mogenet A, Golse B, Mouren-Simeoni MC, Gorwood P, Galli T, Simonneau M, Krebs MO, Robel L. Polymorphisms of coding trinucleotide repeats of homeogenes in neurodevelopmental psychiatric disorders. Psychiatr Genet. 2008;18:295–301. doi:10.1097/YPG.0b013e3283060fa5.
12. Levchenko A, Davtian S, Petrova N, Malashichev Y. Sequencing of five left-right cerebral asymmetry genes in a cohort of schizophrenia and schizotypal disorder patients from Russia. Psychiatr Genet. 2014;24:75–80. doi:10.1097/YPG.0000000000000021.
13. Li H, Yamagata T, Mori M, Momoi MY. Absence of causative mutations and presence of autism-related allele in FOXP2 in Japanese autistic patients. Brain Dev. 2005;27:207–10. doi:10.1016/j.braindev.2004.06.002.
14. Newbury DF, Bonora E, Lamb JA, Fisher SE, Lai CS, Baird G, Jannoun L, Slonims V, Stott CM, Merricks MJ, Bolton PF, Bailey AJ, Monaco AP, International Molecular Genetic Study of Autism C. FOXP2 is not a major susceptibility gene for autism or specific language impairment. Am J Hum Genet. 2002;70:1318–27. doi:10.1086/339931.
15. Roll P, Vernes SC, Bruneau N, Cillario J, Ponsole-Lenfant M, Massacrier A, Rudolf G, Khalife M, Hirsch E, Fisher SE, Szepetowski P. Molecular networks implicated in speech-related disorders: FOXP2 regulates the SRPX2/uPAR complex. Hum Mol Genet. 2010;19:4848–60. doi:10.1093/hmg/ddq415.
16. Tolosa A, Sanjuan J, Dagnall AM, Molto MD, Herrero N, de Frutos R. FOXP2 gene and language impairment in schizophrenia: association and epigenetic studies. BMC Med Genet. 2010;11:114. doi:10.1186/1471-2350-11-114.
17. Wassink TH, Piven J, Vieland VJ, Pietila J, Goedken RJ, Folstein SE, Sheffield VC. Evaluation of FOXP2 as an autism susceptibility gene. Am J Med Genet. 2002;114:566–9. doi:10.1002/ajmg.10415.
18. Zhao Y, Ma H, Wang Y, Gao H, Xi C, Hua T, Zhao Y, Qiu G. Association between FOXP2 gene and speech sound disorder in Chinese population. Psychiatry Clin Neurosci. 2010;64:565–73. doi:10.1111/j.1440-1819.2010.02123.x.
19. Deriziotis P, Graham SA, Estruch SB, Fisher SE. Investigating protein-protein interactions in live cells using bioluminescence resonance energy transfer. J Vis Exp. 2014. doi:10.3791/51438.
20. Deriziotis P, O'Roak BJ, Graham SA, Estruch SB, Dimitropoulou D, Bernier RA, Gerdts J, Shendure J, Eichler EE, Fisher SE. De novo TBR1 mutations in sporadic autism disrupt protein functions. Nat Commun. 2014;5:4954. doi:10.1038/ncomms5954.
21. Estruch SB, Graham SA, Deriziotis P, Fisher SE. The language-related transcription factor FOXP2 is post-translationally modified with small ubiquitin-like modifiers. Sci Rep. 2016;6:20911. doi:10.1038/srep20911.
22. Mizutani A, Matsuzaki A, Momoi MY, Fujita E, Tanabe Y, Momoi T. Intracellular distribution of a speech/language disorder associated FOXP2 mutant. Biochem Biophys Res Commun. 2007;353:869–74. doi:10.1016/j.bbrc.2006.12.130.
23. Wang B, Lin D, Li C, Tucker P. Multiple domains define the expression and regulatory properties of Foxp1 forkhead transcriptional repressors. J Biol Chem. 2003;278:24259–68. doi:10.1074/jbc.M207174200.
24. Li S, Weidenfeld J, Morrisey EE. Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions. Mol Cell Biol. 2004;24:809–22.

