On the identification of FOXP2 gene enhancers and their role in brain development

Proefschrift ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof.dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op
woensdag 26 oktober 2016
om 16:30 uur precies
door

Martin Becker
geboren op 21 maart 1985
te Berlijn (Duitsland)

Promotor

Prof. dr. Simon E. Fisher

Copromotor

Dr. Sonja C. Vernes (Max Planck Instituut voor Psycholinguistiek)

Manuscriptcommissie

Prof. dr. Hans H.L.M. van Bokhoven

Dr. Annette Schenck

Prof. dr. Wolfgang Enard (Ludwig-Maximilians- Universität München, Duitsland)

On the identification of FOXP2 gene enhancers and their role in brain development

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 Wednesday, October 26, 2016
at 16:30 hours
by

Martin Becker
Born on March 21, 1985
in Berlin (Germany)

Supervisor

Prof. dr. Simon E. Fisher

Co-supervisor

Dr. Sonja C. Vernes (Max Planck Institute for Psycholinguistics)

Doctoral Thesis Committee

Prof. dr. Hans H.L.M. van Bokhoven

Dr. Annette Schenck

Prof. dr. Wolfgang Enard (Ludwig-Maximilians- Universität München, Germany)

"Take responsibility for making your own life beautiful."

By Timothy Leary,

Your Brain Is God

Cover design: Antje Märtin, Martin Becker

Printed by: Ipskamp Drukkers, Enschede

Contents

Chapter 1

General introduction

Introduction

The study of genes in health and disease has dramatically increased our understanding of human biology. The identification of mutated genes in diseases of the kidney, liver, muscle or virtually any other human tissue enabled us to study the function of these tissues at molecular and cellular levels. In this dissertation, I describe investigations of the *FOXP2* gene, which has been found to be mutated in people with speech and language problems. The relation between *FOXP2* and human language was first discovered nearly 15 years ago (Lai et al., 2001) and has been described “as a molecular window into speech and language” (Fisher and Scharff, 2009). My aim was to change angles and look through that window at the molecular mechanisms that precede *FOXP2*.

In this chapter I will introduce *FOXP2*, starting with the discovery of this gene in a family with a speech and language disorder. I will describe the expression during development and in adult brains and summarize the current knowledge regarding the downstream molecular and cellular functions of *FOXP2*. Next, I will review the current literature on the upstream processes that may regulate *FOXP2* and show that this aspect of the *FOXP2* story is not well understood. At the end of this chapter, I will formulate the overarching question and aims of this dissertation.

Speech and language problems in people with FOXP2 mutations

In 1990, a large three-generation family was described, of which about 50 percent of the family members presented with primary deficits in speech and language (Gopnik, 1990; Hurst et al., 1990) (Figure 1). A major aspect of the disorder is developmental verbal dyspraxia (DVD), also known as childhood apraxia of speech (CAS) (Hurst et al., 1990). The affected members of this family, generally referred to as the KE family, have severe problems with articulation, as well as poor processing and production of grammatical structures. The articulatory problems involve difficulties in performing the rapid coordinated sequences of oral and facial movements required for speech (Hurst et al., 1990; Vargha-Khadem et al., 1995). In addition to the orofacial motor control problems, the affected members showed problems with language, including

reduced ability in applying grammatical rules governing tenses and plurals (Gopnik and Crago, 1991; Vargha-Khadem et al., 1995) and lower scores on tests for the understanding of grammar (Vargha-Khadem et al., 1995). In terms of general cognition, both unaffected and affected members scored on the lower side of the normal range of the general population. However, a low IQ did not co-segregate with the speech and language disorder (Vargha-Khadem et al., 1995). It has been argued by some that the speech and language problems in this disorder stem from a core deficit affecting the coordination of orofacial motor-movements (Vargha-Khadem et al., 2005).

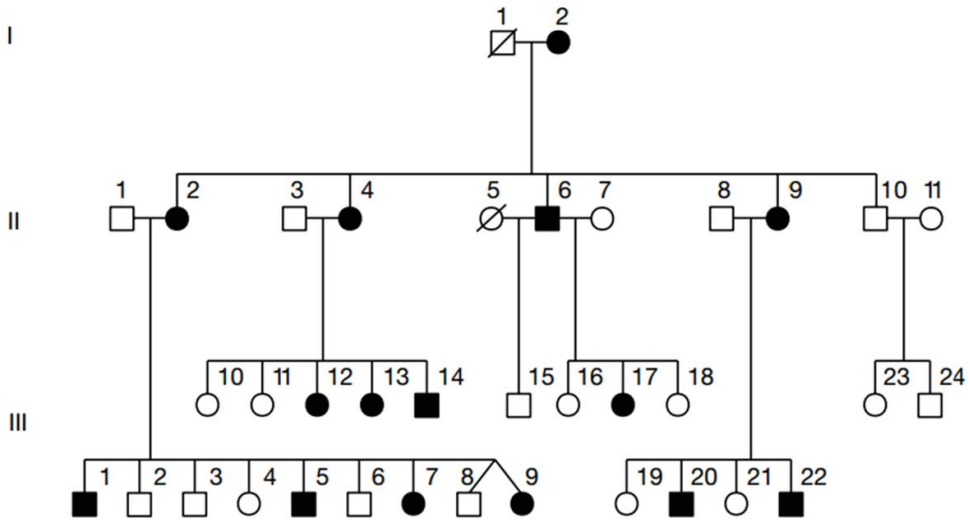

Figure 1: Pedigree of the KE family reproduced from Lai et al. 2001. Affected individuals are indicated by filled symbols. Squares are males, circles are females, and a line through a symbol indicates that the person was deceased in 2001.

Analyses of the phenotype of the KE family suggested that disrupted cognitive and neural motor-control functions underlie the observed problems with speech and language development. To identify affected brain regions in the KE family, brain imaging studies were performed on healthy and affected members. In structural magnetic resonance imaging (MRI) studies the grey matter volumes of the affected family members were compared to the unaffected members and an unrelated control group, using voxel based morphometry (Vargha-Khadem et al., 1998; Watkins et al., 2002; Belton et al., 2003). The volumes of the head of the

caudate nucleus and cerebellar lobule VIIIB were consistently reduced in comparison to control groups, whereas the volume of the putamen was increased. In addition, structural differences were detected in cortical areas involved in motor control and language processing, including increased volumes of the posterior superior temporal gyrus (including Wernicke's area) and angular gyrus, as well as decreased inferior frontal gyrus (including Broca's area) and precentral gyrus. The affected brain structures form cortico-striatal and cortico-cerebellar circuits, which are important for planning and exerting motor movements (Middleton and Strick, 2000). In complementary functional brain studies, the caudate nucleus showed increased activity in PET scans performed during word repetition tasks (Vargha-Khadem et al., 1998) and the putamen reduced activity in fMRI scans during semantic language tasks, where the subjects had to generate verbs in combination to heard nouns (Liegeois et al., 2003; Liegeois et al., 2011). In addition, in these studies, under- and over-activation of cortical areas involved in orofacial motor control, such as the precentral gyrus, and language-related cortical areas, such as the inferior frontal gyrus, were detected (Vargha-Khadem et al., 1998; Liegeois et al., 2003; Liegeois et al., 2011). Thus, structural and functional aberrations overlap in the basal ganglia, cerebellum, motor-control area and language-related cortical areas in the affected KE family members. Based on the observed effects in cortico-striatal and cortico-cerebellar networks, it has been suggested that the speech and language phenotype result from problems in planning and performing sequences of orofacial movements (Vargha-Khadem et al., 2005).

The mode of inheritance in the KE family suggested that a mutation in a single autosomal gene might account for the speech and language disorder of this family. In 1998, the hypothesis of an autosomal dominant locus was confirmed by molecular studies, and the location of the affected gene was narrowed down to a region on chromosome 7 (Fisher et al., 1998). Using genome-wide linkage analysis in the KE family, the authors identified a region on chromosomal band 7q31 that perfectly segregated with affection status. Further clues to the location of the likely damaged gene came from identification of an unrelated case, referred to as CS (Lai et al., 2000), with a strikingly similar phenotype to that seen in the KE family. This child carried a

de novo balanced translocation involving reciprocal exchanges between chromosomes 7 and 5, with one breakpoint directly disrupting a newly identified gene on chromosome 7q31, in the region of linkage identified in the KE family. This novel gene, given the name *FOXP2*, was characterized and subsequently sequenced in the KE family, revealing a non-synonymous mutation in all fifteen affected family members, which was not present in unaffected members or healthy controls (Lai et al., 2001). The disruption of *FOXP2* in CS and the KE family strongly suggested that disruption of this gene was responsible for their speech and language deficits. Indeed, the scientific literature now describes multiple independent cases with speech problems and mutations affecting the coding sequence of *FOXP2,* including deletions of the whole gene (Feuk et al., 2006; Zeesman et al., 2006; Lennon et al., 2007; Palka et al., 2012; Rice et al., 2012; Zilina et al., 2012), chromosomal rearrangements disrupting the gene (Lai et al., 2001; Feuk et al., 2006; Shriberg et al., 2006), missense mutations (Lai et al., 2001; MacDermot et al., 2005), a nonsense mutation (MacDermot et al., 2005) and a two-nucleotide intragenic deletion that yields a frameshift mutation and premature stop-codon (Turner et al., 2013).

Neuronal expression of FOXP2 during development and in adult brains

The expression of the *FOXP2* gene has been described in numerous species. To refer to genes and gene products I will use the standardized nomenclature to distinguish between gene orthologues and gene products (Wain et al., 2002). Genes, including their mRNA, are generally written in *italics* and the proteins are referred to in non-italic style. Human gene symbols are written in all capital letters (eg. *FOXP2* gene, FOXP2 protein*)*. According to standard nomenclature, I will refer to mouse and rat genes with a starting capital letter (eg. *Foxp2),* zebrafish genes with all lower case letters (*foxp2)* and zebra finch genes with a starting and end capital letter (eg. *FoxP2).*

Foxp2 is expressed in multiple tissues including the central nervous system, gut, lung and heart (Table 1) (Shu et al., 2001). Expression starts during embryonic development and

continues postnatally and into adulthood. Since the speech and language deficits in the KE family relate to alterations in the central nervous system (as described above), I focus here on reviewing the expression in the brain. First, I describe the expression pattern during embryonic development and then continue on to detail the findings for postnatal development and in adulthood. In addition I have summarized the findings of the systematic *FOXP2/Foxp2* expression studies in mice and humans in table 1.

The earliest neural expression of *FOXP2/Foxp2* can be detected in the developing medulla oblongata at embryonic day 11.5 (E11.5) in mice and at Carnegie stage 23 (CS23) in humans (Lai et al., 2003). At later embryonic stages the inferior olives of the medulla oblongata can be identified as the *FOXP2/Foxp2* positive nuclei. During further embryonic development *Foxp2* mRNA was detected at E12.5 in the lateral ganglionic eminences (GE) and expression remained in the GE-derived embryonic and postnatal striatum (Ferland et al., 2003). In humans, *FOXP2* mRNA can be detected at comparable developmental stages (CS23) in the developing striatum (Lai et al., 2003). At E14.5 mRNA is detected in the cortical plate and ventricular zone of mice (Ferland et al., 2003; Hisaoka et al., 2010). During embryonic and post-natal development cortical mRNA expression can be detected in the cortical plate and ventricular zone (Ferland et al., 2003; Lai et al., 2003). However, Foxp2 protein expression was only detected in the cortical plate and was absent from the ventricular zone. Similarly, in human prenatal brains *FOXP2* expression can be detected in the subplate and inner cortical plate (Miller et al., 2014). *FOXP2/Foxp2* is not expressed in the hippocampal formation (Ferland et al., 2003; Lai et al., 2003). Additional forebrain regions that showed strong embryonic expression were the developing thalamus and hypothalamus in mouse embryos (E13.5) and human foetuses (Ferland et al., 2003; Lai et al., 2003). Further expression was detected in the amygdala and olfactory bulb and tubercule of mice (Ferland et al., 2003). In addition to forebrain structures, expression was detected in developing mid- and hindbrain regions (Ferland et al., 2003; Lai et al., 2003). Midbrain expression was detected in the substantia nigra and tectum in mice (E16.5) and human at foetal stage 1 (FS1) (Lai et al.,

2003). In the hindbrain *FOXP2/Foxp2* is expressed in the alar plate of the cerebellum in human (CS23) and mice (E13.5) (Lai et al., 2003). With progressing embryonic development the strong cerebellar expression was seen to be specific to the Purkinje cell layer in mice (Fujita and Sugihara, 2012) and human (Lai et al., 2003). The deeper nuclei of the cerebellum show weak *FOXP2/Foxp2* expression (Lai et al., 2003; Fujita and Sugihara, 2012). In addition, *Foxp2* expression was found in interneurons of the developing spinal cord (Morikawa et al., 2009).

Following the developmental expression, *FOXP2/Foxp2* remains expressed during adulthood. The adult cortical expression is mainly limited to the deeper cortical layer VI (Ferland et al., 2003; Tsui et al., 2013). In mice, protein expression was also detected in layer V of the motor and somatosensory cortex (Hisaoka et al., 2010). *Foxp2* expression was absent from the hippocampus and the three-layered paleocortex (Ferland et al., 2003). In the basal ganglia *Foxp2* is highly expressed in the caudate nucleus, putamen and globus pallidus of the striatum (Ferland et al., 2003). Adult expression in the thalamus is strong in the auditory relay nucleus of mice (Horng et al., 2009) and detectable in neuronal subtypes of the principal visual relay nucleus (Iwai et al., 2013). Midbrain expression is detected in the substantia nigra and inferior colliculi of the tectum (Ferland et al., 2003). Adult cerebral expression is specifically strong in the Purkinje cells with weaker expression in deeper cerebellar nuclei (Ferland et al., 2003; Fujita and Sugihara, 2012). *Foxp2* expression in the brainstem is limited to the inferior olive (Ferland et al., 2003; Fujita and Sugihara, 2012).

Table 1: *FOXP2* expression studies in human and mice

	Structure	Detail	Species	Age	Molecule	Ref.
Forebrain/telencephalon	Cortex	inner intermediate zone of the neopallial cortex	mouse	E12.5 to E16.5	mRNA	1
		deeper layers of the cortical plate and subplate	mouse	E14.5 to birth	RNA, protein	2
		cortical plate	mouse	newborn	mRNA	3
		restricted to cortical layer VI	mouse	postnatal	RNA, protein	2
		some layer V expression in very medial and posterior aspects of the cortex	mouse	postnatal	RNA, protein	2
	Basal Ganglia	deep aspects of the ganglionic eminence	mouse	E12.5	RNA, protein	2
		caudate nucleus	mouse/human	E13.5/CS23	mRNA	3
		caudate putamen	mouse/human	E16.5/FS1	mRNA	3
		substantia nigra	mouse	E16.5	mRNA	3
		caudate nucleus	mouse	newborn	mRNA	3
		substantia nigra	mouse	newborn	mRNA	3
		striatum	mouse	adult	RNA, protein	2
		caudate-putamen, substantia nigra and ventral striatum	mouse	adult	RNA, protein	2
	Olfactory system	olfactory bulb	mouse	adult	RNA, protein	2
		anterior olfactory nucleus	mouse	adult	RNA, protein	2
		olfactory tubercule	mouse	adult	RNA, protein	2
		septal nucleus	mouse	adult	RNA, protein	2
	Amygdala	amygdala	mouse	adult	RNA, protein	2
Forebrain/diencephalon	Thalamus	thalamus	mouse/human	E13.5/CS23	mRNA	3
		thalamus	mouse/human	E16.5/FS1	mRNA	3
		habenular nucleus	mouse	E16.5	mRNA	3
		thalamic nuclei	mouse	newborn	mRNA	3
		thalamus (paraventricular, lateral posterior, dorsal thalamic nuclei, habenula, medial and lateral geniculate)	mouse	adult	RNA, protein	2
	Hypothalamus	hypothalamus	mouse/human	E13.5/CS23	mRNA	3
		hypothalamus	mouse/human	E16.5/FS1	mRNA	3
		zona incerta	mouse	E16.5	mRNA	3
		hypothalamus, paraventricular nucleus	mouse	adult	RNA, protein	2
Midbrain	Tectum	inferior colliculus	mouse	E16.5	mRNA	3
		inferior colliculus	mouse	newborn	mRNA	3
		superior and inferior colliculli	mouse	adult	RNA, protein	2
Hindbrain	Cerebellum	alar plate of the cerebellar primordium	mouse/human	E13.5/CS23	mRNA	3
		cerebellum	mouse	E14.5	RNA, protein	2
		developing cerebellum and cerebellar nuclei	mouse/human	E16.5/FS1	mRNA	3
		piriform layer of cerebellum (PCs)	mouse	newborn	mRNA	3

		PCs of the cerebellum	mouse	mature cerebellum	RNA, protein	2
		deep cerebellar nuclei	mouse	mature cerebellum	RNA, protein	2
	Brainstem	precursor of medulla oblongata	mouse	E11.5	mRNA	3
		midline of the hindbrain	human	CS18	mRNA	3
		medullary raphe	mouse/human	E13.5/CS23	mRNA	3
		medulla oblongata	mouse/human	E13.5/CS23	mRNA	3
		medulla oblongata	mouse/human	E16.5/FS1	mRNA	3
		lateral lemniscus nucleus	mouse	E16.5	mRNA	3
		lemniscus nuclei	mouse	newborn	mRNA	3
		inferior olives of the medulla oblongata	mouse	newborn	mRNA	3
		inferior olive	mouse	adult	RNA, protein	2
	Spinal cord	ventral interneurons of the spinal cord	mouse	E12.5 to E16.5	mRNA	1
	non-neuronal tissue	intestine	mouse	E12.5 to E16.5	mRNA	1
		heart	mouse	E12.5 to E16.5	mRNA	1
		lung epithelium	mouse	E12.5 to E16.5	mRNA	1

1: Shu et al. 2001

2: Ferland et al. 2003

3: Lai et al. 2003

The expression of *FOXP2* overlaps with brain regions which have been noted to be structurally and functionally aberrant in the KE family, including the basal ganglia, cerebellum and motor cortical areas (Lai et al., 2003). Thus, the mutated variant of *FOXP2* likely impairs the development and function of these brain regions and associated neuronal networks. *FOXP2* expression is present in additional brain regions, which did not show detectable changes in the KE family. A single functional copy of *FOXP2* may be sufficient for these structures to develop and function normally or the influence of *FOXP2* mutations may have been undetected by the brain imaging studies. In addition, the affected KE family members have no reported problems with the lung, heart or digestive tract in which *FOXP2* is expressed, suggesting that one functional copy of is sufficient in these tissues. In summary, *FOXP2* is expressed in neuronal subpopulations in multiple distinct brain regions, including the striatum, cerebellum and cortex. The affected KE family members show structural and functional aberrations in these brain regions, suggesting that mutations of *FOXP2* impair the development of cortico-striatal and cortico-cerebellar networks.

Protein structure and molecular function of FOXP2

The *FOXP2* gene encodes a transcription factor (TF) that binds to DNA and regulates the expression of other genes (Lai et al., 2001). The FOXP2 protein contains several known protein domains: a glutamine-rich region, a zinc-finger, a leucine-zipper, a forkhead-box (FOX) DNA-binding domain and a C-terminal acidic tail (Figure 2). The glutamine-rich domain of FOXP2 contains a short (10 glutamines) and a long (40 glutamines) stretch of contiguous glutamines; such polyglutamine tracts are believed to mediate interactions with other proteins (Li et al., 2004). The leucine-zipper domain of the FOXP2 protein enables it to form homo- and heterodimers with other closely related TFs and the dimerized form of FOXP2 binds to DNA (Li et al., 2004). Zinc-finger domains can mediate the interaction with a variety of molecules, including DNA, RNA, protein and lipids (Laity et al., 2001). The function of the zinc-finger in FOXP2 has not been studied in detail and, in one study, the deletion of this domain did not affect the regulatory function *in vitro* (Li et al., 2004). The FOX domain is the defining feature of the FOX family of TFs (Katoh and Katoh, 2004; Jackson et al., 2010) and binds to DNA in a sequence-specific manner (Stroud et al., 2006; Vernes et al., 2007; Nelson et al., 2013). The point mutation identified in the KE family affects the FOX domain and replaces an arginine at the amino-acid position 553 with a histidine. The R553H variant of FOXP2 does not bind to the consensus target DNA sequence (Lai et al., 2001; Vernes et al., 2006) or alternative motifs (Nelson et al., 2013). Therefore, the mutation detected in the KE family disturbs the function of the FOXP2 protein.

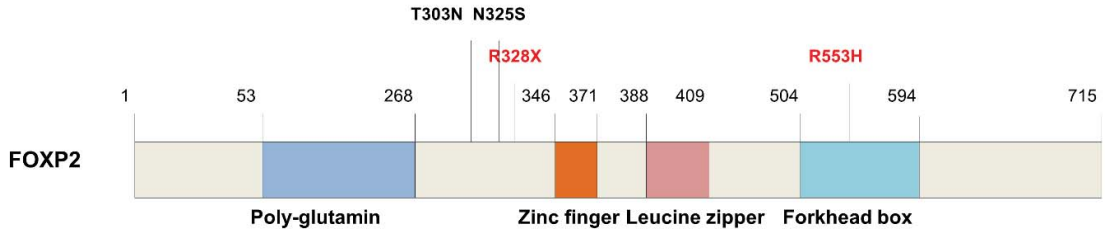

Figure 2: Linear protein structure of FOXP2. Schematic representation of the full-length FOXP2 protein containing the glutamine-rich region, zinc finger, leucine zipper and forkhead-box domains. The amino acid positions of the domains are labelled above the figure. Protein variants, which have been studied in detail, are indicated above the schematic representation. The coding change identified in the KE family changes the arginine (R) at amino acid position 553 within the FOX domain into histamine (H) (Lai et al., 2001). Another aetiological variant, which was identified in another affected family, changes the arginine (R) at position 328 into a stop codon (X) (MacDermot et al., 2005). Two amino acid changes at position 303 (T303N) and 325 (N325S) occurred in the human genome after the split from the common ancestor with chimpanzees (Enard et al., 2002).

The regulatory effect of *FOXP2* on the expression of its direct downstream target genes was initially described as repressive (Li et al., 2004). The repressive effect is mediated through interactions with histone deacetylase complexes (Chokas et al., 2010). However, accumulating evidence shows that FOXP2 protein can act both as repressor and activator of gene expression (Vernes et al., 2011; Devanna et al., 2014). The identification of target genes is a promising strategy to increase our understanding of the genetic underpinnings of speech and language (Vargha-Khadem et al., 1995; Fisher et al., 1998; Lai et al., 2001; Fisher and Scharff, 2009). Several studies have investigated the target genes of *FOXP2* in human neuronal-like cells (Vernes et al., 2007; Roll et al., 2010; Walker et al., 2012; Nelson et al., 2013), neuronal tissue of mice (Vernes et al., 2011) and foetal human (Spiteri et al., 2007). These studies identified hundreds of putative target genes, which partly overlapped between cell type and species. Thus, *FOXP2* regulates hundreds of genes, which may be regulated in a cell-type specific manner. Functional enrichment analysis of the downstream targets identified an overrepresentation of genes known to be involved in neurite outgrowth and axonogenesis (Spiteri et al., 2007; Vernes et al., 2007; Vernes et al., 2011). To analyse the impact on cellular processes, wild-type *Foxp2* was compared to aetiological variants in mutant mice. Differential

analysis of mouse primary neuronal cell cultures, homozygous for the R552H variant, which is equivalent to the human R553H variant, revealed that the mutated variant impaired aspects of neurite outgrowth, such as the number of branching sites and total neurite length (Vernes et al., 2011). The human aetiological variant R328X, which was identified in another family (MacDermot et al., 2005) is equivalent to S321X in mice and does not produce a protein product, likely due to nonsense mediated-decay at the RNA level (Groszer et al., 2008). *In utero* knock-down in the cortex of mouse embryos revealed that *Foxp2* knock-down inhibits neuronal differentiation and migration (Tsui et al., 2013). In addition, overexpression of *FOXP2* in human neuronal-like cells increased neurite outgrowth and reduced cell migration (Devanna et al., 2014). In conclusion, studies thus far suggest that *FOXP2* is able to promote the maturation of neurons through the support of neurite growth and reduced mobility. The aetiological variants fail to regulate *FOXP2* target genes, which results in impaired neuron development.

The functions of FOXP2 orthologues in animal communication

The expression pattern of *FOXP2* is highly conserved among vertebrates and has been described in fish (Bonkowsky and Chien, 2005; Shah et al., 2006; Bonkowsky et al., 2008; Itakura et al., 2008), crocodiles (Haesler et al., 2004), birds (Haesler et al., 2004; Teramitsu et al., 2004; Chen et al., 2013), rodents (Ferland et al., 2003; Takahashi et al., 2003; Campbell et al., 2009), carnivores (Rowell et al., 2010), monkeys (Takahashi et al., 2008b) and humans (Lai et al., 2003). The conserved expression pattern makes animal studies a valuable strategy to investigate the neuronal function of *FOXP2*. *Foxp2* knockout and aetiological variants were studied in mice to investigate the function of *Foxp2* on brain development and behaviour (Shu et al., 2005; French et al., 2007; Fujita et al., 2008; Groszer et al., 2008). Mutant mice have been generated that carry either a conditional *Foxp2* knockout (*Foxp2-Flox)* (French et al., 2007), a complete knock-out (*Foxp2-KO),* a mouse equivalent of the aetiological R553H

mutation generated by a transgenic knock-in strategy (*Foxp2-R552H-KI*) (Fujita et al., 2008) or mouse equivalents generated by chemical mutagenesis (*Foxp2-R552H-ENU* and *Foxp2-S321X)* (Groszer et al., 2008). Homozygotes for aetiological mutations and homozygous knock-out mice showed delayed development, reduced weight gain, severe motor problems and died within 4 weeks after birth (Shu et al., 2005; French et al., 2007; Fujita et al., 2008; Groszer et al., 2008). In addition, R552H and S321X homozygous mouse pups and the *Foxp2-flox,* after global knock-out, presented with reduced cerebellar volume and intact cerebellar cytoarchitecture (Fujita et al., 2008; Groszer et al., 2008). In contrast, the *Foxp2-KO* homozygous mouse pups showed normal cerebellar volume, but abnormal cytoarchitecture of the Purkinje cell layer and the external granular layer of the cerebellum (Shu et al., 2005). The mouse studies demonstrate that complete loss of wild type *Foxp2* is lethal at early post-natal ages. In agreement with this finding, only heterozygous *FOXP2* mutations have been detected in humans.

Heterozygous mice were fully viable in all studies and did not show gross histological anomalies (Shu et al., 2005; French et al., 2007; Fujita et al., 2008; Groszer et al., 2008). Heterozygous *Foxp2-KO* mice presented with mild developmental delay with no differences beyond postnatal day 15 (Shu et al., 2005). Mild developmental delay was also observed in some *Foxp2-R552H-KI* heterozygotes (Fujita et al., 2008). However, the *Foxp2-R552H-ENU* heterozygotes did not show developmental delay, but showed impaired long-term depression in striatal neurons and reduced motor learning skills in comparison to wild-type littermates (Groszer et al., 2008). Because human *FOXP2* mutations affect speech and language, the vocalizations of the mutant mice have been studied in more detail. Mice vocalize in the ultrasonic frequency spectrum and ultrasonic vocalizations of homozygous and heterozygous mice have been analysed in neonatal mouse pups, with debated results (French and Fisher, 2014). The innate nature of neonatal pup calls indicates that differences in mouse pup calls are secondary to general developmental delays (Gaub et al., 2010). In male adult *Foxp2-R552H-ENU* and *Foxp2-S321X* mice it was shown that disruption of *Foxp2* produced altered

ultrasonic vocalizations in response to female/female urine (Gaub et al., 2016). In conclusion, mouse studies indicate that heterozygous disruptions of *Foxp2* have subtle effects on the function and structure of *Foxp2* positive brain regions. Furthermore, the aetiological variants affect vocalization on cognitive and behavioural levels in adult mice.

In addition to the expression pattern, the amino-acid sequence of the protein encoded by *FOXP2* is highly conserved across mammals. The mouse and human orthologues of *FOXP2* vary at three amino acid positions, two of which did arise in the human lineage after the split from the chimpanzee (Enard et al., 2002; Zhang et al., 2002) and are present in Neanderthals (Krause et al., 2007). The two substitutions are located in exon 7 outside of known protein domains (Figure 2). Thus, the impact of the substitutions on protein function is unknown. The presence of two amino-acid substitutions in this highly conserved protein suggested that the substitutions may be of functional importance to human-specific traits (Enard et al., 2002). To test the effect of human *FOXP2* on mouse brain development, transgenic mice were generated that carried the two human amino-acid changes within the endogenous *Foxp2* locus (Enard et al., 2009). Heterozygous and homozygous mice were both healthy and fertile. A set of phenotypic tests revealed effects only on the explorative behaviour of the homozygous mice. In *Foxp2* positive brain regions, the dopamine levels were reduced and medium spiny neurons in the striatum grew longer dendrites and exhibited increased long-term depression. In a follow-up study, *FOXP2* positive neurons in cortical layer VI and the thalamus also showed increased neurite length (Reimers-Kipping et al., 2011). Thus, the human-specific protein-coding changes of *FOXP2* could have improved the connectivity of cortical basal-networks, which may impact on cognition and behaviour. Indeed, it was further shown that the partially humanized mice have improved declarative learning, as compared to wildtype littermates, which was assessed in a T-maze paradigm with spatial cues (Schreiweis et al., 2014). The humanized mice were also faster in switching from declarative to procedural learning. Taken together, the findings suggest that the two amino acid changes, acquired during human evolution, alter the connectivity of neuronal networks and may have improved sensorimotor learning processes.

The conserved expression pattern and protein structure of *FOXP2* suggested that *FOXP2* function is not only involved in human speech but may underlie animal communication. Results supporting this hypothesis were obtained in studies conducted in zebra finch songbirds. Zebra finches learn their songs from a tutor by imitation of the song structure. During the learning process, the structure of the produced song varies during acquisition and stabilizes over time to mimic the tutor song structure. The zebra finch orthologue of *FOXP2,* referred to as *FoxP2,* is strongly expressed in striatal nucleus area X, which is important for song learning (Haesler et al., 2004; Teramitsu et al., 2004). *FoxP2* expression in this area was shown to change during the vocal learning period in juvenile zebra finches (Haesler et al., 2004). Knockdown of *FoxP2* in area X disrupted the ability of juvenile zebra finches to copy the tutor song, resulting in incomplete and inaccurate production of song syllables (Haesler et al., 2007). To study the effects of *FoxP2* in adult zebra finches, the expression of *FoxP2* in area X was compared between males singing to a female (directed singing) and in the absence of a female (undirected singing). The expression was found to be downregulated during undirected singing of male zebra finches, as compared to non-singing males, indicating the regulation of *FoxP2* expression during vocal behaviour (Teramitsu and White, 2006; Miller et al., 2008). In addition, during directed singing the expression levels remained constant, indicating that the regulation of *FoxP2* also depends on the social context (Teramitsu and White, 2006). Analysis of the neuronal song control circuits further indicated that *FoxP2* knockdown alters the signalling between area X and connected brain regions, which has been shown to be caused by decreased dopamine signalling in area X (Murugan et al., 2013). Thus, striatal expression of zebra finch *FoxP2* is dynamically regulated during vocal communication, dependent on the social context, and plays important functional roles in mediating plasticity of song structure.

Mouse studies investigating the functions of *Foxp2* in auditory networks detected differential expression in the auditory nucleus of the thalamus in response to auditory input (Horng et al., 2009). Hence, dynamic expression of *Foxp2* can be detected in mammals. Moreover, aetiological *Foxp2* mutations have been suggested to impair auditory-guided motor learning in

mice (Kurt et al., 2012). Taken together, the dynamic regulation of *Foxp2* in the thalamus may contribute to motor learning in mammals. As no techniques are available to study gene expression in a non-invasive way, the active regulation of *FOXP2* expression in post-natal human brains has not been studied.

Regulation of FOXP2 expression

Signalling pathways upstream of FOXP2

FOXP2/Foxp2 can be used as a molecular marker for specific neuronal populations, such as for cortical layer VI neurons (Chen et al., 2005; Hisaoka et al., 2010) or neurons derived from induced pluripotent stem cells (iPSC) (Bickenbach et al., 2013; Espuny-Camacho et al., 2013; Belinsky et al., 2014; Raitano et al., 2015). Consequently, temporal changes in *Foxp2/FOXP2* expression have been described for these neuronal populations. One study investigated the effect of histamine and expression of histamine type 1 receptor expression on the developing rat cortex (Molina-Hernandez et al., 2013). Histamine-signalling is active during embryonic neurodevelopment and regulates the proliferation of neuronal stem cells (Panula et al., 2014). Panula et al. used *Foxp2* as a marker for postmitotic deep layer cortical neurons and measured the expression in cultured neurons isolated from rat embryos. The authors found that after histamine treatment the proliferation of *Foxp2*-positive cells increased and that this effect was dependent on histamine receptor 1.

Cultured neurons derived from human iPSCs can differentiate into *FOXP2* positive neuronal populations, using specific cell culture conditions and supplements (Bickenbach et al., 2013; Espuny-Camacho et al., 2013; Chen et al., 2014; Raitano et al., 2015). For example, treatment of cultured neurons with Noggin, a signalling protein important for embryonic development, produced neurons with a cortical phenotype and a subpopulation of *FOXP2* positive cortical neurons (Espuny-Camacho et al., 2013). Thus, the upstream signalling pathways that may activate *FOXP2* expression can be studied in cultured human neurons. So far, the studies that

exploit human iPSC-derived neurons use *FOXP2* as a neuronal marker and do not investigate the signalling cascades that lead to *FOXP2* expression. Accordingly, increased *FOXP2* expression has been described in cultured neurons derived from patients with bipolar disorder (Chen et al., 2014) and reduced expression in frontotemporal dementia (Raitano et al., 2015). The nuclear factors and transcription factors that may have mediated the differential expression, and possibly regulate *FOXP2,* were not investigated.

Transcription factors indicated to regulate FOXP2

Gene regulation in response to developmental programs or environmental cues is mediated by TFs. The identification of upstream TFs is, therefore, a valuable approach to identify the mechanisms preceding *FOXP2* expression. The activation of *FOXP2* expression during development is of specific importance to understand the complex expression pattern. Co-expression of TFs, which may indicate positive regulation, has been described in different tissues and species. Table 1 shows a selection of co-expression studies, which include the *FOXP2* gene and other TFs. The co-expression of the listed TFs was determined on mRNA and/or protein levels. The co-expression of two TFs can suggest hypotheses about their function and relation, but experimental evidence for the interaction is required. The direct regulation of *FOXP2/foxp2* has been studied in detail for *pax6* (Bonkowsky et al., 2008) and *lef1* in zebrafish (Coutinho et al., 2011) and for *POU3F2* in human cell lines (Maricic et al., 2013).

Table1: Co-expression of Transcription factors with *FOXP2*

Transcription Factor	Species	Brain structure	Molecule	Publication
Atoh1	mouse	spinal cord	protein	(Miesegaes et al., 2009)
Dlx5	monkey, rat	amygdala, striatum	mRNA	(Kaoru et al., 2010)
Ebf1	monkey, rat	striatum	mRNA	(Kaoru et al., 2010)
emx1	zebrafish	embryonic telencephalon	mRNA	(Shah et al., 2006)
Foxp1	mouse	hypothalamus, subthalamic nuclei	protein	(Skidmore et al., 2008)
Foxp1	monkey, rat	striatum, nucleus accumebens, thalamus	mRNA, protein	(Takahashi et al., 2008b; Kaoru et al., 2010)
FOXP1	songbird	striatum, thalamus, inferior olive	mRNA, protein	(Teramitsu et al., 2004; Mendoza et al., 2015)
Foxp4	rat	ganglionic eminence, embryonic thalamus, amygdala, cortex	mRNA	(Takahashi et al., 2008a)
FOXP4	songbird	adult: Purkinje cells, thalamus, striatum, inferior olive; embryonic: ganglionic eminence, cortical plate, thalamus	mRNA, protein	(Mendoza et al., 2015)
Gli1	mouse	amygdala	protein	(Carney et al., 2010)
lef1	zebrafish	embryonic tectum, mid-hindbrain boundary	mRNA	(Bonkowsky et al., 2008)
Lhx1/Lhx5	mouse	spinal cord	protein	(Miesegaes et al., 2009)
Lmx1b	mouse	hypothalamus, subthalamic nuclei	protein	(Skidmore et al., 2008)
Lmx1b	rat	pons	protein	(Miller et al., 2012)
Meis2	monkey, rat	amygdala, nucleus accumbens, island of calleja, striatum	mRNA	(Takahashi et al., 2008b; Kaoru et al., 2010)
Pax6	monkey, rat	amygdala, striatum	mRNA, protein	(Kaoru et al., 2010)
PBX3	monkey, rat	amygdala, nucleus accumbens, island of caleja, striatum	mRNA	(Takahashi et al., 2008b; Kaoru et al., 2010)
Pitx1	mouse	hypothalamus, subthalamic nuclei	protein	(Skidmore et al., 2008)
POGZ	human	inner cortical plate	protein	(Willsey et al., 2013)
Tbr1	mouse	cortical layer VI	protein	(Hisaoka et al., 2010)
TBR1	human	inner cortical plate	protein	(Willsey et al., 2013)

One study investigated the regulation of *foxp2* by *pax6* in zebrafish (Coutinho et al., 2011). *PAX6* is a TF involved in early neurogenesis and brain patterning. The TF is expressed in the developing neocortex, prethalamus and cerebellum (Osumi, 2001). In the zebrafish genome, the *foxp2* promoter was predicted to contain *pax6* binding sites and Coutinho et al. demonstrated the binding of *pax6* to these sites (Coutinho et al., 2011). The element drove reporter gene expression in *foxp2*-positive forebrain regions of zebrafish embryos. The *in vivo* activity of this enhancer was dependent on *pax6* expression. The results suggested that *foxp2* is a target of *pax6* and may suggest a link to the biological functions of *pax6.* Another study investigated a link between *lef1* and *foxp2* in zebrafish (Bonkowsky et al., 2008). *lef1* is a transcription factor that regulates gene expression in response to WNT signalling. This signalling pathway is important for cell-to-cell communication and the patterning of the central nervous system during embryonic development. The authors characterized the expression of *lef1* and *foxp2,* and detected temporal overlap in the tectum and mid-hindbrain boundary in developing zebrafish embryos, suggesting that *lef1* might regulate *foxp2*. To test if there was a direct regulatory effect, the authors ran in silico searches for *lef1* binding sites near the *foxp2* gene in the zebrafish genome and then determined if these elements drive gene expression in embryonic zebrafish (Bonkowsky et al., 2008). Two of the predicted elements showed *in vivo* enhancer activity and overlapped with *foxp2* expression. These findings suggested that these two enhancers regulated *foxp2.* The fact that *foxp2* is a direct downstream target of *lef1* suggests that *foxp2* may be regulated by WNT-signalling.

Another regulatory relationship was suggested by investigating genetic changes, which occurred in humans during recent evolutionary periods (Krause et al., 2007; Maricic et al., 2013). Parts of the *FOXP2* locus seem to have undergone positive selection after humans split from Neanderthals. The positive selection was determined by the presence of modern human alleles, which are fixed or occur at high frequency in humans and are absent in the Neanderthal genome. Positive selection of these alleles could indicate that they are relevant for *FOXP2* function in humans. One such human-specific substitution, which is located in intron 8 of

FOXP2 and occurred at an evolutionary conserved site, was predicted to reduce the binding of *POU3F2.* Indeed, POU3F2 protein was shown to bind stronger to the Neanderthal allele than to the human allele. In addition, the putative surrounding regulatory element carrying the human allele showed reduced reporter gene expression in response to *POU3F2* overexpression compared to the element carrying the Neanderthal allele. However, the measured reporter gene expression was dependent on an additional viral enhancer, suggesting that the tested regulatory element does not act as an independent enhancer. In accordance to this, the element did not drive gene expression in transgenic mice. The study by Maricic et al. suggests that *POU3F2* regulates the expression of *FOXP2* via an evolutionary conserved site and that this regulatory interaction is reduced in humans.

Genetic variation and the regulation of FOXP2

Regulatory genomic elements control the timing and levels of gene expression. Consequently, genetic changes in regulatory elements may alter gene expression and contribute to normal or pathogenic phenotypic variation. The majority of trait-associated single nucleotide polymorphisms (SNPs) identified in genome-wide association studies (GWAS) are located in non-coding regions of the genome (Welter et al., 2014). Thus, associated SNPs near the *FOXP2* gene may be located within regulatory regions and affect the expression of this gene. GWAS studies have identified associations between SNPs located near the *FOXP2* gene with a number of phenotypes, for example N-glycosylation of immunoglobin (Lauc et al., 2013), Crohn's disease (Julia et al., 2013), lymphoblast cell viability (de With et al., 2015), obesity (Kim et al., 2013) smoking behaviour (Argos et al., 2014; Sung et al., 2015). These phenotypes have no obvious link to motor-learning or communication. Thus, if the associated SNPs tag regulatory regions, they do not seem to underlie *FOXP2* regulation in relation to human speech and language.

Candidate genetic association studies have investigated common genetic variants of *FOXP2* in the context of human traits hypothesized to be related to *FOXP2* function. To improve the

power to detect associations, most studies focused on a limited amount of genetic variants. Studies have investigated associations of *FOXP2* to auditory-visual hallucinations (Sanjuan et al., 2006; McCarthy-Jones et al., 2014), schizophrenia (Tolosa et al., 2010; Spaniel et al., 2011), ADHD (Ribases et al., 2012), autism (Park et al., 2014), dyslexia (Wilcke et al., 2012), brain activation during reading tasks (Pinel et al., 2012) and enhanced language skills (Chandrasekaran et al., 2015). The location of associated variants could indicate the presence of regulatory elements. However, the low density of genetic markers would not allow for the localization of regulatory elements. One study investigated genetic markers at the *FOXP2* locus at high density, looking for association to brain volumes (Hoogman et al., 2014). However, Hoogman et al. did not detect significant associations. In conclusion, candidate association studies so far do not suggest the presence of regulatory SNPs near *FOXP2.* More importantly, candidate association studies could benefit from including SNPs within regulatory regions to improve the identification of genetic variants linked to normal variation in cognitive traits.

Genetic variants in *FOXP2* regulatory elements may also contribute to pathogenic variation. Chromosomal rearrangements in the vicinity of *FOXP2* have been described in clinical reports of people presenting with speech impairments (Feuk et al., 2006; Moralli et al., 2015). These genetic variants leave the protein coding sequence intact but may alter endogenous expression. Chromosomal rearrangements have been detected in cases with speech phenotypes similar to that observed in the KE family (Feuk et al., 2006; Adegbola et al., 2015; Moralli et al., 2015). Moralli et al. described a child with late development of speech, who carries an inversion of chromosome 7 and a breakpoint 200 kb downstream of *FOXP2* (Moralli et al., 2015). Feuk et al. described a patient with developmental delay with a deletion reported to just start at the telomeric end of *FOXP2* (Feuk et al., 2006). However, the authors do not report the genomic location of the breakpoint. The assumed cause of the patient phenotypes is an aberrant *FOXP2* expression, caused by deleted or displaced regulatory elements. However, the putative regulatory elements were not identified. The identification of these

regulatory sequences would help to clinically define these patients and investigate the molecular aetiology of the observed phenotype.

Question and aims of this dissertation

Research into the function and the downstream molecular pathways of the *FOXP2* gene has proven to be highly valuable in investigating cognitive traits from molecule to behaviour. Similarly, the pathways upstream of *FOXP2* promise to increase our understanding of the molecular mechanisms underlying speech and language. Consequently, my aim was to investigate the regulation of *FOXP2* expression in the brain and how the regulatory mechanisms may contribute to human cognitive traits.

One of my aims, which is detailed in **chapter 2,** was to identify enhancers, which may have contributed to the aetiology of delayed speech development in a child with a complex chromosomal rearrangement (Moralli et al., 2015). The chromosomal rearrangement included an inversion, with a breakpoint near the *FOXP2* gene, suggesting this gene was involved in the aetiology of the child's speech phenotype. Using public data and molecular experiments I located an enhancer just downstream of the inversion breakpoint, which would not be able to regulate the *FOXP2* gene in the child. This rare genetic variant allowed me to locate an enhancer, which may regulate *FOXP2* in healthy individuals.

In order for enhancers to regulate their target promoters they get into physical contact with the promoter. Consequently, my aim in **chapter 3** was to identify physical interactions of the *FOXP2* promoter with putative enhancers in *FOXP2* expressing cells. I identified a number of putative enhancers in the vicinity of *FOXP2,* as well as chromatin interactions to the neighbouring gene promoter and the 3'UTR of *FOXP2.*

Enhancers, like promoters, are the integrative hubs for molecular mechanisms to control gene expression during development or in response to environmental stimuli. Thus, my aim in **chapter 4** was to investigate the genetic pathways that regulate the *FOXP2* enhancers. TFs, as the effector proteins of signalling pathways, bind to promoters and enhancers. In a reporter

gene expression assay I determined the effect of TFs on the enhancers and promoters of *FOXP2.*

The aim of **chapter 5** was to characterize the activity of the most promising enhancers in developing and adult brains. I created transgenic mice, carrying the human enhancers to investigate the enhancers' target brain regions and time of activity. In combination with the previous chapters, the results combine aspects of the enhancer's location, the enhancer-interacting proteins, the brain structure in which these interactions may occur and the change in enhancer activity during embryonic and post-natal development.

Following the results obtained in the earlier chapters, my aim in **chapter 6** was to determine how the effect of genetic variation within enhancers relates to normal variation in human traits. To increase the statistical power of the analysis in this chapter, I analysed 296 neurodevelopmental enhancers from a public database, which were characterized with the same method that I used to study the *FOXP2* enhancers in chapter 5. In contrast to the approach in chapter 2, where I investigated the rare genetic variation of a single case, chapter 6 investigated common genetic variants in 13,000 healthy individuals.

In **chapter 7** I will summarize the findings from chapter 2 to 6 and put the results from each chapter into relation to each other. The studies of *FOXP2* enhancers shed light on the upstream mechanism that control this gene and may contribute to the development of *FOXP2* positive neurons and neuronal networks. The combination of the individual chapters enabled me to generate new testable hypotheses on the regulation of human *FOXP2.*

References

Adegbola AA, Cox GF, Bradshaw EM, Hafler DA, Gimelbrant A, Chess A (2015) Monoallelic expression of the human FOXP2 speech gene. Proc Natl Acad Sci U S A 112:6848-6854.

Argos M, Tong L, Pierce BL, Rakibuz-Zaman M, Ahmed A, Islam T, Rahman M, Paul-Brutus R, Rahaman R, Roy S, Jasmine F, Kibriya MG, Ahsan H (2014) Genome-wide association study of smoking behaviours among Bangladeshi adults. Journal of medical genetics 51:327-333.

Belinsky GS, Rich MT, Sirois CL, Short SM, Pedrosa E, Lachman HM, Antic SD (2014) Patch-clamp recordings and calcium imaging followed by single-cell PCR reveal the developmental profile of 13 genes in iPSC-derived human neurons. Stem cell research 12:101-118.

Belton E, Salmond CH, Watkins KE, Vargha-Khadem F, Gadian DG (2003) Bilateral brain abnormalities associated with dominantly inherited verbal and orofacial dyspraxia. Human brain mapping 18:194-200.

Bickenbach JR, Tomanek-Chalkley A, Wiechert S, Winter MC (2013) Human skin keratinocytes can be reprogrammed to express neuronal genes and proteins after a single treatment with decitabine. BioResearch open access 2:217-221.

Bonkowsky JL, Chien CB (2005) Molecular cloning and developmental expression of foxP2 in zebrafish. Developmental dynamics : an official publication of the American Association of Anatomists 234:740-746.

Bonkowsky JL, Wang X, Fujimoto E, Lee JE, Chien CB, Dorsky RI (2008) Domain-specific regulation of foxP2 CNS expression by lef1. BMC developmental biology 8:103.

Campbell P, Reep RL, Stoll ML, Ophir AG, Phelps SM (2009) Conservation and diversity of Foxp2 expression in muroid rodents: functional implications. The Journal of comparative neurology 512:84-100.

Carney RS, Mangin JM, Hayes L, Mansfield K, Sousa VH, Fishell G, Machold RP, Ahn S, Gallo V, Corbin JG (2010) Sonic hedgehog expressing and responding cells generate neuronal diversity in the medial amygdala. Neural development 5:14.

Chandrasekaran B, Yi HG, Blanco NJ, McGeary JE, Maddox WT (2015) Enhanced procedural learning of speech sound categories in a genetic variant of FOXP2. The Journal of neuroscience : the official journal of the Society for Neuroscience 35:7808-7812.

Chen B, Schaevitz LR, McConnell SK (2005) Fezl regulates the differentiation and axon targeting of layer 5 subcortical projection neurons in cerebral cortex. Proc Natl Acad Sci U S A 102:17184-17189.

Chen HM, DeLong CJ, Bame M, Rajapakse I, Herron TJ, McInnis MG, O'Shea KS (2014) Transcripts involved in calcium signaling and telencephalic neuronal fate are altered in induced pluripotent stem cells from bipolar disorder patients. Transl Psychiat 4.

Chen Q, Heston JB, Burkett ZD, White SA (2013) Expression analysis of the speech-related genes FoxP1 and FoxP2 and their relation to singing behavior in two songbird species. The Journal of experimental biology 216:3682-3692.

Chokas AL, Trivedi CM, Lu MM, Tucker PW, Li S, Epstein JA, Morrisey EE (2010) Foxp1/2/4-NuRD interactions regulate gene expression and epithelial injury response in the lung via regulation of interleukin-6. J Biol Chem 285:13304-13313.

Coutinho P, Pavlou S, Bhatia S, Chalmers KJ, Kleinjan DA, van Heyningen V (2011) Discovery and assessment of conserved Pax6 target genes and enhancers. Genome Res 21:1349-1359.

de With SA, Pulit SL, Wang T, Staal WG, van Solinge WW, de Bakker PI, Ophoff RA (2015) Genome-wide association study of lymphoblast cell viability after clozapine exposure. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics 168B:116-122.

Devanna P, Middelbeek J, Vernes SC (2014) FOXP2 drives neuronal differentiation by interacting with retinoic acid signaling pathways. Frontiers in cellular neuroscience 8:305.

Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T, Monaco AP, Paabo S (2002) Molecular evolution of FOXP2, a gene involved in speech and language. Nature 418:869-872.

Enard W et al. (2009) A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. Cell 137:961-971.

Espuny-Camacho I, Michelsen KA, Gall D, Linaro D, Hasche A, Bonnefont J, Bali C, Orduz D, Bilheu A, Herpoel A, Lambert N, Gaspard N, Peron S, Schiffmann SN, Giugliano M, Gaillard A, Vanderhaeghen P (2013) Pyramidal neurons derived from human pluripotent stem cells integrate efficiently into mouse brain circuits in vivo. Neuron 77:440-456.

Ferland RJ, Cherry TJ, Preware PO, Morrisey EE, Walsh CA (2003) Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain. The Journal of comparative neurology 460:266-279.
Feuk L et al. (2006) Absence of a paternally inherited FOXP2 gene in developmental verbal dyspraxia. Am J Hum Genet 79:965-972.
Fisher SE, Scharff C (2009) FOXP2 as a molecular window into speech and language. Trends Genet 25:166-177.
Fisher SE, Vargha-Khadem F, Watkins KE, Monaco AP, Pembrey ME (1998) Localisation of a gene implicated in a severe speech and language disorder. Nat Genet 18:168-170.
French CA, Fisher SE (2014) What can mice tell us about Foxp2 function? Current opinion in neurobiology 28:72-79.
French CA, Groszer M, Preece C, Coupe AM, Rajewsky K, Fisher SE (2007) Generation of mice with a conditional Foxp2 null allele. Genesis 45:440-446.
Fujita E, Tanabe Y, Shiota A, Ueda M, Suwa K, Momoi MY, Momoi T (2008) Ultrasonic vocalization impairment of Foxp2 (R552H) knockin mice related to speech-language disorder and abnormality of Purkinje cells. Proc Natl Acad Sci U S A 105:3117-3122.
Fujita H, Sugihara I (2012) FoxP2 expression in the cerebellum and inferior olive: development of the transverse stripe-shaped expression pattern in the mouse cerebellar cortex. The Journal of comparative neurology 520:656-677.
Gaub S, Fisher SE, Ehret G (2016) Ultrasonic vocalizations of adult male Foxp2-mutant mice: behavioral contexts of arousal and emotion. Genes Brain Behav 15:243-259.
Gaub S, Groszer M, Fisher SE, Ehret G (2010) The structure of innate vocalizations in Foxp2-deficient mouse pups. Genes Brain and Behavior 9:390-401.
Gopnik M (1990) Feature-blind grammar and dysphagia. Nature 344:715.
Gopnik M, Crago MB (1991) Familial aggregation of a developmental language disorder. Cognition 39:1-50.
Groszer M et al. (2008) Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits. Current biology : CB 18:354-362.
Haesler S, Rochefort C, Georgi B, Licznerski P, Osten P, Scharff C (2007) Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. PLoS biology 5:e321.
Haesler S, Wada K, Nshdejan A, Morrisey EE, Lints T, Jarvis ED, Scharff C (2004) FoxP2 expression in avian vocal learners and non-learners. The Journal of neuroscience : the official journal of the Society for Neuroscience 24:3164-3175.
Hisaoka T, Nakamura Y, Senba E, Morikawa Y (2010) The forkhead transcription factors, Foxp1 and Foxp2, identify different subpopulations of projection neurons in the mouse cerebral cortex. Neuroscience 166:551-563.
Hoogman M, Guadalupe T, Zwiers MP, Klarenbeek P, Francks C, Fisher SE (2014) Assessing the effects of common variation in the FOXP2 gene on human brain structure. Frontiers in human neuroscience 8:473.
Horng S, Kreiman G, Ellsworth C, Page D, Blank M, Millen K, Sur M (2009) Differential gene expression in the developing lateral geniculate nucleus and medial geniculate nucleus reveals novel roles for Zic4 and Foxp2 in visual and auditory pathway development. The Journal of neuroscience : the official journal of the Society for Neuroscience 29:13672-13683.
Hurst JA, Baraitser M, Auger E, Graham F, Norell S (1990) An extended family with a dominantly inherited speech disorder. Developmental medicine and child neurology 32:352-355.
Itakura T, Chandra A, Yang Z, Xue X, Wang B, Kimura W, Hikosaka K, Inohaya K, Kudo A, Uezato T, Miura N (2008) The medaka FoxP2, a homologue of human language gene FOXP2, has a diverged structure and function. Journal of biochemistry 143:407-416.
Iwai L, Ohashi Y, van der List D, Usrey WM, Miyashita Y, Kawasaki H (2013) FoxP2 is a parvocellular-specific transcription factor in the visual thalamus of monkeys and ferrets. Cerebral cortex 23:2204-2212.
Jackson BC, Carpenter C, Nebert DW, Vasiliou V (2010) Update of human and mouse forkhead box (FOX) gene families. Human genomics 4:345-352.
Julia A et al. (2013) A genome-wide association study on a southern European population identifies a new Crohn's disease susceptibility locus at RBX1-EP300. Gut 62:1440-1445.
Kaoru T, Liu FC, Ishida M, Oishi T, Hayashi M, Kitagawa M, Shimoda K, Takahashi H (2010) Molecular characterization of the intercalated cell masses of the amygdala: implications for the relationship with the striatum. Neuroscience 166:220-230.

Katoh M, Katoh M (2004) Human FOX gene family (Review). International journal of oncology 25:1495-1500.
Kim HJ, Yoo YJ, Ju YS, Lee S, Cho SI, Sung J, Kim JI, Seo JS (2013) Combined linkage and association analyses identify a novel locus for obesity near PROX1 in Asians. Obesity 21:2405-2412.
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 (2007) The derived FOXP2 variant of modern humans was shared with Neandertals. Current biology : CB 17:1908-1912.
Kurt S, Fisher SE, Ehret G (2012) Foxp2 mutations impair auditory-motor association learning. PloS one 7:e33130.
Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. Nature 413:519-523.
Lai CS, Gerrelli D, Monaco AP, Fisher SE, Copp AJ (2003) FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. Brain : a journal of neurology 126:2455-2462.
Lai CS, Fisher SE, Hurst JA, Levy ER, Hodgson S, Fox M, Jeremiah S, Povey S, Jamison DC, Green ED, Vargha-Khadem F, Monaco AP (2000) The SPCH1 region on human 7q31: genomic characterization of the critical interval and localization of translocations associated with speech and language disorder. Am J Hum Genet 67:357-368.
Laity JH, Lee BM, Wright PE (2001) Zinc finger proteins: new insights into structural and functional diversity. Current opinion in structural biology 11:39-46.
Lauc G et al. (2013) Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers. PLoS Genet 9:e1003225.
Lennon PA, Cooper ML, Peiffer DA, Gunderson KL, Patel A, Peters S, Cheung SW, Bacino CA (2007) Deletion of 7q31.1 supports involvement of FOXP2 in language impairment: clinical report and review. American journal of medical genetics Part A 143A:791-798.
Li S, Weidenfeld J, Morrisey EE (2004) Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions. Molecular and cellular biology 24:809-822.
Liegeois F, Morgan AT, 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 15:283-288.
Liegeois F, Baldeweg T, Connelly A, Gadian DG, Mishkin M, Vargha-Khadem F (2003) Language fMRI abnormalities associated with FOXP2 gene mutation. Nat Neurosci 6:1230-1237.
MacDermot KD, Bonora E, Sykes N, Coupe AM, Lai CS, Vernes SC, Vargha-Khadem F, McKenzie F, Smith RL, Monaco AP, Fisher SE (2005) Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. Am J Hum Genet 76:1074-1080.
Maricic T, Gunther V, Georgiev O, Gehre S, Curlin M, Schreiweis C, Naumann R, Burbano HA, Meyer M, Lalueza-Fox C, de la Rasilla M, Rosas A, Gajovic S, Kelso J, Enard W, Schaffner W, Paabo S (2013) A recent evolutionary change affects a regulatory element in the human FOXP2 gene. Molecular biology and evolution 30:844-852.
McCarthy-Jones S, Green MJ, Scott RJ, Tooney PA, Cairns MJ, Wu JQ, Oldmeadow C, Carr V, Australian Schizophrenia Research B (2014) Preliminary evidence of an interaction between the FOXP2 gene and childhood emotional abuse predicting likelihood of auditory verbal hallucinations in schizophrenia. Journal of psychiatric research 50:66-72.
Mendoza E, Tokarev K, During DN, Retamosa EC, Weiss M, Arpenik N, Scharff C (2015) Differential coexpression of FoxP1, FoxP2, and FoxP4 in the Zebra Finch (Taeniopygia guttata) song system. The Journal of comparative neurology 523:1318-1340.
Middleton FA, Strick PL (2000) Basal ganglia and cerebellar loops: motor and cognitive circuits. Brain research Brain research reviews 31:236-250.
Miesegaes GR, Klisch TJ, Thaller C, Ahmad KA, Atkinson RC, Zoghbi HY (2009) Identification and subclassification of new Atoh1 derived cell populations during mouse spinal cord development. Developmental biology 327:339-351.
Miller JA et al. (2014) Transcriptional landscape of the prenatal human brain. Nature 508:199-206.
Miller JE, Spiteri E, Condro MC, Dosumu-Johnson RT, Geschwind DH, White SA (2008) Birdsong decreases protein levels of FoxP2, a molecule required for human speech. Journal of neurophysiology 100:2015-2025.
Miller RL, Knuepfer MM, Wang MH, Denny GO, Gray PA, Loewy AD (2012) Fos-activation of FoxP2 and Lmx1b neurons in the parabrachial nucleus evoked by hypotension and hypertension in conscious rats. Neuroscience 218:110-125.

Molina-Hernandez A, Rodriguez-Martinez G, Escobedo-Avila I, Velasco I (2013) Histamine up-regulates fibroblast growth factor receptor 1 and increases FOXP2 neurons in cultured neural precursors by histamine type 1 receptor activation: conceivable role of histamine in neurogenesis during cortical development in vivo. Neural development 8:4.

Moralli D, Nudel R, Chan MT, Green CM, Volpi EV, Benitez-Burraco A, Newbury DF, Garcia-Bellido P (2015) Language impairment in a case of a complex chromosomal rearrangement with a breakpoint downstream of FOXP2. Molecular cytogenetics 8:36.

Morikawa Y, Hisaoka T, Senba E (2009) Characterization of Foxp2-expressing cells in the developing spinal cord. Neuroscience 162:1150-1162.

Murugan M, Harward S, Scharff C, Mooney R (2013) Diminished FoxP2 levels affect dopaminergic modulation of corticostriatal signaling important to song variability. Neuron 80:1464-1476.

Nelson CS, Fuller CK, Fordyce PM, Greninger AL, Li H, DeRisi JL (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 Res 41:5991-6004.

Osumi N (2001) The role of Pax6 in brain patterning. The Tohoku journal of experimental medicine 193:163-174.

Palka C, Alfonsi M, Mohn A, Cerbo R, Guanciali Franchi P, Fantasia D, Morizio E, Stuppia L, Calabrese G, Zori R, Chiarelli F, Palka G (2012) Mosaic 7q31 deletion involving FOXP2 gene associated with language impairment. Pediatrics 129:e183-188.

Panula P, Sundvik M, Karlstedt K (2014) Developmental roles of brain histamine. Trends in neurosciences 37:159-168.

Park Y, Won S, Nam M, Chung JH, Kwack K (2014) Interaction between MAOA and FOXP2 in association with autism and verbal communication in a Korean population. Journal of child neurology 29:NP207-211.

Pinel P, Fauchereau F, Moreno A, Barbot A, Lathrop M, Zelenika D, Le Bihan D, Poline JB, Bourgeron T, Dehaene S (2012) Genetic variants of FOXP2 and KIAA0319/TTRAP/THEM2 locus are associated with altered brain activation in distinct language-related regions. The Journal of neuroscience : the official journal of the Society for Neuroscience 32:817-825.

Raitano S, Ordovas L, De Muynck L, Guo W, Espuny-Camacho I, Geraerts M, Khurana S, Vanuytsel K, Toth BI, Voets T, Vandenberghe R, Cathomen T, Van Den Bosch L, Vanderhaeghen P, Van Damme P, Verfaillie CM (2015) Restoration of progranulin expression rescues cortical neuron generation in an induced pluripotent stem cell model of frontotemporal dementia. Stem cell reports 4:16-24.

Reimers-Kipping S, Hevers W, Paabo S, Enard W (2011) Humanized Foxp2 specifically affects cortico-basal ganglia circuits. Neuroscience 175:75-84.

Ribases M, Sanchez-Mora C, Ramos-Quiroga JA, Bosch R, Gomez N, Nogueira M, Corrales M, Palomar G, Jacob CP, Gross-Lesch S, Kreiker S, Reif A, Lesch KP, Cormand B, Casas M, Bayes M (2012) An association study of sequence variants in the forkhead box P2 (FOXP2) gene and adulthood attention-deficit/hyperactivity disorder in two European samples. Psychiatric genetics 22:155-160.

Rice GM, Raca G, Jakielski KJ, Laffin JJ, Iyama-Kurtycz CM, Hartley SL, Sprague RE, Heintzelman AT, Shriberg LD (2012) Phenotype of FOXP2 haploinsufficiency in a mother and son. American journal of medical genetics Part A 158A:174-181.

Roll P, Vernes SC, Bruneau N, Cillario J, Ponsole-Lenfant M, Massacrier A, Rudolf G, Khalife M, Hirsch E, Fisher SE, Szepetowski P (2010) Molecular networks implicated in speech-related disorders: FOXP2 regulates the SRPX2/uPAR complex. Human molecular genetics 19:4848-4860.

Rowell JJ, Mallik AK, Dugas-Ford J, Ragsdale CW (2010) Molecular analysis of neocortical layer structure in the ferret. The Journal of comparative neurology 518:3272-3289.

Sanjuan J, Tolosa A, Gonzalez JC, Aguilar EJ, Perez-Tur J, Najera C, Molto MD, de Frutos R (2006) Association between FOXP2 polymorphisms and schizophrenia with auditory hallucinations. Psychiatric genetics 16:67-72.

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 (2014) Humanized Foxp2 accelerates learning by enhancing transitions from declarative to procedural performance. Proc Natl Acad Sci U S A 111:14253-14258.

Shah R, Medina-Martinez O, Chu LF, Samaco RC, Jamrich M (2006) Expression of FoxP2 during zebrafish development and in the adult brain. The International journal of developmental biology 50:435-438.

Shriberg LD, Ballard KJ, Tomblin JB, Duffy JR, Odell KH, Williams CA (2006) Speech, prosody, and voice characteristics of a mother and daughter with a 7;13 translocation affecting FOXP2. Journal of speech, language, and hearing research : JSLHR 49:500-525.

Shu W, Yang H, Zhang L, Lu MM, Morrisey EE (2001) Characterization of a new subfamily of winged-helix/forkhead (Fox) genes that are expressed in the lung and act as transcriptional repressors. J Biol Chem 276:27488-27497.

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 (2005) Altered ultrasonic vocalization in mice with a disruption in the Foxp2 gene. Proc Natl Acad Sci U S A 102:9643-9648.

Skidmore JM, Cramer JD, Martin JF, Martin DM (2008) Cre fate mapping reveals lineage specific defects in neuronal migration with loss of Pitx2 function in the developing mouse hypothalamus and subthalamic nucleus. Molecular and cellular neurosciences 37:696-707.

Spaniel F, Horacek J, Tintera J, Ibrahim I, Novak T, Cermak J, Klirova M, Hoschl C (2011) Genetic variation in FOXP2 alters grey matter concentrations in schizophrenia patients. Neuroscience letters 493:131-135.

Spiteri E, Konopka G, Coppola G, Bomar J, Oldham M, Ou J, Vernes SC, Fisher SE, Ren B, Geschwind DH (2007) Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. Am J Hum Genet 81:1144-1157.

Stroud JC, Wu Y, Bates DL, Han A, Nowick K, Paabo S, Tong H, Chen L (2006) Structure of the forkhead domain of FOXP2 bound to DNA. Structure 14:159-166.

Sung YJ, de Las Fuentes L, Schwander KL, Simino J, Rao DC (2015) Gene-smoking interactions identify several novel blood pressure loci in the Framingham Heart Study. American journal of hypertension 28:343-354.

Takahashi K, Liu FC, Hirokawa K, Takahashi H (2003) Expression of Foxp2, a gene involved in speech and language, in the developing and adult striatum. Journal of neuroscience research 73:61-72.

Takahashi K, Liu FC, Hirokawa K, Takahashi H (2008a) Expression of Foxp4 in the developing and adult rat forebrain. Journal of neuroscience research 86:3106-3116.

Takahashi K, Liu FC, Oishi T, Mori T, Higo N, Hayashi M, Hirokawa K, Takahashi H (2008b) Expression of FOXP2 in the developing monkey forebrain: comparison with the expression of the genes FOXP1, PBX3, and MEIS2. The Journal of comparative neurology 509:180-189.

Teramitsu I, White SA (2006) FoxP2 regulation during undirected singing in adult songbirds. The Journal of neuroscience : the official journal of the Society for Neuroscience 26:7390-7394.

Teramitsu I, Kudo LC, London SE, Geschwind DH, White SA (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 24:3152-3163.

Tolosa A, Sanjuan J, Dagnall AM, Molto MD, Herrero N, de Frutos R (2010) FOXP2 gene and language impairment in schizophrenia: association and epigenetic studies. BMC medical genetics 11:114.

Tsui D, Vessey JP, Tomita H, Kaplan DR, Miller FD (2013) FoxP2 regulates neurogenesis during embryonic cortical development. The Journal of neuroscience : the official journal of the Society for Neuroscience 33:244-258.

Turner SJ, Hildebrand MS, Block S, Damiano J, Fahey M, Reilly S, Bahlo M, Scheffer IE, Morgan AT (2013) Small intragenic deletion in FOXP2 associated with childhood apraxia of speech and dysarthria. American journal of medical genetics Part A 161A:2321-2326.

Vargha-Khadem F, Gadian DG, Copp A, Mishkin M (2005) FOXP2 and the neuroanatomy of speech and language. Nat Rev Neurosci 6:131-138.

Vargha-Khadem F, Watkins K, Alcock K, Fletcher P, Passingham R (1995) Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. Proc Natl Acad Sci U S A 92:930-933.

Vargha-Khadem F, Watkins KE, Price CJ, Ashburner J, Alcock KJ, Connelly A, Frackowiak RS, Friston KJ, Pembrey ME, Mishkin M, Gadian DG, Passingham RE (1998) Neural basis of an inherited speech and language disorder. Proc Natl Acad Sci U S A 95:12695-12700.

Vernes SC, Spiteri E, Nicod J, Groszer M, Taylor JM, Davies KE, Geschwind DH, Fisher SE (2007) High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders. Am J Hum Genet 81:1232-1250.

Vernes SC, Nicod J, Elahi FM, Coventry JA, Kenny N, Coupe AM, Bird LE, Davies KE, Fisher SE (2006) Functional genetic analysis of mutations implicated in a human speech and language disorder. Human molecular genetics 15:3154-3167.

Vernes SC, Oliver PL, Spiteri E, Lockstone HE, Puliyadi R, Taylor JM, Ho J, Mombereau C, Brewer A, Lowy E, Nicod J, Groszer M, Baban D, Sahgal N, Cazier JB, Ragoussis J, Davies KE,

Geschwind DH, Fisher SE (2011) Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. PLoS Genet 7:e1002145.
Wain HM, Bruford EA, Lovering RC, Lush MJ, Wright MW, Povey S (2002) Guidelines for human gene nomenclature. Genomics 79:464-470.
Walker RM, Hill AE, Newman AC, Hamilton G, Torrance HS, Anderson SM, Ogawa F, Derizioti P, Nicod J, Vernes SC, Fisher SE, Thomson PA, Porteous DJ, Evans KL (2012) The DISC1 promoter: characterization and regulation by FOXP2. Human molecular genetics 21:2862-2872.
Watkins KE, Vargha-Khadem F, Ashburner J, Passingham RE, Connelly A, Friston KJ, Frackowiak RS, Mishkin M, Gadian DG (2002) MRI analysis of an inherited speech and language disorder: structural brain abnormalities. Brain : a journal of neurology 125:465-478.
Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, Parkinson H (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 42:D1001-1006.
Wilcke A, Ligges C, Burkhardt J, Alexander M, Wolf C, Quente E, Ahnert P, Hoffmann P, Becker A, Muller-Myhsok B, Cichon S, Boltze J, Kirsten H (2012) Imaging genetics of FOXP2 in dyslexia. European journal of human genetics : EJHG 20:224-229.
Willsey AJ et al. (2013) Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell 155:997-1007.
Zeesman S, Nowaczyk MJ, Teshima I, Roberts W, Cardy JO, Brian J, Senman L, Feuk L, Osborne LR, Scherer SW (2006) Speech and language impairment and oromotor dyspraxia due to deletion of 7q31 that involves FOXP2. American journal of medical genetics Part A 140:509-514.
Zhang J, Webb DM, Podlaha O (2002) Accelerated protein evolution and origins of human-specific features: Foxp2 as an example. Genetics 162:1825-1835.
Zilina O, Reimand T, Zjablovskaja P, Mannik K, Mannamaa 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:254-256.

Chapter 2

A chromosomal rearrangement in a child with severe speech and language disorder separates FOXP2 from a functional enhancer

This chapter has been published in

Becker M, Devanna P, Fisher SE, Vernes SC (2015) A chromosomal rearrangement in a child with severe speech and language disorder separates FOXP2 from a functional enhancer. *Molecular cytogenetics* 8:69.

Abstract

Mutations of *FOXP2* in 7q31 cause a rare disorder involving speech apraxia, accompanied by expressive and receptive language impairments. A recent report described a child with speech and language deficits, and a genomic rearrangement affecting chromosomes 7 and 11. One breakpoint mapped to 7q31 and, although outside its coding region, was hypothesised to disrupt *FOXP2* expression. We identified an element 2kb downstream of this breakpoint with epigenetic characteristics of an enhancer. We show that this element drives reporter gene expression in human cell-lines. Thus, displacement of this element by translocation may disturb gene expression, contributing to the observed language phenotype.

Mutations and chromosomal rearrangements that disrupt the *FOXP2* coding sequence cause childhood apraxia of speech (CAS) [also known as developmental verbal dyspraxia (DVD)], as well as expressive and receptive deficits in both spoken and written language (Lai et al., 2001; MacDermot et al., 2005; Feuk et al., 2006; Shriberg et al., 2006; Zeesman et al., 2006; Laffin et al., 2012; Palka et al., 2012; Rice et al., 2012; Zilina et al., 2012; Turner et al., 2013). Moralli et al. recently described a child with a complex chromosomal rearrangement affecting chromosome 7 and 11, showing severe speech and language problems, similar to the profile typically seen for *FOXP2* mutation cases (Moralli et al., 2015). For a detailed description of the phenotype we refer to the original clinical report (Moralli et al., 2015).

The rearrangement in this child consists of a pericentric inversion of chromosome 7 (involving 7p15 and 7q31) and a translocation between chromosomes 7 and 11 (involving 7q21 and 11p12) (Moralli et al., 2015). The inversion and translocation breakpoints do not interrupt the sequence of any protein-coding genes (Moralli et al., 2015). It is therefore likely that the observed phenotype is caused by altered expression of nearby genes. *FOXP2* was considered the most promising candidate gene, given that haploinsufficency of this gene is known to cause speech and language disorders with a similar phenotype (Rice et al., 2012). The chromosome 7q31 breakpoint was mapped to a position 205 kb downstream of the *FOXP2* locus and 22 kb upstream of the *MDFIC* gene (Figure 1A). *MDFIC* expression was not significantly different in fibroblasts taken from the proband as compared to those from unaffected relatives. However, Moralli et al. were not able to reliably determine if the breakpoint affected *FOXP2* regulation, because this gene shows very low expression in fibroblasts.

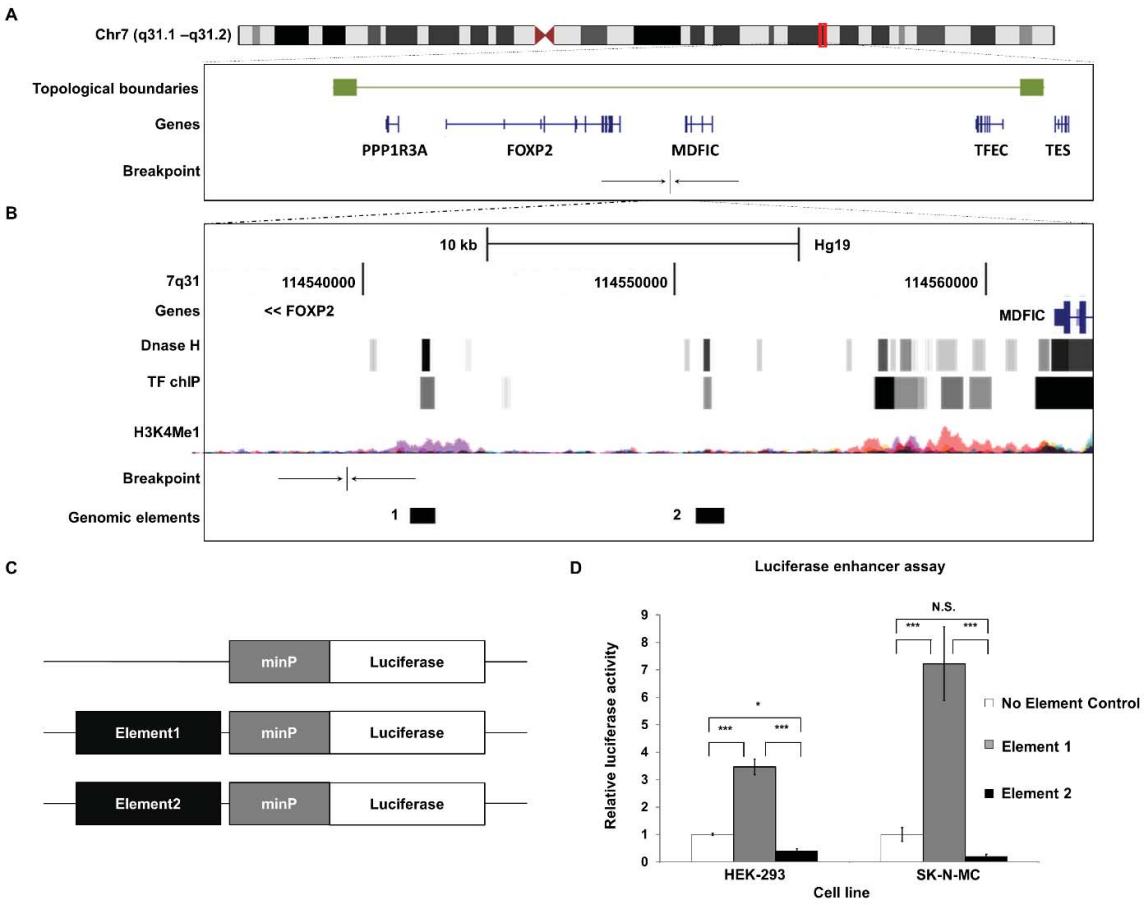

Figure 1: Identification of functional enhancer elements downstream of the inversion breakpoint. (A) The genomic location of the inversion breakpoint on chromosome 7q31 with the boundaries of the topological domain shown by green boxes (Dixon et al., 2012). Genes are depicted in blue. **(B)** Detailed view of the breakpoint with functional annotations from the ENCODE project (ENCODE, 2012) and UCSC genome browser (Kent et al., 2002). The inversion breakpoint was mapped to position 114539340 of chromosome 7 in human genome build hg19 (Moralli et al., 2015) (indicated by arrows). DNase hypersensitivity sites (DNase H) indicates regions of open chromatin. The darkness of the grey boxes is proportional to maximum signal strength of the DNase hypersensitivity in any investigated cell line. TF ChIP indicates sites that were occupied by TFs and the colour intensity is proportional to the amount of different TFs found at one site. The H3K4Me1 track indicates the presence of histone-3 lysine-4 mono-methylation in human cell lines. Pink peaks correspond to those found in NHEK cells (colorectal carcinoma) and orange peaks to those found in GM12878 (EBV transformed lymphoblast) cells. The investigated genomic elements are represented as black bars. Element 1 overlaps a DNase H site detected in 48 cell lines and 4 TFs. Element 2 overlaps a DNase H site detected in 17 cell lines and 1 TF. **(C-D)** Reporter constructs containing no enhancer (no element control), Element 1 (chr7: 114541370 – 114542201, hg19) and Element 2 (chr 7: 114550520 – 114551429, hg19) were transfected into HEK293 and SK-N-MC cells. After 2 days the luciferase activity was determined as described before (Deriziotis et al., 2014). The firefly luciferase activity was normalized to the activity of a co-transfected renilla luciferase control to determine the relative luciferase activity. The luciferase activity driven by the genomic elements was compared to that obtained from the control luciferase plasmid (minP). Element 1, but not Element 2 was able to enhance the activity and therefore the expression of the reporter gene. We performed two independent experiments of each three biological replicates. Significance was determined by an ANOVA followed by post-hoc Tukey test.*p<0.05, ***p<0.001. N.S.: not significant.

We hypothesized that the inversion in the Moralli et al. case would physically separate the *FOXP2* coding region from a genomic element with the potential to regulate expression of this gene. In this letter we identify and characterize a functional regulatory element located >205 kb downstream of *FOXP2*. Our findings suggest a mechanism by which the breakpoint could disrupt regulation of *FOXP2* expression and provide support for the causative nature of this rearrangement.

To determine if the 7q31 breakpoint disrupted a regulatory element we used functional genomics data from the ENCODE project (ENCODE, 2012) to predict possible enhancer regions that would drive gene expression. Although the reported breakpoint did not directly disrupt any predicted enhancers, we identified two possible enhancers that are in close proximity. One element, located 2.5 kb downstream of the breakpoint (Element 1), includes a region of open chromatin (demonstrated via DNase hypersensitivity across multiple cell lines) and carries histone modifications characteristic of an enhancer (H3K4Me1)(Hon et al., 2009; Shlyueva et al., 2014) (Figure 1B). This genomic site has been shown to bind several transcription factors (TF) in a colorectal carcinoma cell line (TF ChIP; Figure 1B), including RNA polymerase II, which is found at transcriptional start sites and active enhancers (Louie et al., 2003; Bonn et al., 2012). A second candidate region, located 12 kb downstream of the breakpoint (Element 2), shows DNase hypersensitivity and TF binding, but no H3K4Me1 modifications (Figure 1B).

To test whether these regions could act as enhancer elements we cloned them into a reporter construct in front of a minimal promoter and luciferase reporter gene (Figure 1C). The resulting constructs were used to measure the ability of each element to drive increased expression of the reporter gene in two human cell-lines; HEK293 and SK-N-MC (Materials and Methods see Chapter 5). Since both cell lines endogenously express *FOXP2* (Schroeder and Myers, 2008)

they likely express TFs that are able to regulate this gene. Element 1, which had multiple chromatin signatures characteristic of an enhancer, was able to act as a functional enhancer in both cell-lines (Figure 1D). In HEK293 cells we observed a 3 fold increase of luciferase expression in comparison to the empty vector control. In SK-N-MC cells the luciferase expression increased nearly 7 fold as compared to the control. Element 2, which lacked histone modifications characteristic of an enhancer, was not able to drive expression in either cell line. Therefore, Element 1 is a functional enhancer capable of driving increased gene expression in both human cell lines.

Thus, following fine-mapping of the breakpoint by Moralli et al. (Moralli et al., 2015), we were able to identify an active enhancer that is displaced by the chromosome 7 inversion. Genome wide structural mapping has shown that there are topological boundaries that regulatory elements are unlikely to cross (Dixon et al., 2012). Enhancers usually regulate genes that lie within the same topological domain, suggesting 4 genes to be potentially regulated by Element 1; *PPP1R3A, FOXP2, MDFIC* and *TFEC.* The inversion separates Element 1 from *PPP1R3A* and *FOXP2.* Given that *PPP1R3A* is a muscle specific gene not thought to be expressed in the brain (Tang et al., 1991), we consider the disrupted regulatory control of *FOXP2* likely to be a contributing factor to the phenotype found in this proband.

In sum, our functional data provide experimental support to the theory posited in Moralli et al. (Moralli et al., 2015) that the chromosome 7 breakpoint carried by this patient contributed to the speech and language phenotype by disrupting the regulation of *FOXP2*.

References

Bonn S, Zinzen RP, Girardot C, Gustafson EH, Perez-Gonzalez A, Delhomme N, Ghavi-Helm Y, Wilczynski B, Riddell A, Furlong EE (2012) Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat Genet 44:148-156.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485:376-380.

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

Feuk L et al. (2006) Absence of a paternally inherited FOXP2 gene in developmental verbal dyspraxia. Am J Hum Genet 79:965-972.

Hon GC, Hawkins RD, Ren B (2009) Predictive chromatin signatures in the mammalian genome. Human molecular genetics 18:R195-201.

Laffin JJ, Raca G, Jackson CA, Strand EA, Jakielski KJ, Shriberg LD (2012) Novel candidate genes and regions for childhood apraxia of speech identified by array comparative genomic hybridization. Genetics in medicine : official journal of the American College of Medical Genetics 14:928-936.

Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. Nature 413:519-523.

Louie MC, Yang HQ, Ma AH, Xu W, Zou JX, Kung HJ, Chen HW (2003) Androgen-induced recruitment of RNA polymerase II to a nuclear receptor-p160 coactivator complex. Proc Natl Acad Sci U S A 100:2226-2230.

MacDermot KD, Bonora E, Sykes N, Coupe AM, Lai CS, Vernes SC, Vargha-Khadem F, McKenzie F, Smith RL, Monaco AP, Fisher SE (2005) Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. Am J Hum Genet 76:1074-1080.

Moralli D, Nudel R, Chan MT, Green CM, Volpi EV, Benitez-Burraco A, Newbury DF, Garcia-Bellido P (2015) Language impairment in a case of a complex chromosomal rearrangement with a breakpoint downstream of FOXP2. Molecular cytogenetics 8:36.

Palka C, Alfonsi M, Mohn A, Cerbo R, Guanciali Franchi P, Fantasia D, Morizio E, Stuppia L, Calabrese G, Zori R, Chiarelli F, Palka G (2012) Mosaic 7q31 deletion involving FOXP2 gene associated with language impairment. Pediatrics 129:e183-188.

Rice GM, Raca G, Jakielski KJ, Laffin JJ, Iyama-Kurtycz CM, Hartley SL, Sprague RE, Heintzelman AT, Shriberg LD (2012) Phenotype of FOXP2 haploinsufficiency in a mother and son. American journal of medical genetics Part A 158A:174-181.

Schroeder DI, Myers RM (2008) Multiple transcription start sites for FOXP2 with varying cellular specificities. Gene 413:42-48.

Shlyueva D, Stampfel G, Stark A (2014) Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet 15:272-286.

Shriberg LD, Ballard KJ, Tomblin JB, Duffy JR, Odell KH, Williams CA (2006) Speech, prosody, and voice characteristics of a mother and daughter with a 7;13 translocation affecting FOXP2. Journal of speech, language, and hearing research : JSLHR 49:500-525.

Tang PM, Bondor JA, Swiderek KM, DePaoli-Roach AA (1991) Molecular cloning and expression of the regulatory (RG1) subunit of the glycogen-associated protein phosphatase. J Biol Chem 266:15782-15789.

Turner SJ, Hildebrand MS, Block S, Damiano J, Fahey M, Reilly S, Bahlo M, Scheffer IE, Morgan AT (2013) Small intragenic deletion in FOXP2 associated with childhood apraxia of speech and dysarthria. American journal of medical genetics Part A 161A:2321-2326.

Zeesman S, Nowaczyk MJ, Teshima I, Roberts W, Cardy JO, Brian J, Senman L, Feuk L, Osborne LR, Scherer SW (2006) Speech and language impairment and oromotor dyspraxia due to deletion of 7q31 that involves FOXP2. American journal of medical genetics Part A 140:509-514.

Zilina O, Reimand T, Zjablovskaja P, Mannik K, Mannamaa 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:254-256.

Chapter 3

Identification of neuronal *FOXP2* enhancer contacts in human cell lines

Abstract

Mutations in the *FOXP2* gene cause a speech and language disorder, involving developmental verbal dyspraxia. The gene is expressed in distinct brain structures, some of which show functional and structural aberrations in individuals with *FOXP2* mutations. In general, gene expression is regulated by genomic elements, such as enhancers, but in the case of *FOXP2* gene, these are mostly unknown.

In this chapter, I describe the identification of enhancers of *FOXP2.* To accomplish this, I used chromatin conformation capture (3C) in human neuronal-like cell lines. I measured the chromatin interactions of the active promoter to distant genomic regions and detected contacts to several genomic regions in cells with and without endogenous *FOXP2* expression. Quantitative analysis revealed that the majority of chromatin interactions were stronger in *FOXP2* positive neuronal-like cell lines. Using epigenetic data from the Roadmap Epigenomics Project, I localized candidate enhancer elements with enhancer specific histone marks in neuronal-derived tissue and cell-samples.

The chromatin interactions of the *FOXP2* promoter strongly indicate that the expression of *FOXP2* is controlled by a number of regulatory genomic regions. I detected promoter-promoter contacts to the *MDFIC* promoter, which suggest that this gene is co-regulated with *FOXP2*. Increased interaction frequencies to the putative enhancers in *FOXP2* expressing cell-lines suggest that the interactions positively influence the *FOXP2* expression. The molecular mechanisms that regulate the putative enhancers and their activity in the brain were investigated in the subsequent chapters.

Introduction

The basic elements of transcription

The human diploid genome carries two copies of each gene, one on each autosomal chromosome. Yet, the cell may contain thousands of RNA copies of each gene, which are further translated into proteins. To achieve this, the genetic information is first converted from DNA to RNA in the process of transcription. This process is tightly controlled to determine precise levels of gene expression during development and in response to environmental cues. In the previous chapter I have identified one enhancer, which may contribute to the expression of *FOXP2.* I have noted that the detection of chromatin interactions between the identified enhancer and the *FOXP2* promoter would be required to validate this link. The focus of this chapter lies on the identification of novel human enhancers of *FOXP2*, using a technique that captures the three-dimensional interactions of chromatin.

In living eukaryotic cells, the genome is packaged in the form of chromatin, which is a complex of DNA associated with proteins and RNA (Cairns, 2009). The functions of chromatin are to condense the DNA molecule so that the genome fits into the cell nucleus, to stabilize and protect the DNA from damage and to control gene expression (Berger, 2007; Bell et al., 2011). At transcriptionally inactive genomic regions the chromatin is tightly packed so that the DNA is inaccessible to the transcription machinery. For transcription to be initiated the chromatin has to be in an open and accessible state. The exposed DNA sequence allows transcription factors (TF), transcription initiation factors and RNA polymerases to bind to the nucleotides (Cairns, 2009; Jiang and Pugh, 2009; Spitz and Furlong, 2012). Transcription is initiated at gene promoters and the first nucleotide that is transcribed into RNA is called the transcriptional start site (TSS). General transcription initiation factors bind to promoter sequences and position the RNA polymerase at the TSS (Coulon et al., 2013). Promoter sequences typically span 2kb around the TSS (Cairns, 2009). The promoters are bound by TFs, which in turn recruit the transcription preinitiation complex and facilitate the positioning of RNA-polymerase

at the TSS. This finally results in an increased transcription rate of the target gene (Coulon et al., 2013).

The *FOXP2* gene is located on the long arm of chromosome 7 on band q31 and is transcribed from the telomeric to the centromeric side of the positive DNA strand (Lai et al., 2001). Four TSSs were previously described for *FOXP2* (Figure 1), which were experimentally identified in human *FOXP2* expressing cell lines and tissues (Bruce and Margolis, 2002; Schroeder and Myers, 2008). The most distal TSS is located at exon S1, the second TSS is 329 kb downstream at exon 1, followed by the third TSS 1.3 kb downstream at exon 1b. The fourth TSS is located another 10kb downstream of TSS3 at exon 2. TSS2 and TSS3 show a tissue-specific activity pattern, with strong activity in the colon, trachea and cervix (Schroeder and Myers, 2008). TSS1 and TSS4 were ubiquitously active in *FOXP2* expressing cell lines and primary human tissue including neuronal tissue (Bruce and Margolis, 2002; Schroeder and Myers, 2008).