25. Chatila TA, Blaeser F, Ho N, Lederman HM, Voulgaropoulos C, Helms C, Bowcock AM. JM2, encoding a fork head-related protein, is mutated in X-linked autoimmunity-allergic dysregulation syndrome. J Clin Invest. 2000; 106:R75–81. doi:10.1172/JCI11679.
26. d'Hennezel E, Bin Dhuban K, Torgerson T, Piccirillo CA. The immunogenetics of immune dysregulation, polyendocrinopathy, enteropathy, X linked (IPEX) syndrome. J Med Genet. 2012;49:291–302. doi:10.1136/jmedgenet-2012-100759.
27. Ferland RJ, Cherry TJ, Preware PO, Morrisey EE, Walsh CA. Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain. J Comp Neurol. 2003;460:266–79. doi:10.1002/cne.10654.
28. Li X, Wang W, Wang J, Malovannaya A, Xi Y, Li W, Guerra R, Hawke DH, Qin J, Chen J. Proteomic analyses reveal distinct chromatin-associated and soluble transcription factor complexes. Mol Syst Biol. 2015;11:775. doi:10.15252/msb.20145504.
29. Takahashi K, Liu FC, Hirokawa K, Takahashi H. Expression of Foxp4 in the developing and adult rat forebrain. J Neurosci Res. 2008;86:3106–16. doi:10.1002/jnr.21770.
30. Bacon C, Rappold GA. The distinct and overlapping phenotypic spectra of FOXP1 and FOXP2 in cognitive disorders. Hum Genet. 2012;131:1687–98. doi:10.1007/s00439-012-1193-z.
31. Sollis E, Graham SA, Vino A, Froehlich H, Vreeburg M, Dimitropoulou D, Gilissen C, Pfundt R, Rappold GA, Brunner HG, Deriziotis P, Fisher SE. Identification and functional characterization of de novo FOXP1 variants provides novel insights into the etiology of neurodevelopmental disorder. Hum Mol Genet. 2015. doi:10.1093/hmg/ddv495.
32. Corominas R, Yang X, Lin GN, Kang S, Shen Y, Ghamsari L, Broly M, Rodriguez M, Tam S, Trigg SA, Fan C, Yi S, Tasan M, Lemmens I, Kuang X, Zhao N, Malhotra D, Michaelson JJ, Vacic V, Calderwood MA, Roth FP, Tavernier J, Horvath S, Salehi-Ashtiani K, Korkin D, Sebat J, Hill DE, Hao T, Vidal M, Iakoucheva LM. Protein interaction network of alternatively spliced isoforms from brain links genetic risk factors for autism. Nat Commun. 2014;5:3650. doi:10.1038/ncomms4650.
33. Rolland T, Tasan M, Charloteaux B, Pevzner SJ, Zhong Q, Sahni N, Yi S, Lemmens I, Fontanillo C, Mosca R, Kamburov A, Ghiassian SD, Yang X, Ghamsari L, Balcha D, Begg BE, Braun P, Brehme M, Broly MP, Carvunis AR, Convery-Zupan D, Corominas R, Coulombe-Huntington J, Dann E, Dreze M, Dricot A, Fan C, Franzosa E, Gebreab F, Gutierrez BJ, Hardy MF, Jin M, Kang S, Kiros R, Lin GN, Luck K, MacWilliams A, Menche J, Murray RR, Palagi A, Poulin MM, Rambout X, Rasla J, Reichert P, Romero V, Ruyssinck E, Sahalie JM, Scholz A, Shah AA, Sharma A, Shen Y, Spirohn K, Tam S, Tejeda AO, Trigg SA, Twizere JC, Vega K, Walsh J, Cusick ME, Xia Y, Barabasi AL, Iakoucheva LM, Aloy P, De Las RJ, Tavernier J, Calderwood MA, Hill DE, Hao T, Roth FP, Vidal M. A proteome-scale map of the human interactome network. Cell. 2014;159:1212–26. doi:10.1016/j.cell.2014.10.050.
34. Sakai Y, Shaw CA, Dawson BC, Dugas DV, Al-Mohtaseb Z, Hill DE, Zoghbi HY. Protein interactome reveals converging molecular pathways among autism disorders. Sci Transl Med. 2011;3:86ra49. doi:10.1126/scitranslmed.3002166.
35. Hildebrand JD, Soriano P. Overlapping and unique roles for C-terminal binding protein 1 (CtBP1) and CtBP2 during mouse development. Mol Cell Biol. 2002;22:5296–307.
36. Stankiewicz TR, Gray JJ, Winter AN, Linseman DA. C-terminal binding proteins: central players in development and disease. Biomol Concepts. 2014;5:489–511. doi:10.1515/bmc-2014-0027.
37. Bergman LM, Morris L, Darley M, Mirnezami AH, Gunatilake SC, Blaydes JP. Role of the unique N-terminal domain of CtBP2 in determining the subcellular localisation of CtBP family proteins. BMC Cell Biol. 2006;7:35. doi:10.1186/1471-2121-7-35.
38. Verger A, Quinlan KG, Crofts LA, Spano S, Corda D, Kable EP, Braet F, Crossley M. Mechanisms directing the nuclear localization of the CtBP family proteins. Mol Cell Biol. 2006;26:4882–94. doi:10.1128/MCB.02402-05.
39. Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T, Monaco AP, Paabo S. Molecular evolution of FOXP2, a gene involved in speech and language. Nature. 2002;418:869–72. doi:10.1038/nature01025.
40. Li G, Wang J, Rossiter SJ, Jones G, Zhang S. Accelerated FoxP2 evolution in echolocating bats. PLoS One. 2007;2:e900. doi:10.1371/journal.pone.0000900.
41. Webb DM, Zhang J. FoxP2 in song-learning birds and vocal-learning mammals. J Hered. 2005;96:212–6. doi:10.1093/jhered/esi025.
42. Krause J, Lalueza-Fox C, Orlando L, Enard W, Green RE, Burbano HA, Hublin JJ, Hanni C, Fortea J, de la Rasilla M, Bertranpetit J, Rosas A, Paabo S. The derived FOXP2 variant of modern humans was shared with Neandertals. Curr Biol. 2007;17:1908–12. doi:10.1016/j.cub.2007.10.008.
43. Enard W, Gehre S, Hammerschmidt K, Holter SM, Blass T, Somel M, Bruckner MK, Schreiweis C, Winter C, Sohr R, Becker L, Wiebe V, Nickel B, Giger T, Muller U, Groszer M, Adler T, Aguilar A, Bolle I, Calzada-Wack J, Dalke C, Ehrhardt N, Favor J, Fuchs H, Gailus-Durner V, Hans W, Holzlwimmer G, Javaheri A, Kalaydjiev S, Kallnik M, Kling E, Kunder S, Mossbrugger I, Naton B, Racz I, Rathkolb B, Rozman J, Schrewe A, Busch DH, Graw J, Ivandic B, Klingenspor M, Klopstock T, Ollert M, Quintanilla-Martinez L, Schulz H, Wolf E, Wurst W, Zimmer A, Fisher SE, Morgenstern R, Arendt T, de Angelis MH, Fischer J, Schwarz J, Paabo S. A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. Cell. 2009;137:961–71. doi:10.1016/j.cell.2009.03.041.
44. Reimers-Kipping S, Hevers W, Paabo S, Enard W. Humanized Foxp2 specifically affects cortico-basal ganglia circuits. Neuroscience. 2011;175:75–84. doi:10.1016/j.neuroscience.2010.11.042.
45. Schreiweis C, Bornschein U, Burguiere E, Kerimoglu C, Schreiter S, Dannemann M, Goyal S, Rea E, French CA, Puliyadi R, Groszer M, Fisher SE, Mundry R, Winter C, Hevers W, Paabo S, Enard W, Graybiel AM. Humanized Foxp2 accelerates learning by enhancing transitions from declarative to procedural performance. Proc Natl Acad Sci U S A. 2014;111:14253–8. doi:10.1073/pnas.1414542111.
46. Enard W. FOXP2 and the role of cortico-basal ganglia circuits in speech and language evolution. Curr Opin Neurobiol. 2011;21:415–24. doi:10.1016/j.conb.2011.04.008.
47. Butland SL, Devon RS, Huang Y, Mead CL, Meynert AM, Neal SJ, Lee SS, Wilkinson A, Yang GS, Yuen MM, Hayden MR, Holt RA, Leavitt BR, Ouellette BF. CAG-encoded polyglutamine length polymorphism in the human genome. BMC Genomics. 2007;8:126. doi:10.1186/1471-2164-8-126.
48. Riley BE, Orr HT. Polyglutamine neurodegenerative diseases and regulation of transcription: assembling the puzzle. Genes Dev. 2006;20:2183–92. doi:10.1101/gad.1436506.
49. Bruce HA, Margolis RL. FOXP2: novel exons, splice variants, and CAG repeat length stability. Hum Genet. 2002;111:136–44. doi:10.1007/s00439-002-0768-5.
50. Faux N. Single amino acid and trinucleotide repeats: function and evolution. Adv Exp Med Biol. 2012;769:26–40.
51. Bonkowsky JL, Chien CB. Molecular cloning and developmental expression of foxP2 in zebrafish. Dev Dyn. 2005;234:740–6. doi:10.1002/dvdy.20504.
52. Itakura T, Chandra A, Yang Z, Xue X, Wang B, Kimura W, Hikosaka K, Inohaya K, Kudo A, Uezato T, Miura N. The medaka FoxP2, a homologue of human language gene FOXP2, has a diverged structure and function. J Biochem. 2008;143:407–16. doi:10.1093/jb/mvm235.
53. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, Van Vooren S, Moreau Y, Pettett RM, Carter NP. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet. 2009;84:524–33. doi:10.1016/j.ajhg.2009.03.010.
54. Baris I, Arisoy AE, Smith A, Agostini M, Mitchell CS, Park SM, Halefoglu AM, Zengin E, Chatterjee VK, Battaloglu E. A novel missense mutation in human TTF-2 (FKHL15) gene associated with congenital hypothyroidism but not athyreosis. J Clin Endocrinol Metab. 2006;91:4183–7. doi:10.1210/jc.2006-0405.
55. Beysen D, Moumne L, Veitia R, Peters H, Leroy BP, De Paepe A, De Baere E. Missense mutations in the forkhead domain of FOXL2 lead to subcellular mislocalization, protein aggregation and impaired transactivation. Hum Mol Genet. 2008;17:2030–8. doi:10.1093/hmg/ddn100.
56. Brice G, Mansour S, Bell R, Collin JR, Child AH, Brady AF, Sarfarazi M, Burnand KG, Jeffery S, Mortimer P, Murday VA. Analysis of the phenotypic abnormalities in lymphoedema-distichiasis syndrome in 74 patients with FOXC2 mutations or linkage to 16q24. J Med Genet. 2002;39:478–83.
57. Sen P, Yang Y, Navarro C, Silva I, Szafranski P, Kolodziejska KE, Dharmadhikari AV, Mostafa H, Kozakewich H, Kearney D, Cahill JB, Whitt M, Bilic M, Margraf L, Charles A, Goldblatt J, Gibson K, Lantz PE, Garvin AJ, Petty J, Kiblawi Z, Zuppan C, McConkie-Rosell A, McDonald MT, Peterson-Carmichael SL, Gaede JT, Shivanna B, Schady D, Friedlich PS, Hays SR, Palafoll IV, Siebers-Renelt U, Bohring A, Finn LS, Siebert JR, Galambos C, Nguyen L, Riley M, Chassaing N, Vigouroux A, Rocha G, Fernandes S, Brumbaugh J, Roberts K, Ho-Ming L, Lo IF, Lam S, Gerychova R, Jezova M, Valaskova I, Fellmann F, Afshar K, Giannoni E, Muhlethaler V, Liang J, Beckmann JS, Lioy J, Deshmukh H, Srinivasan L, Swarr DT, Sloman M, Shaw-Smith C, van Loon RL, Hagman C, Sznajer Y, Barrea C, Galant C, Detaille T, Wambach JA, Cole FS, Hamvas A, Prince LS, Diderich KE, Brooks AS, Verdijk RM, Ravindranathan H, Sugo E, Mowat D, Baker ML, Langston C, Welty S, Stankiewicz P. 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. 2013;34:801–11. doi:10.1002/humu.22313.
58. Reuter MS, Riess A, Moog U, Briggs TA, Chandler KE, Rauch A, Stampfer M, Steindl K, Glaser D, Joset P, Study DDD, Krumbiegel M, Rabe H, Schulte-