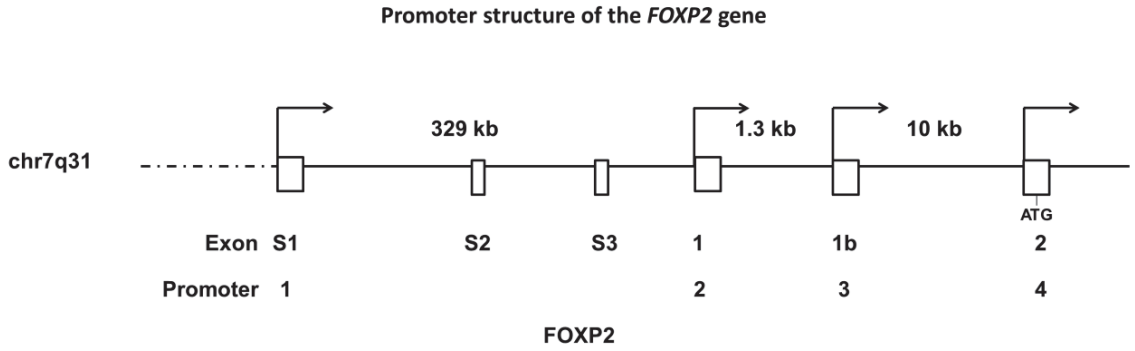

Figure 1: Schematic representation of the *FOXP2* promoter structure. *FOXP2* has four promoter regions, each containing one transcriptional start site (TSS). The promoters are indicated by arrows. The distal promoter 1 contains TSS1 and is located at the start of exon S1 at nucleotide 113,726,365 of chromosome 7 (hg19). The middle promoters 2 and 3 are located at exon 2 and exon 1b, respectively. Promoter 4 contains TSS4 and is located at exon 2. The distance between neighbouring promoters is given in kilobases (kb). Exon 2 contains the first translation start codon.

Enhancers, like promoters, are regulatory sequences that are bound by TFs and increase the transcription rate of their target genes, but may be located up to 1 million base pairs distant to the promoter (Shlyueva et al., 2014). Enhancers facilitate the assembly of the transcription complex by establishing physical contacts with the promoter, which creates a loop of

interspacing DNA (Visel et al., 2009; Ong and Corces, 2011). Enhancer contacts to *FOXP2* promoters have not been described in the literature. Following *in silico* predictions, two studies report the identification of enhancer regions near *Foxp2/foxp2* in zebrafish (Bonkowsky et al., 2008; Coutinho et al., 2011) and mouse (Coutinho et al., 2011). In addition, I have identified a human enhancer near *FOXP2* in the previous chapter 2 (Becker et al., 2015). However, it has not been shown that these enhancers make contact with *FOXP2*.

Challenges and possibilities to identify regulatory enhancer elements

The identification of enhancer elements is challenging for several reasons. Enhancer elements are relatively small with a size of a hundred to a few thousand base pairs. TFs bind to short sequence motifs of about 6-10 nucleotides within the enhancer sequence (Shlyueva et al., 2014). The TF binding site (TFBS) for an individual TF is degenerate and usually represented by one consensus motif. Therefore random sequence motifs that match a TFBS occur frequently in the genome without being functional (Wasserman and Sandelin, 2004). The functionality of any one enhancer is dependent on the combination of TFBSs. Thus, enhancers consist of clusters of short redundant binding motifs, of unknown combinatorial complexity. Additionally, enhancers can be distant from the genes that they regulate and the spacing DNA between enhancer and gene can be of magnitudes bigger than the size of the enhancer sequence itself (Ong and Corces, 2011; Spitz and Furlong, 2012). In short, the problem in predicting enhancer sequences is that they are short, redundant and complex clusters of sequence motifs, located at unknown distances from the protein-coding gene that they regulate. Finally, enhancers are highly context-specific, being active in specific tissues, developmental periods and in response to specific environmental stimuli. Therefore, the prediction of enhancers is error prone and relies substantially on the incorporation of additional data from molecular experiments or evolutionary conservation (Hardison and Taylor, 2012). For example, an algorithm for whole-genome identification of enhancers incorporated evolutionary constrained and clustered motif occurrence to achieve a predictive rate of only

13% (Blanchette et al., 2006). As a consequence, functional molecular data are critical to robustly identify enhancer elements.

Distinctive molecular properties enable the experimental identification of active enhancers. One property is that active enhancer regions are in an open chromatin state and accessible to TFs and chromatin remodelling complexes (Bell et al., 2011). Because these regions are open, they are accessible to DNA degrading enzymes and can be identified as DNase hypersensitivity sites (DHS). Another property of active enhancers is that the surrounding chromatin is marked by specific histone modifications. Different chemical histone modifications mark chromatin with specific functions, such as active enhancers, promoters or transcribed regions (Roh et al., 2005; Heintzman et al., 2009; Hardison and Taylor, 2012). Thus, genome-wide identification of enhancer-specific histone marks can be used to predict enhancer regions. The third property is that active enhancers are in physical contact with the promoter of their target genes. The genome-wide identification of active enhancers incorporating these experimental data achieves a predictive power of 40 to 80% (Roh et al., 2005; Heintzman et al., 2009; Cao et al., 2010; Hardison and Taylor, 2012). To reliably define regulatory elements of *FOXP2* I determined the physical contact of the *FOXP2* promoter to distant genomic regions as a proxy to identify enhancers.

Chromatin conformation at the FOXP2 gene locus

The interaction of distant genomic regions can be measured by chromatin conformation capture (3C) (Figure 2) (Dekker et al., 2002). With this technique, the chromatin is chemically fixated in its native state in living cells (Figure 2B). Interactions of distant genomic regions are cross-linked via intermediate proteins. The chromatin is then isolated from the cells and digested with restriction enzymes that recognize short cleavage sites (Figure 2C). The digested chromatin is re-ligated within an increased volume to favour proximity ligation of DNA strands within one chromatin complex (Figure 2D). The enhancer and promoter sequences will form one linear strand of DNA (Figure 2E). As a consequence of this procedure the

enhancer and promoter are spaced only by a few nucleotides instead of thousands of kilobases. The sequence of the ligation products can be predicted to design specific PCR primers, which anneal to the enhancer and promoter sequences. The ligation product can finally be detected by regular PCR (Figure 2A, E). In this chapter, I described the use of 3C to screen non-coding regions at the *FOXP2* locus to identify long-distant enhancer loops at this gene.

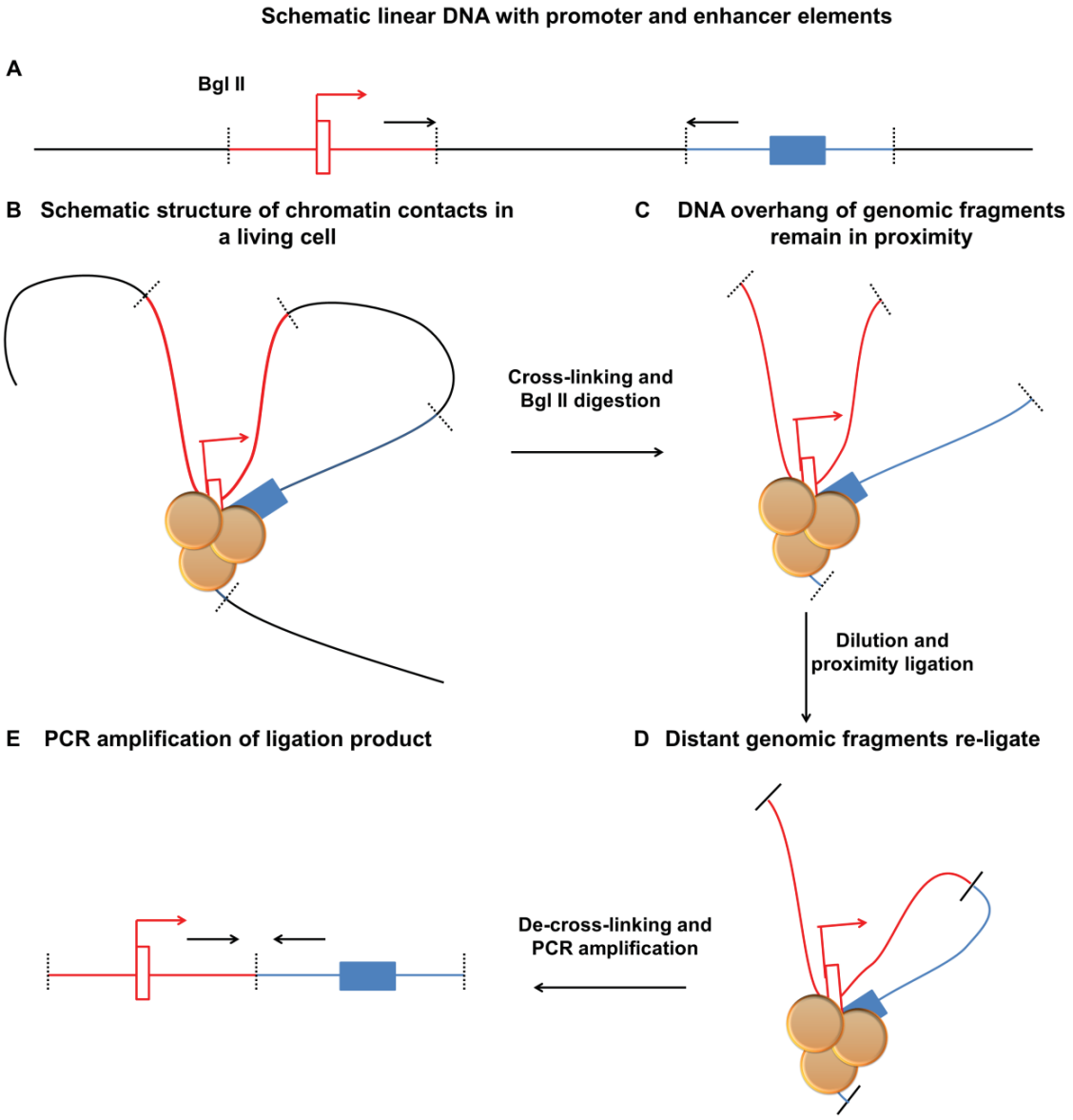

Figure 2: Overview of chromatin conformation capture (3C). A) Regulatory enhancer elements (blue block) are distant to their target promoters (red block and arrow) on the linear genome. Specific target primers (black arrows) are designed at BglII restriction sites (black dotted lines). B) In the living cells the genome is folded by proteins (brown spheres) and active enhancers contact their target promoters. The chromatin packed genome is cross-linked in this state and digested with BglII restriction enzymes. B) The genomic fragments and DNA overhangs remain in physical proximity. The DNA is diluted to favor proximity ligations between these genomic fragments. D) Genomic fragments that were distant in the linear genome are ligated to form one short DNA ligation product. E) The primers that were distant in genome amplify from the ligation product.

The aims in this chapter were to i) measure the *FOXP2* expression status in human neuronal-like cell lines, ii) determine, which active *FOXP2* promoter should act as the anchor point for the 3C design, iii) analyse long-range chromatin interactions of this promoter and iv) narrow down enhancer elements that can be further studied for their regulatory roles (as described in

chapters 4 and 5). The enhancer elements are the regulatory hubs that determine when, where and how *FOXP2* is expressed in the living organism. The identification of *FOXP2* enhancers enabled me to study the genetic pathways that regulate *FOXP2* expression (Chapter 4), as well as the brain structures and developmental periods in which these enhancers may act (Chapter 5).

Materials and Methods

Cell Culture

HEK293, SH-SY5Y, IMR-32, SK-N-AS and KELLY cell lines were purchased from HPA Culture Collections, SK-N-MC and PFSK-1 were purchased from ATCC. GM22647, GM22671, GM22731 and GM22737 were obtained from Coriell Cell Repositories.

HEK293 cells were grown in DMEM (Invitrogen), supplemented with 10% fetal bovine serum (FBS; Invitrogen) and 1% penicillin/streptavidin (P/S; Invitrogen). SK-N-MC and IMR-32 cells were grown in MEM (Invitrogen), supplemented with 10% FBS, 1% P/S and 2 mM L-glutamine (Invitrogen). SH-SY5Y and SK-N-AS were grown in DMEM:F12 (Invitrogen) supplemented with 10% FBS, 1% P/S, 2 mM L-glutamine and 1% non-essential amino acids (NEAA; Invitrogen). KELLY and PFSK-1 cell lines were grown in RPMI-160 (Invitrogen) medium supplemented with 10% FBS and 1% P/S and 2mM glutamine. EBV-cell lines GM22647, GM22671, GM22731 and GM22737 were grown in RPMI 1640 supplemented with 15% FBS, 1% P/S and 2mM L-glutamine.

Cells were cultured at 5% CO2 and 37°C. The adherent cells HEK293, SK-N-MC, PFSK-1, SH-SY5Y, KELLY, IMR-32 and SK-N-AS were passaged at 90-100% confluency. To passage cells, they were washed with PBS (Invitrogen) and detached from the cell culture surface using trypsin (Invitrogen). Cells were spun at 300g for 4 min to remove the old medium and re-suspended in fresh medium. Cells were routinely seeded into culture flasks for passaging or culture dishes for experiments.

Expression analysis

Cells were grown to 80-90 % confluency in 6 cm cell culture dishes and lysed in 1 ml of cold TRIzol (Life technologies). After repeatedly pipetting the TRIzol over the plate, the solution was transferred into Eppendorf tubes and 200 µl of chloroform was added. The mixture was

shaken vigorously for 2 min at RT and spun at 12,000g for 5 min at 4°C to separate the aqueous from the organic phase. The aqueous phase was transferred to a clean Eppendorf tube and mixed with 600 µl of 70% ethanol. The solution was then transferred to an RNeasy Spin Column (Qiagen) and the RNA cleaned according to the Qiagen RNeasy Spin Colum protocol. To remove genomic DNA on-column DNase digestion (Qiagen) was performed first, followed by a membrane washing step and elution in 30 µl of RNase-free water. The amount of RNA was quantified using Nanodrop (ThermoScientific) and the RNA quality determined on a 1% agarose gel.

The reverse transcription PCR was performed on 2000ng of RNA using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). In short, the Master Mix (2 µl RT buffer, 0.8 µl 25x dNTP, 2 µl RT random primer, 1 µl MultiScribe reverse transcriptase and 4.2 µl water) was prepared, 2000 ng of RNA added and the volume filled to 20 µl with RNase-free water. The used RT-PCR program was 10 min at 25°C followed by 120 min at 37°C and 5 min at 85°C.

To determine the relative amount of cDNA using real-time quantitative PCR (qPCR) the sample was diluted 1:10. 2 µl of the cDNA solution (approx. 20 ng of cDNA) was added to 18 µl of SYBR green (BioRad) Master Mix (10 µl 2x SYBR green, 8.3 µl water, 0.6 µl 10 µM forward primer, 0.6 µl 10 µM reverse primer). The primer sequences and coordinates are listed in appendix 1, table 1. The 2-step real-time PCR protocol consisted of an initial 3 min 95°C denaturation and enzyme activation step, followed by 40 cycles of 95°C denaturation for 15 sec and annealing/extension at 60°C for 30 sec. After 40 cycles the melt curve was recorded in a range from 55°C to 95°C in 0.5°C increments of 10 sec each. The PCR reaction was performed in the CFX96 PCR thermocycler (BioRad) and the fluorescence produced by the PCR amplification was recorded using the CFX manager software (BioRad). All real-time PCR reactions were performed in duplicate. The software automatically corrected for the background fluorescence and determined the threshold fluorescence value to derive the threshold cycle (Ct), at which logarithmic amplification was reached. The Ct values were

recorded for each sample and the Ct values of a GAPDH control PCR were subtracted to derive the delta Ct values.

Chromatin immunoprecipitation

Cells were grown to 80-90% confluency in T-75 cell culture flasks and trypsinized for 1-2 min. The living cells were counted (Biorad TC-10) and diluted to 2.5 x 10^6 cells per ml to a total volume of 10 ml (25 x 10^6 cells total) using fresh medium. 135 µl of 37% formaldehyde was added to cross-link cells for 10 min using 0.5% final formaldehyde concentration. The cells were shaken for 10 min at room temperature (RT), after which the cross-linking reaction was quenched by adding 125 mM freshly dissolved glycine. The solution was shaken for another 2 min to fully stop the reaction. The cells were spun down at 1600 rpm for 5 min at 4°C and the supernatant removed. To clean the cell pellet, 10 ml of cold PBS were added and the cross-linked cells re-suspended. The cells were spun again at 1600 rpm for 5 min at 4°C and the supernatant decanted afterwards. The pellet was re-suspended in 10 ml lysis buffer (20 mM HEPES pH7.6; 10 mM EDTA; 0.5 mM EGTA, 0.5% triton-x100) and the solution shaken for 10 min at 4°C. Following lysis, the cells were spun down at 1800 rpm for 5 min at 4°C and the suspension was decanted. The cell pellet was re-suspended in 10 ml of stop buffer solution (50 mM HEPES pH7.6; 150 mM NaCl; 1 mM EDTA; 0.5 mM EGTA) and shaken for 10 min at 4°C. The solution was spun down again at 1800 rpm for 5 min at 4°C and the suspension was carefully removed. The pellet was finally resuspended in 2.1 ml of incubation buffer (10 mM Tris pH8.0; 150 mM NaCl; 1 mM EDTA; 0.5 mM EGTA; 0.15% SDS; 1% Triton X-100) and pipetted into aliquots of 300 µl. The aliquots were sonicated in 1.5 ml Eppendorf tubes in the Bioruptor (Diagenode) set to “High” frequency power with cycles of 30 sec on and 30 sec off. The number of sonication cycles was determined for each cell type to result in fragments of 300-800 bp in size.

Before setting up the chromatin immunoprecipitation, 30 µl of protein G magnetic beads (Invitrogen) were washed three times with incubation buffer, using magnetic racks. After removing the supernatant of the third washing step, 40 µl of 5 x incubation buffer was added to the beads. To block for unspecific binding 3 µl of a 10mg/µl bovine serum albumin solution and to avoid protein degradation 10 µl protease inhibitor complex was added. 2 µl of Pol2RA (Diagenode, AC-055-100) and molecular grade water was added to a final volume of 200 µl. After adding 100 µl of sonicated chromatin, the immunoprecipitation was rotated overnight at 4°C.

The following day the samples were washed six times in total, using each 500 µl of different washing buffers; two times using wash buffer 1 (10 mM Tris pH 8.0; 150 mM NaCl; 1 mM EDTA; 0.5 mM EGTA; 0.1% SDS; 0.1% DOC; 1% Triton X-100), once using wash buffer 2 (10 mM Tris pH 8.0; 500 mM NaCl; 1 mM EDTA; 0.5 mM EGTA; 0.1% SDS; 0.1% DOC; 1% Triton X-100), once using wash buffer 3 (10 mM Tris pH 8.0; 1 mM EDTA; 0.5 mM EGTA; 0.25 M LiCl; 0.5% DOC; 0.5% NP-40) and two times using wash buffer 4 (10 mM Tris pH 8.0; 1 mM EDTA; 0.5 mM EGTA). To elute the remaining DNA from the magnetic beads, 200 µl of fresh elution buffer (1% SDS, 0.1 M NaHCO3) was added and the samples were rotated for 20 min at RT. The eluate was taken off the magnetic beads and phenol extracted using 200 µl of phenol: chloroform: isoamyl alcohol (SigmaAldrich). The DNA was further purified using ethanol precipitation.

Target DNA fragments were amplified using the iQ SYBR Green Supermix (BioRad). 2 µl of the pulled-down DNA sample was added to 18 µl of SYBR Green (BioRad) Master Mix (10 µl 2x SYBR Green, 8.3 µl water, 0.6 µl 10 µM forward primer, 0.6 µl 10 µM reverse primer). Primer sequences for each target fragment are listed in appendix 1, table 2. The 2-step real-time PCR protocol consisted of an initial 3 min 95°C denaturation and enzyme activation step, followed by 40 cycles of 95°C denaturation for 15 sec and annealing/extension at 60°C for 30 sec. After 40 cycles the melt curve was recorded in a range from 55°C to 95°C in 0.5°C increments of each 10 sec. The PCR reaction was performed in the CFX96 PCR thermocycler

(BioRad) and the fluorescence produced by the PCR amplification was recorded using the CFX manager software (BioRad). All real-time PCR reactions were performed in duplicate. The delta Ct values were derived as described before. Input DNA samples were taken from the sonicated aliquots, left at 4°C during the chromatin immunoprecipitation and included again in the elution step. The delta Ct values were normalized to the promoter region of the GAPDH housekeeping gene.

Chromatin conformation capture

The anchor point primer and TaqMan probe were designed to match the restriction fragment that contains the first transcriptional start site (TSS) of the *FOXP2* gene. Detection primers were designed to be complementary to the 5' end of restriction fragments so that the amplicons of the ligation products were not bigger than 250 bp. Primers were designed to match genomic fragments up to 106 kb upstream of TSS1 and 1,391 kb downstream of TSS1, spanning a total of 1,497 kb around the *FOXP2* gene locus. Within this genomic region, there are 428 restriction fragments produced by BglII digestion. Primers were designed for 50 fragments. The quality of the primers was assessed using standard curves of the PCR reaction on 3C samples with standardized concentrations (50ng, 125ng, 250ng, 500ng, 750ng and 1,000ng). 45 primers passed this quality control step, and they covered fragments of a total size of 218 kb. All primers are listed in appendix 1, table 3.

Cells were grown to 80-90% confluency in T175 cell culture flasks. Cells were trypsinized and the number of living cells was determined using TC-10 (BioRad) cell counter. 10^6 cells were transferred into a clean 50 ml falcon tube and fresh medium was added to a total volume of 10 ml. 135 µl of 37% formaldehyde was added to cross-link cells at a final concentration of 0.5% formaldehyde. The cross-linking reaction was set to RT. After 10 min the reaction was quenched by adding 1.41 ml of freshly dissolved 1 M glycine. The solution was inverted six times. The cells were spun down at 1600 rpm for 5 min at 4°C and the supernatant decanted.

The cell pellet was re-suspended in 2 ml of cold lysis buffer (50 mM tris, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1% triton X-100 and 1x protease inhibitor complex) and set to shake for 10 min at 4°C. To improve cell lysis the cell solution was transferred to a dounce tissue grinder and homogenized on ice with 2 times 15 piston strokes. The cell solution was again transferred to a 50 ml falcon tube and pelleted at 1800 rpm for 5 min at 4°C. The supernatant was carefully removed and the cell nuclei resuspended in 500 µl 1x restriction buffer 3 (NEB). The solution was transferred to 1.5 ml Eppendorf tubes and spun at 600g, for 5 min at 4°C. The supernatant was carefully removed and the pellet dissolved in 360 µl H_2O and 60 µl of 10x restriction buffer 3. To loosen the chromatin structure, 15 µl of 10% was added and incubated for 10 min at 65°C, shaking at 900 rpm on a Thermomixer (Eppendorf). The SDS was quenched by adding 150 µl of 10% triton X-100 and incubated for 1h at 37°C, shaking at 900 rpm. To control the digestion efficiency, 5 µl of the solution was separated as undigested control. To digest the chromatin, 400 units of BglII (8 µl of 50,000 U/ml; NEB) were added and the digestion reaction was incubated overnight at 37°C, shaking at 900 rpm. The next morning, another 200 units of BglII restriction enzyme were added for another incubation period of 1h at 37°C to guarantee proper digestion. Digestion efficiency was controlled on a 0.6% agarose gel. A barely visible genomic band and high molecular weight smear were indicative of proper genomic digestion.

Before starting the re-ligation of genomic fragments, the restriction enzyme was inhibited by adding 80 µl of 10% SDS and incubating for 30 min at 37°C. The solution was transferred to a 50 ml tube and 4,860 µl of H_2O and 700 µl of 10X ligation buffer were added. The ligation reaction was done in a 10X higher volume to favour ligation events between cross-linked DNA strands. The SDS was quenched with 750 µl of 10% triton X-100 and incubated for 30 min at 37°C. Finally, 50 Units of T4 Ligase (10 µl of 5 U/µl; Roche) were added and incubated at 16°C overnight.

To de-crosslink the samples 30µl Prot K (10 mg/ml) was added and incubated overnight at 65°C. The next morning, remaining RNA was digested by adding 30 µl of 10 mg/ml RNAse A (Invitrogen) and incubating for 45 min at 37°C. The DNA was subsequently purified by phenol

extraction using 50 ml PhaseLock Tubes (5 Prime). After ethanol precipitation, the DNA pellets were recovered in 500 µl TE to give the final 3C library DNA. To determine the concentration of the samples, a quantitative real-time PCR was performed, using SYBR green (BioRad). The calibration curve was made using human genomic DNA in a range of concentrations from 0.5 to 250 ng per PCR reaction. The amplified region in the calibration curve is a genomic fragment in the promoter region of GAPDH that does not contain a BglII restriction site.

The TaqMan real-time PCR of 3C ligation products was done using SsoFast Probes Supermix (BioRad). Each PCR reaction mixture contained 2,000 ng of 3C library DNA, 1.5 µM TaqMan probe, 0.5 µM anchor-point primer and 0.5 µM test primer in a 20 µl total reaction volume. All real-time PCR reactions were performed in triplicates. Primer sequences are listed in appendix 1, table 3. The sequence of the MGB-TaqMan probe is 5'-GATCTCTTAAACCACTGGGAATTCA-3' and matches the sequence on chromosome 7 from nucleotide 113,732,166 to 113,732,190 (reference genome hg19). The probe was ordered from Applied Biosystems. The 2-step real-time PCR protocol consisted of an initial 3 min 95°C denaturation and enzyme activation step, followed by 50 cycles of 95°C denaturation for 15sec and annealing/extension at 60°C for 30 sec. The PCR reaction was performed in the CFX96 PCR thermocycler (BioRad) and the fluorescence produced by the PCR amplification was recorded using the CFX manager software (BioRad).

The software automatically corrected for the background fluorescence and determined the threshold fluorescence value to derive Ct. The Ct values were recorded for all ligation products. The delta Ct value for each ligation product was calculated by subtracting the Ct value from the average Ct value of same ligation product in the EBV-lymphocytes. In case that the ligation product could not be detected in the EBV cell lines I subtracted a value of 40, which is equal to the lowest detected ligation product in the EBV cell lines. From the Ct values, I calculated the relative amount of starting ligation products by raising the negative Ct value to the power of two, which is a measure of the interaction frequency. To normalize across the different cell lines, I determined the ligation product with the lowest variation across all cell

lines (primer at -10,574 bp) and normalized to the interaction frequency of this ligation product. The final value is the relative interaction frequency.

Statistical analysis

Statistical analysis was performed using SPSS software. Statistical significance of the RNA polymerase II ChIP-qPCR was assessed using pairwise ANOVA and posthoc Tukey test. Statistical significance of the difference of 3C crosslinking frequencies to LBV-lymphoblast cell crosslinking frequencies was assessed using two-tailed student t-test. P-values were corrected for the amount of investigated interactions (45) using Bonferroni correction. To determine interactions with statistically significantly increased crosslinking frequency in FOXP2(+) cell lines, pairwise ANOVA was used between neuronal-like cell lines followed by post-hoc LSD test.

Results

FOXP2 is expressed in neuronal-like cell lines

Enhancers positively influence the expression of their target genes. Therefore, I expected to *FOXP2* enhancers to be active in cells that express this gene. I determined the expression of this gene in human neuronal-like cell lines. Human cell lines were used to study enhancers in the human genome. Also, cell lines were required to obtain sufficient amounts of chromatin for subsequent 3C analysis. I used neuronal-like cells to identify enhancer elements that are likely active in neuronal tissue, as enhancer interactions have been shown to be retained in cell lines derived from the corresponding tissue (Jager et al., 2015). I screened seven cell lines (SK-N-MC, PFSK-1, KELLY, IMR-32, SK-N-AS, SH-SY5Y and HEK293 cells), which are commonly used in molecular studies of neuronal genes (Vernes et al., 2007; Schroeder and Myers, 2008; Deriziotis et al., 2014). SK-N-AS, SH-SY5Y, IMR-32 and KELLY are neuroblastoma cell lines that were derived from different sites of metastasis. SK-N-AS and SH-SY5Y were isolated from the bone marrow; IMR-32 from an abdominal mass and for KELLY the site of isolation is unpublished (Schwab et al., 1983). SK-N-MC and PFSK-1 are malignant primitive neuroectodermal cancer cell lines. SK-N-MC was derived from the supra-orbital area of a patient with neuroepithelioma (Biedler et al., 1973; Biedler et al., 1978; Dunn et al., 1994) and PFSK-1 from a cerebral hemisphere (Fults et al., 1992). HEK293 cells have been obtained from human embryonic kidney tissue using adenovirus infection (Graham et al., 1977) and have previously been used for the study of *FOXP2* and other related neuronal genes (Vernes et al., 2006; Schroeder and Myers, 2008; Deriziotis et al., 2014). HEK293, SK-N-MC and PFSK-1 cells have previously been shown to express *FOXP2* (Schroeder and Myers, 2008).

To measure endogenous *FOXP2* expression, I harvested RNA from cultured cells and performed reverse transcription using an oligo-d(T) primer. The produced cDNA was amplified by quantitative PCR. The expression levels were first normalized to *GAPDH* and compared to

expression levels in HEK293, as this cell lined was repeatedly shown to express *FOXP2* (Bruce and Margolis, 2002; Schroeder and Myers, 2008). SK-N-MC cells show comparable *FOXP2* expression levels to HEK293, whereas the expression in PFSK-1 is significantly lower (Figure 3). No mRNA could be detected in KELLY, IMR-32, SK-N-AS and SH-S-Y5Y cells. This result shows that three cell lines, which have previously been shown to express *FOXP2,* are indeed *FOXP2* positive (FOXP2(+)). The other four cell lines are *FOXP2* negative (FOXP2(-)). Enhancer regions that increase *FOXP2* expression are expected to be active in the FOXP2(+) cell lines.

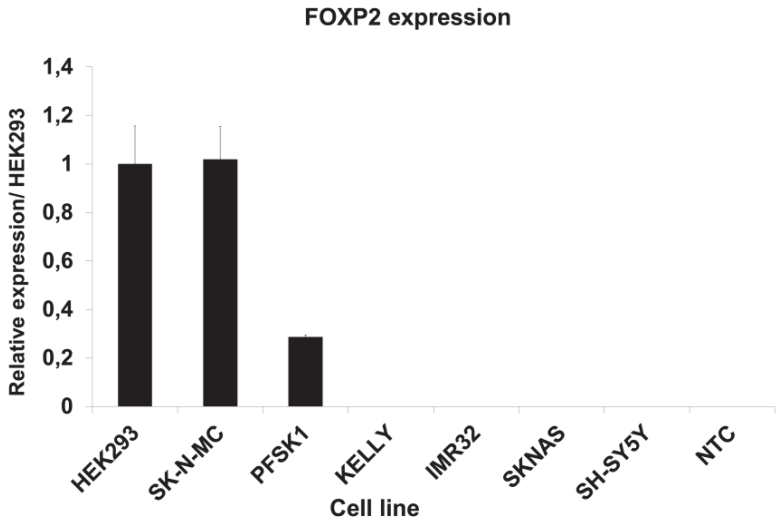

Figure 3: Endogenous *FOXP2* expression. Endogenous *FOXP2* mRNA expression levels in HEK, SK-N-MC, PFSK-1, KELLY, IMR-32, SK-N-AS and SH-SY5Y cells as determined by RT-qPCR. All expression levels have been normalized to HEK, which has been repeatedly shown to express *FOXP2*. NTC: no template control.

The promoter at TSS1 is active in FOXP2(+) cell lines

The *FOXP2* gene has four promoter regions (Figure 1), which could be transcriptionally active (Schroeder and Myers, 2008). In order to determine the promoter, to which I would measure the physical contacts in the 3C experiment, I first investigated the transcriptional activity of the four promoters. RNA Polymerase II (PolII) binding occurs at transcriptionally active promoters (Bonn et al., 2012; Core et al., 2012; Le Martelot et al., 2012). Therefore, I measured the PolII binding at the promoters, using chromatin immunoprecipitation (ChIP). Three primer pairs

were designed around promoter 1 (Figure4A: Primer 1-3) and promoter 4 (Figure4A: Primer 8-10) and four pairs around promoter 2 and 3 (Figure 4A: Primer 4-7). The primers were designed to anneal just upstream of the TSS as well as 1-2 kb upstream and downstream. At active TSSs, the strongest PolII binding can be detected a few nucleotides upstream of the TSS (Core et al., 2012). Thus, using multiple primer pairs per promoter I expected a peak shape signal around an active TSS.

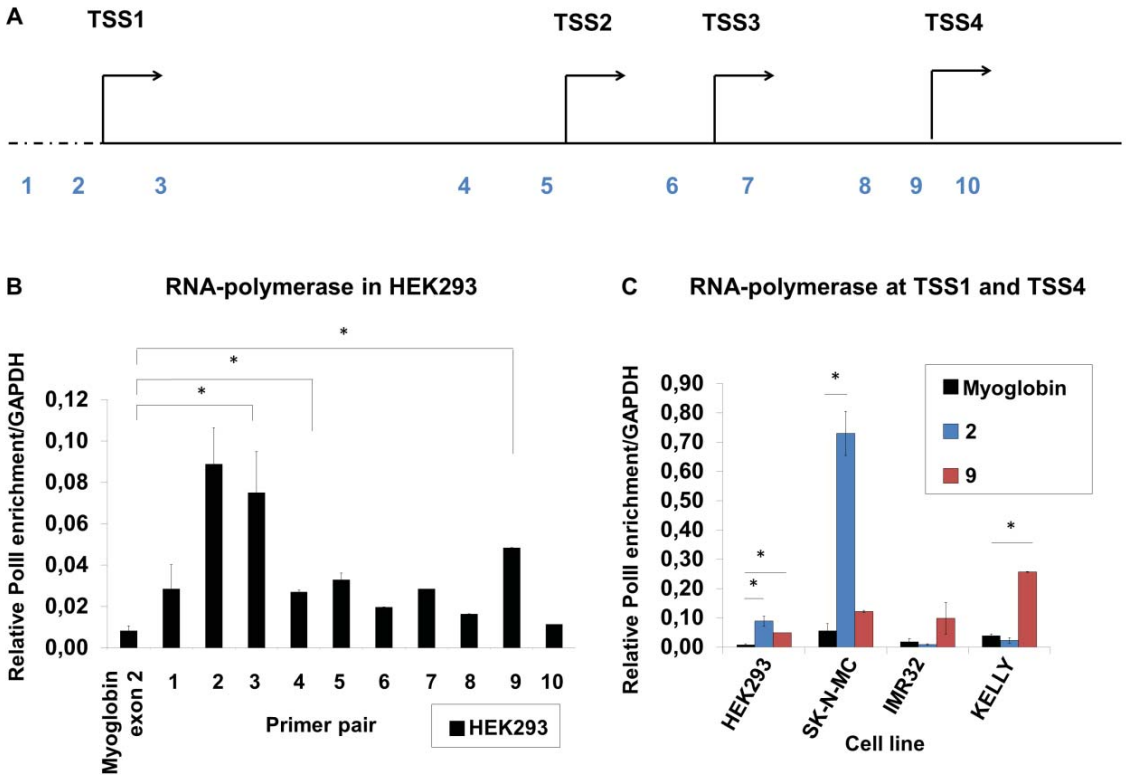

Figure 4: RNA polymerase occupancy at FOXP2 promoters. A) Schematic representation of the FOXP2 promoter regions and transcriptional start sites (TSS). The relative location of the qPCR primer pairs is indicated by blue lines. The primer coordinates are listed with the primer sequences in appendix1. B) RNA polymerase II (PolII) occupancy at the promoter regions of FOXP2 measured by ChIP-qPCR. The relative PolII occupancy in HEK293 cells is shown for each primer pair 1-10 and the myoglobin exon 2, which is used as a negative control. The RNA-PolII occupancy is normalized to the positive control region at the GAPDH promoter. C) The normalized PolII occupancy is shown for the cell lines HEK293, SK-N-MC, IMR32 and KELLY for TSS1 (primer pair 2), TSS4 (primer pair 9) and myoglobin exon 2. The qPCR was performed in duplicate. Significance was determined in comparison to the negative control by an ANOVA followed by post-hoc Tukey test.*$p < 0.05$.

First, I mapped the PolII binding at the promoter regions in the FOXP2(+) HEK293 cells. I used the promoter region of the *GAPDH* housekeeping gene, which is active across different cell-types to normalize the PolII signal between cell lines (Carter et al., 2010). As a negative control I included a region at exon 2 of the myoglobin gene, which is inactive in tissue other than muscle cells (Paschos et al., 2012; Tarrant-Elorza et al., 2014). I detected significant enrichment of PolII with primer pair 2 and 3 (Figure 4B), which are located 130 bp upstream and 1094 bp downstream of TSS1, respectively. No primer pair at TSS 2 or 3 was enriched for PolII (Figure 4B: primers 4-7). I also observed enriched PolII binding at primer pair 9, which is located 152bp upstream of TSS4 (Figure 4B). Thus, polymerase enrichment was detected just upstream of TSS 1 and TSS 4, which indicates that they are active in HEK293. I further determined PolII binding at promoter 1 and 4 in additional FOXP2(+) and FOXP2(-) cell lines, using the primer pairs 2 and 9 as proxy for TSS1 and 4, respectively. At TSS1 -primer 2- the two FOXP2(+) cell lines HEK293 and SKNMC showed, compared to the negative myoglobin control, 9x and 12x enrichment of PolII binding, respectively (Figure 4C). The FOXP2(-) cell lines IMR-32 and KELLY were not enriched for PolII at this promoter. At promoter 4, I detected significant PolII enrichment in HEK293 and KELLY. Thus, PolII binding occurs at promoter 4 in both FOXP2(+) and FOXP2(-) cell lines. PolII occupancy at promoter 1 was only detected in FOXP2(+) cells. This finding is in agreement with previous studies, which detected expression from this promoter in HEK293 and SK-N-MC (Schroeder and Myers, 2008). In conclusion, two FOXP2(+) cell lines show transcriptional activity at promoter 1. Therefore, I examined the chromatin interactions to this promoter in the following 3C experiment.

The FOXP2 promoter is in contact with coding and non-coding regions in HEK293 cells

I designed the 3C experiment to capture chromatin contacts of promoter 1 –the anchor-point- to distant genomic regions and thus determine enhancers that interact with this promoter. In the 3C protocol, the chromatin is chemically fixed in living cells and subsequently digested

with a site-specific restriction enzyme to produce genomic fragments (Figure 2). *In silico* digestion of the genome predicts the produced restriction fragments, including their size and sequence. Digestion sites within known functional elements could introduce a bias in the digestion and ligation reactions (Naumova et al., 2012). Therefore, I chose the restriction enzyme BglII, which produces a digestion fragment on which promoter 1 is located centrally. I designed one primer, further referred to as anchor-point primer, to the 3'end of the promoter-containing fragment (Figure 5B). In addition, I designed a TaqMan probe to match the sequence between the anchor primer and the BglII digestion site. Thus, both the anchor-point primer and TaqMan probe anneal on the promoter-containing restriction fragment. The use of a fluorescent TaqMan probe allowed me to perform high sensitive PCR reactions (Navarro et al., 2015).

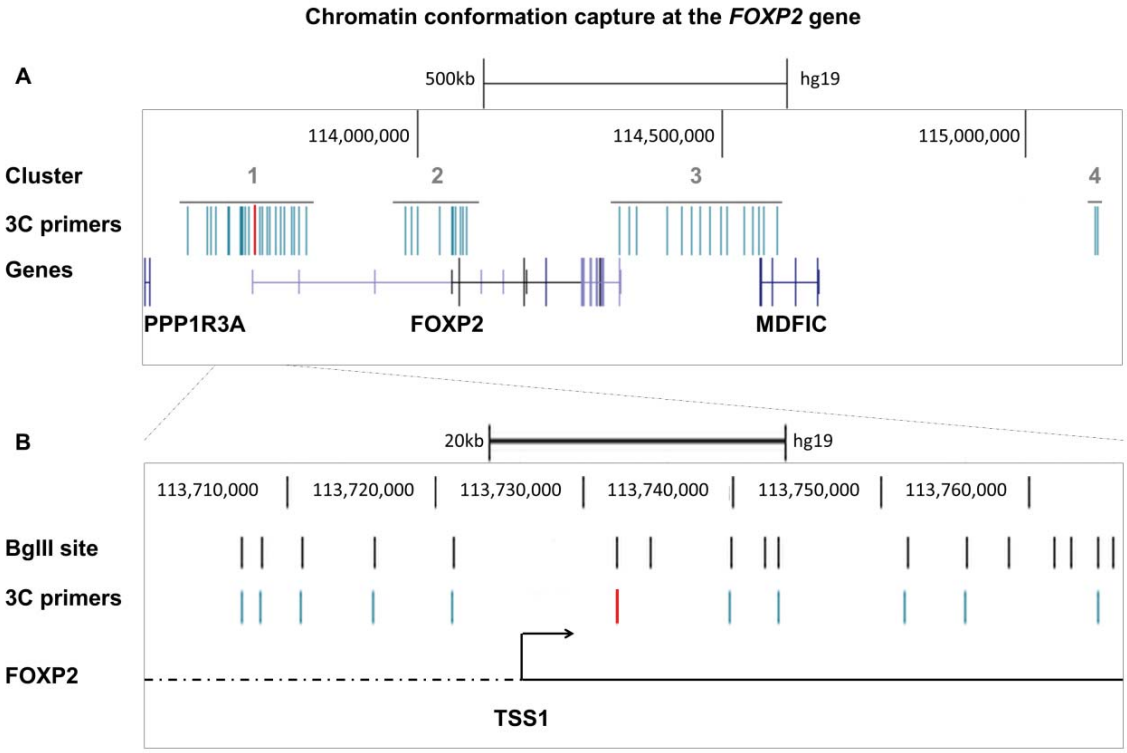

Figure 5: 3C design at the *FOXP2* gene locus. A) The genomic location of the *FOXP2* gene. The gene structure of *FOXP2* and the neighboring genes *PPP13RA* and *MDFIC* are shown in dark blue. The anchor point primer is located at the distal promoter 1 and shown in red. The 3C target primers, depicted in light blue, were designed to cluster at four locations of the gene locus: 1. around the anchor point promoter 1; 2. clustered around the downstream promoters 2-4; 3. at the downstream intergenic region between the *FOXP2* 3'UTR and *MDFIC*, 4. far downstream of the *MDFIC* gene. B) *FOXP2* TSS1 is located at nucleotide 113,726,365 of the human genome (hg19). The BglII restriction sites are depicted as black bars. The 3C primers (blue bars) have been designed just upstream of the restriction sites. The viewpoint primer (red) is located on the restriction fragment that contains TSS1. Note that not all BglII restriction sites are tagged by a 3C primer.

Target primers were designed to match the 3' ends of the predicted restriction fragments. To enable primer design with specific sequence recognition, the 3' sequence was required to be unique. On a genome-wide scale, the frequency of enhancer occurrences declines with the distance to gene promoters in a symmetrical distribution (Stranger et al., 2007; Mifsud et al., 2015). Therefore, I targeted candidate regions around the *FOXP2* promoters and the adjacent *MDFIC* promoter and designed several target primers to the genomic regions. One cluster of target primers was designed in a 200kb window around promoter 1, spanning the intergenic region between *PPP1R3A* and the first intron of *FOXP2* (Figure 5A). The second cluster of

target primers was designed in a 100kb window around the three downstream promoters of *FOXP2.* The third cluster was designed to the intergenic region between *FOXP2* and *MDFIC,* including the 3'-end of *FOXP2*. This cluster includes the functional enhancer that I have identified in the previous chapter (Chapter 2). In addition, two target primers were designed to an intronic region 460kb downstream of the *MDFIC* gene, targeting an unpublished deletion detected in a child with speech problems (personal communication). This patient has problems in expressive and receptive language, shows delays in communication skills and carries a 430kb deletion at the tagged genomic site, which does not overlap coding genes. In total, I designed primers that successfully target 45 digestion fragments (Figure 5A).

To normalize the interactions between cell lines I used lymphoblast cell lines as a control. In contrast to neuronal cells, which are derived from the ectoderm, these cells originate from the mesoderm (Seremetis et al., 1989). Therefore, I did not expect neuronal enhancer contacts within lymphoblast cells. First, I investigated the interaction landscape in HEK293 cells. In HEK293 cells I detected nine genomic fragments with significantly increased cross-linking frequencies compared to EBV lymphoblast cells (Figure 6B). Eight of nine significant interactions were found at local maxima, where the genomic target fragment has a higher cross-linking frequency than the neighbouring target fragments. The fragment at the local maximum likely contains the enhancer element that mediated the chromatin interaction. The eight local maxima in HEK293 cells are located at -37, 70, 330, 346, 604, 706, 772 and 843 kb relative to *FOXP2* promoter 1. This result shows that there are chromatin interactions between these regions and the *FOXP2* promoter in HEK293 cells, which are not present in lymphoblast cells.

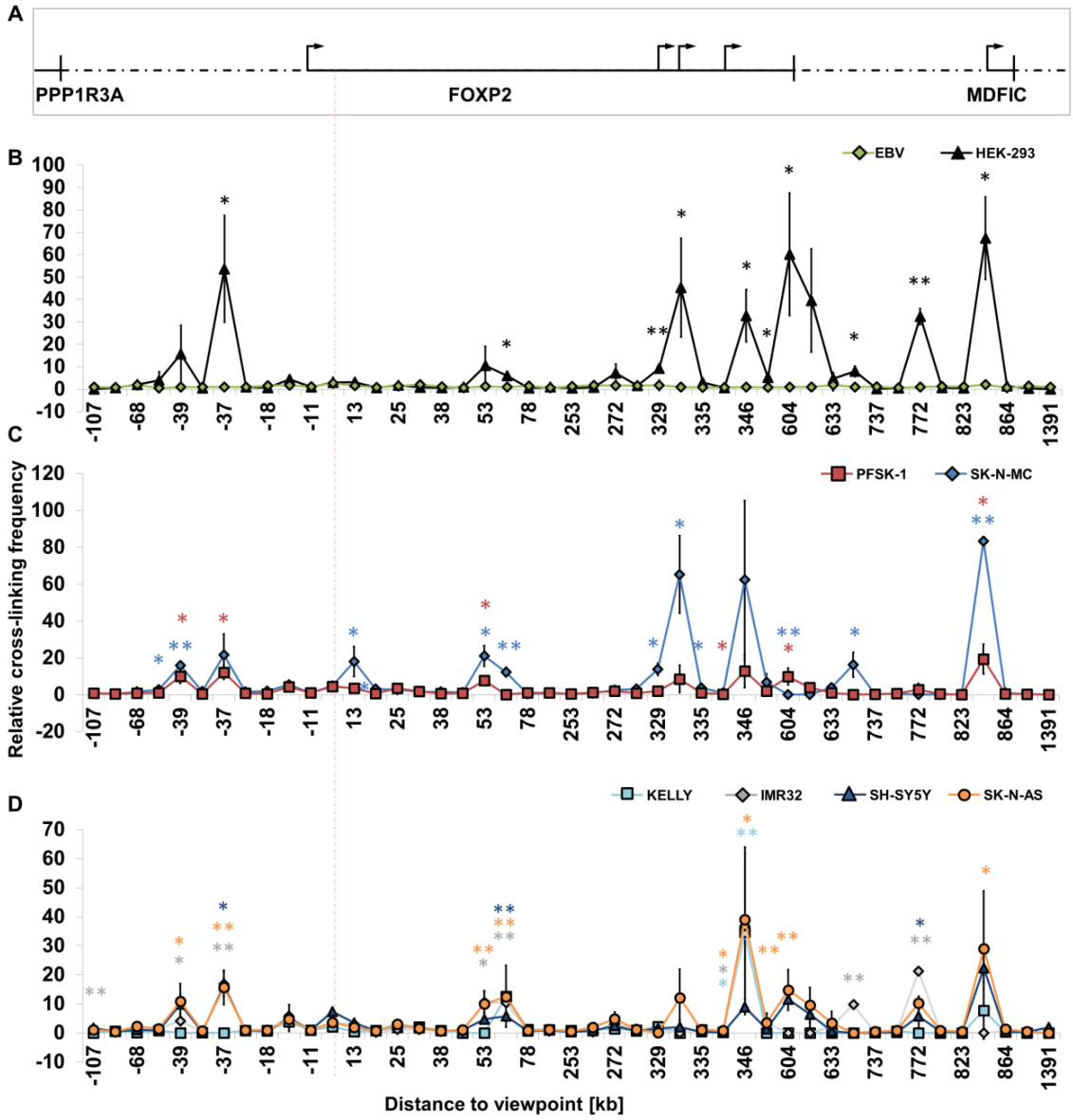

Figure 6: 3C interaction landscape at the *FOXP2* gene locus. Chromatin interactions at the *FOXP2* gene measured by chromatin conformation capture (3C). A) Schematic representation of the genomic locus of *PPP1R3A*, *FOXP2* and *MDFIC*. The indicated promoters (arrows) and 3'UTR (black bar) align with the data points of the following 3C graphs. B-D) The relative interaction frequencies of genomic fragments to the promoter viewpoint are shown according to their distance to the TSS. B) Chromatin interactions in HEK293 and EBV cell lines. C) Interactions in neuronal-like FOXP2(+) cell lines SK-N-MC and PFSK1. D) Interactions in neuronal-like FOXP2(-) cell lines KELLY, IMR32, SH-SY5Y and SKNAS. The cross-linking frequencies for each target primer were normalized to the cross-linking frequency of the target primer at –11kb (internal normalization). The cross-linking frequency for each target primer was normalized to the average cross-linking frequency of three measurements each in two EBV lymphoblast cell lines (between sample normalization). All interactions were determined in triplicate. Significance was calculated using two-tailed student t-test. P-values were corrected for the amount of tested genomic fragments using Bonferroni correction. * p-value < 0.05; ** p-value < 0.001

FOXP2(+) and (-) neuronal-like cells share regions of increased chromatin conformation

HEK293 cells were derived by viral transformation of embryonic kidney tissue, but their morphology and transcriptome suggest a neuronal cell of origin (Shaw et al., 2002). They are commonly used as a cellular system to study the function of neuronal genes. To obtain the interaction landscape of additional FOXP2(+) cell lines, which originate from neuronal tissue, I analysed the chromatin conformation in SK-N-MC and PFSK-1. In SK-N-MC cells I detected ten genomic fragments with significantly increased cross-linking frequencies, of which five mark local maxima at -39, 13, 53, 330, 706 and 843 kb (Figure 6C). All but one maximum at 13 kb were previously detected in HEK293 cells. The other significantly increased interactions occur at the slope of these peaks at neighbouring genomic regions. Additionally, I detected a significantly decreased cross-linking frequency at 604 kb, which showed increased interaction in HEK293. In PFSK1 cells I detected four genomic regions with a significant increase in cross-linking frequency (Figure 6B). Two of these mark local maxima that were detected in SK-N-MC and HEK293 at 53 and 843 kb. The genomic region at 604 kb, which was significantly decreased in SK-N-MC, is significantly increased in PFSK-1. I detected suggestive local maxima at genomic regions that were significant in the other two FOXP2(+) cell lines. No chromatin interaction was detected between promoter 1 and the gnomic fragment at 823kb distance, which contains a previously identified enhancer (Chapter 2).

To determine differences in chromatin conformation between neuronal-like cells of different *FOXP2* expression status I measured the chromatin conformation in FOXP2(-) neuronal-like cell lines KELLY, IM32, SH-SY5Y and SK-N-AS (Figure 6C). Eleven target fragments show significantly increased cross-linking frequencies to the anchor-point fragment in one or several cell lines. Eight of these interactions are located at local maxima and three interactions at neighbouring genomic fragments. The maxima were located at -39, -37, 70, 346, 604, 706, 772 and 843 kb. This result shows that chromatin contacts between these regions and the *FOXP2* promoter are present in the absence of active transcription. The interactions of these

regions were previously detected in FOXP2(+) cells, which indicates that chromatin contacts are shared between FOXP2(+) and (-) neuronal-like cell lines.

Chromatin interactions are stronger in FOXP2 expressing cell lines

In total I identified eleven local maxima at which genomic regions contact the *FOXP2* promoter (Figure 7M). Two maxima, at 13 and 330 kb, were exclusively detected in FOXP2(+) cells and nine were not shared with FOXP2(-) cell lines. To determine differences in chromatin interactions between FOXP2(+) and (-) cells I compared the relative crosslinking frequencies between the neuronal-like cell lines (Figure 7A-L). For each local maximum I calculated the pairwise difference between cell lines and determined maxima that were significantly stronger in FOXP2(+) cell lines. As expected, the cross-linking frequency at 13 kb is significantly stronger in SK-N-MC cells than in any of the FOXP2(-) cells (Figure 7C). Similarly, the interaction at region 330 kb is significantly stronger in HEK293 and SK-N-MC (Figure 7F). In total, this analysis identified eight genomic regions at -37, 13, 53, 330, 604, 706, 772 and 843 kb with significantly increased chromatin interactions in at least one *FOXP2* expressing cell line (Figure 7M). These genomic fragments likely contain *FOXP2* regulatory elements with enhancer functions.

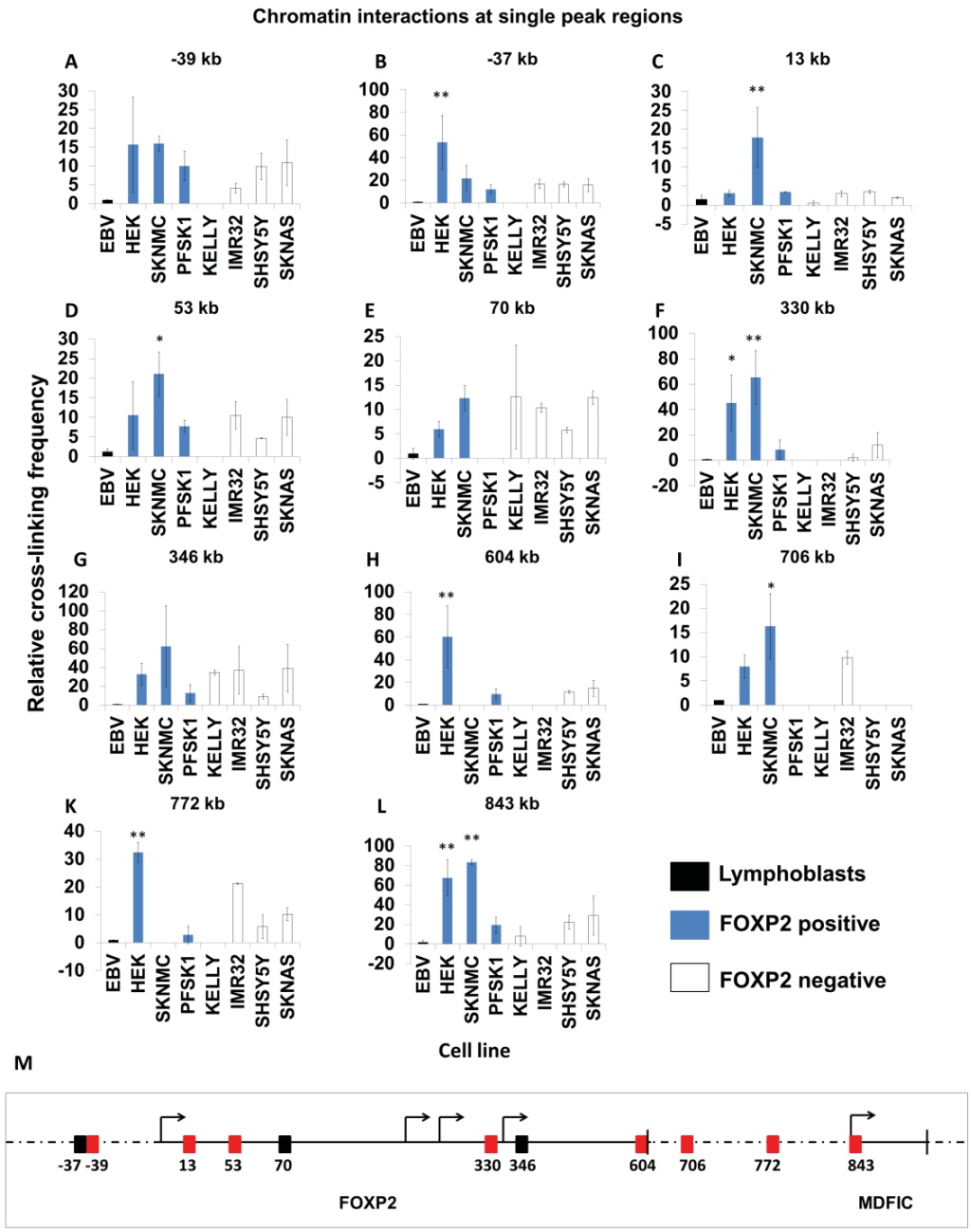

Figure 7: Chromatin interactions at single peak regions. The chromatin interactions of the *FOXP2* promoter to the local maxima were re-plotted from figure 7B-D. The graphs compare the cross-linking frequencies for individual maxima across the investigated cell lines. The cross-linking frequencies for EBV lymphoblast cells are shown as black bars. The FOXP2(+) cell lines are shown as blue bars and the FOXP2(-) cell lines as white bars. Peak regions are labelled according to their distance to promoter 1: A) -37 kb, B) -39 kb, C) 13 kb, D) 68 kb, E) 70 kb, F) 330 kb, G) 346 kb, H) 604 kb, I) 706 kb, K) 772 kb, L) 843 kb. Pairwise statistical difference was determined using univariate ANOVA followed by post-hoc LSD test. Statistical significant local maxima show stronger interaction in one FOXP2(+) cell lines compared to all FOXP2(-) cell lines. The reported p-value represents the least significant difference between a FOXP2(+) and FOXP2(-) cell line. * p-value < 0.05; ** p-value < 0.001. M) The peak regions are depicted in a schematic representation of the *FOXP2* genomic locus. Red: Regions with significantly higher interaction frequency in one or more FOXP2(+) cell lines.

Local maxima map to regions with neuronal epigenetic marks

I detected eight chromatin interactions with increased cross-linking frequencies in *FOXP2* expressing cell lines. The interactions occurred between promoter 1 and genomic fragments located at a distance of -37, 13, 53, 330, 604, 706, 772 and 843 kb (Figure 7M). The restriction digestion in the 3C method produced genomic fragments of 1 to 10 kb. The fragments likely contain enhancer sequences, which have a size of several hundred nucleotides. To follow up the enhancer function in molecular assays, it was necessary to define candidate enhancer elements on the genomic fragments. The next step was to narrow down the enhancer elements using histone marks.

The Roadmap Epigenomics Project mapped histone modifications in human tissue, including samples from seven adult brain regions, two foetal brains (male and female), the foetal brain germinal matrix, two neurosphere cultures derived from embryonic tissue (cortex and ganglionic eminence), two neuronal progenitor cultures (derived from H1 and H9 embryonic stem cells) and one neuron culture (derived from H9 embryonic stem cells). Twelve types of epigenetic mark (eleven histone modifications and DNase hypersensitivity) were analysed genome-wide and the combination of marks was used to predict the function of genomic regions in the respective tissue. I used the predicted states in neuronal tissues to analyse the genomic fragments with increased interaction frequencies and looked specifically for enhancer marks. This allowed me to link the predictive properties of neuron-specific histone marks to the enhancer interactions at the *FOXP2* promoter.

I analysed the genomic regions that potentially contain enhancer elements (Figure 11-15). The genomic fragment at -37 kb contains a site that is predicted to exert different functions in the 15 analysed neuronal samples. In cultured neurospheres, derived from the ganglionic eminence and tissue derived from the cingulate gyrus, angular gyrus and dorsolateral prefrontal cortex this site was predicted to function as an enhancer (Figure 8: GE derived neurosphere, cingulate gyrus, angular gyrus, dorsolateral prefrontal cortex). All predictions

overlap at one site, which is located at the 5'-digestion site of the -37kb genomic fragment. Hence, this region is located on the restriction fragment that was tagged by the 3C primer located at -37 kb. The marked region potentially mediated the observed interaction and was regarded as a candidate enhancer for follow-up experiments (Table 1).

Table 1: Candidate enhancer elements

Target 3C primer	Element	Size	Start (hg19)	End (hg19)	Mean PhastCons
-37	**-37**	774	113,688,009	113,688,782	0.023
13	**13**	886	113,732,364	113,733,250	0.203
53	**53**	1,094	113,781,175	113,782,269	0.045
330	**330**	1,801	114,056,845	114,058,646	0.891
604	**604**	2,779	114,327,034	114,329,813	0.692
706	**700**	3,645	114,424,203	114,427,847	0.146
706	**704**	4,087	114,427,887	114,431,974	0.192
772	**772**	2,243	114,495,963	114,498,206	0.048
843	**843**	3,958	114,568,454	114,572,411	0.111

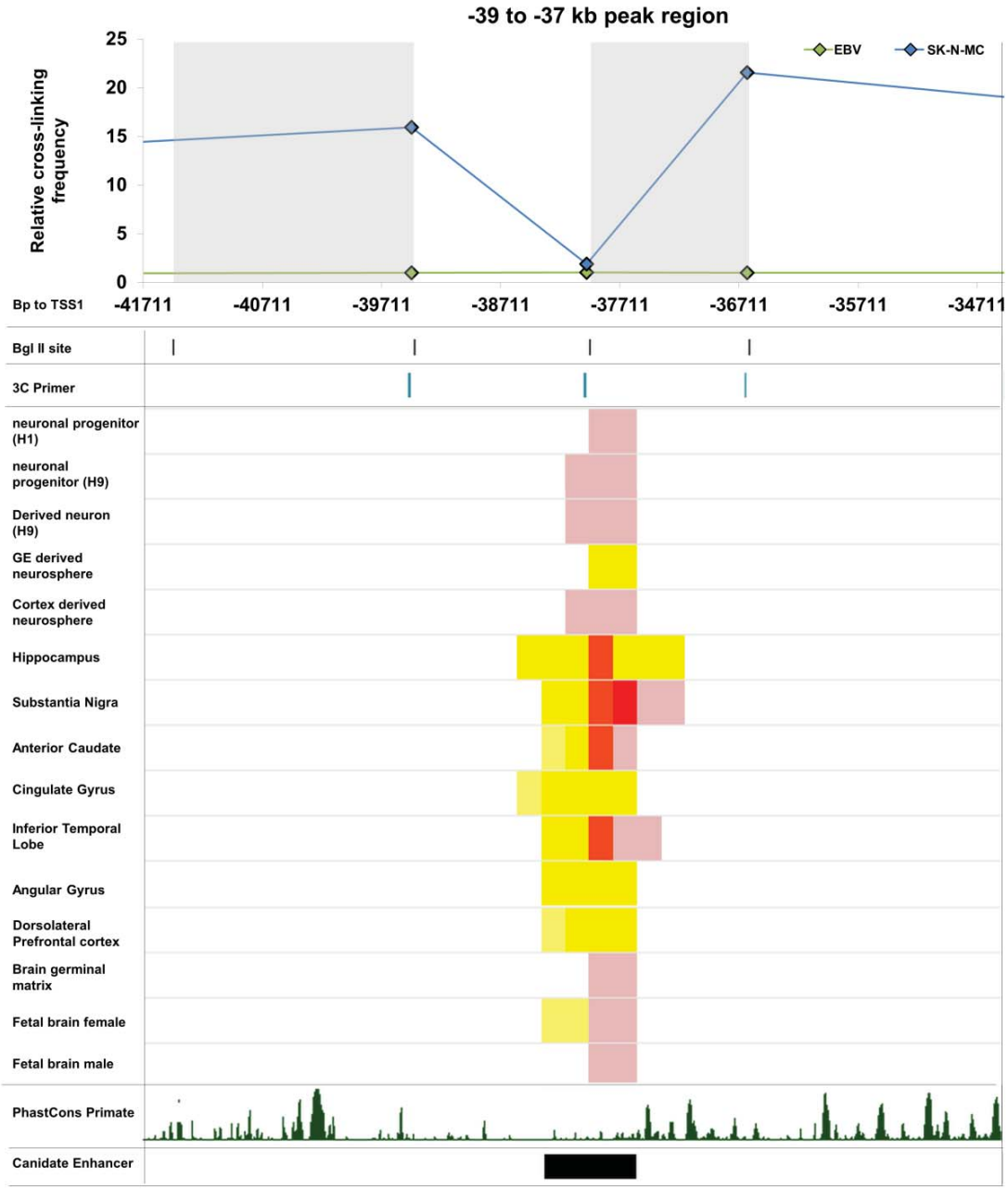

Figure 8: Detail of -39 to -37 kb peak region. The top panel shows the chromatin interactions in EBV-lymphoblast and the representative SK-N-MC cell line. The alternative grey and white fields indicate the location of genomic fragments produced by BglII digestion. The individual data points represent the interactions of whole genomic fragments, as meassured with the respective target primers. The x-axis shows the distance of the region to the *FOXP2* TSS1. Below the graph are annotations aligned to the genomic position (chr7:113,684,653-113,691,878 [hg19]). The BglII digestion sites (black lines) and 3C target primers (blue lines) are shwon in the first two rows. The predicted functional states from the Epigenome Roadmaps are shown for 5 human derived primary neuronal cell-types and 10 neuronal tissues. (light red = poised promoter, red = active TSS, yellow = enhancer activity). The PhastCons Primate track shows the evolutionary conservation between 10 primate genomes. The candidate enhancer annotation (black box) shows the suggested enhancer element at this peak region.

In total, I analysed eight genomic fragments for the presence of enhancer elements (Figure 8-15) and identified nine candidate enhancers (Table 1). In addition to candidate enhancer elements, two interacting fragments contained elements of known function. The genomic region at 604 kb contains the 3'-UTR of *FOXP2* (Figure 12) and the fragment at 843 kb spans the promoter of the adjacent *MDFIC* gene (Figure 15). Therefore, the *FOXP2* promoter seems to interact with gene elements that typically function in transctiption initiation and mRNA stability, but may also contribute to transcription regulation via long-distance interactions (Jash et al., 2012; Sahlen et al., 2015). The evolutionary conservation, which may indicate functional importance, of the candidate enhancers varies. The enhancer at 330kb is highly conserved (Mean Primate PhastConservation Score: 0.891, Figure 11, Table 1) and may be functional in a wide range of related species. The enhancer at -37kb shows low conservation (Mean Primate Phast Cons: 0.023, Figure 8, Table 1) and the function may be more specific to humans or primates. Thus, the 3C genomic fragments contain promising candidate enhancers of variable evolutionary conservation, which may regulate the expression of *FOXP2.*

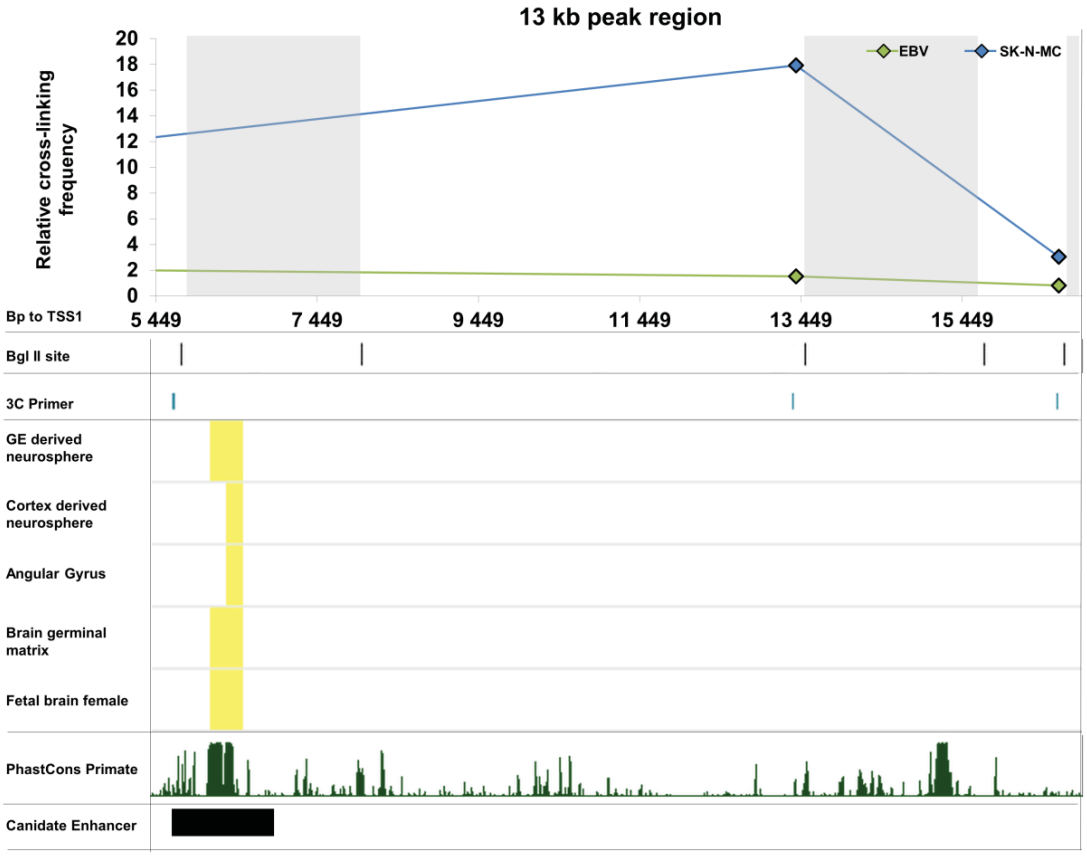

Figure 9: Detail of 13 kb peak region. See figure 8. Genomic position (chr7:113731832-113743326 [hg19]).

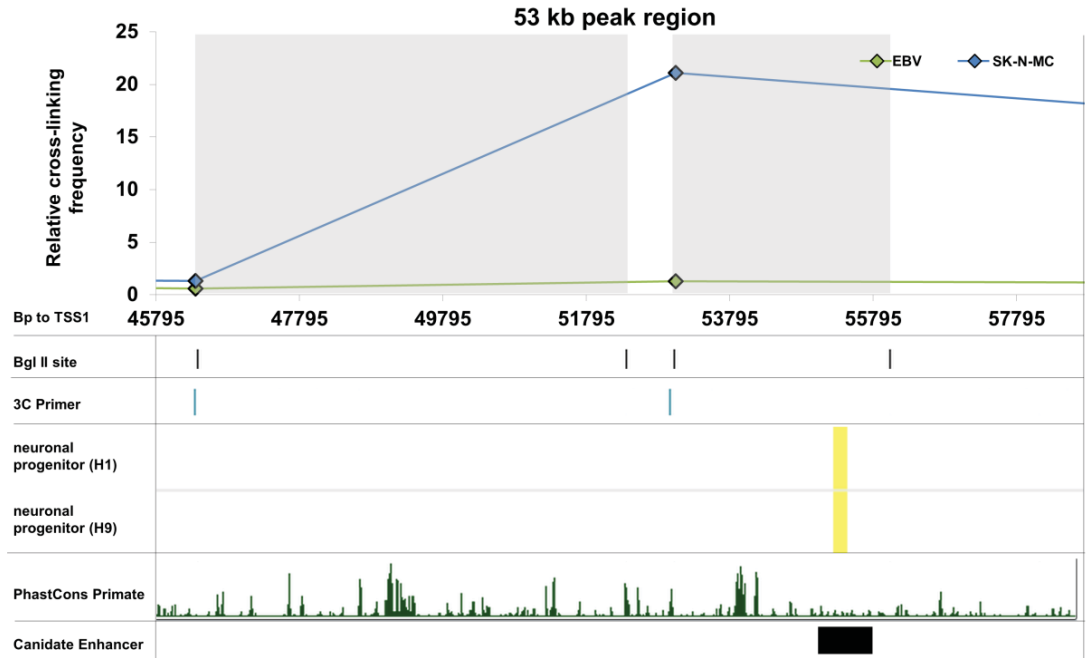

Figure 10: Detail of 53 kb peak region. See figure 8. Genomic position (chr7:113,772,159-113,785,107 [hg19]).

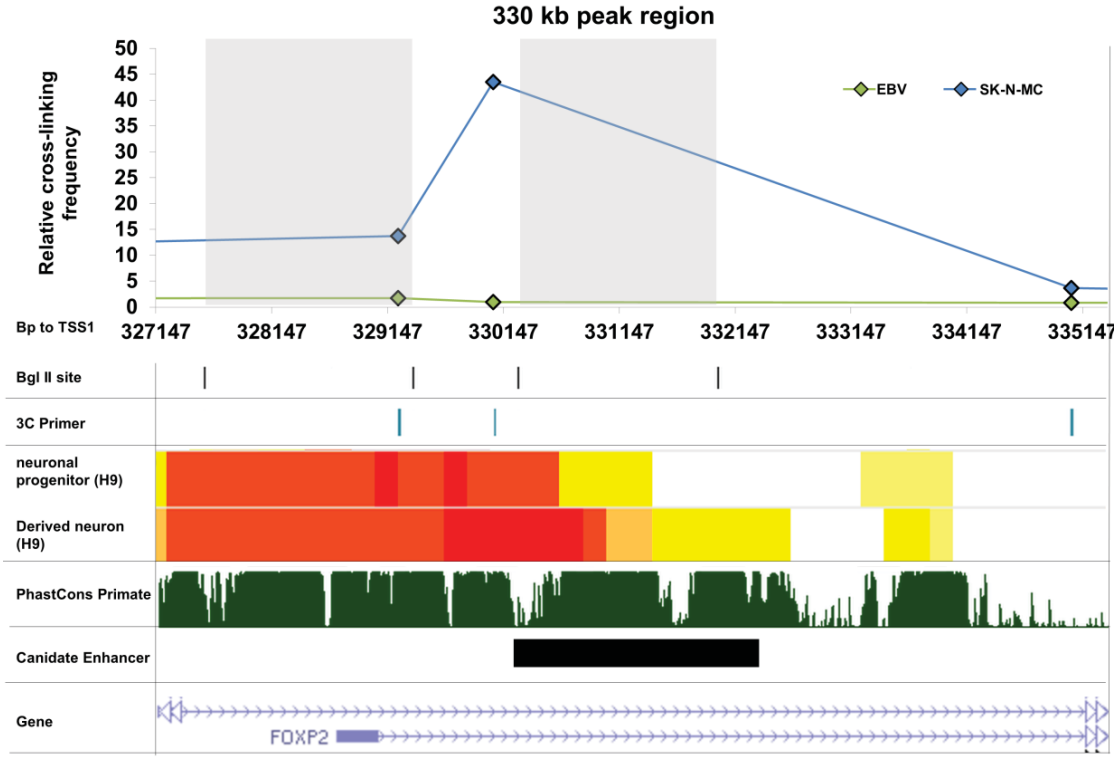

Figure 11: Detail of 330 kb peak region. See figure 8. Genomic position (chr7:114,053,511-114,061,726 [hg19]). The structure *FOXP2* transcripts are shown below, indicating exon 1 at promoter 2 at which the transcription of a short isoform starts.

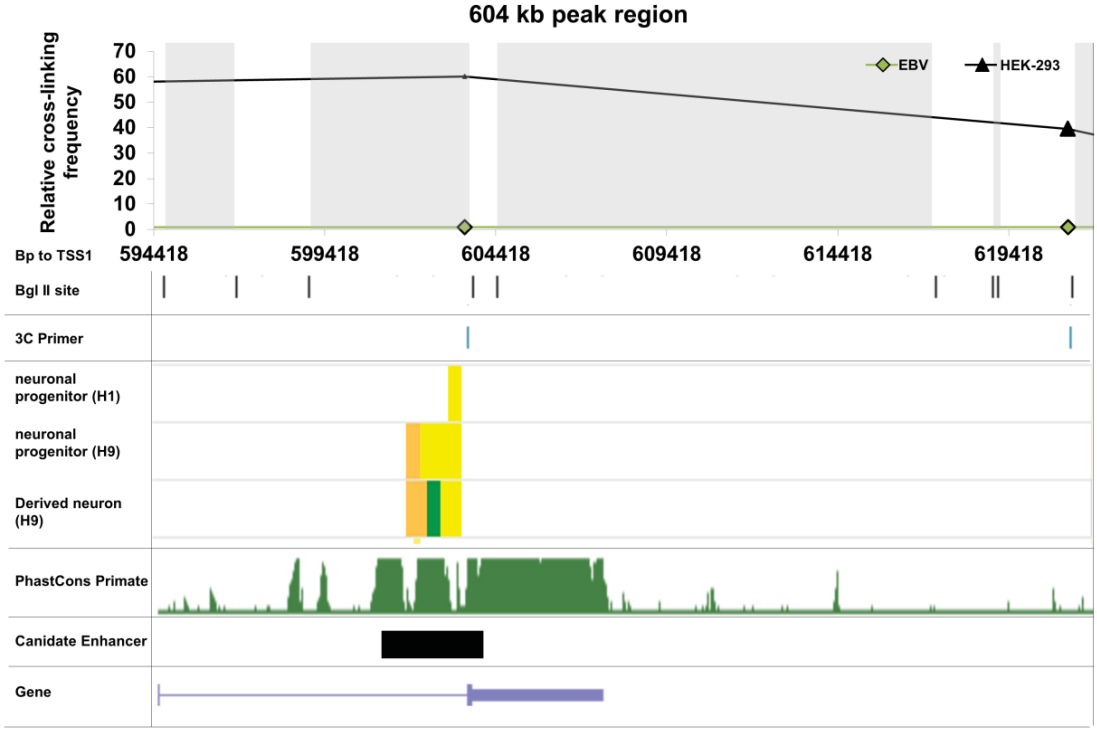

Figure 12: Detail of 604 kb peak region. See figure 8. The crosslinking frequencies of the representative HEK293 cell lines are shown. Genomic position (chr7:114320782-114348266 [hg19]). The structure of the *FOXP2* transcript is shown below, indicating the 3'-UTR at exon 17.

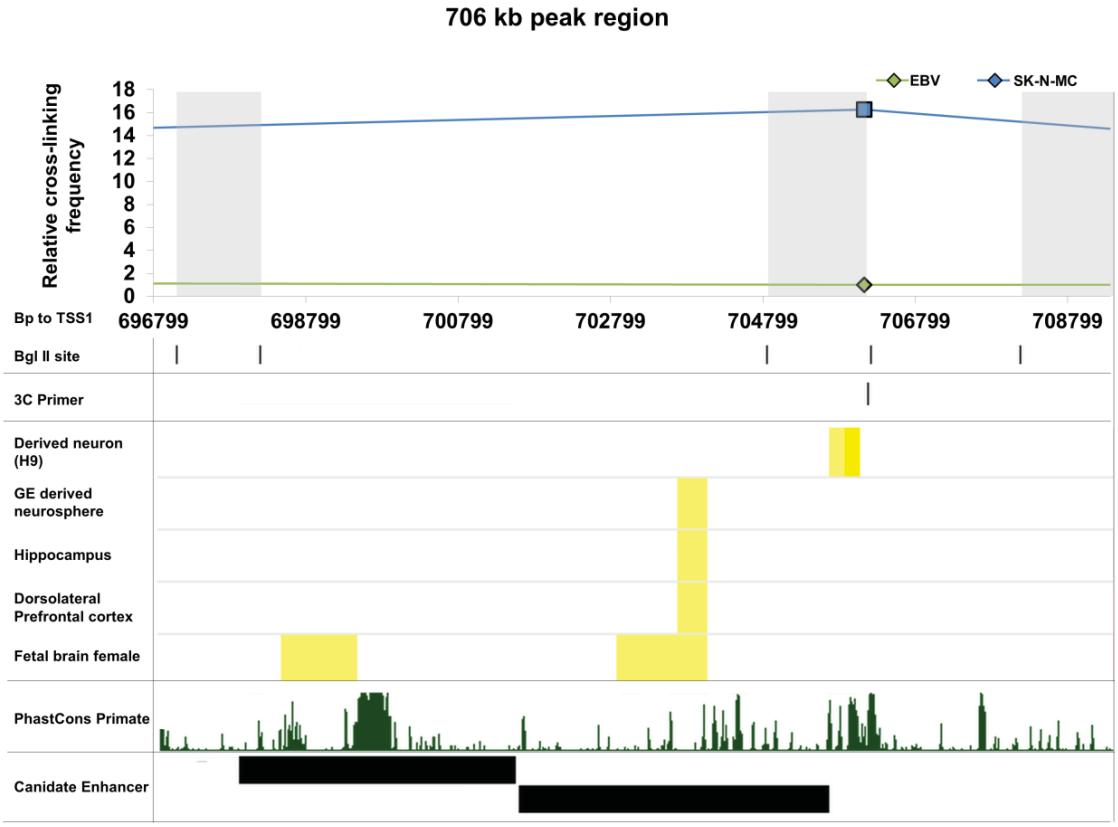

Figure 13: Detail of 706 kb peak region. See figure 8. Genomic position (chr7:114,423,163-114,435,725 [hg19]).

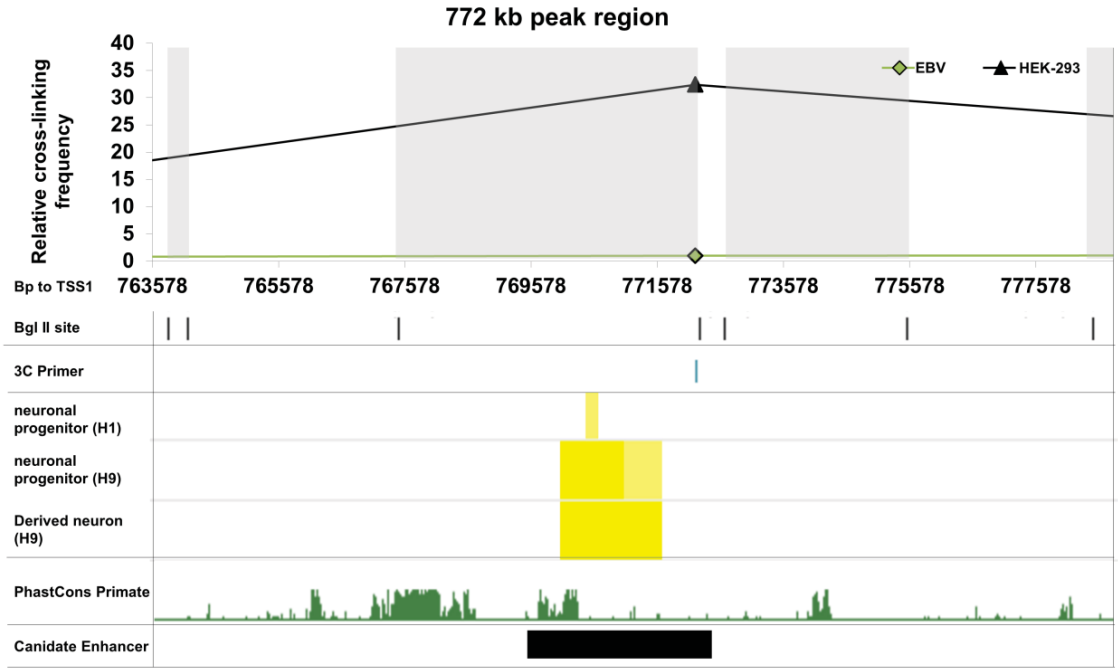

Figure 14: Detail of 772 kb peak region. See figure 8. The crosslinking frequencies of the representative HEK293 cell lines are shown. Genomic position (chr7:114,489,942-114,505,170 [hg19]).

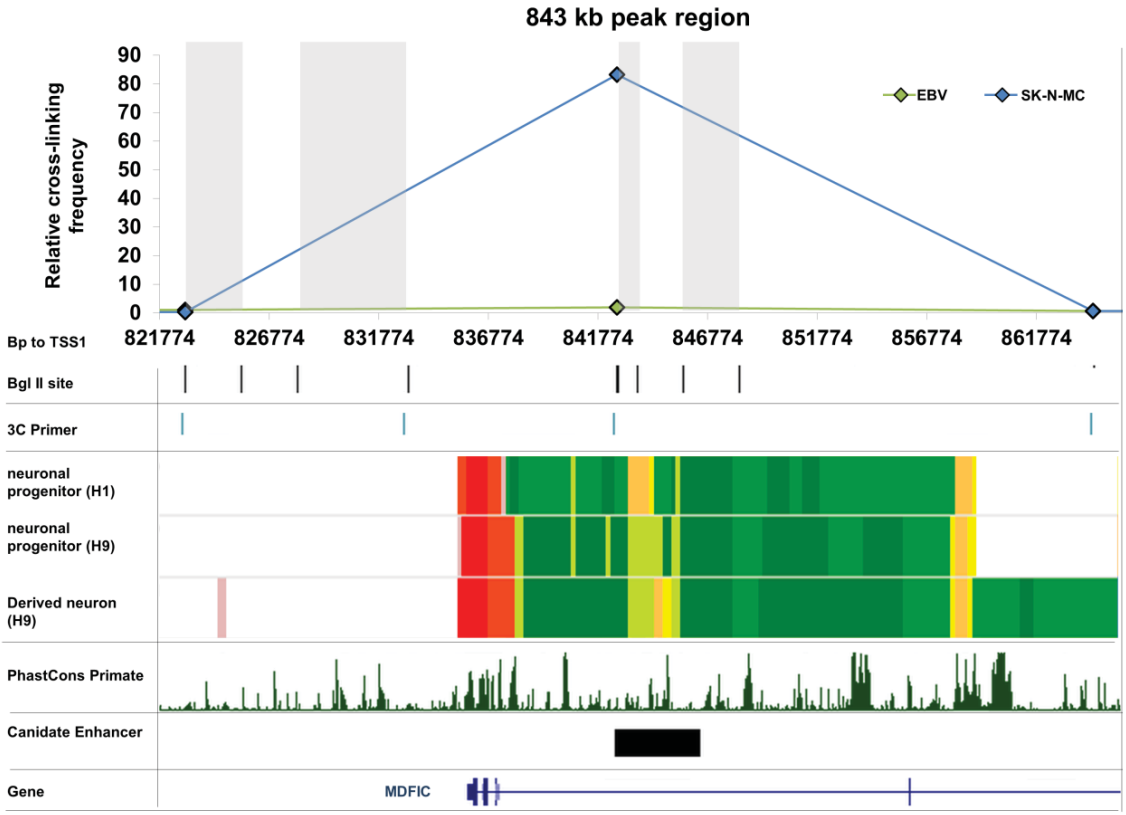

Figure 15: Detail of 843 kb peak region. See figure 8. Genomic position (chr7:114,548,138-114,592,061 [hg19]). The structure of the *MDFIC* transcript is shown below.

In summary, I detected eleven long-distance interactions to the *FOXP2* promoter 1, which is active in *FOXP2* expressing neuronal-like cell lines. Eight interactions were significantly stronger in *FOXP2* expressing cells, suggesting that the underlying mechanisms increase *FOXP2* expression. Finally, I identified nine candidate *FOXP2* enhancer elements on these genomic fragments, which show typical histone marks of enhancers in neuronal tissue. To validate the enhancer function, it is necessary to test elements in molecular assays, as described in chapter 4 of this thesis.

Discussion

Four transcriptional start sites have been identified for *FOXP2* and their activity has been determined by a set of human cell lines and tissue samples (Schroeder and Myers, 2008). Transcription from TSS1 and TSS4 was shown to be active in neuronal-like cell lines and primary neuronal tissue, whereas TSS2 and TSS3 were active in the colon, trachea and cervix (Schroeder and Myers, 2008). In agreement with this, I detected PolII occupancy at TSS1 and/or TSS4 in neuronal-like cell lines. PolII enrichment at TSS4 was detected in HEK293 cells, as well as in the FOXP2(-) KELLY cells. The PolII enrichment in FOXP2(-) cells could indicate a poised polymerase, which may become active in response to external stimuli (Gaertner et al., 2012). The published neuronal activity and potentially poised polymerase make TSS4 an interesting target for future 3C studies. The promoter of TSS1 was occupied exclusively in the investigated *FOXP2* expressing cell lines. Therefore, I expected clear differences for this promoter between FOXP2(+) and (-) cell lines and prioritized the identification of long-range interactions with this promoter.

I detected an interaction between the *FOXP2* promoter and a genomic fragment in the proximity of the *MDFIC* promoter. Promoter-promoter interactions have been identified between a substantial amount of co-regulated neighbouring genes in prior genome-wide studies (Li et al., 2012; Sahlen et al., 2015). Thus, the promoter-promoter interaction to the *MDFIC* gene suggests that this gene could be co-regulated with *FOXP2*. The *MDFIC* gene encodes the MyoD family inhibitor domain containing protein, which modulates the transcriptional regulation of the WNT and JNK/SAPK pathways (Kusano and Raab-Traub, 2002; Reiss-Sklan et al., 2009). Strong expression of this gene was described in T-cells (Thebault et al., 2000), but the expression of this gene has not been studied in other tissues. The Allen Brain Atlas indicates ubiquitous expression of *Mdfic* in adult mouse brains (Lein et al., 2007) and the gene likely overlaps *Foxp2* expression, which shows strong expression in the basal ganglia, thalamus, cerebellum and neocortex (Chapter 1). Detailed studies of *MDFIC/Mdfic* expression in adult and developing brain tissue would be required to identify

neuronal co-regulation with *FOXP2/Foxp2* in mouse or human. The co-regulation of genes can be investigated in published weighted gene co-expression network analysis (WGCNA) data. WGCNA data summarizes the co-regulation of genes across many samples of the same origin (Fuller et al., 2007). One WGCNA study indicated that *MDFIC* and *FOXP2* are co-regulated in the human and primate frontal lobe (Konopka et al., 2012), which may suggest a functional interaction in this tissue. *MDFIC* has established roles in WNT-signalling and prior research has shown that downstream target genes of *FOXP2* are enriched for function in the WNT-signalling pathway (Vernes et al., 2007; Vernes et al., 2011). In addition, *foxp2* in zebrafish is likely regulated by *lef1,* a TF that regulates target gene expression in response to activation of WNT-signalling. Thus, the co-regulation could further implicate *FOXP2* in this pathway.

In this chapter I described how the chromatin interactions were investigated in cell lines of neuronal origin. Enhancer interactions have been shown to be retained in cell lines derived from the corresponding tissue (Jager et al., 2015). This suggests that the identified interactions occur in neuronal tissue. The interactions occurred between genomic elements that show histone marks of active enhancers in one or several neuronal tissues (Roadmap Epigenomics et al., 2015), further indicating that the identified enhancers show neuronal activity. The target tissue and developmental period of the enhancer activity remain to be determined.

The Roadmap Epigenomics Consortium determined the histone marks in human derived cultured neuronal progenitors, derived neurons, neurospheres, whole foetal brains and adult neuronal tissue (Ernst and Kellis, 2015). I used these marks to determine regions of overlapping function in several neuronal tissues. The example of the region at -37 kb (Figure 8) shows that histone marks overlap at one narrow region at this locus. Follow-up experiments are important to validate the enhancers' function and to determine tissue-specificity, as discussed in subsequent chapters of this thesis. Transgenic mice are commonly used to study tissue-specificity during development (Visel et al., 2007; Boyd et al., 2015) and in chapter 5 I

used this approach to determine the developmental pattern and adult tissue-specificity of the enhancer elements at -37 kb and 330 kb.