Mattler U, Bauer P, Beck-Wodl S, Kohlhase J, Reis A, Zweier C. FOXP2 variants in 14 individuals with developmental speech and language disorders broaden the mutational and clinical spectrum. J Med Genet. 2016. doi:10.1136/jmedgenet-2016-104094.
59. Stessman HA, Bernier R, Eichler EE. A genotype-first approach to defining the subtypes of a complex disease. Cell. 2014;156:872–7. doi:10.1016/j.cell.2014.02.002.
60. Meredith LJ, Wang CM, Nascimento L, Liu R, Wang L, Yang WH. The key regulator for language and speech development, FOXP2, is a novel substrate for SUMOylation. J Cell Biochem. 2015. doi:10.1002/jcb.25288.
61. Usui N, Co M, Harper M, Rieger MA, Dougherty JD, Konopka G. Sumoylation of FOXP2 regulates motor function and vocal communication through Purkinje cell development. Biol Psychiatry. 2016. doi:10.1016/j.biopsych.2016.02.008.
62. Kuppuswamy M, Vijayalingam S, Zhao LJ, Zhou Y, Subramanian T, Ryerse J, Chinnadurai G. Role of the PLDLS-binding cleft region of CtBP1 in recruitment of core and auxiliary components of the corepressor complex. Mol Cell Biol. 2008;28:269–81. doi:10.1128/MCB.01077-07.
63. Clabough EB, Zeitlin SO. Deletion of the triplet repeat encoding polyglutamine within the mouse Huntington's disease gene results in subtle behavioral/motor phenotypes in vivo and elevated levels of ATP with cellular senescence in vitro. Hum Mol Genet. 2006;15:607–23. doi:10.1093/hmg/ddi477.
64. Nasu M, Yada S, Igarashi A, Sutoo D, Akiyama K, Ito M, Yoshida N, Ueda S. Mammalian-specific sequences in pou3f2 contribute to maternal behavior. Genome Biol Evol. 2014;6:1145–56. doi:10.1093/gbe/evu072.
65. Zhao L, Ng ET, Davidson TL, Longmuss E, Urschitz J, Elston M, Moisyadi S, Bowles J, Koopman P. Structure-function analysis of mouse Sry reveals dual essential roles of the C-terminal polyglutamine tract in sex determination. Proc Natl Acad Sci U S A. 2014;111:11768–73. doi:10.1073/pnas.1406666111.
66. Zhao Y, Liu X, Sun H, Wang Y, Yang W, Ma H. Contactinassociated proteinlike 2 expression in SHSY5Y cells is upregulated by a FOXP2 mutant with a shortened polyglutamine tract. Mol Med Rep. 2015;12:8162–8. doi:10.3892/mmr.2015.4483.

NEDERLANDSE SAMENVATTING

Taal is een onderscheidende eigenschap van de mens. Taal gaat verder dan enkel het overbrengen van simpele en praktische informatie zoals andere diersoorten doen, het geeft ons de mogelijkheid om gedachten uit te wisselen, gevoelens uit te drukken en te praten over de meest abstracte zaken die we kunnen bedenken. Ondanks de complexiteit van taal, is het iets dat ons van nature gemakkelijk af gaat. Ons genoom bevat alle instructies voor het vormen van hersenen die voorbereid zijn voor taal, waarin veel verschillende genen aan en uit gezet worden om de hersencircuits te ontwikkelen die ervoor zorgen dat we gesproken taal kunnen opnemen uit onze omgeving. Genen bevatten de informatie om moleculen te vormen die bijna elke eigenschap in elke cel in ons lichaam bepalen.

Om te ontrafelen welke genen bijdragen aan onze taalvaardigheden, kunnen onderzoekers kijken naar mutaties in het genoom van personen met een taalstoornis. Op deze manier is *FOXP2* geïdentificeerd als het eerste gen dat in verband gebracht kon worden met een spraak- en taalstoornis. In een grote familie van meerdere generaties (de KE familie) was de helft van de familieleden gediagnostiseerd met een spraak- en taalprobleem. Alleen in de familieleden met het spraak- en taalprobleem werd een mutatie in *FOXP2* gevonden. Onderzoek wees uit dat deze mutatie de functie van het FOXP2 eiwit aantastte. Inmiddels zijn meer mutaties in FOXP2 ontdekt in mensen met een spraak- en taalstoornis, hoewel in de meeste gevallen de oorzaak van een spraak- en taalstoornis nog niet gevonden kan worden. Studies naar de functie van FOXP2 in cel- en diermodellen hebben tot op heden de meeste informatie verschaft over de moleculaire mechanismen van taal.