The evolutionary conservation of the candidate enhancers varies from an element that is well conserved across vertebrates (Figure 11: 330 kb) to an element that shows low conservation even across primates (Figure 8: -37 kb). The presence of conservation at a genomic site can be regarded as the result of evolutionary constraint on sequence variety due to its functional impact (Ureta-Vidal et al., 2003; Miller et al., 2004). Hence, highly conserved enhancers likely drive target gene expression across related species. High conservation of *FOXP2* enhancers is in agreement with the conserved expression pattern of this gene across different vertebrate species (Enard et al., 2002; Ferland et al., 2003; Lai et al., 2003; Teramitsu et al., 2004; Li et al., 2007; Campbell et al., 2009). However, I also identified an evolutionary non-conserved enhancer in the human gene locus, which may suggest that humans gained regulatory control mechanisms. This may indicate that additional brain structures gained *FOXP2* expression or that the spatio-temporal dynamics of *FOXP2* regulation in evolutionarily shared brain structures has changed. The evolutionary conserved expression pattern would suggest that regulation is more likely to change within shared structures. To explore this possibility, the temporal expression changes of *FOXP2* need to be studied in more detail across species. Examples of dynamic adult expression changes have been found in the thalamus of mice (Horng et al., 2009) and striatum of songbirds (Teramitsu and White, 2006). The evolutionary conservation of these dynamic expression changes may indicate the presence of absence of species-dependent regulatory enhancers.

On a molecular level, enhancer elements consist of nucleotide sequences that are bound by TFs. Knowing the location and sequence of the potential enhancer elements, it is possible to investigate TFs that bind to these enhancers. I investigated TF binding in chapter 4 and discuss this aspect of the identified enhancer there.

The aim of the current chapter was to identify long-distance interactions, which may exert an enhancing effect on *FOXP2* expression. However, in some cases long-distance interactions have been reported to exert a repressive effect (Ogbourne and Antalis, 1998; Maston et al., 2006). Therefore, the 3C approach could potentially also detect repressing elements, called silencers. In comparison to enhancers, I would expect these elements to show strong interaction in FOXP2(-) cell lines. However, I did not detect chromatin interactions that were stronger in the absence of *FOXP2* expression. Silencers may overlap with enhancer sequences and the composition of TF binding determines the effect on transcription (Perissi et al., 2004). I can therefore not exclude that the identified elements may repress transcription under certain conditions. The interaction of the chromatin fragment at 604 kb was significantly weaker in the FOXP2(+) SK-N-MC cells compared to the control EBV-lymphoblast cells. As SK-N-MC cells express *FOXP2*, one explanation for the reduced interaction of this region could be that a silencing element was actively repelled from the active promoter. However, the same interaction was significantly increased in the FOXP2(+) HEK293 cells. Thus, the expression status does not distinguish the presence or depletion of this interaction. The fragment at 604 kb covers parts of the *FOXP2* 3'-UTR. At the RNA level, 3'-UTRs are known to mediate mRNA transcript stability and translational efficiency (He and Hannon, 2004). However, on the DNA level, 3'UTRs have also been found to establish long-distance interactions with their gene promoter. The effect of this interaction on gene expression can be repressive (Le Cam and Legraverend, 1995; Paul et al., 1998) or activating (Salerno et al., 2000; Jash et al., 2012). My 3C results show that the 3'-UTR of *FOXP2* interacts with the promoter both in the absence and presence of expression. To determine the function of this element on transcription, it will be necessary to study this element in isolation, using reporter gene assays.

3C is a low-throughput method that requires the manual design for every investigated interaction. Therefore, only a limited amount of genomic fragments at the *FOXP2* gene locus could be studied. As a consequence, long-range interactions to additional enhancers might

have been missed. Recent research has shown that the genome is partitioned into topological domains (TD) (Dixon et al., 2012). Dynamic enhancer contacts have been shown to occur within TDs and not across the topological borders (TB). TBs are reported to be evolutionarily conserved (Vietri Rudan et al., 2015) and stable during development (Dixon et al., 2012). Therefore, the TBs determined in one cell-type limit enhancer interactions in all or most differentiated tissues. All investigated genomic fragments and identified interactions were located within the TD that contains the *FOXP2* gene. However, the TD of *FOXP2* further covers the upstream *PPP1R3A* gene and the downstream genes *MDFIC* and *TFIC*. An extension of the 3C design, which covers the complete TD of *FOXP2,* may identify further enhancer elements.

Currently, a broad variety of 3C-based technologies are being developed to identify enhancer interactions with higher specificity and sensitivity on a genome-wide scale (Dostie et al., 2006; Dekker et al., 2013; Jin et al., 2013; Lefebvre et al., 2013; Nagano et al., 2013). For example, capture Hi-C combines high-throughput sequencing technologies with the sensitivity of targeted 3C approaches (Jager et al., 2015; Mifsud et al., 2015). Future experiments investigating *FOXP2* chromatin contacts could aim to identify additional or indeed all long-distance enhancer contacts within a tissue. Additionally, one 3C based method is capable of determining the genome-wide chromatin interaction landscape of single cells (Nagano et al., 2013). This circumvents the requirement to use cultured cell lines, which I used to obtain sufficient amount of chromatin. These approaches have so far focused on animal tissue, or non-neuronal human tissue (Dekker et al., 2013; Jin et al., 2013; Nagano et al., 2013; Jager et al., 2015; Mifsud et al., 2015; Vietri Rudan et al., 2015). Thus, future studies may identify all enhancers of *FOXP2* in a cell type-specific manner.

Enhancer elements have been shown to contain genetic variants that cause, or increase the susceptibility to, human disease. For example, rare genetics variants in regulatory regions have been linked to autism spectrum disorders (Poitras et al., 2010; Chen et al., 2014). The investigation of the here identified enhancer sequences in people with speech and language

disorders may reveal rare genetic variants underlying their deficits. In addition, common genetic variants within these enhancers may be linked to normal variation in brain structure and function (Chapter 6, (Becker et al.)). The catalogue of published genome-wide association studies (GWAS) of the National Human Genome Research Institute (NHGRI) collects genome-wide associations of common variants from published studies (Welter et al., 2014). According to this catalogue associations have been found near the *FOXP2* gene locus for traits such as smoking behaviour (Argos et al., 2014; Sung et al., 2015), N-glycosylation of immunoglobin (Lauc et al., 2013), Crohn's disease (Julia et al., 2013), lymphoblast cell viability (de With et al., 2015) and obesity (Kim et al., 2013). The associated single-nucleotide polymorphisms (SNPs), however, are not located within the candidate *FOXP2* enhancers or the tagged genomic fragments. Thus, so far, no trait-associated SNP has been detected within the enhancers. However, the identified enhancers are candidate regions for future association studies of common genetic variants in language-related traits.

This chapter represents the first study investigating long-distance enhancer interactions at the *FOXP2* gene locus. The identified interactions occurred in neuronal-like cell lines and may provide information about the regulatory mechanisms contributing to *FOXP2* expression in the human brain. The interactions occur between the active *FOXP2* promoter to the 3'-UTR of *FOXP2*, the promoter of *MDFIC* and candidate enhancer elements. The identification of candidate enhancers facilitates the study of *FOXP2* regulation and enables the identification of regulatory genetic variants underlying normal or pathogenic variation.

References

Argos M, Tong L, Pierce BL, Rakibuz-Zaman M, Ahmed A, Islam T, Rahman M, Paul-Brutus R, Rahaman R, Roy S, Jasmine F, Kibriya MG, Ahsan H (2014) Genome-wide association study of smoking behaviours among Bangladeshi adults. Journal of medical genetics 51:327-333.

Becker M, Devanna P, Fisher SE, Vernes SC (2015) A chromosomal rearrangement in a child with severe speech and language disorder separates FOXP2 from a functional enhancer. Molecular cytogenetics 8:69.

Bell O, Tiwari VK, Thoma NH, Schubeler D (2011) Determinants and dynamics of genome accessibility. Nat Rev Genet 12:554-564.

Berger SL (2007) The complex language of chromatin regulation during transcription. Nature 447:407-412.

Biedler JL, Helson L, Spengler BA (1973) Morphology and growth, tumorigenicity, and cytogenetics of human neuroblastoma cells in continuous culture. Cancer research 33:2643-2652.

Biedler JL, Roffler-Tarlov S, Schachner M, Freedman LS (1978) Multiple neurotransmitter synthesis by human neuroblastoma cell lines and clones. Cancer research 38:3751-3757.

Blanchette M, Bataille AR, Chen X, Poitras C, Laganiere J, Lefebvre C, Deblois G, Giguere V, Ferretti V, Bergeron D, Coulombe B, Robert F (2006) Genome-wide computational prediction of transcriptional regulatory modules reveals new insights into human gene expression. Genome Res 16:656-668.

Bonkowsky JL, Wang X, Fujimoto E, Lee JE, Chien CB, Dorsky RI (2008) Domain-specific regulation of foxP2 CNS expression by lef1. BMC developmental biology 8:103.

Bonn S, Zinzen RP, Girardot C, Gustafson EH, Perez-Gonzalez A, Delhomme N, Ghavi-Helm Y, Wilczynski B, Riddell A, Furlong EE (2012) Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat Genet 44:148-156.

Boyd JL, Skove SL, Rouanet JP, Pilaz LJ, Bepler T, Gordan R, Wray GA, Silver DL (2015) Human-Chimpanzee Differences in a FZD8 Enhancer Alter Cell-Cycle Dynamics in the Developing Neocortex. Current biology : CB.

Bruce HA, Margolis RL (2002) FOXP2: novel exons, splice variants, and CAG repeat length stability. Human genetics 111:136-144.

Cairns BR (2009) The logic of chromatin architecture and remodelling at promoters. Nature 461:193-198.

Campbell P, Reep RL, Stoll ML, Ophir AG, Phelps SM (2009) Conservation and diversity of Foxp2 expression in muroid rodents: functional implications. The Journal of comparative neurology 512:84-100.

Cao Y, Yao Z, Sarkar D, Lawrence M, Sanchez GJ, Parker MH, MacQuarrie KL, Davison J, Morgan MT, Ruzzo WL, Gentleman RC, Tapscott SJ (2010) Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. Dev Cell 18:662-674.

Carter CL, Lin C, Liu CY, Yang L, Liu ZR (2010) Phosphorylated p68 RNA helicase activates Snail1 transcription by promoting HDAC1 dissociation from the Snail1 promoter. Oncogene 29:5427-5436.

Chen CH, Huang CC, Cheng MC, Chiu YN, Tsai WC, Wu YY, Liu SK, Gau SS (2014) Genetic analysis of GABRB3 as a candidate gene of autism spectrum disorders. Molecular autism 5:36.

Core LJ, Waterfall JJ, Gilchrist DA, Fargo DC, Kwak H, Adelman K, Lis JT (2012) Defining the status of RNA polymerase at promoters. Cell Rep 2:1025-1035.

Coulon A, Chow CC, Singer RH, Larson DR (2013) Eukaryotic transcriptional dynamics: from single molecules to cell populations. Nat Rev Genet 14:572-584.

Coutinho P, Pavlou S, Bhatia S, Chalmers KJ, Kleinjan DA, van Heyningen V (2011) Discovery and assessment of conserved Pax6 target genes and enhancers. Genome Res 21:1349-1359.

de With SA, Pulit SL, Wang T, Staal WG, van Solinge WW, de Bakker PI, Ophoff RA (2015) Genome-wide association study of lymphoblast cell viability after clozapine exposure. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics 168B:116-122.

Dekker J, Marti-Renom MA, Mirny LA (2013) Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat Rev Genet 14:390-403.

Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. Science 295:1306-1311.

Deriziotis P, O'Roak BJ, Graham SA, Estruch SB, Dimitropoulou D, Bernier RA, Gerdts J, Shendure J, Eichler EE, Fisher SE (2014) De novo TBR1 mutations in sporadic autism disrupt protein functions. Nature communications 5:4954.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485:376-380.

Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, Rubio ED, Krumm A, Lamb J, Nusbaum C, Green RD, Dekker J (2006) Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res 16:1299-1309.

Dunn T, Praissman L, Hagag N, Viola MV (1994) ERG gene is translocated in an Ewing's sarcoma cell line. Cancer genetics and cytogenetics 76:19-22.

Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T, Monaco AP, Paabo S (2002) Molecular evolution of FOXP2, a gene involved in speech and language. Nature 418:869-872.

Ernst J, Kellis M (2015) Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. Nat Biotechnol 33:364-U374.

Ferland RJ, Cherry TJ, Preware PO, Morrisey EE, Walsh CA (2003) Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain. The Journal of comparative neurology 460:266-279.

Fuller TF, Ghazalpour A, Aten JE, Drake TA, Lusis AJ, Horvath S (2007) Weighted gene coexpression network analysis strategies applied to mouse weight. Mammalian genome : official journal of the International Mammalian Genome Society 18:463-472.

Fults D, Pedone CA, Morse HG, Rose JW, McKay RD (1992) Establishment and characterization of a human primitive neuroectodermal tumor cell line from the cerebral hemisphere. Journal of neuropathology and experimental neurology 51:272-280.

Gaertner B, Johnston J, Chen K, Wallaschek N, Paulson A, Garruss AS, Gaudenz K, De Kumar B, Krumlauf R, Zeitlinger J (2012) Poised RNA polymerase II changes over developmental time and prepares genes for future expression. Cell Rep 2:1670-1683.

Graham FL, Smiley J, Russell WC, Nairn R (1977) Characteristics of a human cell line transformed by DNA from human adenovirus type 5. The Journal of general virology 36:59-74.

Hardison RC, Taylor J (2012) Genomic approaches towards finding cis-regulatory modules in animals. Nat Rev Genet 13:469-483.

He L, Hannon GJ (2004) MicroRNAs: small RNAs with a big role in gene regulation. Nat Rev Genet 5:522-531.

Heintzman ND et al. (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 459:108-112.

Horng S, Kreiman G, Ellsworth C, Page D, Blank M, Millen K, Sur M (2009) Differential gene expression in the developing lateral geniculate nucleus and medial geniculate nucleus reveals novel roles for Zic4 and Foxp2 in visual and auditory pathway development. The Journal of neuroscience : the official journal of the Society for Neuroscience 29:13672-13683.

Jager R, Migliorini G, Henrion M, Kandaswamy R, Speedy HE, Heindl A, Whiffin N, Carnicer MJ, Broome L, Dryden N, Nagano T, Schoenfelder S, Enge M, Yuan Y, Taipale J, Fraser P, Fletcher O, Houlston RS (2015) Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. Nature communications 6:6178.

Jash A, Yun K, Sahoo A, So JS, Im SH (2012) Looping mediated interaction between the promoter and 3' UTR regulates type II collagen expression in chondrocytes. PloS one 7:e40828.

Jiang C, Pugh BF (2009) Nucleosome positioning and gene regulation: advances through genomics. Nat Rev Genet 10:161-172.

Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen CA, Schmitt AD, Espinoza CA, Ren B (2013) A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature 503:290-294.

Julia A et al. (2013) A genome-wide association study on a southern European population identifies a new Crohn's disease susceptibility locus at RBX1-EP300. Gut 62:1440-1445.

Kim HJ, Yoo YJ, Ju YS, Lee S, Cho SI, Sung J, Kim JI, Seo JS (2013) Combined linkage and association analyses identify a novel locus for obesity near PROX1 in Asians. Obesity 21:2405-2412.

Konopka G, Friedrich T, Davis-Turak J, Winden K, Oldham MC, Gao F, Chen L, Wang GZ, Luo R, Preuss TM, Geschwind DH (2012) Human-specific transcriptional networks in the brain. Neuron 75:601-617.

Kusano S, Raab-Traub N (2002) I-mfa domain proteins interact with Axin and affect its regulation of the Wnt and c-Jun N-terminal kinase signaling pathways. Molecular and cellular biology 22:6393-6405.

Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. Nature 413:519-523.

Lai CS, Gerrelli D, Monaco AP, Fisher SE, Copp AJ (2003) FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. Brain : a journal of neurology 126:2455-2462.

Lauc G et al. (2013) Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers. PLoS Genet 9:e1003225.

Le Cam A, Legraverend C (1995) Transcriptional repression, a novel function for 3' untranslated regions. European journal of biochemistry / FEBS 231:620-627.

Le Martelot G, Canella D, Symul L, Migliavacca E, Gilardi F, Liechti R, Martin O, Harshman K, Delorenzi M, Desvergne B, Herr W, Deplancke B, Schibler U, Rougemont J, Guex N, Hernandez N, Naef F, Cycli XC (2012) Genome-wide RNA polymerase II profiles and RNA accumulation reveal kinetics of transcription and associated epigenetic changes during diurnal cycles. PLoS biology 10:e1001442.

Lefebvre C, Vachon F, Grimault S, Thibault J, Guimond S, Peretz I, Zatorre RJ, Jolicoeur P (2013) Distinct electrophysiological indices of maintenance in auditory and visual short-term memory. Neuropsychologia 51:2939-2952.

Lein ES et al. (2007) Genome-wide atlas of gene expression in the adult mouse brain. Nature 445:168-176.

Li G, Wang J, Rossiter SJ, Jones G, Zhang S (2007) Accelerated FoxP2 evolution in echolocating bats. PloS one 2:e900.

Li G et al. (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell 148:84-98.

Maston GA, Evans SK, Green MR (2006) Transcriptional regulatory elements in the human genome. Annual review of genomics and human genetics 7:29-59.

Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, Wingett SW, Andrews S, Grey W, Ewels PA, Herman B, Happe S, Higgs A, LeProust E, Follows GA, Fraser P, Luscombe NM, Osborne CS (2015) Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet 47:598-606.

Miller W, Makova KD, Nekrutenko A, Hardison RC (2004) Comparative genomics. Annual review of genomics and human genetics 5:15-56.

Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, Laue ED, Tanay A, Fraser P (2013) Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature 502:59-64.

Naumova N, Smith EM, Zhan Y, Dekker J (2012) Analysis of long-range chromatin interactions using Chromosome Conformation Capture. Methods 58:192-203.

Navarro E, Serrano-Heras G, Castano MJ, Solera J (2015) Real-time PCR detection chemistry. Clinica chimica acta; international journal of clinical chemistry 439:231-250.

Ogbourne S, Antalis TM (1998) Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. The Biochemical journal 331 (Pt 1):1-14.

Ong CT, Corces VG (2011) Enhancer function: new insights into the regulation of tissue-specific gene expression. Nat Rev Genet 12:283-293.

Paschos K, Parker GA, Watanatanasup E, White RE, Allday MJ (2012) BIM promoter directly targeted by EBNA3C in polycomb-mediated repression by EBV. Nucleic Acids Res 40:7233-7246.

Paul C, Simar-Blanchet AE, Ro HS, Le Cam A (1998) Characterization of three transcriptional repressor sites within the 3' untranslated region of the rat serine protease inhibitor 2.3 gene. European journal of biochemistry / FEBS 254:538-546.

Perissi V, Aggarwal A, Glass CK, Rose DW, Rosenfeld MG (2004) A corepressor/coactivator exchange complex required for transcriptional activation by nuclear receptors and other regulated transcription factors. Cell 116:511-526.

Poitras L, Yu M, Lesage-Pelletier C, Macdonald RB, Gagne JP, Hatch G, Kelly I, Hamilton SP, Rubenstein JL, Poirier GG, Ekker M (2010) An SNP in an ultraconserved regulatory element affects Dlx5/Dlx6 regulation in the forebrain. Development 137:3089-3097.

Reiss-Sklan E, Levitzki A, Naveh-Many T (2009) The complex regulation of HIC (Human I-mfa domain containing protein) expression. PloS one 4:e6152.

Roadmap Epigenomics C et al. (2015) Integrative analysis of 111 reference human epigenomes. Nature 518:317-330.

Roh TY, Cuddapah S, Zhao K (2005) Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. Genes Dev 19:542-552.

Sahlen P, Abdullayev I, Ramskold D, Matskova L, Rilakovic N, Lotstedt B, Albert TJ, Lundeberg J, Sandberg R (2015) Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution. Genome Biol 16:156.

Salerno MS, Mordvinov VA, Sanderson CJ (2000) Binding of octamer factors to a novel 3'-positive regulatory element in the mouse interleukin-5 gene. J Biol Chem 275:4525-4531.

Schroeder DI, Myers RM (2008) Multiple transcription start sites for FOXP2 with varying cellular specificities. Gene 413:42-48.

Schwab M, Alitalo K, Klempnauer KH, Varmus HE, Bishop JM, Gilbert F, Brodeur G, Goldstein M, Trent J (1983) Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. Nature 305:245-248.

Seremetis S, Inghirami G, Ferrero D, Newcomb EW, Knowles DM, Dotto GP, Dalla-Favera R (1989) Transformation and plasmacytoid differentiation of EBV-infected human B lymphoblasts by ras oncogenes. Science 243:660-663.

Shaw G, Morse S, Ararat M, Graham FL (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 16:869-871.

Shlyueva D, Stampfel G, Stark A (2014) Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet 15:272-286.

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

Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, Ingle CE, Dunning M, Flicek P, Koller D, Montgomery S, Tavare S, Deloukas P, Dermitzakis ET (2007) Population genomics of human gene expression. Nat Genet 39:1217-1224.

Sung YJ, de Las Fuentes L, Schwander KL, Simino J, Rao DC (2015) Gene-smoking interactions identify several novel blood pressure loci in the Framingham Heart Study. American journal of hypertension 28:343-354.

Tarrant-Elorza M, Rossetto CC, Pari GS (2014) Maintenance and replication of the human cytomegalovirus genome during latency. Cell host & microbe 16:43-54.

Teramitsu I, White SA (2006) FoxP2 regulation during undirected singing in adult songbirds. The Journal of neuroscience : the official journal of the Society for Neuroscience 26:7390-7394.

Teramitsu I, Kudo LC, London SE, Geschwind DH, White SA (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 24:3152-3163.

Thebault S, Gachon F, Lemasson I, Devaux C, Mesnard JM (2000) Molecular cloning of a novel human I-mfa domain-containing protein that differently regulates human T-cell leukemia virus type I and HIV-1 expression. J Biol Chem 275:4848-4857.

Ureta-Vidal A, Ettwiller L, Birney E (2003) Comparative genomics: genome-wide analysis in metazoan eukaryotes. Nat Rev Genet 4:251-262.

Vernes SC, Spiteri E, Nicod J, Groszer M, Taylor JM, Davies KE, Geschwind DH, Fisher SE (2007) High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders. Am J Hum Genet 81:1232-1250.

Vernes SC, Nicod J, Elahi FM, Coventry JA, Kenny N, Coupe AM, Bird LE, Davies KE, Fisher SE (2006) Functional genetic analysis of mutations implicated in a human speech and language disorder. Human molecular genetics 15:3154-3167.

Vernes SC, Oliver PL, Spiteri E, Lockstone HE, Puliyadi R, Taylor JM, Ho J, Mombereau C, Brewer A, Lowy E, Nicod J, Groszer M, Baban D, Sahgal N, Cazier JB, Ragoussis J, Davies KE, Geschwind DH, Fisher SE (2011) Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. PLoS Genet 7:e1002145.

Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DT, Tanay A, Hadjur S (2015) Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell Rep 10:1297-1309.

Visel A, Rubin EM, Pennacchio LA (2009) Genomic views of distant-acting enhancers. Nature 461:199-205.

Visel A, Minovitsky S, Dubchak I, Pennacchio LA (2007) VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res 35:D88-92.

Wasserman WW, Sandelin A (2004) Applied bioinformatics for the identification of regulatory elements. Nat Rev Genet 5:276-287.

Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, Parkinson H (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 42:D1001-1006.

Chapter 4

Upstream regulatory mechanisms acting at the promoters and enhancers of *FOXP2*

Abstract

Enhancers and promoters of *FOXP2* control the expression of this gene during development and in adult tissue. Transcription factors (TF) bind to such enhancers and promoters and consequently regulate their activity. Thus, the TFs acting upstream of *FOXP2* determine the developmental and tissue-specific expression patterns of the gene. However, the TFs that regulate the promoters and enhancers of human *FOXP2* are largely unknown.

In this chapter I investigated the *in vitro* activity of *FOXP2* promoters and candidate enhancers in human cell lines using a reporter gene assay. One promoter and three enhancers increased reporter gene expression in human neuronal-like cell lines. I further tested the effect of TFs that had been linked to *FOXP2* in prior literature. Using overexpression plasmids in combination with the promoter and enhancer reporter constructs I detected regulation by *FOXP1*, *FOXP4*, *POU3F2*, *LEF1*, *PAX6*, *SOX5* and *TBR1,* as well as auto-regulation by *FOXP2*.

The presented results demonstrate that *FOXP2* is regulated by known TFs and suggest upstream regulatory pathways. Auto-regulation of *FOXP2* may explain the haploinsuficiency effect observed for mono-allelic gene mutations and gives new insight into possible disease mechanisms. One enhancer of interest (enhancer 330) is the target of TFs important for cortical development and may contribute to *FOXP2* expression during cortical development.

Introduction

Transcription factors (TFs) bind to gene promoters and enhancers to activate target gene expression. Promoters and enhancers of *FOXP2* have been identified (as described in Chapter 3), but the TFs that bind to them are unknown. Importantly, TFs are known to be effector proteins of signalling pathways and modulate target gene expression in response to developmental stimuli or environmental signals. Thus, the response of *FOXP2* expression to developmental programs and cellular processes is mediated by TFs. The aim of the current chapter was validate candidate enhancers identified in the previous chapter 3 and to identify TFs that regulate the enhancers and promoters of *FOXP2* to help reveal the genetic pathways acting upstream of this gene

Promoters and enhancers of FOXP2 are the target sites for transcription factor binding

Two promoters of *FOXP2*, promoter 1 and 4, are active in neuronal cell lines and neuronal tissue (Bruce and Margolis, 2002; Schroeder and Myers, 2008) (Chapter 3, Figure 1). In the prior work of this thesis I identified one enhancer downstream of an inversion breakpoint of a complex chromosomal rearrangement in a child with speech and language problems matching those seen in cases of *FOXP2* mutation (Moralli et al., 2015). As described in Chapter 2, the inversion separates the enhancer from the rest of the *FOXP2* locus, suggesting that disrupted regulation by this enhancer contributes to the observed phenotype in the affected child (Becker et al., 2015) (Chapter 2). In addition, in Chapter 3 I used chromatin confirmation capture to identify genomic regions, which interacted with promoter 1 in neuronal-like cell lines, and could thus be considered as candidates for novel enhancers (Chapter 3). Enhancers and promoters function independently from their genomic context. Their activity does not depend on their orientation, distance to the target promoter and are thus able to regulate the expression of reporter genes *in vitro* (Pennacchio et al., 2013). The activity of these elements can therefore be studied in cultured cell lines (Deriziotis et al., 2014).

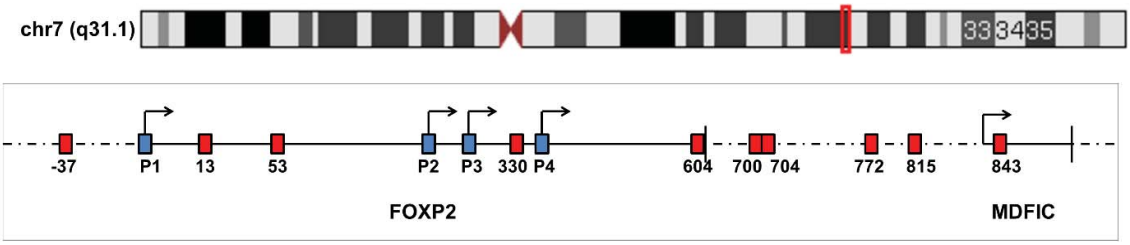

Figure 1: Schematic of *FOXP2* promoter and potential enhancer elements. Schematic representation of the *FOXP2* promoters and enhancers in the human genome on chromosome 7 band q31.1. The figure shows position 113,500,000 to 114,600,000 in reference human genome (build hg19). The candidate enhancers, identified in Chapter 2 and 3 of this thesis, are depicted as red blocks and the promoters are shown in blue. The *FOXP2* and *MDFIC* genes are shown in black.

Prioritizing putative upstream TFs of FOXP2

The study of enhancers in cell cultures further allows the investigation of their response to transcription factors (TFs). TFs are expressed or activated under specific conditions and fulfil varying functions. A group of TFs function in the general regulatory machinery of the cell, such as *EP300* (Chan and La Thangue, 2001) or *CTCF* (Ong and Corces, 2014). In contrast, some TFs are expressed at specific developmental periods and regulate, for example, the body patterning during embryonic development (Wellik, 2007). Other TFs are active in response to environmental cues (Sheng and Greenberg, 1990) or environmental stress (DiDomenico et al., 1982; Kothary et al., 1989) and regulate the rapid response to the changed conditions. Certain TFs are expressed in a cell-type specific manner to establish the identity of differentiated cells. The TFs that regulate *FOXP2* can therefore shed light on the developmental programs upstream of *FOXP2* expression. However, the prediction of TFs that regulate a gene is challenging, as the genome encodes about a thousand TFs, of which approximately 300 are expressed in each foetal and neuronal tissue (Fulton et al., 2009; Vaquerizas et al., 2009). The combination of TFs found in every cell is variable and determines the final regulatory output. In the same way, the promoter and enhancer activity depends on the combination of binding TFs. The co-expression of TFs with *FOXP2* is a limiting factor for positive regulatory interactions and has been described for a number of TFs (Introduction:

Table1). I selected a number of TFs that were linked to *FOXP2* in the literature to test their regulation of the promoters and enhancers that I had identified.

Regulatory TFs described in the literature

In zebrafish embryos *lef1* is co-expressed with *foxp2* in the tectum, mid-hindbrain boundary and hindbrain (Bonkowsky et al., 2008). Bonkowsky et al. further showed that knock down of *lef1* abolishes *foxp2* expression in the tectum and mid-hindbrain boundary, suggesting that *foxp2* is regulated by *lef1*. *LEF1* is the effector TF of the WNT-signalling pathway, which is involved in neurogenesis and regulates neural differentiation in the neocortex, cerebral cortex, hippocampus, olfactory bulb (OB) and thalamus (Oliva et al., 2013). In mouse embryos *Lef1* is expressed in the mesencephalon and diencephalon and involved in brain patterning (van Genderen et al., 1994). Strong adult expression of *Lef1* is detected in the thalamus, hypothalamus, hippocampus and OB (Shimogori et al., 2004; Oliva et al., 2013). If human *FOXP2* enhancers are regulated by *LEF1,* this could indicate the regulation of *FOXP2* by WNT-signalling pathways and associated biological processes. In addition, prior studies in human and mouse foetal tissue indicate that *FOXP2/Foxp2* regulates gene if the WNT-signalling pathway (Spiteri et al., 2007; Vernes et al., 2011). Thus, the expression of *FOXP2* may regulate the balance of WNT-signalling pathway.

Pax6 co-expression with *Foxp2* was described in the amygdala and striatum of non-human primates and rats (Kaoru et al., 2010). In zebrafish *pax6* is co-expressed with *foxp2* in the embryonic telencephalon and knock-down of *pax6* reduces levels of f*oxp2* in this brain region (Coutinho et al., 2011). In addition, Coutinho et al. detected reduced *Foxp2* expression in the telencephalon of mouse embryos carrying a *Pax6* null mutation, suggesting that this regulatory interaction is evolutionary preserved. *PAX6* has a dual role in maintaining neural stem cell identity and promoting neural differentiation (Osumi, 2001; Osumi et al., 2008) and is expressed in the developing and adult neocortex, amygdala, basal ganglia, thalamus, OB and

cerebellum (Duan et al., 2013). In humans, *PAX6* promotes the transition from embryonic stem cells to the neuronal lineage (Zhang et al., 2010).

Another TF has been proposed to directly regulate *FOXP2* expression in humans: *POU3F2*. The regulation of *FOXP2* by *POU3F2* was tested on a putative regulatory element of *FOXP2*. This element is located in an intron, within a bigger genomic region that carries a significant excess of genetic variants that arose in the human genome after the split from a common ancestor with Neanderthals (Maricic et al., 2013). The element did not show *in vitro* activity in human cell lines or *in vivo* activity in transgenic mouse embryos (Chapter 1). However, a human-specific genetic variant, which is conserved among vertebrates, was predicted to decrease the binding of *POU3F2.* In agreement with the prediction, the human variant decreased enhancer activity in response to *POU3F2* overexpression, in comparison to the Neanderthal variant. Thus, the study by Maricic et al. suggests that vertebrate *FOXP2/Foxp2,* including that of the Neanderthals, was regulated by *POU3F2* and this regulatory interaction is reduced in humans. *POU3F2* has been shown to regulate other TFs in retinoic acid-treated embryonic stem cells and was suggested to be an upstream effector of the retinoic acid signalling pathway (Urban et al., 2015). Similarly, *FOXP2* can promotes neural differentiation in combination with retinoic acid signalling in human cell lines (Devanna et al., 2014). Thus, it is conceivable that *FOXP2* could be one of the *POU3F2* target genes that promote retinoic acid induced neural differentiation.

Another candidate TF, which may regulate *FOXP2,* is *TBR1.* This TF has been shown to either activate or repress the expression of its target genes (Hevner et al., 2001; Bedogni et al., 2010; Han et al., 2011). *TBR1* expression overlaps with *FOXP2* expression in post-migratory neurons of the embryonic cortical plate in humans (Willsey et al., 2013) and mice (Hisaoka et al., 2010). In the adult mouse cortex *Tbr1* is co-expressed with *Foxp2* in cortical layer VI (Hevner et al., 2001; Bedogni et al., 2010; Hisaoka et al., 2010). *FOXP2* and *TBR1* physically interact with each other and may co-regulate target genes in a cooperative manner (Deriziotis et al., 2014). In the cortex of *Tbr1* knock-out mice the expression of *Foxp2* is significantly

decreased in the caudal area of layer VI (Bedogni et al., 2010). This could suggest that *TBR1* enhances *FOXP2* expression in the cortex. Another TF that is co-expressed with *TBR1* in the cortex is *SOX5* (Rouaux and Arlotta, 2010). During embryonic development *SOX5* is expressed in the ganglionic eminence and cortical plate and regulates cortical neuron subtype specification (Lai et al., 2008; Rouaux and Arlotta, 2010). Similar to *TBR1, SOX5* can both repress (Kwan et al., 2008) and activate target gene expression (Hao et al., 2014). *SOX5* and *TBR1* together determine the subtype-specification of cortico-fugal neurons of cortical layer VI (Lai et al., 2008; Greig et al., 2013). *TBR1* and *SOX5* are therefore promising candidates for contributing to the layer specific expression of *FOXP2*.

FOXP2 autoregulation and overlapping expression with FOXP family members

The persistent expression of *FOXP2* in adult tissue could suggest that *FOXP2* positively regulates its expression in a direct auto-regulatory loop (Takahashi et al., 2003; Alon, 2007). Direct positive autoregulation is a common mechanism for TFs that maintain cellular identity of differentiated cells (Crews and Pearson, 2009). Aberrant auto-regulation in a heterozygous mutant could contribute to haploinsufficiency effects (Cook et al., 1998; Bhatia et al., 2013), which is interesting in light of suggestions that *FOXP2*-associated language disorder may be a haploinsufficiency syndrome (Zeesman et al., 2006; Palka et al., 2012; Rice et al., 2012).

The FOXP2 protein forms hetero-dimers with FOXP1 and FOXP4 proteins (Li et al., 2004; Sin et al., 2015) and the combination of FOXPs may determine the directionality of the regulatory output (Sin et al., 2015). Co-expression of *FOXP1* and *FOXP2* is well described in the thalamus, hypothalamus and basal ganglia of mice, rats, songbirds and monkeys (Skidmore et al., 2008; Takahashi et al., 2008b; Kaoru et al., 2010; Mendoza et al., 2015). In the cortex, *FOXP1* shows strong expression in layer III to layer V, whereas *FOXP2* is most strongly expressed in cortical layer VI (Ferland et al., 2003). However, some neurons at the transition zone express both TFs in mice (Hisaoka et al., 2010). *FOXP4* expression is strong during development and overlaps with *FOXP2* in the ganglionic eminence, cortical plate and

thalamus of rats (Takahashi et al., 2008a). In the adult brain the co-expression between *FOXP2* and *FOXP4* was described in Purkinje cells, thalamus, the inferior olive and the striatum of rats and songbirds (Takahashi et al., 2008a; Mendoza et al., 2015). Thus, in addition to the hypothesis that *FOXP2* may autoregulate its own expression, I hypothesize that the co-expressed *FOXP* family members also influence *FOXP2* expression.

In silico prediction of TF motifs in enhancer and promoter sequences

TFs typically recognize motifs of 6 to 10 nucleotides via their DNA-binding domains (Shlyueva et al., 2014). The DNA-binding domain has the capacity to bind motifs of variable sequence compositions. Therefore, a TF binds to a range of short motifs of similar nucleotide sequence composition. The set of sequence motifs and the frequency of each nucleotide per position can be experimentally determined, often by genome-wide binding assays using chromatin immunoprecipitation based techniques (Heinz et al., 2010; Zambelli et al., 2013). The TF motifs can then be depicted as a consensus motif or position-weight matrix (PWM). The consensus motif contains the most frequent nucleotide per nucleotide position. The PWM provides a more informative representation of the redundant motif, because it includes the frequencies of the four DNA nucleotides A, T, G and C per nucleotide position (Stormo, 2000). In experimental studies, TF binding sites (TFBS) are commonly detected at hundreds to thousands of regions in the genome. The high number of binding sites enables computational algorithms to detect statistically significant motifs, which are represented as PWMs (Heinz et al., 2010). The motifs of hundreds of TFs have been characterized in molecular experiments and the PWMs are collected in databases, such as Transfac (Wingender et al., 2001), HOCOMOCO (Kulakovskiy et al., 2013), Jaspar (Mathelier et al., 2014) and HT-SELEX (Jolma et al., 2013).

The prediction of TFBS in the promoters and enhancers of *FOXP2* may validate the presence of TF motifs or identify novel TFBS. A number of computational methods aim to predict TF motifs in DNA sequences (Wasserman and Sandelin, 2004; Hardison and Taylor, 2012).

However, it is estimated that in a genome-wide scan 1,000 false positive predictions would occur for every 1 functional TF binding event (Wasserman and Sandelin, 2004) and the predictive power of these methods is dependent on additional information, such as epigenetic marks (Roh et al., 2005; Visel et al., 2009), evolutionary conservation (Woolfe et al., 2005; Stark et al., 2007; Visel et al., 2008) or prior molecular experiments (Grice et al., 2005; Hardison and Taylor, 2012). The enhancers that I identified at the *FOXP2* locus were narrowed down by chromatin conformation capture and histone marks, such that the real positive prediction rate is expected to be increased in these sequences. However, it remains crucial to determine the direct interaction of the TF and the sequences.

The aim of this chapter was to identify the upstream regulatory TFs and pathways that regulate *FOXP2*. To achieve this i) I characterized the basal *in vitro* activity of promoters and candidate enhancers, defined in prior chapters, ii) tested the regulation of selected TFs from the literature, and iii) and identified TF motifs in the enhancer and promoter sequences.

Materials and Methods

Molecular cloning of FOXP2 regulatory sequences

I designed primers to amplify the regulatory sequences of *FOXP2,* as indicated in chapter 3. Primer sequences are listed in appendix 2, table 1. The position and size of the cloned sequences are listed in table 1 and table 2. For a detailed description of the included sequences, conservation and functional annotations, see chapter 3. All primers were designed using the Primer 3 software (Untergasser et al., 2012) and the hg19 reference genome sequence. Promoter sequences were designed around the transcriptional start sites. Downstream of the transcription start site (TSS), I included the 5'untranslated region of the adjacent exon excluding the annotated start codon or splicing site. Upstream of the TSS, I included evolutionarily conserved stretches, as a proxy for promoter proximal regulatory sequences. .

The regulatory sequences were cloned from healthy human genomic DNA (Novagen) using Advantage 2 Polymerase kit (Clontech). The PCR program started with a 95˚C activation step and was followed by 10 cycles of denaturation (95˚C – 15 sec) and annealing/elongation (64˚C) and another 25 cycles of denaturation (95˚C – 15 sec) and annealing/elongation (68˚C). The annealing/elongation times were calculated and changed for each PCR product (1min extension per 1 kb). The PCR program included a final elongation step of 68˚C for 7 min. The PCR samples were run on a 1% agarose gel to identify the sizes of the products. PCR products that had the correct nucleotide size were excised from the gel and extracted using the Wizard SV Gel and PCR Clean-Up System (Promega). The DNA was eluted in 20 µl water and 1 µl of the product was subsequently used in the TOPO TA Cloning Kit (Invitrogen) to insert the PCR product into the pCR2.1-TOPO vector. After leaving the TOPO reaction for 1h at room temperature, the plasmids were transformed into subcloning-efficient DH5α competent cells (Invitrogen). The cells were incubated over night at 37˚C. Individual clones were picked the next day and grown in liquid cultures for at last 12 h at 37˚C. The plasmids were harvested

from the liquid cultures using the PureYield Plasmid MiniPrep System (Promega). At this stage all inserts were sequenced, using the Sanger method, to confirm the sequence identity and to select clones without amplification errors.

The primers were designed to add KpnI and XhoI restriction enzyme sites to the 5' and 3'ends of each PCR product, respectively. The regulatory sequence-containing pCR2.1-TOPO plasmids and the pGL4.23 destination vector (Promega) were double-digested with KpnI and XhoI high-fidelity digestion enzymes (New England Biolabs) for 1 h at 37˚C. The digestion reactions were run on a 1% agarose gel to identify the DNA fragments with the desired product sizes. The fragments were excised from the gel and extracted using the Wizard SV Gel and PCR Clean-Up System (Promega). The digested destination vector and genomic fragments were mixed for a ligation reaction using 25 ng of the destination plasmid and the insert fragment in a 1:3 or 1:6 ratio. For the ligation I used 0.5 µl of 5U/µl T4 ligase (Fermentas) and incubated the reaction over night at 16˚C. The next day 1 µl of ligation product was used to transform subcloning-efficient DH5α cells (Invitrogen). The cells were incubated on LB agar plates overnight at 37˚C and the next day single clones were picked and transferred into liquid LB medium. The liquid cultures were incubated for at least 12 h at 37˚C and subsequently harvested. Small liquid cultures (2ml) were harvested using the PureYield Plasmid MiniPrep System (Promega) and big cultures (125 ml) were harvested using the PureYield Plasmid MaxiPrep System (Promega). The eluates of each mini and maxi prep were sequenced to confirm the sequence identity of the pGL4.23-cis-regulatory-element plasmids.

Luciferase assay of enhancer and promoter elements

HEK293 or SK-N-MC cells were seeded in 96-well cell-culture plates. Cell-culture conditions and growth medium for both cell lines were as described in chapter 2. One day after seeding the cells in the 96-well plates were transfected using Genejuice transfection reagent (Merck Millipore). Each well was transfected with 48 ng of pGL4.23-CRE construct and 6 ng of pGL4-hRLuc-TK control plasmid (Promega), which expresses renilla luciferase under the control of

a herpes-simplex-virus thymidine kinase promoter element. After 2 days the cells were washed once with PBS and lysed in 20 µl Passive Lysis buffer (Promega) and shaken for 20 min at room temperature. Luciferase enzymatic activity reactions and fluorescence recordings were performed in a TECAN Infinite 2002 (TECAN) coupled with two software controlled injector pumps using the iControl (TECAN) software package. The program was set up to inject 50 µl of freshly prepared luciferase reaction reagent II (Promega), shake for 1 sec, wait for 1 sec and then measure firefly luciferase fluorescence for 10 sec. To measure the control renilla fluorescence the program was set up to inject 50 µl of freshly prepared Stop & Glo reagent (Promega), shake for 1 sec, wait for 1 sec and measure fluorescence for 10 sec. The automated filter settings of the software program were used to avoid detection overflow. The fluorescence of the firefly luciferase was divided by the fluorescence of the renilla luciferase to normalize the measurement and derive the relative luciferase activity. The promoter and enhancer activities were each detected in triplicate.

Co-transfection experiments of pGL4.23-CRE and transcription factor over-expression plasmids were done in HEK cells according to the above described protocol. Each well was transfected with 2 ng of pGL4.23-CRE construct, 2 ng of pGL4-hRLuc-TK control plasmid (Promega) and 10 ng of the individual TF plasmid. *SOX5* was cloned into pcDNA4. pcYFP-*POU3F2*, pcDNA4-*TBR1* and pYFP-*CASK* plasmids were kindly made available by Pelagia Deriziotis. pcDNA4-*PAX6* was kindly made available by Dario Gajewski. pcDNA4-*FOXP2*, pcDNA4-KE, pcDN4-*FOXP1* and pcDNA4-*FOXP4* were kindly made available by Sonja Vernes (Vernes et al., 2006). pBABE-*LEF1* was obtained from Addgene (Nguyen et al., 2009).

In silico motif scanning

I downloaded the PWMs from HOCOMOCO (Kulakovskiy et al., 2013), Transfac (Wingender et al., 2000; Wingender et al., 2001), Jaspar (Mathelier et al., 2014) and HT-SELEX (Jolma et al., 2013) and filtered those that were annotated to exist in humans. I retrieved the regulatory sequences from the UCSC genome browser and loaded them into the MotifLab program

environment (Klepper and Drablos, 2013). I scanned the regulatory sequences using MotifScanner (Coessens et al., 2003). This algorithm was developed to identify TFBS in single sequences. The algorithm calculates the probability that a TF motif is present above a sequence background model. I used nucleotide-triplet frequencies generated from all human promoter sequences as background model. Hence, I penalized motifs that bind frequently at unrelated regulatory regions and favoured motifs that occur infrequently. The motif scan was run, using a prior of 0.1 to allow for some degeneracy in the retrieved motifs. The prior can be set between 0 and 1, where 0 does not allow for any variation from the consensus motif. The predicted binding sites, including the motif short name, binding score and binding position within the regulatory elements were extracted for analysis.

Statistical analysis

Statistical analysis was performed in SPSS software. To determine the basal activities, the relative luciferase activities of the promoter/enhancer elements were compared to the luciferase activity of the empty (minP) construct within the same cell lines. Statistical significance of the basal enhancer and promoter activities was assessed using pairwise ANOVA and posthoc LSD test. The luciferase activity after TF overexpression was compared to luciferase activity after co-transfection with the empty pcDNA4 overexpression vector. The difference between empty plasmid and TF overexpression was assessed per individual element and the statistical significance was assessed using two-way ANOVA and post-hoc LSD test.

Results

Promoter 1 drives reporter gene expression in neuronal-like cell lines

To study the activity of *FOXP2* promoters, I cloned the promoter sequences into reporter gene plasmids directly upstream of the firefly luciferase reporter gene (Figure 2). The cloning strategy included promoter sequences up- and downstream of each TSS. Downstream of each TSS, I included the 5'-untranslated region excluding the following start codon or splice site. Upstream of each TSS, I cloned 2-4 kb to include regions of high evolutionarily conservation (primate PhastCons >0.9). I could not obtain promoter 4 as amplification was unsuccessful. The cloned elements are listed in table 1.

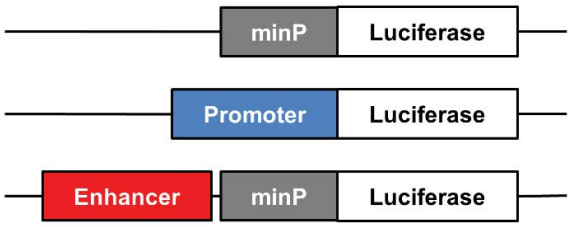

Figure 2: Cloning strategy for promoter and enhancer luciferase constructs. The promoter elements were cloned to replace the minimal promoter (minP) of the luciferase construct. The enhancer elements were cloned upstream of the minimal promoter.

Table 1: Cloned promoter sequences

Element	Size [bp]	Start (hg19)	End (hg19)
Promoter 1	1,791	113,724,817	113,726,609
Promoter 2	4,104	114,051,220	114,055,324
Promoter 3	1,006	114,055,454	114,056,459

The promoter-luciferase constructs were transfected into HEK293 and SK-N-MC cell lines. Both cell lines endogenously express *FOXP2* and show RNA-polymerase II binding at promoter 1, which is an indicator of transcriptional activity (Chapter 3). Therefore I expected promoter 1 to be active. Two days after transfection, I quantified the firefly luciferase expression using a dual-luciferase system. The relative luciferase expression was normalized to the expression from a minimal TATA-box containing promoter. Promoter 2 and 3 did not

show increased luciferase activity in HEK293 or SK-N-MC (Figure 3A), suggesting they are inactive. This is in line with previous PolII binding assays (Chapter3: Figure 4). Promoter 1 significantly increased luciferase expression in HEK293 (5-fold) and in SK-N-MC cells (80-fold). The stronger signal in SK-N-MC cells is in agreement with the earlier finding that PolII binding at promoter 1 was more enriched in SK-N-MC (Chapter 3: Figure 4). Thus, the luciferase results demonstrate that promoter 1 is active in both cell lines and may be more active in SK-N-MC.

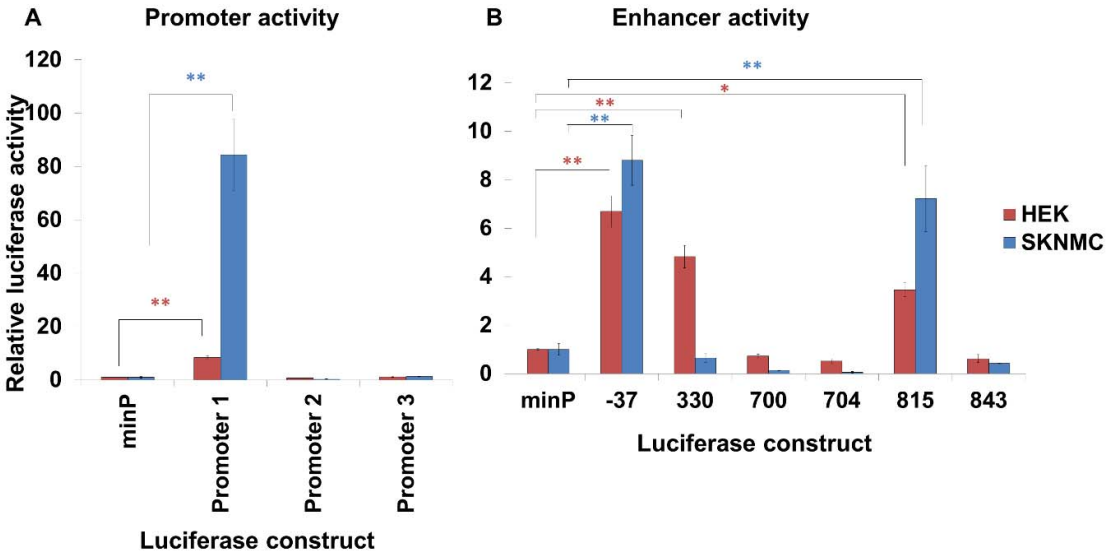

Figure 3: Basal promoter and enhancer activity in HEK293 and SK-N-MC. Relative luciferase activity of A) promoter and B) enhancer elements driving firefly luciferase gene expression in HEK293 and SK-N-MC cells. The promoter and enhancer constructs were each transfected with a control element expressing the renilla luciferase under the control of a viral promoter. The firefly luciferase signal was divided by the renilla signal to derive the relative luciferase activity. The activity for each luciferase construct was normalized for the activity observed for the minimal promoter (minP). The constructs were measured in two independent experiments in a total of six biological replicates, with the exception of elements 346, 700, 704 and 843 in SK-N-MC cells, which were measured in one experiment in three biological replicates. The significance of the difference between each construct and minP was determined with two-way ANOVA and post-hoc LSD testing. *p<0.05, ** p< 0.001. The significance in HEK293 is labeled with red asterisks; the significance in SK-N-MC is labeled with blue asterisks.

Enhancer elements increase luciferase reporter gene expression

To test the candidate enhancers of *FOXP2* in the dual-luciferase assay I cloned the candidate enhancer sequences upstream of the minimal promoter (Figure 2, Table 2). I cloned the candidate enhancers located 37, 330, 700, 704 and 843 kb distance from promoter 1 (Chapter 3). Due to time restrictions, the candidate enhancers at 13, 53, 604 and 772 kb were not included. Element 815 was defined in chapter 2 and there referred to as “enhancer 1”. The data for enhancer 815 were collected in the experiments for chapter 2 and are also included here to enable direct comparison of enhancer activities. The plasmids were transfected into HEK293 and SK-N-MC cells and reporter gene expression was measured. In HEK293 cells the enhancers at 37, 330 and 815 kb significantly increased the reporter gene expression, compared to the minimal promoter (Figure 3B). The other elements did not show significant effects. In the SK-N-MC cell line I detected increased reporter gene expression for the elements at 37 and 815 kb. The luciferase results demonstrate that elements at 37, 330 and 815 kb are functional enhancers that drive gene expression *in vitro*. The elements 700, 704 and 843 show no enhancer activity in HEK293 and SK-N-MC. The activities of elements at 13, 53, 604 and 772 kb remain to be determined.

Table 2: Cloned enhancer sequences

Element[a]	Size [bp]	Start (hg19)	End (hg19)
Enhancer 37	774	113,688,009	113,688,782
Enhancer 330	1,801	114,056,845	114,058,646
Enhancer 700	3,645	114,424,203	114,427,847
Enhancer 704	4,087	114,427,887	114,431,974
Enhancer 815	832	114,541,370	114,542,201
Enhancer 843	3,958	114,568,454	114,572,411

[a]The Name of the enhancer element indicates the distance to P1 in kb

FOXP2 overexpression indicates autoregulation at promoters and enhancers

To test the autoregulation of *FOXP2*, I co-transfected the luciferase constructs with a plasmid overexpressing the *FOXP2* gene. As a control, the luciferase plasmids were co-transfected with an empty pcDNA4 plasmid. Co-transfection experiments were conducted in HEK293 cells because the transfection efficiency rate in SK-N-MC was below 20%, which did not guarantee co-transfection of luciferase and expression plasmids in the same cells. FOXP2 overexpression increased the activity of promoter1, element 37 and 815 (Figure 4). Also, the previously inactive promoter 2 and element 700 show significant increases of luciferase expression. To determine if the increased enhancer activity is mediated by the direct binding of FOXP2 to DNA, I tested two aetiological variants of *FOXP2* that have been found in families with monogenic forms of speech and language disorder (Lai et al., 2001; MacDermot et al., 2005). The R553H variant of FOXP2 was found in the large multigenerational KE family and has an amino-acid substitution in the DNA-binding domain (Lai et al., 2001). The R553H variant is unable to bind to the consensus binding motif (Vernes et al., 2006). The other variant is R328X, which was found in a smaller family and lacks the DNA-binding FOX domain due to a premature stop codon (MacDermot et al., 2005) and cannot bind to DNA (Vernes et al., 2006). As expected, the two variants fail to increase the activity of promoter 1, promoter 2 and enhancers 37, 700, 815. The effects at promoter 1 and enhancer 37 are severely reduced for the R328X variant, but remain significant in comparison to control. The reduced promoter and enhancer activities suggest that the enhancing effects observed for wild-type *FOXP2* were mediated by direct DNA-binding. In agreement with this finding, a *FOXP2* binding site within promoter 1 has been detected by chromatin immunoprecipitation in human neuronal-like cell lines SK-N-MC and PFSK-1 (Nelson et al., 2013). The aetiological variants increased the activity of enhancer 330. The increase could be mediated by protein-protein interactions to other DNA-binding factors, which in turn regulate the enhancer. In sum the luciferase results suggest that *FOXP2* is capable to auto-regulate via direct DNA-binding to its regulatory elements.

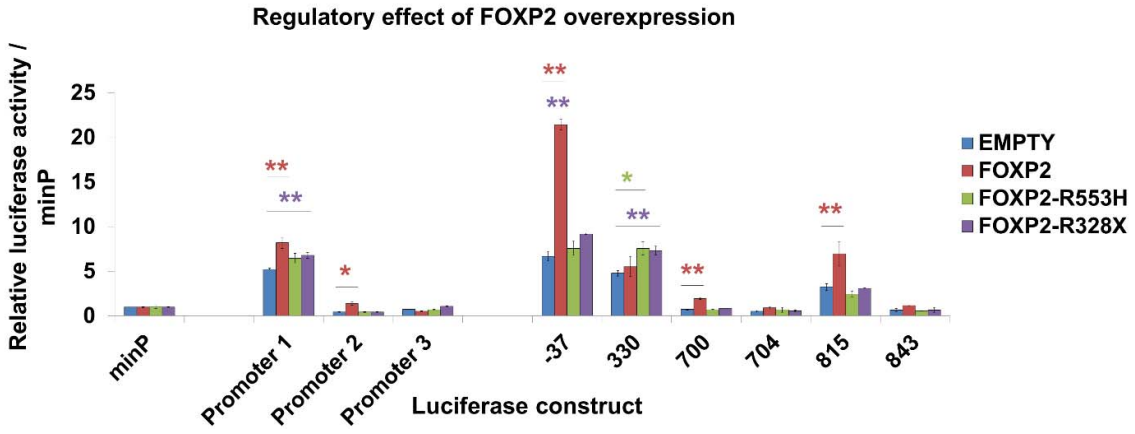

Figure 4: Assessing auto-regulation of FOXP2 via promoter and enhancer constructs. Relative luciferase activity of promoter and enhancer elements after co-transfection with FOXP2, FOXP2-R553H or FOXP2-R328X in HEK293. The promoter and enhancer constructs were co-transfected with a control construct expressing the renilla luciferase. The cells were additionally co-transfected with an empty pcDNA4 control, or FOXP2 overexpression plasmid. The firefly luciferase signal was divided by the renilla signal to derive the relative luciferase activity. The activity for each luciferase construct was normalized for the activity observed for the minimal promoter (minP). Each combination was tested in three biological replicates. The statistical significant of the TF overexpression effect was determined with two-way ANOVA and post-hoc LSD testing for each construct. P-value: *<0.05, ** < 0.001.

The FOXP family members regulate FOXP2 enhancers and promoters

FOXP2 can form hetero-dimers with the family members FOXP1 and FOXP4 to regulate target gene expression (Li et al., 2004; Sin et al., 2015). Because FOXP2 regulates its enhancers and promoters, it is conceivable that FOXP1 and FOXP4 have similar effects at the same elements. After co-transfection of the luciferase constructs with a pcDNA4-FOXP1 overexpression plasmid, I detected increased activity for promoter 1, enhancer 37 and element 815 (Figure 5), which were also up-regulated by *FOXP2* (Figure 4). Overexpression of FOXP4 increased the activity of enhancer 37 and decreased enhancer 330 (Figure 5). Thus, the three FOXP family members increase enhancer 37. Promoter 1 and enhancer 815 are up-regulated by both *FOXP1* and *FOXP2.* The activity of enhancer 330, which was increased by the aetiological *FOXP2* variants, was repressed by *FOXP4.* The results suggest that the FOXPs show regulatory effects at the same set of enhancers and the promoter 1.

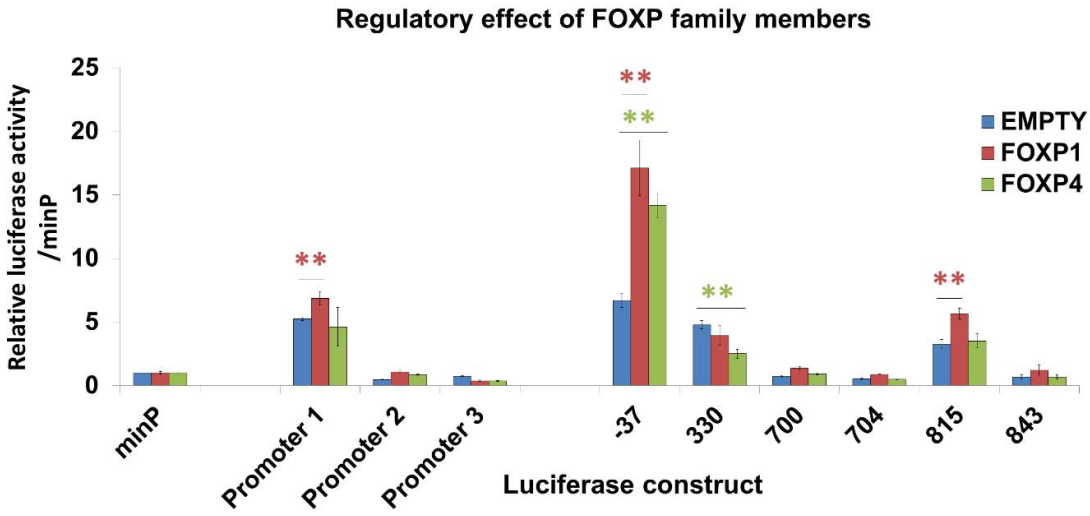

Figure 5: Regulatory effect of FOXP1 and FOXP4 on promoter and enhancer constructs. Relative luciferase activity of promoter and enhancer elements in HEK293 cells. The promoter and enhancer constructs were each transfected with a control construct expressing the renilla luciferase under the control of a viral promoter. The cells were also transfected with a control, FOXP1 or FOXP4 overexpression plasmid. The firefly luciferase signal was divided for the renilla signal to derive the relative luciferase activity. The activity for each luciferase construct was normalized for the activity observed for the minimal promoter (minP). Each combination was tested in three biological replicates. The statistical significant of the TF overexpression effect was determined with two-way ANOVA and post-hoc LSD testing for each construct. P-value: *<0.05, ** < 0.001.

The transcription factors LEF1 and PAX6, but not POU3F2, regulate human FOXP2 promoter and enhancers in vitro

I tested the promoter and enhancer constructs in combination with an overexpression vector of *LEF1*. The WNT-signalling pathway is active under normal culture conditions in HEK293 cells (Oloumi et al., 2006). Therefore, I expected LEF1 to be active in the absence of further stimulation of the WNT-pathway. I co-transfected the regulatory elements with a pBABE-LEF1 overexpression plasmid and determined the luciferase activity. I observed significantly increased activity of promoter1, promoter 3, enhancer 37 and enhancer 330 (Figure 6). Thus, *LEF1* activates regulatory elements of *FOXP2.* Since LEF1 is the downstream effector of the WNT-signalling pathway, these findings suggest that the promoter and enhancer activities might be controlled by WNT-signalling.

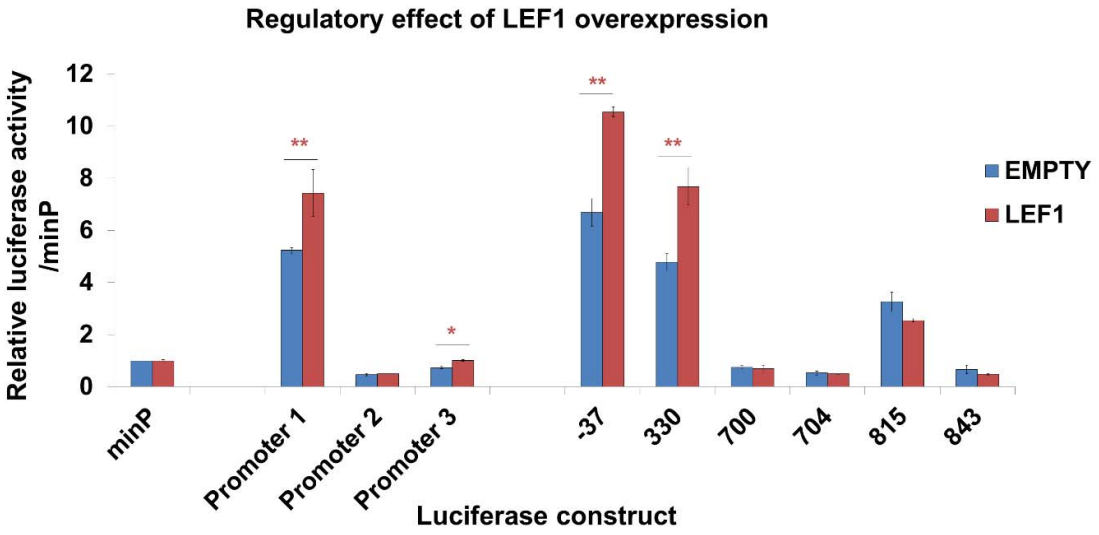

Figure 6: Regulatory effect of the WNT-signalling TF LEF1 on promoter and enhancer constructs. Relative luciferase activity of promoter and enhancer elements in HEK293 cells. The promoter and enhancer constructs were each transfected with a control construct expressing the renilla luciferase under the control of a viral promoter. The cells were also transfected with a pcDNA4, or LEF1 overexpression plasmid. The firefly luciferase signal was divided by the renilla signal to derive the relative luciferase activity. The activity for each luciferase construct was normalized for the activity observed for the minimal promoter (minP). Each combination was tested in three biological replicates. The statistical significance of the TF overexpression effect was determined with two-way ANOVA and post-hoc LSD testing for each construct. P-value: *p<0.05, **p< 0.001.

Next, I tested the overexpression of PAX6, which reduced the activity of promoter 1 and element 815 (Figure 7A). The luciferase activity remained stable for the other regulatory elements. The luciferase results of promoter 1 and element 815 indicate a repressive effect of *PAX6* on *FOXP2* expression. This is inconsistent with published results from mouse and zebrafish embryos, where *Pax6/pax6* increased *Foxp2/foxp2* expression (Coutinho et al., 2011). Despite the difference in direction of effect, my luciferase findings do support the hypothesis that *FOXP2* is under control of *PAX6* in humans.

I went on to investigate a prior suggested link between *POU3F2* and *FOXP2* by testing the regulatory elements that I had identified, in combination with *POU3F2* overexpression. I co-transfected the luciferase constructs with pYFP-POU3F2 and an empty pcDNA4 plasmid as

control. The *in vitro* activity of all promoters and enhancers remained constant (Figure 7B), indicating that *POU3F2* does not regulate the tested *FOXP2* enhancers and promoters.

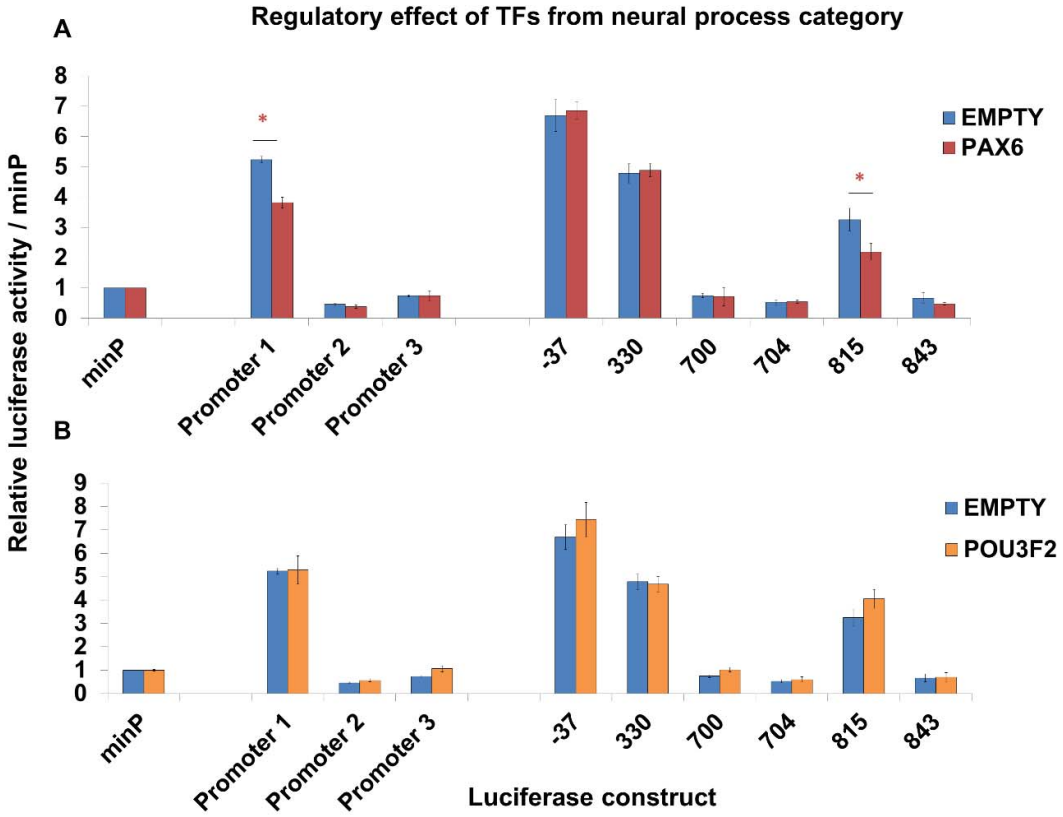

Figure 7: Testing regulatory effects of PAX6 and POU3F2 on promoter and enhancer constructs. Relative luciferase activity of promoter and enhancer elements in HEK293 cells for experiments with A) PAX6, and B) POU3F2. The promoter and enhancer constructs were each transfected with a control construct expressing the renilla luciferase under the control of a viral promoter and the relevant TF overexpression construct or a pcDNA4 control plasmid. The firefly luciferase signal was divided by the renilla signal to derive the relative luciferase activity. The activity for each luciferase construct was normalized for the activity observed for the minimal promoter (minP). Each combination was tested in three biological replicates. The statistical significance for each construct was assessed using two-way ANOVA and post-hoc LSD testing. P-value: * $p<0.05$, ** $p< 0.001$.

SOX5 and combinations of TBR1 with CASK regulate FOXP2 promoter and enhancer activity

Next, I tested the regulation of the *FOXP2* regulatory elements by *TBR1*. Co-transfection with TBR1 reduced the activity of enhancer 330 and 37. *TBR1* has been shown to exert both

activating and repressive effects on target gene expression (Bedogni et al., 2010; Han et al., 2011), which may dependent on the co-activator *CASK* (Hsueh et al., 2000; Wang et al., 2004). I therefore co-transfected the promoter and enhancer constructs with TBR1, CASK or a combination of both. The combination increased the activity of the enhancer 37 and promoter 1 (Figure 7). Overexpression of CASK alone had a repressive effect on enhancer 37. As predicted, the activating effect of TBR1 was detectable at promoter 1 and enhancer 37 only in combination with the co-activator CASK. As CASK does not bind DNA, the repressive effect, observed in the absence of TBR1, must be mediated via protein-protein interactions to another DNA-binding factor.

Finally, I tested the regulatory elements in combination with a pcDNA4-*SOX5* overexpression plasmid. Promoter 1, elements 37 and 704 increased luciferase gene expression in response to SOX5 overexpression (Figure 5C). The activity of enhancer 330 however decreased. Thus, *SOX5* exerts repressive and activating effects. Repressive and activating effects of SOX5 have been reported before (Kwan et al., 2008; Aza-Carmona et al., 2011). This suggests that SOX5 may enhance or decrease *FOXP2* expression levels, depending on the regulatory element. To my best knowledge, opposing regulatory effects of single TFs at regulatory elements of the same target gene have not been reported.

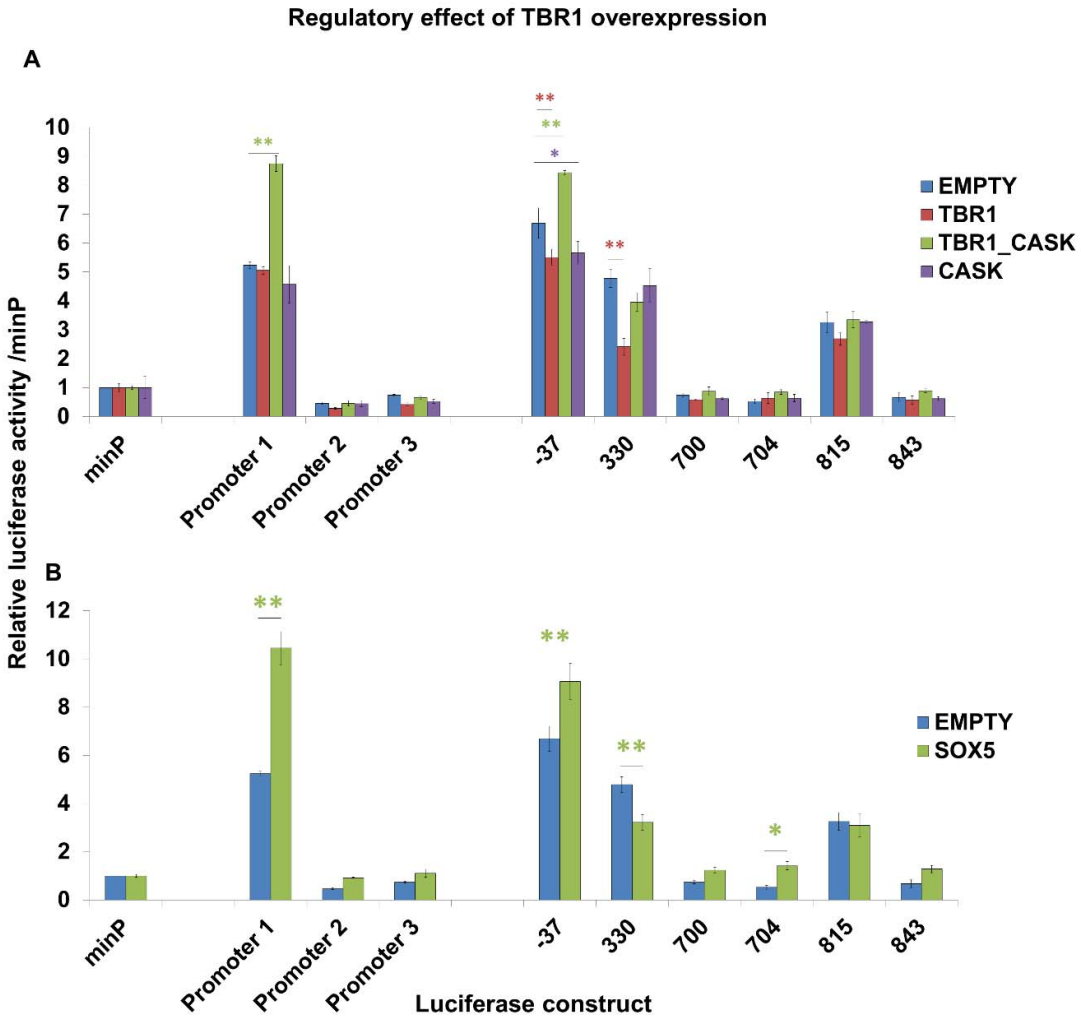

Figure 7: Regulatory effect of TBR1 and SOX5 on promoter and enhancer constructs. A, B) Relative luciferase activity of promoter and enhancer elements in HEK293 cells. The promoter and enhancer constructs were each transfected with a control construct expressing the renilla luciferase under the control of a viral promoter. The cells were also transfected with a control, A) TBR1, CASK or a combination of TBR1 and CASK overexpression plasmids and B) a SOX5 overexpression plasmid. The firefly luciferase signal was divided by the renilla signal to derive the relative luciferase activity. The activity for each luciferase construct was normalized for the activity observed for the minimal promoter (minP). Each combination was tested in three biological replicates. The statistical significance of the TF overexpression effect was determined with two-way ANOVA and post-hoc LSD testing for each construct. P-value: *p<0.05, **p< 0.001.

The motif scan locates motifs of tested TFs and predicts neuronal TF motifs

To identify TF binding motifs I scanned the cloned sequences for the presence of published PWMs. This allowed me to locate the binding motifs of the tested TFs, as well as to predict the binding of untested TFs. I obtained 2,161 PWMs from the four databases Transfac

(Wingender et al., 2000), HOCOMOCO (Kulakovskiy et al., 2013), HT-Selex (Jolma et al., 2013) and Jaspar (Mathelier et al., 2014). These PWMs were used within the MotifScanner algorithm to scan the cloned sequences (Aerts et al., 2003; Coessens et al., 2003). The databases included the PWMs for all investigated TFs (FOXP1, FOXP2, LEF1, PAX6, POU3F2, TBR1 and SOX5), except FOXP4. Therefore, I could not identify potential FOXP4 binding sites in the enhancer and promoter sequences.

The motif scan predicted between 373 (Enhancer 815) to 855 (Enhancer 330) motifs in the 9 analysed sequences (3 promoters, 6 enhancers) (Table 3, Appendix 2: Table S2). The predictions located putative motifs of the previously tested TFs (Table 4). *FOXP2* motifs were predicted within seven cloned elements (promoter 1-3, enhancer 330, enhancer 700, enhancer 815 and enhancer 843). However, in response to *FOXP2* overexpression, I only detected changes in reporter gene expression at five regulatory elements (Promoter 1, promoter 2, enhancer 37, enhancer 700 and enhancer 815). Therefore, the results of the motif prediction and luciferase are concordant at four elements (Promoter1, promoter 2, enhancer 700 and enhancer 815). The *FOXP2* motifs at promoter 3, enhancer 330 and enhancer 843 represent false positive predictions, possibly caused by non-functional binding sites. In addition, no *FOXP2* motif was predicted for enhancer 37. Explanations for the false negative prediction may be a too stringent detection threshold in the motif scan or that the observed regulation is due to an indirect interaction. In total I identified 19 TF motifs at concordant promoters and enhancers (Table 5). To verify that changes in luciferase expression were mediated by direct TF - DNA interactions it would be necessary to mutate or delete the predicted TF motifs.

Table3: Summary of predicted TFs

Element	Length [bp]	# of predicted motifs	motifs/bp
Promoter 1	1,791	444	0.25
Promoter 2	4,104	741	0.18
Promoter 3	1,006	428	0.43
Enhancer 37	774	384	0.50
Enhancer 330	1,801	855	0.47
Enhancer 700	4,087	449	0.11
Enhancer 704	832	520	0.63
Enhancer 815	3,645	373	0.10
Enhancer 843	3,958	493	0.12
Total	21,998	4,687	0.21

Table 4: Predicted TF motifs per promoter/enhancer

Elements	FOXP2		FOXP2-R328X		FOXP2-R553H		FOXP1		FOXP4		LEF1		PAX6		POU3F2		TBR1		TBR1/CASK		CASK		SOX5	
	observed effect[a]	# of predicted motifs	observed effect[a]	# of predicted motifs	observed effect[a]	# of predicted motifs	observed effect	# of predicted motifs	observed effect	# of predicted motifs	observed effect	# of predicted motifs	observed effect	# of predicted motifs	observed effect	# of predicted motifs	observed effect	# of predicted motifs	observed effect	# of predicted motifs	observed effect	# of predicted motifs	observed effect	# of predicted motifs
P1	+	1	+	NA		NA	+	0		NA	+	0	-	0		0		0	+	NA		NA	+	1
P2	+	2		NA		NA		2		NA		4		1		3		0		NA		NA		2
P3		1		NA		NA		1		NA	+	3		1		2		1		NA		NA		0
E37	+	0	+	NA		NA	+	0	+	NA	+	1		0		1	-	0	+	NA	-	NA	+	2
E330		1	+	NA	+	NA		1	-	NA	+	2		3		4	-	0		NA		NA	-	1
E700	+	1		NA		NA		0		NA		0		0		4		1		NA		NA		2
E704		0		NA		NA		0		NA		5		0		1		0		NA		NA	+	3
E815	+	2		NA		NA	+	1		NA		0	-	0		2		0		NA		NA		1
E843		1		NA		NA		0		NA		4		1		2		0		NA		NA		0

green = concordant luciferase and prediction; yellow = false negative prediction; red = false positive prediction

[a] + = increase of luciferase expression; - = decrease of luciferase expression

Table 5: Location of concordant TF motifs

Element	TF	Start [hg18]	End [hg18]	PWM ID	Motif
Promoter 1	FOXP2	113,513,165	113,513,175	MA0593	GTTGTTTACAT
Promoter 1	SOX5	113,513,220	113,513,227	HM00343	GAACAATA
Promoter 2	FOXP2	113,841,749	113,841,759	MA0593	ATTGTTTACTT
Promoter 2	FOXP2	113,840,403	113,840,413	MA0593	ATGTAAACAGA
Promoter 3	LEF1	113,843,051	113,843,056	M00805	GTTTGA
Promoter 3	LEF1	113,843,147	113,843,161	HT00135	AAACATGAAAGGAGT
Promoter 3	LEF1	113,843,177	113,843,183	HM00191	CTTTGTA
Enhancer 37	LEF1	113,475,432	113,475,438	HM00191	CTTTGAT
Enhancer 37	SOX5	113,475,604	113,475,611	HM00343	CATTGTTG
Enhancer 300	LEF1	113,844,283	113,844,289	HM00191	CTTTGTA
Enhancer 300	LEF1	113,844,761	113,844,770	M01022	CACTTTGAAA
Enhancer 300	LEF1	113,844,980	113,844,994	HT00135	AGTGCTTTCATCTTT
Enhancer 300	SOX5	113,844,505	113,844,512	HM00343	CATTGTTC
Enhancer 704	SOX5	114,215,241	114,215,250	M00042	TTAACAATAT
Enhancer 704	SOX5	114,216,589	114,216,598	M00042	CAATTGTTAG
Enhancer 704	SOX5	113,513,220	113,513,227	HM00343	GAACAATA
Enhancer 815	FOXP2	114,328,727	114,328,737	MA0593	CAGTAAACACA
Enhancer 815	FOXP2	114,328,816	114,328,824	HM00111	ATGTTTACT
Enhancer 815	FOXP1	114,328,724	114,328,738	MA0481	TATCAGTAAACACAC

A promising approach to select novel TFs that may regulate *FOXP2* is to prioritize TFs that are expressed in the brain. Therefore, I filtered the list of predicted TFs using the DAVID bioinformatics web tool to filter for TFs within the brain expressed category (UniProt_Tissue: brain) (Huang da et al., 2009) and detected 145 TFs within the brain expressed category (Appendix 2: Table S3). As a complementary alternative, I prioritized TFs that are collected in the gene ontology category of neuron differentiation (Appendix 2: Table S4) because *FOXP2* function in neuron differentiation is well established (Vernes et al., 2006; Vernes et al., 2011; Tsui et al., 2013; Devanna et al., 2014). The “neuron differentiation” list contained 86 TFs, of which 54 overlap with the “brain expressed” category. The absence of some “neuron differentiation” genes from the “brain expressed” list may be caused by incorrect manual annotation of the GO-terms. An alternative explanation may be the absence of TFs, which are expressed during brain development in the “brain expressed” list. Both lists include several TFs that I studied experimentally in this chapter, such as *SOX5*, *POU3F2*, and *PAX6*. In addition the neuron differentiation list includes TFs which have been shown to be co-expressed with *FOXP2*, such as *ATOH1*, *DLX5*, *EMX1*, *LMX1B* (as described in Chapter 1).

The motif scan could be used to prioritize additional TFs for further luciferase experiments to investigate the regulation of *FOXP2* by these TFs.

Discussion

FOXP2 enhancer and promoters are active in neuronal-like cell lines

Promoter 1 is transcriptionally active in HEK293 and SK-N-MC cells, which is in agreement with the report that endogenous *FOXP2* expression is driven by this promoter in these cell lines (Schroeder and Myers, 2008). HEK293 and SK-N-MC cells express comparable levels of *FOXP2* mRNA, yet the promoter activity in HEK293 cells was substantially lower than in SK-N-MC. Similarly, I previously detected lower RNA polymerase II binding at this promoter in the HEK293 cell line (Chapter 3). These findings suggest that regulation of the promoter does not solely determine endogenous *FOXP2* expression in HEK293 and that enhancers increase expression levels. In accordance with this hypothesis, three enhancers were active in HEK293 cells. One of the enhancers, enhancer 330, was inactive in SK-N-MC, which may contribute to the similarity in expression levels of *FOXP2* in HEK293 and SK-N-MC. However, additional regulatory elements, such as promoter 4 and four untested candidate enhancers, may further contribute to the total endogenous *FOXP2* expression (Chapter 3).

Promoter 2 and 3 did not drive luciferase expression in SK-N-MC or HEK293 cells, which is in accordance with the previous observation that these promoters are not bound by RNA polymerase II in these cells (Chapter 3). In addition, this is in agreement with a previous study, which did not detect mRNA transcripts originating from promoter 2 or 3 in these cell lines (Schroeder and Myers, 2008). However, Schroeder and Myers detected tissue-specific expression in non-neuronal tissue, such as the cervix, liver and trachea. Luciferase experiments in cell lines derived from these tissues may reveal promoter 2 and/or promoter 3 activities. Because I wanted to gain insight into the neuronal function of *FOXP2,* its regulation in non-neuronal tissue was outside the scope of this thesis.

I demonstrated that the enhancers at 37 and 330 kb increase reporter gene expression *in vitro*. The results support the hypothesis that the measured chromatin interactions represent functional promoter-enhancer loops (Chapter3). The basal activity of the elements was tested

in the same FOXP2(+) cell lines which had been used to study the chromatin interactions. The activity of enhancer 330 was detected in HEK293 and absent in SK-N-MC. However, the chromatin interactions were comparable between both cell lines (see Chapter 3: Figure 5). The chromatin interaction is therefore not proportional to the *in vitro* activity of the enhancer element. The absence of enhancer 330 activity in SK-N-MC cells suggests that the enhancer may be "poised" in this cell line. Poised enhancers regulate rapid changes in target gene expression in response to developmental or environmental stimuli (Zentner et al., 2011; Jin et al., 2013; Shlyueva et al., 2014). As I observed repressing effects on enhancer 330 by *TBR1*, *SOX5* and *FOXP4*, the inactivity of this enhancer in SK-N-MC could be caused by the presence of these TFs. The knock-down of these TFs in SK-N-MC cells might potentially activate enhancer 330. An alternative explanation could be the absence of an activating TF. A promising candidate could be *LEF1,* which increased enhancer 330 activity in HEK293 cells. As *LEF1* activity is dependent on WNT-signalling, this hypothesis could be tested by activation of this pathway in SK-N-MC cells.

Positive direct auto-regulation of FOXP2 likely contributes to neuronal expression

I hypothesized that FOXP2 could auto-regulate via its promoters or enhancers, and detected increased activity of five regulatory elements in response to *FOXP2* overexpression. Positive auto-regulatory loops enable rapid amplification of the TF and to maintain expression at stable plateau levels (Bateman, 1998). Auto-regulation has been described for developmental (Bateman, 1998; di Gennaro et al., 2013; Mead et al., 2013) and neurodevelopmental TFs (Meredith et al., 2009). Heterozygous mutations in positive auto-regulatory TFs could decrease the expression of the wildtype protein from the healthy allele. Consequently, the expression of wildtype protein would be severely reduced and below 50% compared to the non-mutated situation. Thus, heterogeneous mutations of positive auto-regulatory TFs may lead to haploinsufficiency, where the absence of one healthy allele is sufficient to cause a disease phenotype (Lamb et al., 2012; di Gennaro et al., 2013; Mead et al., 2013). Mutations of *FOXP2* or the loss of one *FOXP2* allele would be predicted to result in reduced activation

of *FOXP2* expression and subsequently reduce the expression of the healthy allele. Indeed, heterogeneous whole-gene deletions of *FOXP2* have been detected in people with speech impairments, strongly indicating that haploinsufficiency underlies the molecular aetiology of the *FOXP2*-related speech phenotype in some cases (Zeesman et al., 2006; Palka et al., 2012; Rice et al., 2012). In addition, studies in mouse models of aetiological *FOXP2* mutations indicate that heterozygous mice are overtly normal with specific effects on cortico-striatal neural networks (Groszer et al., 2008; Gaub et al., 2016). Positive auto-regulation could explain the occurrence of the haploinsufficiency effect in a subset of *FOXP2*-positive tissues, as this effect would be present only in target tissue of the enhancers.

I have shown that both FOXP1 and FOXP4 can regulate the enhancers and promoters of *FOXP2 in vitro*, suggesting that they may also enhance *FOXP2* expression *in vivo*. The FOXP family members have been shown to form functional hetero-dimers (Li et al., 2004; Sin et al., 2015; Spaeth et al., 2015) and hetero-dimerization could influence the regulation of *FOXP2*. HEK293 cells do express *FOXP1*, *FOXP2* and *FOXP4* (Gierman et al., 2007; Uhlen et al., 2015). It is thus possible that endogenous expression of these TFs influenced the luciferase results and that the observed effects are caused by FOXP dimers. Co-transfection of *FOXP* family members may further investigate the role of hetero- and homodimers in the regulation of *FOXP2* enhancers.

Regulatory TFs of promoters/enhancers suggest upstream pathways for FOXP2 expression

I detected that *LEF1* is capable of increasing the activity of promoter 1, promoter 3 and two enhancer elements. In prior work *LEF1* has been shown to activate expression of target genes (Arce et al., 2006), which is consistent with the observed effects in my experiments. Moreover, the increase of enhancer and promoter activity after *LEF1* overexpression is consistent with previously described effects of *Lef1* in zebrafish, where knockdown of *lef1* abolishes *foxp2* expression in the tectum, mid-hindbrain boundary and hindbrain (Bonkowsky et al., 2008). Thus, the luciferase results indicate that *LEF1* regulates *FOXP2* also in human cells. It remains

to be determined if regulation by *LEF1* occurs in the same regions as observed for the zebrafish. The activities of promoter 1, promoter 3, enhancer 37 and enhancer 330 during development and adulthood should be characterized to address this question. As the protein structures of TFs are evolutionary more stable than enhancer sequences (Domene et al., 2013; Rubinstein and de Souza, 2013) the human enhancers are likely recognized by orthologous TFs in related species. Thus, the human enhancer activities could be tested in transgenic mice (Visel et al., 2008), which I did for the most promising enhancers in chapter 5.

I found that in human cells PAX6 overexpression decreased promoter 1 and enhancer 815 activity. A previous report showed in zebrafish that *pax6* knockdown reduced *foxp2* expression in the developing telencephalon, suggesting that it activates *foxp2* expression (Coutinho et al., 2011). The conflicting results between my data and the study by Coutinho et al. could be caused by inter-species differences. Also, my data was obtained within neuronal-like cell lines, which are unlikely to resemble the transcriptional program of embryonic telencephalon. However, my luciferase data indicates that FOXP2 is a target gene of *PAX6* in humans. Co-expression of both TFs was shown in the amygdala and striatum of non-human primates and rats (Kaoru et al., 2010) and the embryonic (E12) cortical plate of mice (Tsui et al., 2013). In the developing cortex of *Pax6* knock-out mice, cortico-fugal axons show aberrant migratory pathways (Jones et al., 2002). *FOXP2* is expressed specifically in cortico-fugal neurons of cortical layer VI (Hisaoka et al., 2010) and regulates neurite outgrowth (Vernes et al., 2011). Thus, aberrant *Foxp2* regulation in *Pax6* knock-out mice may contribute to the defect in axon pathfinding.

In a previous report, *POU3F2* has been suggested to regulate the expression of *FOXP2* via a putative enhancer element in intron 8 (Maricic et al., 2013). Maricic et al. suggested that the regulation by *POU3F2* was reduced in modern humans, due to a genetic variant that arose in humans after the split from a common ancestor with Neanderthals. However, the element proposed by Maricic et al. did not show independent *in vitro* or *in vivo* activity in their study

and was not included here. In my experiments, *POU3F2* did not increase or decrease the activity of the promoters and enhancers that I tested. The presented results cannot exclude the possibility that *POU3F2* is capable of regulating *FOXP2* expression via untested regulatory elements or indirect interactions.