Het *FOXP2* gen komt tot expressie in specifieke hersengebieden en codeert voor een transcriptiefactor, een type eiwit dat de activiteit van veel andere genen reguleert. Als de functie van *FOXP2* verstoord is, kan het niet langer genexpressie reguleren. FOXP2 is onderdeel van de FOXP familie van eiwitten, samen met FOXP1 en FOXP4, die ook in het centrale zenuwstelsel tot expressie komen en overlappende, maar ook afzonderlijke functies hebben tijdens

de hersenontwikkeling. FOXP1 lijkt het meest op FOXP2 en is betrokken bij een syndroom met verstandelijke beperking en taalproblemen

Om de biologische mechanismen van spraak en taal verder op te helderen, is het grootste deel van dit proefschrift gewijd aan onderzoek naar de moleculaire functies van de FOXP familieleden. In hoofdstuk 2 tot en met 4 heb ik mij gericht op de identificatie en karakterisatie van eiwitten die interacteren met FOXPs. Het FOXP interactoom was nauwelijks in kaart gebracht en informatie hierover zou waardevolle inzichten kunnen geven over hoe deze eiwitten de hersen- en taalontwikkeling reguleren, en helpen om nieuwe taalgerelateerde kandidaatgenen te identificeren.

In hoofdstuk 2 heb ik de interactie tussen FOXP2 en PIAS eiwitten onderzocht. PIAS eiwitten zijn enzymen die SUMOylatie van andere eiwitten reguleren. SUMOylatie is een dynamische en omkeerbare vorm van posttranslationele modificatie, waarin SUMO moleculen aan een eiwit bevestigd worden, wat invloed kan hebben op de functie van het eiwit. Ik heb ontdekt dat PIAS1 met FOXP2 interacteert om de SUMOylatie van FOXP2 te reguleren, de eerste en enige posttranslationele modificatie van FOXP2 die tot nu toe bekend is. FOXP2 SUMOylatie vindt plaats op een specifieke plek in het eiwit die door de evolutie heen niet veranderd is, wat suggereert dat deze posttranslationele modificatie al vroeg in de evolutie een rol heeft gespeeld. FOXP1 en FOXP4 hebben eenzelfde SUMOylatie positie, en FOXP1 en FOXP4 interacteren ook met SUMO moleculen en PIAS eiwitten, wat er op duidt dat FOXP1 en FOXP4 ook geSUMOyleerd worden. De etiologische mutatie in FOXP2, ontdekt in de KE familie, zorgt voor een duidelijk verminderde SUMOylatie van het FOXP2 eiwit. In verschillende experimenten met celmodellen heb ik geen aanwijzingen kunnen vinden dat SUMOylatie aspecten van de functie van het FOXP2 eiwit verandert. Deze effecten zijn echter zo subtiel, dat ze wellicht alleen waargenomen kunnen worden in de context van een ontwikkelend organisme.

In hoofdstuk 3 heb ik de gegevens van eerdere experimenten van onze onderzoeksgroep over potentiële FOXP2 eiwitinteracties opnieuw geëvalueerd, en heb ik twee kandidaat FOXP2 interactoren, NONO en SFPQ, verder onderzocht. NONO en SFPQ zijn twee transcriptiefactoren die tot expressie komen in het centrale zenuwstelsel en met elkaar interacteren. De interactie tussen FOXP2 en NONO/SFPQ kon echter niet worden aangetoond in verschillende eiwitinteractie-experimenten. Deze resultaten suggereren dat deze eiwitten hoogstwaarschijnlijk geen fysiologische FOXP2 interactiepartners zijn en dat de methodolo-

gie van de eerdere experimenten gevoelig geweest zou kunnen zijn voor valspositieve resultaten.

Daarom heb ik in hoofdstuk 4 een nieuwe screening opgezet om eiwitinteracties van FOXP2, FOXP1 en FOXP4 te identificeren. Ik heb zeven nieuwe FOXP-interacterende transcriptiefactoren ontdekt (NR2F1, NR2F2, SATB1, SATB2, SOX5, YY1 en ZMYM2), waarvan van vijf bekend is dat ze een rol spelen in hersenontwikkeling. Deze transcriptiefactoren komen samen tot expressie met FoxP2 in de hersenschors en het cerebellum, wat betekent dat ze daar mogelijk genexpressie reguleren in vivo. Een aantal van deze interacties worden ernstig verstoord door etiologische mutaties in *FOXP1* en *FOXP2* die ontwikkelingsstoornissen met spraak- en taalproblemen veroorzaken. Dit zouden nieuwe potentiële moleculaire mechanismen kunnen zijn die deze ontwikkelingsstoornissen veroorzaken.

De FOXPs zijn niet de enige genen die gelinkt zijn aan spraak en taal. De genetische achtergrond van zulke eigenschappen is complex en veel meer andere genen – de meeste nog onbekend – dragen ook bij aan spraak en taal. In hoofdstuk 5 focus ik op een nieuw taal-gerelateerd gen, *BCL11A*. Mutaties in dit gen veroorzaken een syndroom met een verstandelijke beperking en taalproblemen. BCL11A is ook een transcriptiefactor en tot voor kort was de functie van dit gen enkel onderzocht in het bloed. De functies van BCL11A in de hersenen waren nog nauwelijks bestudeerd. De identificatie van *BCL11A* mutaties in kinderen met een ontwikkelingsstoornis belichtte de rol van dit gen in de hersenontwikkeling. In hoofdstuk 5 heb ik de effecten van drie mutaties in *BCL11A* op de functies van BCL11A in de cel onderzocht. Ik heb ontdekt dat deze drie mutaties de lokalisatie van BCL11A in de cel ernstig verstoren, een aantal BCL11A eiwitinteracties aantasten en de transcriptionele regulerende activiteit beïnvloeden. Deze resultaten bevestigen de etiologische rol van deze mutaties in ontwikkelingsstoornissen.