TF important for cortical neuron subtype specification may contribute to FOXP2 expression

I identified regulatory effects on *FOXP2* promoters and enhancers by *SOX5,* which is expressed in post-migratory neurons of the developing cortex and is involved in subtype specification (Kwan et al., 2008; Lai et al., 2008). Similar to SOX5, TBR1 is involved in the development and maturation of cortical neurons (Hevner et al., 2001; Han et al., 2011). *FOXP2* is involved in embryonic cortical development (Tsui et al., 2013) and restricted to cortical layer VI (Lai et al., 2003). Thus, in the adult cortex expression of *TBR1* and *SOX5* overlaps with *FOXP2* expression in cortical layer VI. I showed that TBR1, in absence of the co-activator CASK, represses the activity of enhancer 330, while in presence of CASK it promotes activity of enhancer 37 and promoter 1. CASK is expressed throughout the cortex with the highest expression in cortical layers III and IV (Kristiansen et al., 2010). SOX5 and TBR1 in combination with CASK increase the activity of enhancer 37. This could indicate that enhancer 37 is active in cortical layer VI, where *FOXP2* is co-expressed with these TFs. However, enhancer 330 activity is decreased by TBR1 and SOX5. It is therefore unlikely that both enhancer 37 and 330 are active in the same neurons. The enhancer activity pattern and an assessment of the overlap with *FOXP2* in the brain are thus important to fully evaluate the effects observed in my experiments. As there are no methods to study enhancer activity in human brains it is required to test the enhancers in transgenic animals (Visel et al., 2007).

Mutations in components of the upstream regulatory network reported in disease phenotypes

Some TFs that regulate the *FOXP2* enhancers and promoters may cause disease phenotypes if mutated. Rare *de novo* mutations of *TBR1* have been implicated in the aetiology of autism spectrum disorder (ASD) (O'Roak et al., 2012b; O'Roak et al., 2012a). Patients with ASD have impaired communication skills and language disruption is a common phenotype shared between patients with *TBR1* and *FOXP2* mutations (Deriziotis et al., 2014). Haploinsufficiency of SOX5 causes intellectual disability with prominent speech delay (Lamb et al., 2012; Schanze et al., 2013). Mutations of *PAX6* and chromosomal rearrangements at the *PAX6* gene have been shown to cause aniridia, a disorder that is characterized by the absence of the iris in affected cases (Prosser and van Heyningen, 1998). In rare cases mutations of *PAX6* have been found in patients with intellectual disability (ID) and autism (Ticho et al., 2006; Graziano et al., 2007; Davis et al., 2008). The possibility of an underlying molecular link between language phenotypes observed in for mutations in *FOXP2, PAX6, TBR1* or *SOX5* need to be addressed in future studies.

In this chapter, I included enhancer 815, which I have identified in chapter 2 downstream of a *de novo* inversion breakpoint found in a child with a complex rearrangement and language deficits (Becker et al., 2015; Moralli et al., 2015). The proximity to FOXP2 suggested that the de-regulation of this gene might contribute to the observed phenotype in this child. In the present chapter, I determined that this element is responsive to the overexpression of *PAX6*, *FOXP1* and *FOXP2*. *FOXP1* mutations have been found in cases of autism spectrum disorders (Bacon et al., 2015), which are partly characterized by communication problems. The language problems of people with *FOXP1* and *PAX6* mutations could be secondary to major problems in unrelated molecular pathways. *PAX6* mutations commonly cause developmental problems, which are unrelated to cognition and behaviour (Prosser and van Heyningen, 1998). It will be important to identify the patterns of tissue activity of element 815 to determine whether and how these overlap with sites of expression of *FOXP1* and *PAX6*. The brain regions that are affected in the child described by Moralli et al. have not been reported. The overlap of affected brain regions and the expression of *FOXP1* or *PAX6*, as well

as the activity of element 815 could provide an explanation for the molecular aetiology of the observed phenotype.

HEK293 cells have been shown to express *FOXP1*, *FOXP4*, *SOX5*, *CASK*, *LEF1* and *PAX6* (Gierman et al., 2007; Uhlen et al., 2015). In an alternative experiment the endogenous TF levels could be knocked down, using RNA silencing techniques, to assess the effects on promoters and enhancers. Ideally, follow-up experiments should employ cells obtained from tissue and developmental periods in which the regulatory effect is expected to occur. It will be useful to switch to more advanced cellular systems, such as primary neurons from mouse embryos or induced pluripotent stem (iPS) cells. In developmental protocols of iPS cells, *FOXP2* is used as a marker for development stages (Molina-Hernandez et al., 2013; Belinsky et al., 2014). Therefore the longitudinal expression pattern of *FOXP2* has been described for this system (Belinsky et al., 2014). This makes iPS cells an ideal system to manipulate the expression of potential regulatory TFs and to analyse the effect on *FOXP2* expression during their development.

In this chapter I demonstrated the *in vitro* activity of previously identified enhancer elements and one promoter of *FOXP2*. The enhancer and promoter activities respond to the overexpression of several TFs that had previously been linked to *FOXP2* in the literature. However, changes of endogenous *FOXP2* mRNA in response to TFs still need to be determined. *In silico* motif scan partially predicted the DNA binding motifs in the responding regulatory sequences and suggested additional TFs that may regulate the enhancers/promoters. The activity patterns of the enhancer and promoter elements need to be characterized to address the developmental period and target tissue of the observed regulatory effects, which I describe in the following chapter. Overall, the results of the current chapter indicated the influence of known developmental programs and signalling pathways on the expression of *FOXP2* and open new directions to investigate the biological role of this TF in normal and pathogenic development.

References

Aerts S, Thijs G, Coessens B, Staes M, Moreau Y, De Moor B (2003) Toucan: deciphering the cis-regulatory logic of coregulated genes. Nucleic Acids Res 31:1753-1764.

Alon U (2007) Network motifs: theory and experimental approaches. Nat Rev Genet 8:450-461.

Arce L, Yokoyama NN, Waterman ML (2006) Diversity of LEF/TCF action in development and disease. Oncogene 25:7492-7504.

Aza-Carmona M, Shears DJ, Yuste-Checa P, Barca-Tierno V, Hisado-Oliva A, Belinchon A, Benito-Sanz S, Rodriguez JI, Argente J, Campos-Barros A, Scambler PJ, Heath KE (2011) SHOX interacts with the chondrogenic transcription factors SOX5 and SOX6 to activate the aggrecan enhancer. Human molecular genetics 20:1547-1559.

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

Bateman E (1998) Autoregulation of eukaryotic transcription factors. Progress in nucleic acid research and molecular biology 60:133-168.

Becker M, Devanna P, Fisher SE, Vernes SC (2015) A chromosomal rearrangement in a child with severe speech and language disorder separates FOXP2 from a functional enhancer. Molecular cytogenetics 8:69.

Bedogni F, Hodge RD, Elsen GE, Nelson BR, Daza RA, Beyer RP, Bammler TK, Rubenstein JL, Hevner RF (2010) Tbr1 regulates regional and laminar identity of postmitotic neurons in developing neocortex. Proc Natl Acad Sci U S A 107:13129-13134.

Belinsky GS, Rich MT, Sirois CL, Short SM, Pedrosa E, Lachman HM, Antic SD (2014) Patch-clamp recordings and calcium imaging followed by single-cell PCR reveal the developmental profile of 13 genes in iPSC-derived human neurons. Stem cell research 12:101-118.

Bhatia S, Bengani H, Fish M, Brown A, Divizia MT, de Marco R, Damante G, Grainger R, van Heyningen V, Kleinjan DA (2013) Disruption of Autoregulatory Feedback by a Mutation in a Remote, Ultraconserved PAX6 Enhancer Causes Aniridia. Am J Hum Genet 93:1126-1134.

Bonkowsky JL, Wang X, Fujimoto E, Lee JE, Chien CB, Dorsky RI (2008) Domain-specific regulation of foxP2 CNS expression by lef1. BMC developmental biology 8:103.

Bruce HA, Margolis RL (2002) FOXP2: novel exons, splice variants, and CAG repeat length stability. Human genetics 111:136-144.

Chan HM, La Thangue NB (2001) p300/CBP proteins: HATs for transcriptional bridges and scaffolds. J Cell Sci 114:2363-2373.

Coessens B, Thijs G, Aerts S, Marchal K, De Smet F, Engelen K, Glenisson P, Moreau Y, Mathys J, De Moor B (2003) INCLUSive: A web portal and service registry for microarray and regulatory sequence analysis. Nucleic Acids Res 31:3468-3470.

Cook DL, Gerber LN, Tapscott SJ (1998) Modeling stochastic gene expression: Implications for haploinsufficiency. Proc Natl Acad Sci U S A 95:15641-15646.

Coutinho P, Pavlou S, Bhatia S, Chalmers KJ, Kleinjan DA, van Heyningen V (2011) Discovery and assessment of conserved Pax6 target genes and enhancers. Genome Res 21:1349-1359.

Crews ST, Pearson JC (2009) Transcriptional autoregulation in development. Current biology : CB 19:R241-246.

Davis LK, Meyer KJ, Rudd DS, Librant AL, Epping EA, Sheffield VC, Wassink TH (2008) Pax6 3' deletion results in aniridia, autism and mental retardation. Human genetics 123:371-378.

Deriziotis P, O'Roak BJ, Graham SA, Estruch SB, Dimitropoulou D, Bernier RA, Gerdts J, Shendure J, Eichler EE, Fisher SE (2014) De novo TBR1 mutations in sporadic autism disrupt protein functions. Nature communications 5:4954.

Devanna P, Middelbeek J, Vernes SC (2014) FOXP2 drives neuronal differentiation by interacting with retinoic acid signaling pathways. Frontiers in cellular neuroscience 8:305.

di Gennaro A, Spadaro O, Baratta MG, De Felice M, Di Lauro R (2013) Functional analysis of the murine Pax8 promoter reveals autoregulation and the presence of a novel thyroid-specific DNA-binding activity. Thyroid : official journal of the American Thyroid Association 23:488-496.

DiDomenico BJ, Bugaisky GE, Lindquist S (1982) The heat shock response is self-regulated at both the transcriptional and posttranscriptional levels. Cell 31:593-603.

Domene S, Bumaschny VF, de Souza FS, Franchini LF, Nasif S, Low MJ, Rubinstein M (2013) Enhancer turnover and conserved regulatory function in vertebrate evolution. Philosophical transactions of the Royal Society of London Series B, Biological sciences 368:20130027.

Duan D, Fu Y, Paxinos G, Watson C (2013) Spatiotemporal expression patterns of Pax6 in the brain of embryonic, newborn, and adult mice. Brain structure & function 218:353-372.

Ferland RJ, Cherry TJ, Preware PO, Morrisey EE, Walsh CA (2003) Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain. The Journal of comparative neurology 460:266-279.

Fulton DL, Sundararajan S, Badis G, Hughes TR, Wasserman WW, Roach JC, Sladek R (2009) TFCat: the curated catalog of mouse and human transcription factors. Genome Biol 10:R29.

Gaub S, Fisher SE, Ehret G (2016) Ultrasonic vocalizations of adult male Foxp2-mutant mice: behavioral contexts of arousal and emotion. Genes Brain Behav 15:243-259.

Gierman HJ, Indemans MH, Koster J, Goetze S, Seppen J, Geerts D, van Driel R, Versteeg R (2007) Domain-wide regulation of gene expression in the human genome. Genome Res 17:1286-1295.

Graziano C, D'Elia AV, Mazzanti L, Moscano F, Guidelli Guidi S, Scarano E, Turchetti D, Franzoni E, Romeo G, Damante G, Seri M (2007) A de novo nonsense mutation of PAX6 gene in a patient with aniridia, ataxia, and mental retardation. American journal of medical genetics Part A 143A:1802-1805.

Greig LC, Woodworth MB, Galazo MJ, Padmanabhan H, Macklis JD (2013) Molecular logic of neocortical projection neuron specification, development and diversity. Nat Rev Neurosci 14:755-769.

Grice EA, Rochelle ES, Green ED, Chakravarti A, McCallion AS (2005) Evaluation of the RET regulatory landscape reveals the biological relevance of a HSCR-implicated enhancer. Human molecular genetics 14:3837-3845.

Groszer M et al. (2008) Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits. Current biology : CB 18:354-362.

Han W, Kwan KY, Shim S, Lam MM, Shin Y, Xu X, Zhu Y, Li M, Sestan N (2011) TBR1 directly represses Fezf2 to control the laminar origin and development of the corticospinal tract. Proc Natl Acad Sci U S A 108:3041-3046.

Hao H, Li Y, Tzatzalos E, Gilbert J, Zala D, Bhaumik M, Cai L (2014) Identification of a transient Sox5 expressing progenitor population in the neonatal ventral forebrain by a novel cis-regulatory element. Developmental biology 393:183-193.

Hardison RC, Taylor J (2012) Genomic approaches towards finding cis-regulatory modules in animals. Nat Rev Genet 13:469-483.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell 38:576-589.

Hevner RF, Shi L, Justice N, Hsueh Y, Sheng M, Smiga S, Bulfone A, Goffinet AM, Campagnoni AT, Rubenstein JL (2001) Tbr1 regulates differentiation of the preplate and layer 6. Neuron 29:353-366.

Hisaoka T, Nakamura Y, Senba E, Morikawa Y (2010) The forkhead transcription factors, Foxp1 and Foxp2, identify different subpopulations of projection neurons in the mouse cerebral cortex. Neuroscience 166:551-563.

Hsueh YP, Wang TF, Yang FC, Sheng M (2000) Nuclear translocation and transcription regulation by the membrane-associated guanylate kinase CASK/LIN-2. Nature 404:298-302.

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

Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen CA, Schmitt AD, Espinoza CA, Ren B (2013) A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature 503:290-294.

Jolma A, Yan J, Whitington T, Toivonen J, Nitta KR, Rastas P, Morgunova E, Enge M, Taipale M, Wei G, Palin K, Vaquerizas JM, Vincentelli R, Luscombe NM, Hughes TR, Lemaire P, Ukkonen E, Kivioja T, Taipale J (2013) DNA-binding specificities of human transcription factors. Cell 152:327-339.

Jones L, Lopez-Bendito G, Gruss P, Stoykova A, Molnar Z (2002) Pax6 is required for the normal development of the forebrain axonal connections. Development 129:5041-5052.

Kaoru T, Liu FC, Ishida M, Oishi T, Hayashi M, Kitagawa M, Shimoda K, Takahashi H (2010) Molecular characterization of the intercalated cell masses of the amygdala: implications for the relationship with the striatum. Neuroscience 166:220-230.

Klepper K, Drablos F (2013) MotifLab: a tools and data integration workbench for motif discovery and regulatory sequence analysis. BMC bioinformatics 14:9.

Kothary R, Clapoff S, Darling S, Perry MD, Moran LA, Rossant J (1989) Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. Development 105:707-714.

Kristiansen LV, Bakir B, Haroutunian V, Meador-Woodruff JH (2010) Expression of the NR2B-NMDA receptor trafficking complex in prefrontal cortex from a group of elderly patients with schizophrenia. Schizophrenia research 119:198-209.

Kulakovskiy IV, Medvedeva YA, Schaefer U, Kasianov AS, Vorontsov IE, Bajic VB, Makeev VJ (2013) HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. Nucleic Acids Res 41:D195-202.

Kwan KY, Lam MM, Krsnik Z, Kawasawa YI, Lefebvre V, Sestan N (2008) SOX5 postmitotically regulates migration, postmigratory differentiation, and projections of subplate and deep-layer neocortical neurons. Proc Natl Acad Sci U S A 105:16021-16026.

Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. Nature 413:519-523.

Lai CS, Gerrelli D, Monaco AP, Fisher SE, Copp AJ (2003) FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. Brain : a journal of neurology 126:2455-2462.

Lai T, Jabaudon D, Molyneaux BJ, Azim E, Arlotta P, Menezes JR, Macklis JD (2008) SOX5 controls the sequential generation of distinct corticofugal neuron subtypes. Neuron 57:232-247.

Lamb AN et al. (2012) Haploinsufficiency of SOX5 at 12p12.1 is associated with developmental delays with prominent language delay, behavior problems, and mild dysmorphic features. Human mutation 33:728-740.

Li S, Weidenfeld J, Morrisey EE (2004) Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions. Molecular and cellular biology 24:809-822.

MacDermot KD, Bonora E, Sykes N, Coupe AM, Lai CS, Vernes SC, Vargha-Khadem F, McKenzie F, Smith RL, Monaco AP, Fisher SE (2005) Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. Am J Hum Genet 76:1074-1080.

Maricic T, Gunther V, Georgiev O, Gehre S, Curlin M, Schreiweis C, Naumann R, Burbano HA, Meyer M, Lalueza-Fox C, de la Rasilla M, Rosas A, Gajovic S, Kelso J, Enard W, Schaffner W, Paabo S (2013) A recent evolutionary change affects a regulatory element in the human FOXP2 gene. Molecular biology and evolution 30:844-852.

Mathelier A, Zhao X, Zhang AW, Parcy F, Worsley-Hunt R, Arenillas DJ, Buchman S, Chen CY, Chou A, Ienasescu H, Lim J, Shyr C, Tan G, Zhou M, Lenhard B, Sandelin A, Wasserman WW (2014) JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Res 42:D142-147.

Mead TJ, Wang Q, Bhattaram P, Dy P, Afelik S, Jensen J, Lefebvre V (2013) A far-upstream (-70 kb) enhancer mediates Sox9 auto-regulation in somatic tissues during development and adult regeneration. Nucleic Acids Res 41:4459-4469.

Mendoza E, Tokarev K, During DN, Retamosa EC, Weiss M, Arpenik N, Scharff C (2015) Differential coexpression of FoxP1, FoxP2, and FoxP4 in the Zebra Finch (Taeniopygia guttata) song system. The Journal of comparative neurology 523:1318-1340.

Meredith DM, Masui T, Swift GH, MacDonald RJ, Johnson JE (2009) Multiple transcriptional mechanisms control Ptf1a levels during neural development including autoregulation by the PTF1-J complex. The Journal of neuroscience : the official journal of the Society for Neuroscience 29:11139-11148.

Molina-Hernandez A, Rodriguez-Martinez G, Escobedo-Avila I, Velasco I (2013) Histamine up-regulates fibroblast growth factor receptor 1 and increases FOXP2 neurons in cultured neural precursors by histamine type 1 receptor activation: conceivable role of histamine in neurogenesis during cortical development in vivo. Neural development 8:4.

Moralli D, Nudel R, Chan MT, Green CM, Volpi EV, Benitez-Burraco A, Newbury DF, Garcia-Bellido P (2015) Language impairment in a case of a complex chromosomal rearrangement with a breakpoint downstream of FOXP2. Molecular cytogenetics 8:36.

Nelson CS, Fuller CK, Fordyce PM, Greninger AL, Li H, DeRisi JL (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 Res 41:5991-6004.

Nguyen DX, Chiang AC, Zhang XH, Kim JY, Kris MG, Ladanyi M, Gerald WL, Massague J (2009) WNT/TCF signaling through LEF1 and HOXB9 mediates lung adenocarcinoma metastasis. Cell 138:51-62.

O'Roak BJ et al. (2012a) Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 485:246-250.

O'Roak BJ et al. (2012b) Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science 338:1619-1622.
Oliva CA, Vargas JY, Inestrosa NC (2013) Wnts in adult brain: from synaptic plasticity to cognitive deficiencies. Frontiers in cellular neuroscience 7:224.
Oloumi A, Syam S, Dedhar S (2006) Modulation of Wnt3a-mediated nuclear beta-catenin accumulation and activation by integrin-linked kinase in mammalian cells. Oncogene 25:7747-7757.
Ong CT, Corces VG (2014) CTCF: an architectural protein bridging genome topology and function. Nat Rev Genet 15:234-246.
Osumi N (2001) The role of Pax6 in brain patterning. The Tohoku journal of experimental medicine 193:163-174.
Osumi N, Shinohara H, Numayama-Tsuruta K, Maekawaa M (2008) Concise review: Pax6 transcription factor contributes to both embryonic and adult neurogenesis as a multifunctional regulator. Stem cells 26:1663-1672.
Palka C, Alfonsi M, Mohn A, Cerbo R, Guanciali Franchi P, Fantasia D, Morizio E, Stuppia L, Calabrese G, Zori R, Chiarelli F, Palka G (2012) Mosaic 7q31 deletion involving FOXP2 gene associated with language impairment. Pediatrics 129:e183-188.
Pennacchio LA, Bickmore W, Dean A, Nobrega MA, Bejerano G (2013) Enhancers: five essential questions. Nat Rev Genet 14:288-295.
Prosser J, van Heyningen V (1998) PAX6 mutations reviewed. Human mutation 11:93-108.
Rice GM, Raca G, Jakielski KJ, Laffin JJ, Iyama-Kurtycz CM, Hartley SL, Sprague RE, Heintzelman AT, Shriberg LD (2012) Phenotype of FOXP2 haploinsufficiency in a mother and son. American journal of medical genetics Part A 158A:174-181.
Roh TY, Cuddapah S, Zhao K (2005) Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. Genes Dev 19:542-552.
Rouaux C, Arlotta P (2010) Fezf2 directs the differentiation of corticofugal neurons from striatal progenitors in vivo. Nat Neurosci 13:1345-1347.
Rubinstein M, de Souza FS (2013) Evolution of transcriptional enhancers and animal diversity. Philosophical transactions of the Royal Society of London Series B, Biological sciences 368:20130017.
Schanze I, Schanze D, Bacino CA, Douzgou S, Kerr B, Zenker M (2013) Haploinsufficiency of SOX5, a member of the SOX (SRY-related HMG-box) family of transcription factors is a cause of intellectual disability. European journal of medical genetics 56:108-113.
Schroeder DI, Myers RM (2008) Multiple transcription start sites for FOXP2 with varying cellular specificities. Gene 413:42-48.
Sheng M, Greenberg ME (1990) The regulation and function of c-fos and other immediate early genes in the nervous system. Neuron 4:477-485.
Shimogori T, VanSant J, Paik E, Grove EA (2004) Members of the Wnt, Fz, and Frp gene families expressed in postnatal mouse cerebral cortex. The Journal of comparative neurology 473:496-510.
Shlyueva D, Stampfel G, Stark A (2014) Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet 15:272-286.
Sin C, Li H, Crawford DA (2015) Transcriptional regulation by FOXP1, FOXP2, and FOXP4 dimerization. Journal of molecular neuroscience : MN 55:437-448.
Skidmore JM, Cramer JD, Martin JF, Martin DM (2008) Cre fate mapping reveals lineage specific defects in neuronal migration with loss of Pitx2 function in the developing mouse hypothalamus and subthalamic nucleus. Molecular and cellular neurosciences 37:696-707.
Spaeth JM, Hunter CS, Bonatakis L, Guo M, French CA, Slack I, Hara M, Fisher SE, Ferrer J, Morrisey EE, Stanger BZ, Stein R (2015) The FOXP1, FOXP2 and FOXP4 transcription factors are required for islet alpha cell proliferation and function in mice. Diabetologia 58:1836-1844.
Spiteri E, Konopka G, Coppola G, Bomar J, Oldham M, Ou J, Vernes SC, Fisher SE, Ren B, Geschwind DH (2007) Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. Am J Hum Genet 81:1144-1157.
Stark A et al. (2007) Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures. Nature 450:219-232.
Stormo GD (2000) DNA binding sites: representation and discovery. Bioinformatics 16:16-23.
Takahashi K, Liu FC, Hirokawa K, Takahashi H (2003) Expression of Foxp2, a gene involved in speech and language, in the developing and adult striatum. Journal of neuroscience research 73:61-72.
Takahashi K, Liu FC, Hirokawa K, Takahashi H (2008a) Expression of Foxp4 in the developing and adult rat forebrain. Journal of neuroscience research 86:3106-3116.

Takahashi K, Liu FC, Oishi T, Mori T, Higo N, Hayashi M, Hirokawa K, Takahashi H (2008b) Expression of FOXP2 in the developing monkey forebrain: comparison with the expression of the genes FOXP1, PBX3, and MEIS2. The Journal of comparative neurology 509:180-189.

Ticho BH, Hilchie-Schmidt C, Egel RT, Traboulsi EI, Howarth RJ, Robinson D (2006) Ocular findings in Gillespie-like syndrome: association with a new PAX6 mutation. Ophthalmic genetics 27:145-149.

Tsui D, Vessey JP, Tomita H, Kaplan DR, Miller FD (2013) FoxP2 regulates neurogenesis during embryonic cortical development. The Journal of neuroscience : the official journal of the Society for Neuroscience 33:244-258.

Uhlen M et al. (2015) Proteomics. Tissue-based map of the human proteome. Science 347:1260419.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3--new capabilities and interfaces. Nucleic Acids Res 40:e115.

Urban S, Kobi D, Ennen M, Langer D, Le Gras S, Ye T, Davidson I (2015) A Brn2-Zic1 axis specifies the neuronal fate of retinoic-acid-treated embryonic stem cells. J Cell Sci 128:2303-2318.

van Genderen C, Okamura RM, Farinas I, Quo RG, Parslow TG, Bruhn L, Grosschedl R (1994) Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in LEF-1-deficient mice. Genes Dev 8:2691-2703.

Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM (2009) A census of human transcription factors: function, expression and evolution. Nat Rev Genet 10:252-263.

Vernes SC, Nicod J, Elahi FM, Coventry JA, Kenny N, Coupe AM, Bird LE, Davies KE, Fisher SE (2006) Functional genetic analysis of mutations implicated in a human speech and language disorder. Human molecular genetics 15:3154-3167.

Vernes SC, Oliver PL, Spiteri E, Lockstone HE, Puliyadi R, Taylor JM, Ho J, Mombereau C, Brewer A, Lowy E, Nicod J, Groszer M, Baban D, Sahgal N, Cazier JB, Ragoussis J, Davies KE, Geschwind DH, Fisher SE (2011) Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. PLoS Genet 7:e1002145.

Visel A, Minovitsky S, Dubchak I, Pennacchio LA (2007) VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res 35:D88-92.

Visel A, Prabhakar S, Akiyama JA, Shoukry M, Lewis KD, Holt A, Plajzer-Frick I, Afzal V, Rubin EM, Pennacchio LA (2008) Ultraconservation identifies a small subset of extremely constrained developmental enhancers. Nat Genet 40:158-160.

Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457:854-858.

Wang GS, Hong CJ, Yen TY, Huang HY, Ou Y, Huang TN, Jung WG, Kuo TY, Sheng M, Wang TF, Hsueh YP (2004) Transcriptional modification by a CASK-interacting nucleosome assembly protein. Neuron 42:113-128.

Wasserman WW, Sandelin A (2004) Applied bioinformatics for the identification of regulatory elements. Nat Rev Genet 5:276-287.

Wellik DM (2007) Hox patterning of the vertebrate axial skeleton. Developmental dynamics : an official publication of the American Association of Anatomists 236:2454-2463.

Willsey AJ et al. (2013) Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell 155:997-1007.

Wingender E, Chen X, Hehl R, Karas H, Liebich I, Matys V, Meinhardt T, Pruss M, Reuter I, Schacherer F (2000) TRANSFAC: an integrated system for gene expression regulation. Nucleic Acids Res 28:316-319.

Wingender E, Chen X, Fricke E, Geffers R, Hehl R, Liebich I, Krull M, Matys V, Michael H, Ohnhauser R, Pruss M, Schacherer F, Thiele S, Urbach S (2001) The TRANSFAC system on gene expression regulation. Nucleic Acids Res 29:281-283.

Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, Smith SF, North P, Callaway H, Kelly K, Walter K, Abnizova I, Gilks W, Edwards YJ, Cooke JE, Elgar G (2005) Highly conserved non-coding sequences are associated with vertebrate development. PLoS biology 3:e7.

Zambelli F, Pesole G, Pavesi G (2013) Motif discovery and transcription factor binding sites before and after the next-generation sequencing era. Briefings in bioinformatics 14:225-237.

Zeesman S, Nowaczyk MJ, Teshima I, Roberts W, Cardy JO, Brian J, Senman L, Feuk L, Osborne LR, Scherer SW (2006) Speech and language impairment and oromotor dyspraxia due to deletion of 7q31 that involves FOXP2. American journal of medical genetics Part A 140:509-514.

Zentner GE, Tesar PJ, Scacheri PC (2011) Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res 21:1273-1283.

Zhang X, Huang CT, Chen J, Pankratz MT, Xi J, Li J, Yang Y, Lavaute TM, Li XJ, Ayala M, Bondarenko GI, Du ZW, Jin Y, Golos TG, Zhang SC (2010) Pax6 is a human neuroectoderm cell fate determinant. Cell stem cell 7:90-100.

Chapter 5

Activity of *FOXP2* enhancers during development and in adult brains

Abstract

Mutations in *FOXP2* affect the structure and function of the striatum, cortex and cerebellum. During development *FOXP2* is expressed in these brain areas, as well as other neural sites, and it remains expressed during adulthood. The mechanisms that initiate developmental *FOXP2* expression in these different brain structures are unknown. Enhancers of *FOXP2* and their upstream regulatory processes have been identified (as described in prior chapters of this thesis), but their tissue-specific activity has not been investigated.

In this chapter I aimed to characterize the neural sites of action of a number of human *FOXP2* enhancer elements defined from my earlier work. I created transgenic mice that expresses the *LacZ* reporter gene under control of the enhancer 37 and enhancer 330, and analysed the reporter gene expression in embryonic and adult mouse brains. Adult mice carrying enhancer 37 showed *LacZ* expression in the hippocampus, the ventricles and the olfactory bulb (OB). The ventricular expression of *LacZ* started at early post-natal age (P7) and remained during adulthood. Expression in OB began during late embryogenesis (E18.5) and faded in adult animals (P280+). By contrast, activity of enhancer 330 in adulthood was detected in the cortex and the Purkinje cell (PC) layer of the cerebellum. The enhancer activity in the cortex started in the ventricular zone during embryonic development and was found in all cortical layers at early post-natal ages (P8) and early adult ages (P88). During adulthood the enhancer activity weakened in the deeper cortical layers (P150).

The tissue-specific activity of enhancer 37 may link *FOXP2* to adult neurogenesis. The *in vivo* activity of enhancer 330 suggests that it activates *FOXP2* early during cortical development. The regulatory mechanisms of this enhancer may contribute to the cortical layer specific expression of *FOXP2.*

Introduction

The expression of human and mouse *FOXP2/Foxp2* starts during embryonic development in sub-structures of the forebrain, midbrain and hindbrain, including the thalamus, basal ganglia, cortex and cerebellum (Ferland et al., 2003; Lai et al., 2003). This complex expression pattern persists into adulthood within developed brain structures (Ferland et al., 2003). I have identified enhancer elements which may regulate the expression of human *FOXP2* during development or adulthood (Becker et al., 2015) (Chapter 3). However, the tissue-specific activity of the enhancers needs to be identified, and this is the focus of the current chapter.

To characterize enhancer tissue-specific activity, transgenic mice can be created that carry a certain enhancer sequence and express a reporter gene under the control of the enhancer (Visel et al., 2007; Boyd et al., 2015). The enhancer sequence is commonly cloned upstream of a minimal promoter and a reporter such as the beta-galactosidase (*LacZ*) gene (Kothary et al., 1989), which can be visualized by enzymatic staining in tissue samples to identify the tissue-specific activity of the enhancer (Visel et al., 2007). The preferred minimal promoter for these studies is typically derived from the mouse heat-shock 68 (*Hsp68*) gene and inactive under normal conditions (Kothary et al., 1989). Therefore, the reporter gene expression is dependent on the upstream enhancer and induced in endogenous tissue relevant for the activity of the enhancer (Visel et al., 2007; Visel et al., 2008; Visel et al., 2009; Visel et al., 2013).

The study of human *FOXP2* enhancers in mice is a promising approach to characterize human enhancer activity in homologous mammalian brain regions (Visel et al., 2013; Boyd et al., 2015). Comparative studies of TFs and enhancers strongly suggest that enhancers are subject to evolutionary changes whereas the protein structure of TFs and their binding motifs remain more stable during evolution (Rubinstein and de Souza, 2013; Villar et al., 2015). Furthermore, the expression patterns of many TFs also tend to change little during evolution (Rubinstein and de Souza, 2013). Similarly, the expression pattern of *FOXP2/Foxp2* is conserved between humans and mice (Ferland et al., 2003; Lai et al., 2003). One example of the cross-species

activity of gene enhancers can be seen for two mouse enhancers of the *Pomc* gene, which drive gene expression in the hypothalamus (Young et al., 1998). The *Pomc* enhancer sequences evolved *de novo* in mammals, but are capable to drive reporter gene expression in the hypothalamus of zebrafish (Domene et al., 2013). Thus, the tissue-specific activity of enhancers can be conserved across species and the activity of human enhancers in transgenic mice is assumed to mimic the activity pattern in humans.

The hypothesis for the work carried out in this chapter was that the *FOXP2* enhancers would drive gene expression within known regions of *FOXP2/Foxp2* expression. The expression of human and mouse *FOXP2/Foxp2* is detailed in the general introduction (Chapter 1). In short, *FOXP2/Foxp2* is mainly expressed in the neocortex, thalamus, hypothalamus, basal ganglia, amygdala, olfactory system, midbrain, cerebellum and medulla oblongata (Figure 1, 2) (Shu et al., 2001; Ferland et al., 2003; Lai et al., 2003). The expression in the cortex of mice starts at embryonic day 12.5 (E12.5) (Shu et al., 2001). In the developed cortex the *Foxp2* expression is mostly restricted to the deep layer VI (Figure 2) (Ferland et al., 2003) with some expression extending into layer V (Campbell et al., 2009; Hisaoka et al., 2010). The expression in the thalamus and hypothalamus was detected at E13.5 in mice and Carnegie Stage (CS) 23 in human (Figure 1C) (Lai et al., 2003). The ganglionic eminence, the developmental origin of the basal ganglia, expresses *FOXP2/Foxp2* at E12.5 in mice (Ferland et al., 2003) and CS23 in human (Figure 1D) (Lai et al., 2003). Midbrain expression of *FOXP2/Foxp2* is detected in the superior and inferior colliculi of the tectum and substantia nigra of mice (Ferland et al., 2003) and human (Figure 1B) (Lai et al., 2003). The cerebellar primordium shows *FOXP2/Foxp2* expression at CS23/E14.5 (Figure 1E, F) and the precursor of the medulla oblongata shows the first expression of *Foxp2* at E11.5 in mice and CS18 in human (Figure 1C) (Lai et al., 2003). Adult cerebellar expression is restricted to the Purkinje cells (PCs) and weak expression is detected in some cerebellar nuclei in mice (Ferland et al., 2003; Fujita and Sugihara, 2012). Expression in the medulla oblongata is specific to the inferior olive (Figure 1G, H) (Lai et al., 2003; Fujita et al., 2012). I focused on the neuronal expression, as mutations in *FOXP2* affect

the structure and function of neurons and neuronal networks, yielding disorders of speech and language (Vargha-Khadem et al., 1998; Watkins et al., 2002; Lai et al., 2003; Vernes et al., 2011).

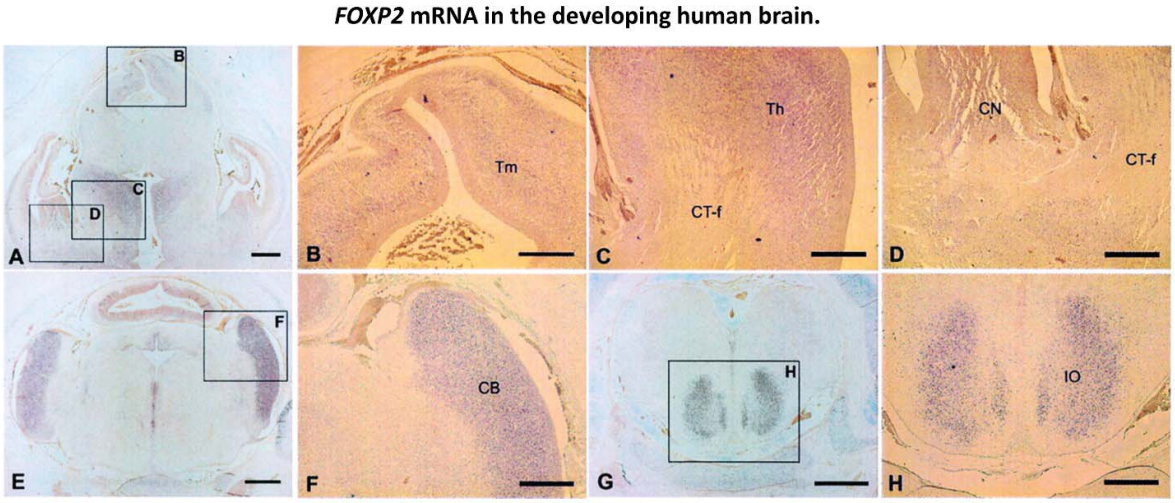

Figure 1: Reproduced from Lai et al. 2003, *Brain*. *FOXP2* mRNA expression in the developing human brain. *In situ* hybridization to *FOXP2* mRNA in transverse sections of human fetal brain slices (~9 weeks post-fertilization). Boxed areas B, C, D, F and H in pictures A, E and G are magnified in adjacent pictures. *FOXP2* expression is shown in the tectum (B), thalamus (C), caudate nucleus (D), cerebellum (F) and inferior olives (H). CN: caudate nucleus, Th: thalamus, CB: cerebellum, IO: inferior olivary complex, Tm: tectum, CT-f: cortico-thalamic tract. Scale bars: (A,E and G) 1mm, (B-D, F and H) 0.5 mm.

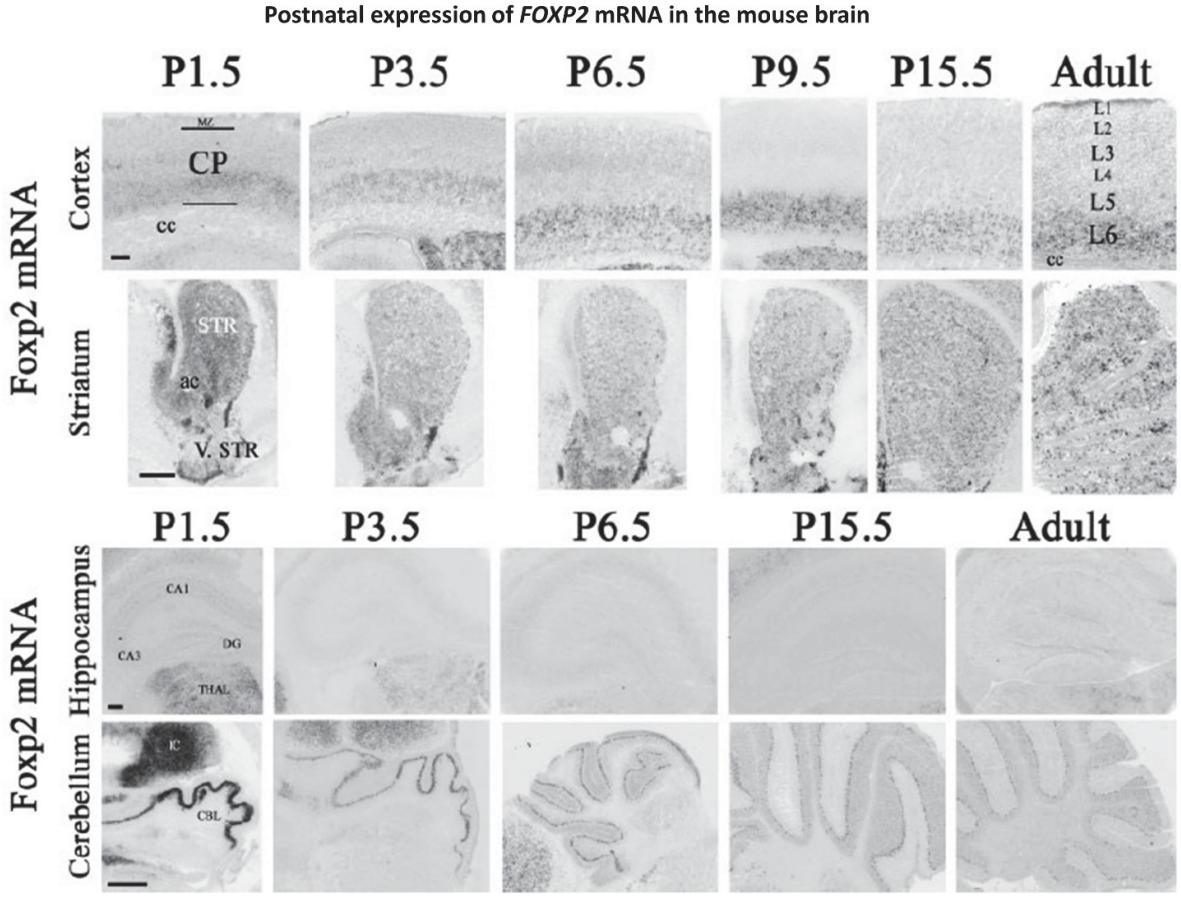

Figure 2: Reproduced from Ferland et al. 2003, *J Comp Neurol*. Post-natal expression of *Foxp2* mRNA in the mouse brain. Representative photomicrographs of *in situ* hybridizations to *Foxp2* mRNA in mouse brain slices at different postnatal time points (P1.5, P3.5, P6.5, P9.5, P15.5 and adult). *Foxp2* expression is visualized as dark staining and series of photomicrographs are shown for the cortex, striatum, hippocampus and cerebellum. L1-6: cortical layer 1-6, MZ: marginal zone, STR: striatum, V.STR: ventral striatum, THAL: thalamus.

In earlier work described in this thesis, I identified enhancer elements of the human *FOXP2* gene (Chapter 3) and TFs that may regulate the activity of these enhancers (Chapter 4). The enhancer elements were identified based on the demonstration that they contact the active *FOXP2* promoter in neuronal-like cells using chromatin conformation capture (3C). In addition two of the enhancers, enhancer 37 and enhancer 330, activated the expression of a luciferase reporter gene *in vitro* (Chapter 4). Activity of enhancer 37 was increased by overexpression of *LEF1*, *SOX5*, *FOXP1*, *FOXP2*, *FOXP4* and *TBR1* in the presence of the *CASK* co-activator. Enhancer 330 was increased by *LEF1* and repressed by *TBR1* alone, *SOX5* and *FOXP4*.

However, the enhancer's tissue-specific activities (i.e. the sites where they are active) are unknown. The *in vivo* characterization can therefore link the previously identified regulatory interactions to the brain regions and developmental periods they occur in.

In a genome-wide analysis of ultra-conserved non-coding regions, defined by long stretches (>200 bp) of identical nucleotide sequences between human and mouse, the tissue activity of hundreds of human enhancers were characterized in transgenic mice and the data made publicly available in the VISTA Enhancer Browser (Visel et al., 2008). The tissue activity was determined in E11.5 mouse embryos using the *Hsp68-LacZ* reporter gene construct (Visel et al., 2008), the same construct I used for analyses in this chapter. The regions that were tested in the VISTA project included enhancer 330 (VISTA ID: hs218), because of the high evolutionary conservation across vertebrate species (Figure 3A). Enhancer 330 is annotated in the VISTA Browser as being active in the forebrain. Here I followed the activity between early post-natal (P8) and adult mice (P200).

The aim of this chapter was to characterize the neural regions of *FOXP2* enhancer 330 (Figure 3A) and enhancer 37 activity (Figure 3B). First, I determined the tissue-specificity activity of the enhancers in brains of adult animals. The aim was to then follow the enhancer activity during late embryonic and early post-natal development. In combination with the previously described *FOXP2/Foxp2* expression pattern, these results will help us understand the regulatory mechanisms that guide expression of *FOXP2* in specific brain regions.

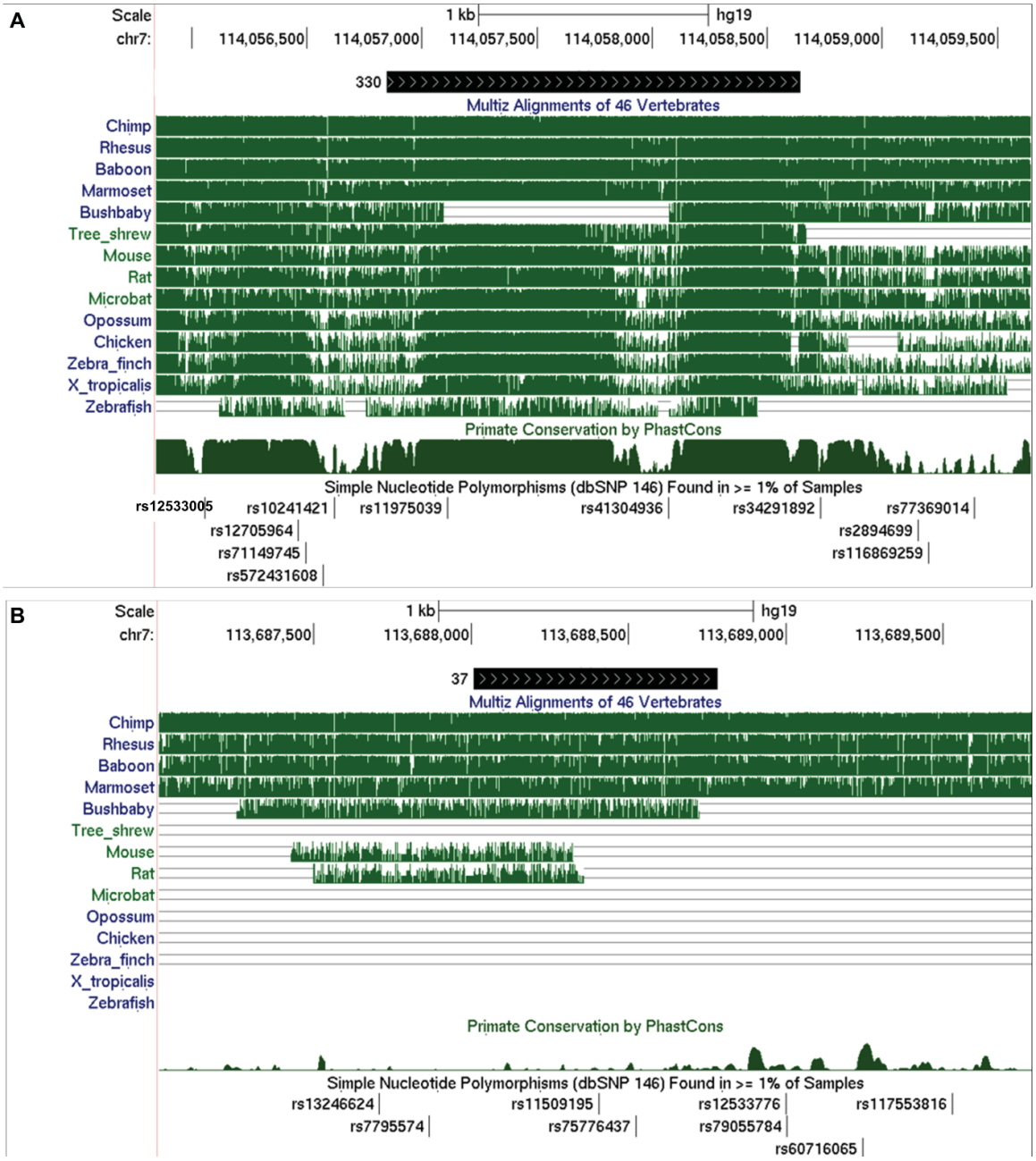

Figure 3: Evolutionary and common human genetic variation of *FOXP2* enhancer elements. The alignment of human genomic sequences to primate, mammal and genomes of evolutionary more distant species is shown in the below the enhancer. Evolutionary conservation of each nucleotide across primate species is shown within the Primate Conservation by PhastCons track. In addition, common single/simple nucleotide polymorphisms (SNPs), which are found in more than 1% of the human population, are shown in the last track. The A) Genomic position of enhancer 330 (chr7: 114056845-114058646, hg19) and 1kb of flanking sequences. The enhancer sequence is highly conserved across vertebrates. The enhancer, including the flanking sequences contains 12 common human SNPs B) Genomic position of enhancer 37 (ch7: 113688009- 113688782, hg19) and 1kb of flanking sequence. The enhancer sequence shows low conservation across primates and is mostly absent from non-primate vertebrates. The enhancer, including flanking sequences contains 8 common human SNPs.