Het werk dat ik beschrijf in mijn proefschrift heeft de kennis vergroot over hoe bepaalde genen de taal- en hersenontwikkeling reguleren. De karakterisatie van nieuwe FOXP interactiepartners geeft nieuwe inzichten in de transcriptieprocessen in de ontwikkelende hersenschors, en geeft een nieuwe kijk op de moleculaire mechanismen die de FOXP functies reguleren in spraak- en taalontwikkeling. De functionele karakterisatie van de *BCL11A* mutaties bevestigt dat dit gen betrokken is in een nieuwe monogene ontwikkelingsstoornis en wijst op een mogelijke bijdrage van dit gen in de ontwikkeling van spraak en taal.

SUMMARY

Language is a defining feature of the human condition. Human language goes beyond transmitting simple and practical information like other species do, it also enables us to share our thoughts, express our feelings and talk about even the most abstract matters we can imagine. Given its high complexity, it is quite extraordinary that language comes so naturally to us. The instructions to build a language-ready brain are encoded in our genome, where a myriad of different genes will be switched on and off in order to develop the brain circuits that allow us to soak up language from our environment. Genes contain the information to create the molecules that determine almost every single feature of each cell in our body.

To decipher which genes contribute to our language abilities scientist can look for mutations in the genomes of individuals with language impairment. This is how *FOXP2* was identified as the first gene implicated in human speech and language. A mutation in *FOXP2* was found only in the affected individuals of a large multigenerational family (the KE family) in which half of their members presented speech and language difficulties. Following investigations revealed that the mutation disrupted FOXP2 protein function. Since then multiple other cases of FOXP2 mutations leading to speech and language impairment have been discovered, although most cases of speech and language problems remains unexplained. The study of FOXP2 function in cellular and animal models has provided to date the vast majority of knowledge on the molecular mechanisms underlying language function.

The *FOXP2* gene is expressed in specific areas of the brain and encodes for a transcription factor, a type of protein which regulates the switching on and off of many genes. When disrupted, *FOXP2* can no longer regulate gene expression. FOXP2 belongs to the FOXP family of proteins together with FOXP1 and FOXP4, which are also expressed in the central nervous system and play overlapping yet distinct roles in brain development. FOXP1 is the most similar protein to FOXP2 and has been implicated in an intellectual disability syndrome that includes language impairment.

To shed light on the biological foundations of speech and language, the major part of this dissertation further investigates the molecular function of the FOXP family of proteins. In Chapters 2 to 4 I sought to identify and characterize which proteins interact with the FOXPs. The FOXP interactome had been scarcely explored and promised to reveal valuable insights into how these proteins function to regulate brain and language development, as well as to pinpoint novel language-related candidate genes.

In Chapter 2, I studied the interaction between FOXP2 and the PIAS family of proteins, which are enzymes that mediate the SUMOylation of other proteins. SUMOylation is a dynamic and reversible post-translational modification that consists in the physical attachment of SUMO molecules to proteins in order to finely modulate protein function. I discovered that PIAS1 interacts with FOXP2 to mediate its SUMOylation which is the first and only post-translational modification ever reported for FOXP2. I found that FOXP2 SUMOylation takes places at a specific site of the protein that is evolutionarily conserved, suggesting an ancient role for this post-translational modification. The SUMOylation site is also present in FOXP1 and FOXP4, for which I found that they also interact with SUMO molecules and PIAS proteins, suggesting that they are also subject to SUMOylation. Finally I observed that the etiological mutation in *FOXP2* found in the KE family leads to substantially reduced SUMOylation of the encoded protein. Using a range of diverse cellular assays, I did not observe that SUMOylation altered any aspect of FOXP2 protein function, however the effects of SUMOylation are often so subtle that they may only detectable in the context of a developing organism.

In Chapter 3, I re-evaluated a dataset of FOXP2 putative protein interactors from prior experiments performed in our group and selected two promising-candidate FOXP2 interactors, NONO and SFPQ for targeted investigation. NONO and SFPQ are two neural transcription factors that interact with each other. The interaction between FOXP2 and NONO/SFPQ was consistently not detected in any of the several protein-protein interactions assays I employed. This suggests that these proteins are unlikely to be true physiological FOXP2 interaction partners and that the methodology used in the previous experiments may have been vulnerable to false positive results.

Therefore, in Chapter 4, I set up fresh screens to identify protein-protein interactions of FOXP2, FOXP1 and FOXP4. I discovered and characterized seven novel FOXP-interacting transcription factors (NR2F1, NR2F2, SATB1, SATB2,

SOX5, YY1 and ZMYM2), five of which have well-established roles in brain development. I found that these transcription factors are co-expressed with FoxP2 in the brain cortex and in the cerebellum, suggesting that they may regulate gene expression *in vivo* in these neural sites of co-expression. Several of these interactions were severely disrupted by etiological mutations in *FOXP1* and *FOXP2* known to cause neurodevelopmental disorders that include speech and language, which reveals novel putative molecular mechanisms underlying these disorders.

The FOXPs are not the only genes implicated in speech and language. The genetic architecture underlying these traits is complex and many other genes – although most of them unkown – also contribute to this trait. In Chapter 5, I focus on a novel candidate language-related gene, *BCL11A*, mutations in which cause an intellectual disability syndrome that includes language impairment. BCL11A is also a transcription factor and, until recently, it had been primarily studied by its role in the blood. Its function in the brain had been rarely explored. The discovery of mutations in *BCL11A* in children with neurodevelopmental disorders highlighted a role for this gene in brain development. In Chapter 5, I investigated how three mutations in *BCL11A* affect its cellular functions. I discovered that the three mutations severely disrupt BCL11A localization inside the cell, some of its protein interactions and also its transcriptional regulatory activity, therefore confirming its etiological role in neurodevelopmental disorder.

The work in this dissertation has expanded a bit more the knowledge on how certain genes regulate language and brain development. The characterization of novel FOXP-interactors revealed novel insights into the transcriptional architecture underlying cortical development and provided fresh understanding of the molecular mechanisms that regulate FOXP function in language and speech development. The functional characterization study of *BCL11A* mutations confirmed its involvement in a novel monogenic neurodevelopmental disorder and highlighted its putative contributions to speech and language.

ACKNOWLEDGEMENTS

Just as transcription factors wouldn't be able to function without interacting with other proteins, the work presented in this dissertation would not have been possible without all the humans I've interacted with during these past years.

I've had the luxury of having two great thesis supervisors. Simon, *thank you* for giving me the opportunity to pursue my PhD project in your incredible department, and for all your valuable help during these years, your feedback always hit the nail on the head. Sarah, I've learnt so much from you I can't thank you enough. I will never be able again to do an experiment that doesn't include a hundred control conditions, at least, and I will forever be fearful of "scary capitals" and messy labs with no tags. *Thank you* for everything.

My collaborators during these years have been also crucial to complete this dissertation. Christina and Darren made it very easy and efficient to collaborate in the BCL11A project. *Thanks* for the synergy. Martí and Raymond welcomed me in their lab and taught me how to purify transcription factor complexes, identify its components and helped interpreting the results from mass spectrometry, a key piece of work in this thesis. *Bedankt*, *Gràcies*!

I've also been lucky to work in the Protein Group, which has been like a family to me. Pela, ευχαριστώ πολύ for always having my back, for being my career and life coach, for having the best sense of humor in the lab and for all the raspberry brownies and calzones. Ary, *Grazie mile* for always making me laugh with your stories, for all the baking and for all your help in the lab. Elliot, we've walked the same path side by side and I couldn't be happier about that. *Thank you*, for being a wonderful companion inside and outside the lab. Fabian, Lot, Joery, Swathi and Moritz, *Bedankt*, *Danke*, *Thanks "a Lot a Lot"* to you too for adding even more good vibes to the Protein Group and for all the insightful scientific discussions!

I want also to give a huge *Thanks* to the other people in the lab, Pedro, Paolo, Past, Present and Future Kai, Midas, Martin, Jasper, Jurgen, Janine, Laura, and Joses. *Thanks*, *Grazie*, *Gracias*, *Bedankt* and *Danke* for all the coffee, surreal conversations and all the never-ending fun! And, *sí*, Paolo, *Grazie mile* for always

making me think about the bigger picture, but even more for introducing me to Battista and Pocket Coffee, they've fueled the work in this thesis.

Thanks to the rest of the Language & Genetics department too! I started my PhD knowing little about bioinformatics, I am happy that after four years of biweekly department meetings I understand 20% of your talks and that, after Amaia has fixed all the errors for me, I can even press the "run" button on R! Speaking of which, Amaia, it wasn't friendship at first sight but you've become one of my closest friends and my best scientific colleague. Coffee breaks with a hint of science should be mandatory for any researcher that wants to keep himself mentally sane. *Eskerrik asko* for all the priceless *xerrameca*. A special *Danke* to Martina for taking care of all of us in the department, coordinating everyone's agendas and making sure all the myriad of non-scientific tasks were done in time so I could complete this dissertation.

It's been a pleasure to work in such an interdisciplinary institute. Most of the times I felt out of my comfort zone, but this has enriched me as a scientist, it's opened my mind and made me think outside the box. *Thanks*, also to all the friends I've made in the MPI. Suz & Lisa, *Dank je wel* and *Merci beacoup* for being my go-to duo when I needed a cocktail of fun, sarcasm and dark humor (perhaps in a rooftop). Ewelina, I just want to say one thing to you *Jesteś z zelaza*. Sara, *Grazie* for filling my last year with your positive energy. But also Gaby, Valeria, Amie, Harald, Francisco, Nietzsche, Ashley and many others, *Thanks* for all the amusing moments we've shared. Nijmegen have surpassed my initial expectations thanks to these guys but, specially, thanks to Varun and Amaia. *Eskerrik asko* for all the food, movies, walks in the woods, trips and your warm company. It has been *orixi bera*!

I won the lottery with my friends from Barcelona. They're all crazy in their own particular way and that's why I love them. They are now spread around the world, I feel the long distances but we keep the closest friendships. And I want to thank them all because they've shaped my personality and always boost my confidence. Laia was there when I was 4, she still is and forever will be, *Gràcies*, *my person*. Martí, you were my first scientific collaborator in our high school research project on polyploid strawberries, who would have thought we'd be collaborating in our PhDs too! It's a dream come true. Marta, we met one day, we started talking and we haven't stop since, *Gràcies* for always listening to me. Núria, spending time with you just feels naturally good, *Gràcies* for always being honest. Albert, it's all about feeling and we have an incredible one, *Grà-*

cies. Xavi, it may not look like it when we're talking but after our usually long conversations I always have the clearest thoughts, *Gràcies*. *Gràcies també* to the *neosímies* Cris and Alexita for your fresh friendship and to Victor, my mentor, supervisor and friend during my Master's thesis. You gave me the final push to move to the Netherlands to pursue this PhD. Estel, Mireia i Lluis, *Gràcies* for all the evening walks in Irvine, I really needed them after long days of writing while taking care of a baby.

Moltes Gràcies to my family back home. My parents set no limits for me, they are the most easygoing persons I know and I am here because of them, *Mil gràcies*. Anna, Toni, Cristina, Miquel, Mar, Pol, Ramon, Marian iaia Maria, avi Jordi, iaia Marta, avi Xavier i *els incontables* Estruchs, *no podia haver crescut rodejada de millor gent*. Familia Lobo, *Gràcies* for always making me feel like at home and for all your support and encouragement.

Finally, my best friend, "my team", my partner in life, *Gràcies amb majúscules*, Joan. And *Gràcies també* to little Ona, my most successful experiment, pure perfection.

To everyone who made this work possible, Thank You!

Sara Busquets Estruch
Irvine, California, April 2018

BIOGRAPHY

Sara Busquets Estruch was born on the 16th March of 1988 in Barcelona (Spain). She studied Biology at the University of Barcelona where she also received her Msc. degree in Biomedical Sciences in 2012. She carried out her master thesis in the Institute of Biomedicine of the University of Barcelona (IBUB) where she characterized protein-protein interactions of the neural transcription factor NR2E1.

After graduating in 2012 she decided to move to the Netherlands to carry out a PhD at the Language and Genetics department in the Max Planck Institute for Psycholinguistics in Nijmegen (the Netherlands) under the supervision of Prof. Simon Fisher and Dr. Sarah Graham. There she continued performing proteomic studies involving neural transcription factors, this time focusing in the FOXP family of proteins and also specialized in functional genetics by investigating the effect of genetic mutations that lead to neurodevelopmental disorders on protein function. The research during her PhD has led to this dissertation entitled "Characterization of transcription factors in monogenic disorders of speech and language".

LIST OF PUBLICATIONS

Estruch S.B., Quevedo-Calero M., Graham S.A., Vino A., Sollis E., Deriziotis P., Fisher S.E. (2018). Proteomic analysis of FOXP proteins reveals interactions between cortical transcription factors associated with neurodevelopmental disorders. Human Molecular Genetics. doi:10.1093/hmg/ddy035.

Estruch S.B.§, Graham S.A.§, Chinnappa S.M., Deriziotis P., Fisher S.E. (2016). Functional characterization of rare FOXP2 variants in neurodevelopmental disorder. Journal of Neurodevelopmental Disorders, 8: 44. doi:10.1186/s11689-016-9177-2.

Dias C.§, **Estruch S.B.**§, Graham S.A., McRae J., Sawiak S.J., Hurst J.A. , Joss S.K., Holder S.E., Morton J.E.V., Turner C., Thevenon J., Mellul K., Sanxhez-Andrade G., Ibarra-Soria X., Deriziotis P., Santos R.F., Lee S., Faivre L., Kleefstra T., Liu P., Hurles M.E., DDD Study, Fisher S.E., Logan D.W. (2016). BCL11A haploinsufficiency causes an intellectual disability syndrome and dysregulates transcription. The American Journal of Human Genetics, 99(2), 253-274. doi:10.1016/j.ajhg.2016.05.030.

Estruch S.B., Graham S.A., Deriziotis P., Fisher S.E. (2016). The language-related transcription factor FOXP2 is post-translationally modified with small ubiquitin-like modifiers. Scientific reports, 6:20911. 10.1038/srep20911.

Deriziotis P., O'Roak B. J., Graham S. A., **Estruch S. B.**, Dimitropoulou D., Bernier R. A., Gerdts J., Shendure J., Eichler E. E., Fisher S. E. (2014). De novo TBR1 mutations in sporadic autism disrupt protein functions. Nature Communications, 5: 4954. doi:10.1038/ncomms5954.

Deriziotis P.§, Graham S. A.§, **Estruch S. B.**, Fisher S. E. (2014). Investigating protein-protein interactions in live cells using Bioluminescence Resonance En-

§ indicates equal contribution

ergy Transfer. Journal of visualized experiments, 87: e51438. doi:10.3791/51438.

Estruch S.B., Buzón V., Carbó L. R., Schorova L., Lüders J., Estébanez-Perpiñá E. (2012). The Oncoprotein BCL11A Binds to Orphan Nuclear Receptor TLX and Potentiates its Transrepressive Function. PLoS one, 7(6), e37963. doi:10.1371 /journal.pone.0037963.

Carbó L.R., Buzón-Redorta V., **Estruch S.B.**, Fletterick R.J., Estébanez-Perpiñá E. (2012). A Surface on the Androgen Receptor that Allosterically Regulates Coactivator Binding. Molecular and Cellular Endocrinology, 348(2), 394-402. doi:10.1016/j.mce.2011.08.012.

MPI SERIES IN PSYCHOLINGUISTICS

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing. *Miranda van Turennout*
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-theore-tical 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 elec-tro-encephalography, 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 Dingemanse*

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. lncrementality 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 Gisladottir*

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, morphosyntatic, 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 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 (LSA). *Elisabeth Manrique*

120. Talking Sense: the behavioural and neural correlates of sound symbolism *Gwylm 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. *William 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 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: Studies on sentence production and comprehension in Tagalog and German. *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*