Materials and Methods

Molecular cloning of LacZ constructs

I created DNA constructs that contain the enhancer sequence in a cassette with a minimal promoter (*Hsp68*) and the beta-galactosidase (*LacZ*) gene (Kothary et al., 1989). Enhancer 37 and 330 were amplified from genomic DNA of a healthy individual and inserted into TOPO vectors, as described before (Chapter 3). The enhancer sequences were further amplified using primers, tagged with attB gateway-cloning sequences. The resulting PCR products were run on a 1% agarose gel and extracted from the gel using Wizard SV Geld and PCR Clean-Up System (Promega). The attB-tagged PCR product together with pDONR 221 vector was mixed with a BP clonase to insert the PCR product into the plasmid (Gateway BP cloning reaction). The resulting pDONR/Enhancer plasmids were incubated with pHsp68-LacZ.Gateway and an LR clonase to produce the final pHsp68-LacZ.Enhancer37 and pHsp68-LacZ.Enhancer330 plasmids (Gateway LR cloning reaction).

Pronuclear injection of enhancer constructs

The Enhancer constructs were linearized using SalI restriction enzymes (NEB) in 2h digestion at 37˚C. The digestion sample was run on a 0.6% agarose gel (TAE buffer), loading 200 ng of product on individual lanes. The gel was run in TAE buffer and the desired product was extracted from the gel using QIAEX II (QIAGEN). This was necessary to remove the plasmid backbone, which reduces the success rate of the genomic integration. The final products were eluted and stored at 4˚C.

The transgenic mice were generated in the laboratory of Friedemann Kiefer, Max Planck Institute for Molecular Biomedicine, Münster, Germany using the C57BL/6 mouse strain. Pronuclear injection (PNI) was done as described before (Ittner and Gotz, 2007). With this technique the enhancer-hsp68-LacZ construct is injected into fertilized mouse oocytes at the one-cell stage where it integrates randomly into the mouse genome. Upon successful integration, the enhancer-reporter cassette is present in all cells of the developing mouse.

To validate genomic integration, I genotyped the mouse pups for the presence of the LacZ reporter gene. Mouse tails were lysed in 100 µl lysis buffer containing 50 µg/ml fresh Proteinase K for 10 min at 100˚C. The samples were constantly shaken at 1000RPM and vortexed for several seconds after 5 min of incubation and at the end of the incubation. The samples were spun down for 5 min at 12,000g. The supernatant was transferred to a clean Eppendorf tube and the eluted DNA precipitated with ice-cold ethanol. After ethanol precipitation the pelleted DNA was dissolved in 20 µl water. To amplify the LacZ gene, 1µl was used in a standard titanium PCR reaction (Clontech) and a LacZ specific primer pair (Forward: 5'- CGATGAGCGTGGTGGTTATGCC-3'; Reverse: 5'- GACGATTCATTGGCACCATGC-3'). The band pattern was analysed in a 1% agarose gel. The expected band size for presence of the LacZ reporter gene was 440 bp.

Positively genotyped mice were transported from the Max Planck Institute for Molecular Biomedicine in Münster, Germany to the central animal facility (Centraal Dierenlaboratorium) in Nijmegen, The Netherlands according to regulations. The mice were housed for 1-2 months in quarantine to ensure that the mice were pathogen free.

Staining of mouse brain tissue

Whole brains were dissected from sacrificed mice and washed in PBS. The brains were first incubated in 4% formalin (4% formaldehyde, 10% methanol) for 30 min at RT and then cut. All brains were cut at the sagittal midline. Embryonic day 14 brains were not cut further. One hemisphere of embryonic day 18/19 and early post-natal brains were cut at two coronal planes; one cut through the frontal cortex and ganglionic eminences and another cut at the dorsal end of the cortex. Adult (P15+) brains were cut at an additional coronal plane at the medial cortical line. The brain blocks were incubated in 4% formalin for another 30 min at RT to further fixate the tissue. The brain was then transferred into staining buffer (10 mM phosphate buffer pH 7, 150 mM NaCl, 1 mM $MgCl_2$, 0.1 % deoxycholate, 0.02% Igepal and 0.01 % SDS) with fresh added 50 mg/ml X-Gal, 200 mM ferrocyanide and 200 mM ferricyanide. The tissue was

incubated over night at 37˚C. The protein product of LacZ, beta-galactosidase, is enzymatically active and able to hydrolyze X-Gal (5-Brom-4-chlor-3-indoxyl-β-D-galactopyranosid) to galactose and a blue water-insoluble indigo-dye (Juers et al., 2012). The next morning the tissue was washed three times in PBS and further fixed in 4% formaldehyde for 1 day. Finally the tissue was transferred to 0.05% azine in PBS solution and stored at 4˚C. The staining was analysed under the S2000C (Zeiss) stereo microscope and documented using the powershot G7 (Canon) digital camera at 3.2x magnification.

Results

Enhancer 330 drives cortical and cerebellar expression

In order to investigate the *in vivo* activity of the enhancer elements I cloned the sequences previously used in chapter 4, into a plasmid that contains the E.coli derived *LacZ* gene downstream of the *Hsp68* minimal promoter. The enhancer constructs integrated randomly into the mouse genome (Ittner and Gotz, 2007). I generated several mice to search for reproducible reporter gene expression across random insertion events. According to the VISTA enhancer browser, a brain structure with reporter gene expression can be defined as "reproducibly staining brain region" if it is found in at least three founder animals (Visel et al., 2007). In addition to this criterion, I defined a region as reproducibly staining if it is found in 50% or more of the analysed founder animals. Reproducible staining is believed to represent the endogenous enhancer activity.

Pronuclear injections (PNI) of enhancer 330 yielded 108 mice, of which seven mice were positively genotyped for *LacZ*. One of the seven mice was excluded from the analysis because it showed staining across the whole brain, which is a potential source of false-positive replication. The initial assessment of reporter gene expression was done on adult brains, because, the transgenic mice were generated in a separate institute and for ethical reasons, the pregnant foster mothers could not be transported to our facility. In the adult (P108+) mice I detected reproducible staining in the cortex (4 of 6 mice) and cerebellum (3 of 6 mice) (Figure 4). I further detected non-reproducible staining in the thalamus (2 of 6 mice). Two founder mice did not show any staining. The cortical staining was distributed in patches with no recurrent pattern (Figure 5). The staining was detectable in all cortical layers and the brain with the strongest overall staining showed strong staining in the upper layers and weak staining in the deeper layers (Figure 5 E, F; Figure 4A, B). Prior work has established that *FOXP2/Foxp2* cortical expression is largely limited to the deeper cortical layer VI. Therefore, it is surprising that the enhancer shows a tendency to stain the upper cortical layers. Positive *LacZ* staining was strong within the cell bodies of the PC layer (Figure 4C, D; Figure 6). One brain showed

strong staining within all lobules (Figure 6A), whereas two brains showed more prominent staining in the lobules VIII-X (Figure 6A, C). *FOXP2/Foxp2* shows strong expression in PCs within most lobules, with moderate to weak expression in lobule X of the mouse cerebellum (Fujita and Sugihara, 2012). The observed enhancer activity thus overlaps known expression patterns for *FOXP2/Foxp2* in the cerebellum. Immunohistochemistry (IHC) and/or in situ hybridization (ISH) experiments would be necessary to determine the exact overlap. The reporter gene expression suggests that enhancer 330 regulates human *FOXP2* expression in the cortex and PCs.

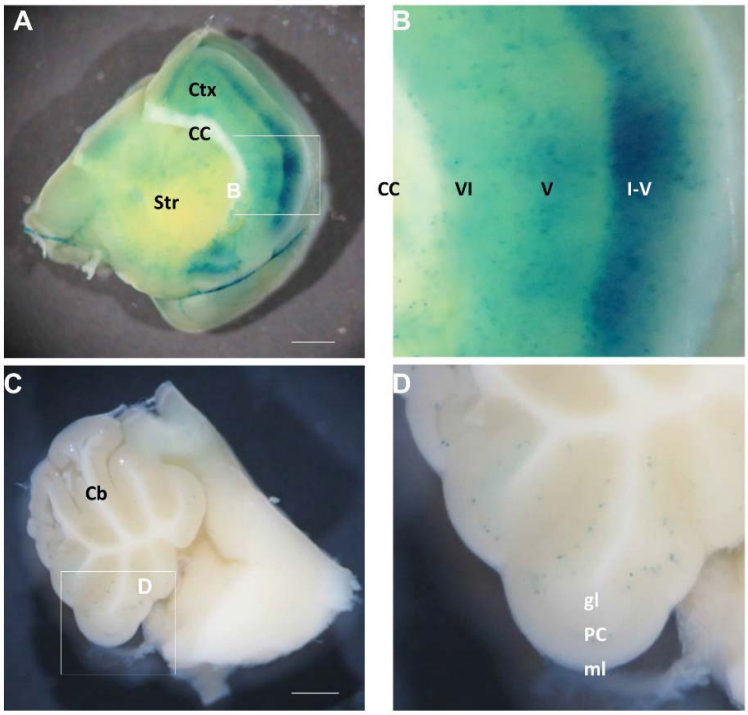

Figure 4: Tissue activity of enhancer 330 in adult mouse brain. Representative stereomicroscopic pictures of adult (P108) mouse brains stained for *LacZ* expression. Boxed areas B and D in pictures A and C are manual zooms of adjacent pictures. A,B) The upper cortical layers of the neocortex show strong staining and scattered positive cells are visible in the lower layers; C-E) The Purkinje cell layer in the cerebellum shows strong and specific staining. CC: corpus callosum, Ctx: cortex, I-VI: cortical layer I-VI, Str: striatum, Cb: cerebellum, gl: granular layer, PC: Purkinje cell layer, ml: molecular layer. Scale bars: 1 mm.

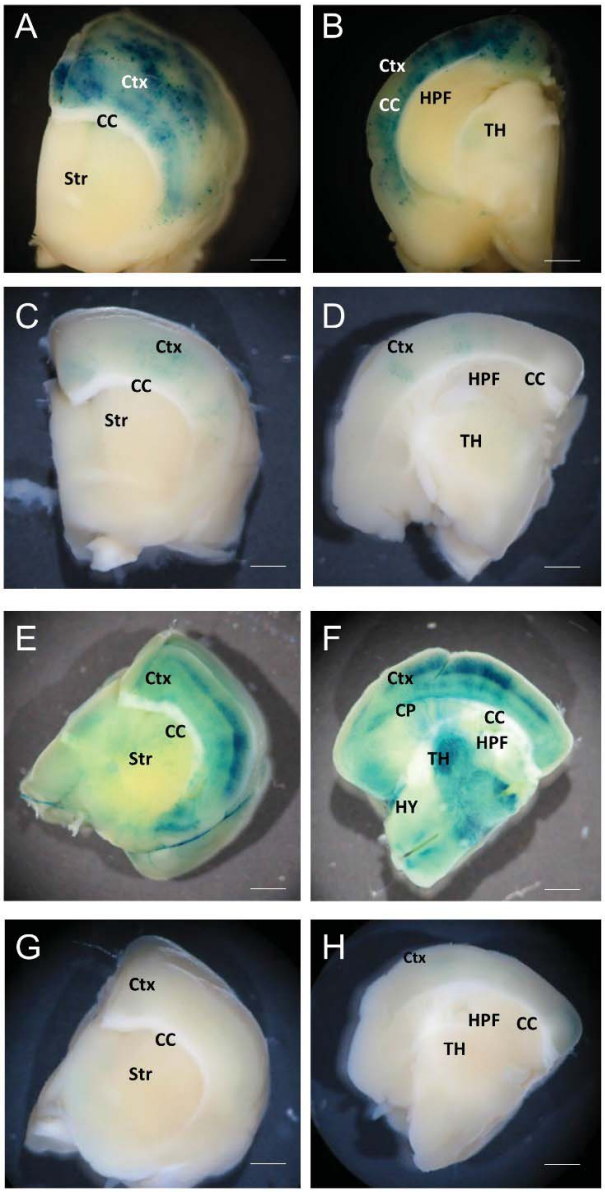

Figure 5: Cortical staining of enhancer 330. Stereomicroscopic pictures of cortical sliced tissue blocks of mouse brains stained for *LacZ* expression. A,B) Mouse #64 (P200+) with patchy staining across the cortex. C,D) Mouse #112 (P108) shows weak patchy staining. E,F) Mouse #113 (P108) with strong cortical staining in the upper cortical layers and weak uneven staining in the rest of the cortex. This mouse also showed strong thalamic staining. G,H) #117 (P108) shows weak and patchy staining in the upper cortical layers. Ctx: cortex, CC: corpus callosum, Str: Striatum, TH: thalamus, HPF: hippocampal formation, CP: caudate-putamen, HY: hypothalamus. Scale bar: 1 mm.

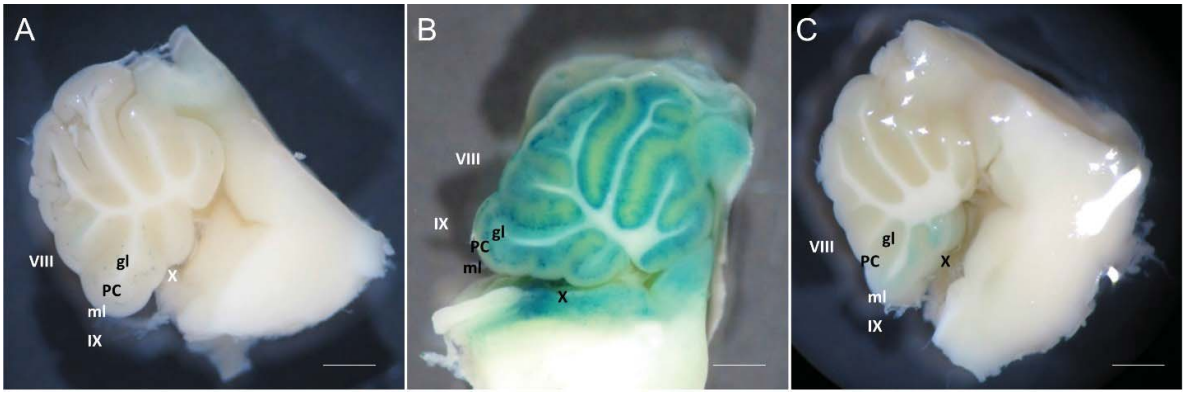

Figure 6: Cerebellar staining of enhancer 330. Stereomicroscopic pictures of the cerebellum mouse brains stained for *LacZ* expression. A) Mouse #112 shows specific staining in the cell body of the PCs. The majority of stained PCs are located in lobule IX. B) Mouse #113 shows strong staining in the PC layer of the whole cerebellum, C) Mouse #117 shows weak staining in lobules X and IX. gl: granular layer, ml: molecular layer, PC: Purkinje cell layer. Scale bar: 1 mm.

Cortical activity of enhancer 330 during development overlaps published FOXP2 expression sites

The most reproducible staining for enhancer 330 was detected in the neocortex, which is in agreement with data from the VISTA enhancer browser (Visel et al., 2007). According to the prior VISTA data, enhancer 330 is active in the neocortex of the forebrain at E11.5 (Figure 7A-C). In the cortical sections from the VISTA project the staining was present in the ventricular zone of the pallium, which is the developmental precursor of the neocortex, (Figure 7B, C) along the anterior to posterior line (Figure 7A). These findings are consistent with the hypothesis that the enhancer activates *Foxp2* in the embryonic cortex. However, the VISTA enhancer data is limited to the early embryonic development and further enhancer mapping was necessary to determine the activity in post-natal and adult cortices.

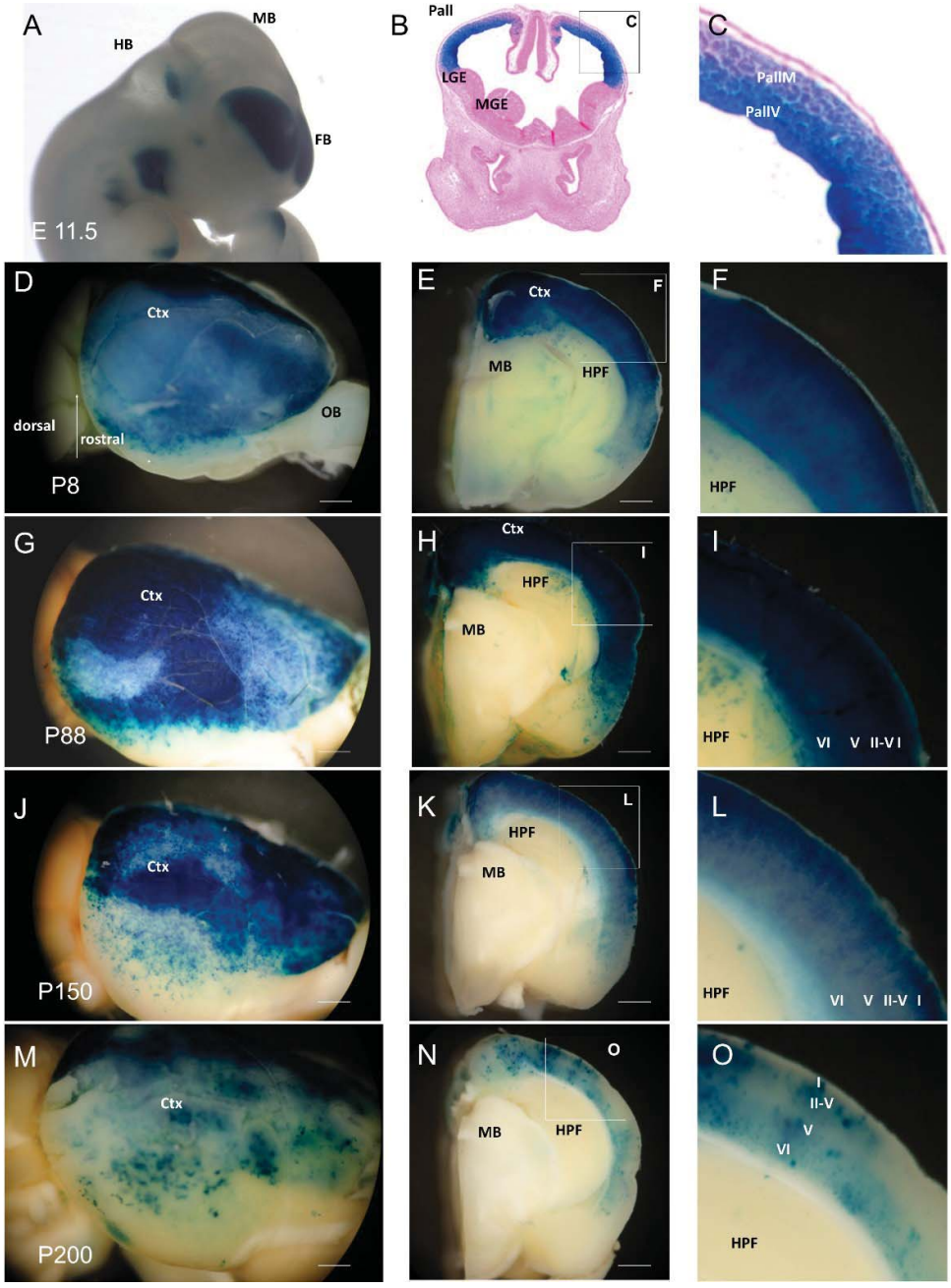

Figure 7: Embryonic, neo-natal and adult cortical staining of enhancer 330. Enhancer 330 tissue activity in mouse brains, visualized by *LacZ* staining. Pictures C,F,I ,L and O are manual zooms of the adjacent pictures. A-C) Reproduced and adapted pictures of enhancer 330 in E11.5 mouse embryos downloaded from the VISTA Enhancer Browser (http://enhancer.lbl.gov/cgi-bin/imagedb3.pl?form=presentation&show=1&experiment_id=218&organism_id=1). A) Lateral view of the embryo and B-C) Microscopic picture of *LacZ* staining in cortical slices of the same embryo. D-O) Mouse brains from mouse line #64. (D, G, J, M) lateral view of mouse brains, showing intense staining across the forebrain and (E, F, H, I, K, L, N, O) cortical sliced brain tissue blocks, showing *LacZ* staining across the cortical layers. HB: hindbrain, MB: midbrain, FB: forebrain, Pal: pallium, LGE: lateral ganglionic eminence: MGE: medial ganglionic eminence, PallV: ventricular zone of pallium, PallM: mantle zone of pallium, Ctx: cortex, OB: olfactory bulb, HPF: hippocampal formation, I-VI: cortical layer I-VI. Scale bar: 1mm, no scale bars available for pictures A-C.

I established a mouse line of founder mouse #64, chosen because it showed strongest neocortical staining. To establish this mouse line, I crossed the founder mouse with a WT C57BL/6 mouse to produce heterozygote offspring. The heterozygotes were further crossed with WT or heterozygote offspring of the same founder animal. I did not observe differences in staining intensity or staining pattern for the homozygous of heterozygotes state. Due to limited access to the experimental mice, timed mating was not possible and I determined only post-natal reporter gene expression of this mouse line. I analysed enhancer activity at post-natal day 8 (P8), in adult mice (P88 and P150) and at a late adult time point (P200). At P8 the cortical layers of the neocortex are established and the enhancer activity in all layers appeared equally strong (Figure 7E, F). In the coronal plane the cortical staining was present along the rostral-ventral line in the medial and lateral parts of cortex. At age P150 the staining decreased in the deeper layers of the cortex (Figure 7J-L) and at age P200 the staining intensities further decreased in both the deeper and upper cortical layers (Figure 7M-O). The activity pattern along the rostral-ventral line remained stable through adulthood, for the different ages that I studied. Some patches of weaker staining were visible within the stained cortical structures, but the location of these patches was inconsistent across post-natal time points. At all developmental stages the reporter gene staining was absent from the dorsal side of the cortex, which forms the three-layered paleocortex. In summary, enhancer 330 is activated in the ventricular zone of the pallium, remains active across the cortical plate in early post-natal mice and in adult mice the enhancer activity decreases. I could not study the cerebellar activity of enhancer 330, as the mouse line did not show cerebellar staining.

Enhancer 37 is active in regions both with and without known FOXP2 expression

PNI of the enhancer 37 construct yielded 87 mice, of which 13 mice had positive genotypes for the reporter gene. The mice were sacrificed at adult age (P145+) and the brains were dissected for LacZ staining. Two mice showed unspecific whole-brain staining and were excluded from the analysis as potential source for false-positive replication. In sum, I analysed the brain staining pattern of eleven mice (Mouse #2, 49, 41, 38, 45, 48, 57 58, 59, 32 and 33).

I detected LacZ staining in six brain regions. Reproducible staining was observed in only two regions the hippocampus (7/11) and ventricles (6/11). In the hippocampus I observed staining in the pyramidal neurons of the dentate gyrus and CA1 (Figure 8A, B; Figure 9). Expression was visible in the dentate gyrus of 5 brains (Figure 9A, D-G), whereas weak CA1 expression was detected only in three brains. Thus, according to my definition the staining in individual pyramidal regions of the hippocampus is not reproducible. The ventricular staining was found in a thin layer of the lateral ventricles (Figure 8C, D; Figure 10) and the third ventricle (Figure 8C, E). The staining seems to appear in the outermost cell layer, which consist of specialist ependymal cells. IHC of ISH for cell-type specific markers would be required to validate the cell identity. The staining pattern in hippocampus and ventricles was surprising as *FOXP2* has not been reported to be expressed in these regions in either the mouse or human.

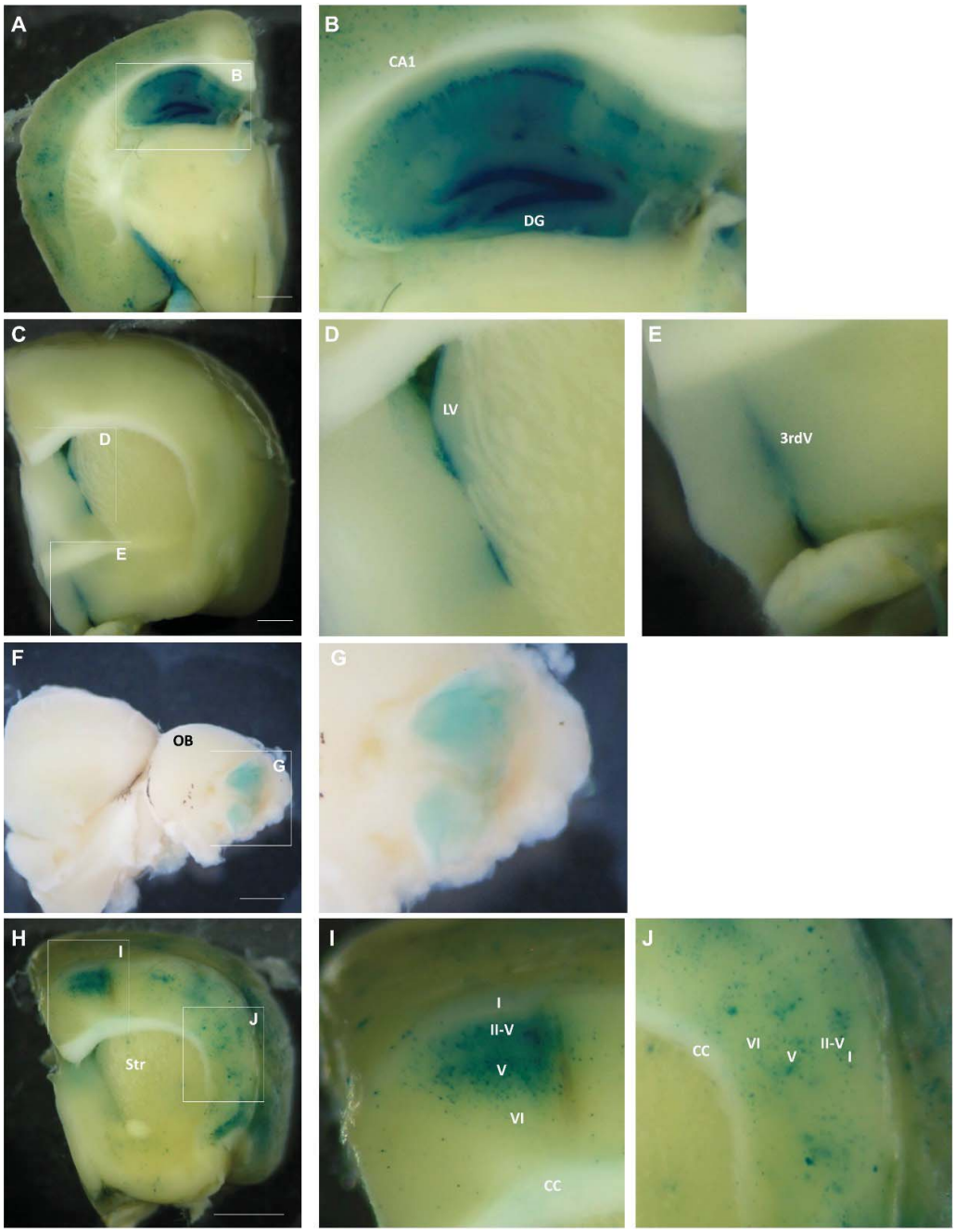

Figure 8: Tissue activity of enhancer 37 in adult mouse brain at P200+. Representative stereomicroscopic pictures of mouse brains stained for *LacZ* expression. Boxed areas B, D, E, G, I and J in pictures A, C, F and H are manual zooms of adjacent pictures. *LacZ* staining shown A,B) in the pyramidal neurons of the hippocampus; C-E) in the ventricular system of the left (D) and third ventricle (E); E,G) in the olfactory bulb and H-J) in the neocortex. DG: dentate gyrus; CA1: cornu ammonis area 1; LV: left ventricle; 3rdV; third ventricle; OB: olfactory bulb, Str: striatum, CC: corpus callosum, I-VI: cortical layer I-VI.

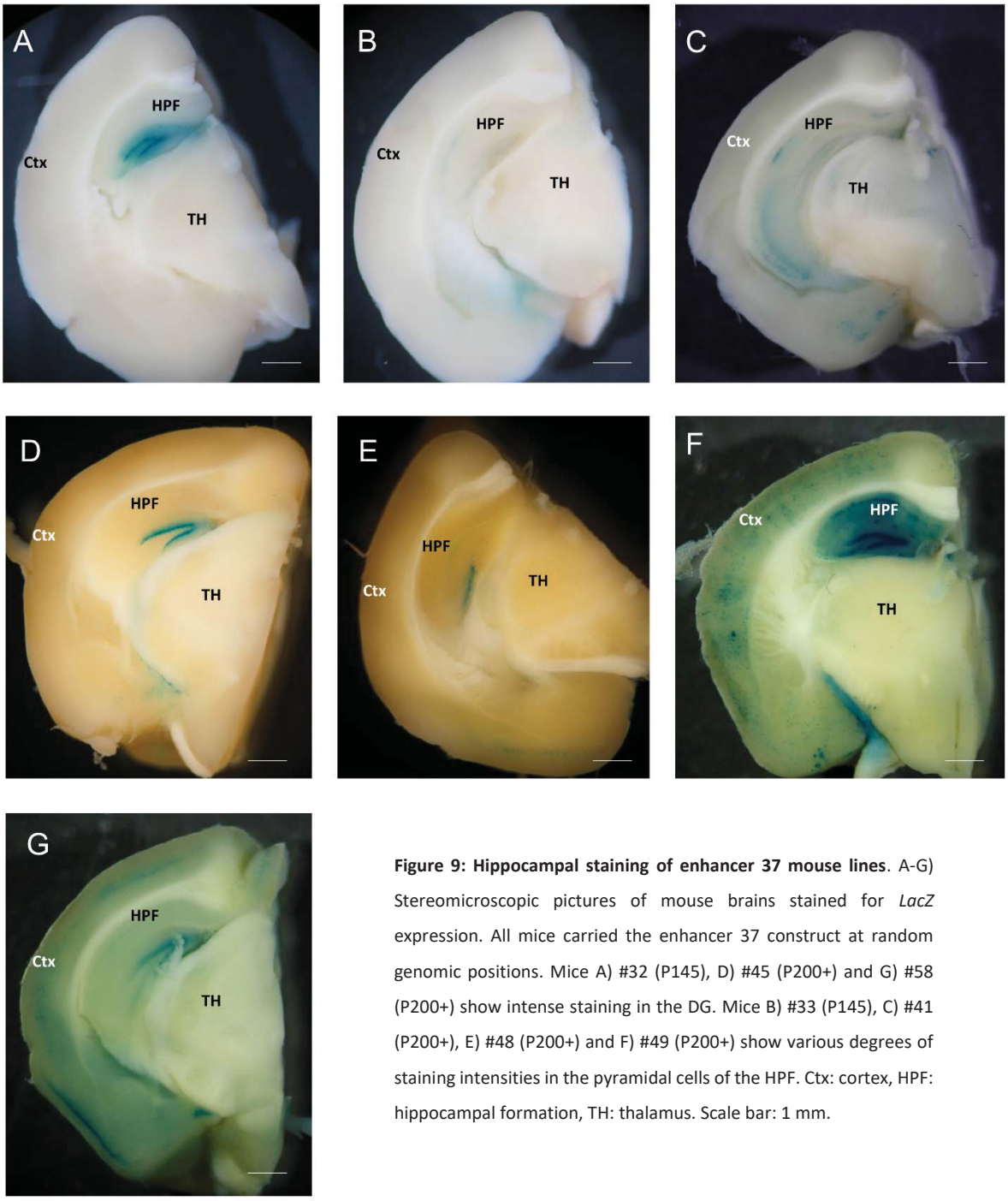

Figure 9: Hippocampal staining of enhancer 37 mouse lines. A-G) Stereomicroscopic pictures of mouse brains stained for *LacZ* expression. All mice carried the enhancer 37 construct at random genomic positions. Mice A) #32 (P145), D) #45 (P200+) and G) #58 (P200+) show intense staining in the DG. Mice B) #33 (P145), C) #41 (P200+), E) #48 (P200+) and F) #49 (P200+) show various degrees of staining intensities in the pyramidal cells of the HPF. Ctx: cortex, HPF: hippocampal formation, TH: thalamus. Scale bar: 1 mm.

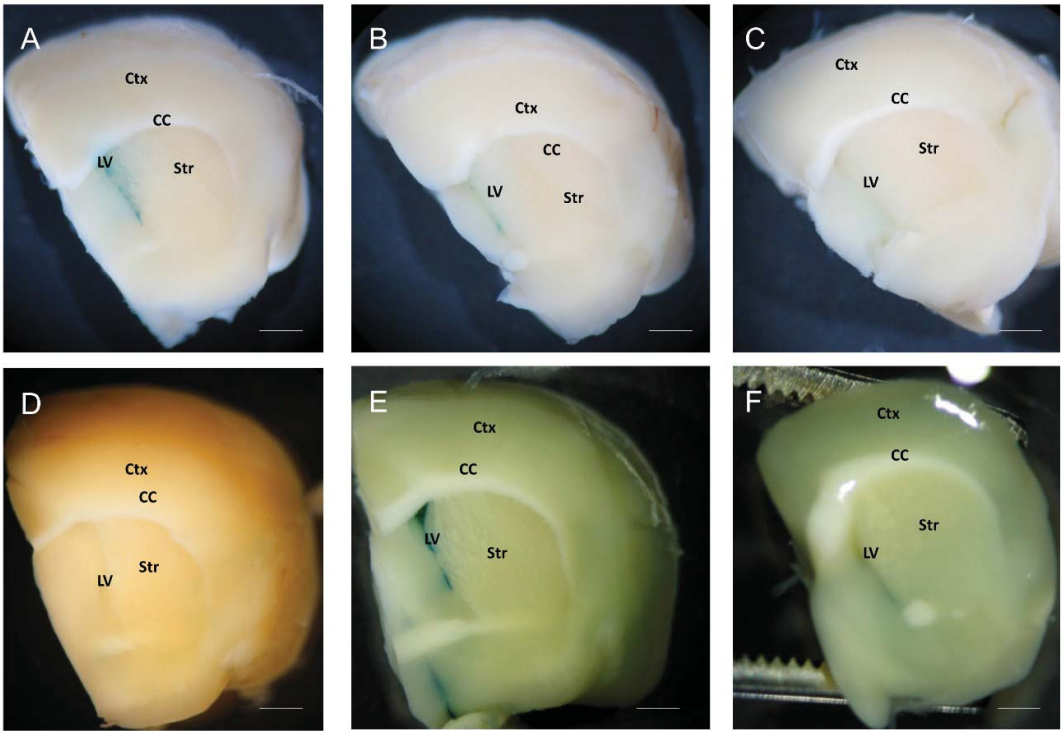

Figure 10: Lateral ventricle staining of enhancer 37 mouse lines. A-F) Stereomicroscopic pictures of mouse brains stained for *LacZ* expression. All mice carried the enhancer 37 construct at random genomic positions. Mouse A) #2 (P129), B) #32 (P145), C) #33 (P145), D) #45 (P200+), E) #48 (P200+) and F) #57 (P200+). A,B,E) Strong staining in the LV. C,D,F) weak staining in the LV. CC: corpus callosum, Ctx: cortex, Str: striatum, LV: lateral ventricle. Scale bar: 1 mm

The OB and neocortex stained positive in five out of eleven founder mice (Figure 11 and 12). These staining meet the criteria of being stained in more or equal to 3 mice, but do not meet the strict criteria for reproducible staining in 50% of the analysed animals. Because *FOXP2/Foxp2* is expressed in these regions I determined their activity pattern. In the OB I saw consistent staining in a lateral region (Figure 8F, G; Figure 11), which does not overlap with a defined region from the Allen Brain Reference Atlas (Hawrylycz et al., 2012). The OB staining could suggest overlap with *FOXP2* expression. The cortical staining was distributed in clusters with dense staining in the upper cortical layers and 4 brains showed staining in the allocortex (Figure 8 H-J). However, the allocortex, as well as the upper layers of the neocortex, do not express *Foxp2 (Ferland et al., 2003).* The staining therefore suggests that the activity of the enhancer does not overlap with known *FOXP2* expression patterns in the cortex.

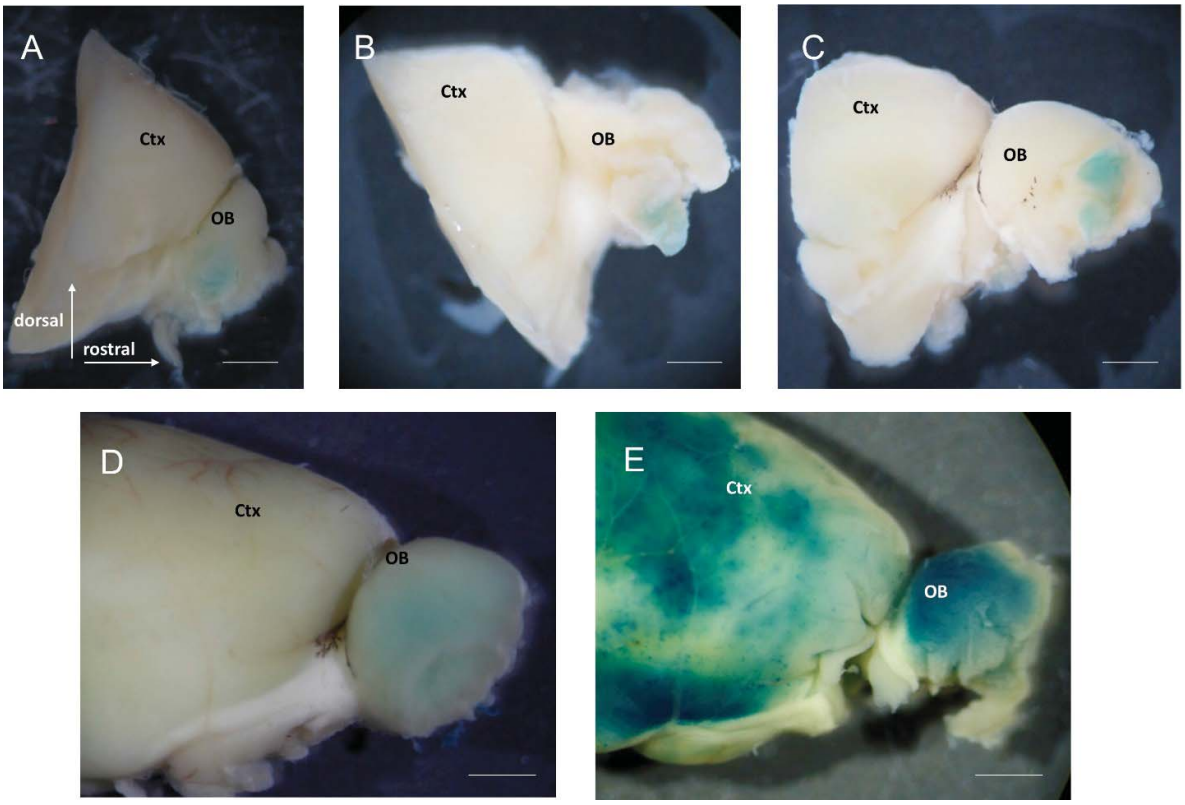

Figure 11: Olfactory bulb staining of enhancer 37 mouse lines. A-E) Stereomicroscopic pictures of mouse brains stained for *LacZ* expression. All mice carried the enhancer 37 construct at random genomic positions. Mice A) #2 (P200+), B) #32 (P145), C) #33 (P135), D) #41 (P200+) and E) #49 (P200+) show staining of a ventral region in the olfactory bulb. Ctx: cortex, OB: olfactory bulb. Scale bars: 1 mm

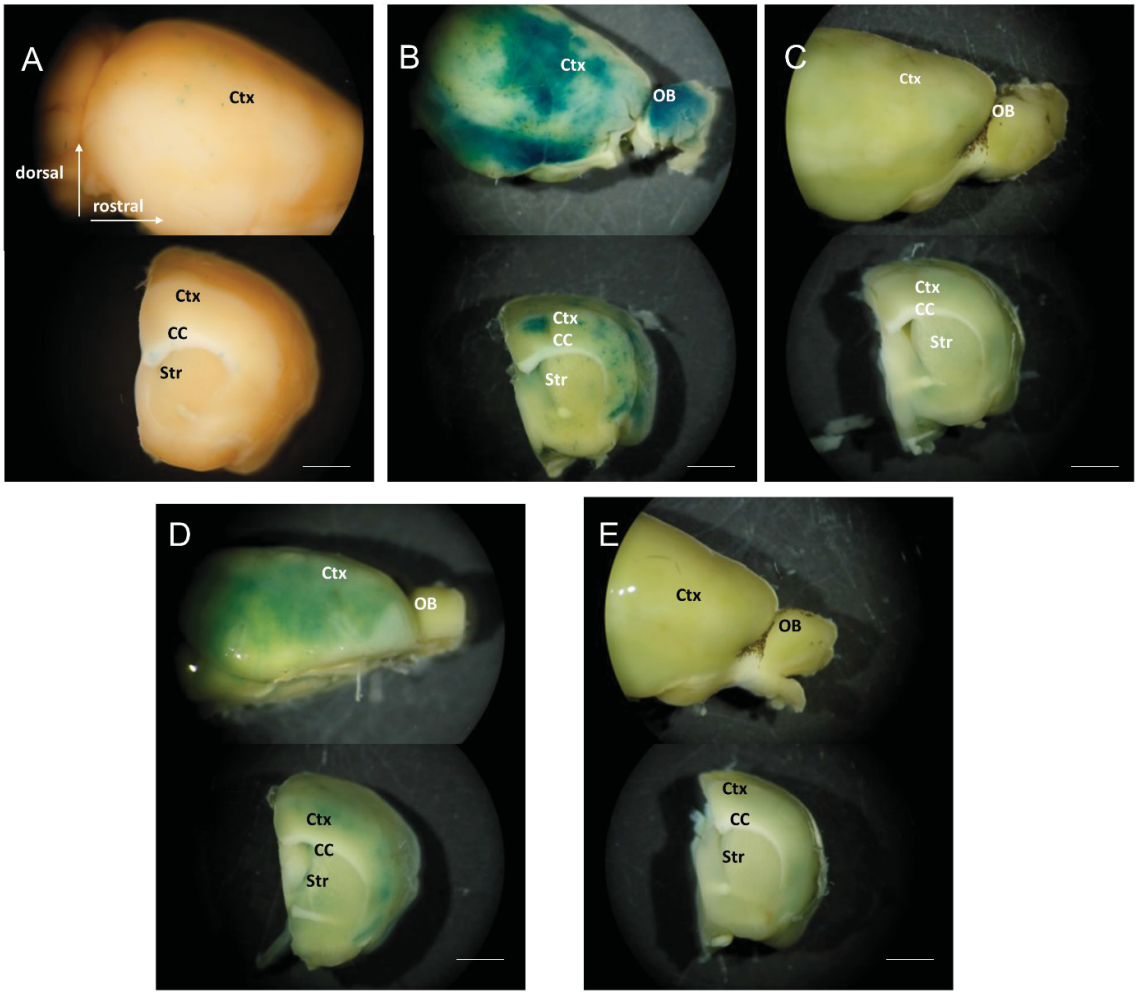

Figure 12: Neocortical staining of enhancer 37 mouse lines. A-E) Stereomicroscopic pictures of mouse brains stained for *LacZ* expression. All mice carried the enhancer 37 construct at random genomic positions. A) Mouse #38 (P200+) shows scattered positive stained cells across the cortex. B) Mouse #49 (P200+) shows patchy staining across the cortex and OB. C) Mouse #57 (P200+), D) #58 (P200+) and E) #59 (P200+) with weak and patchy staining across the cortex. CC: corpus callosum, Ctx: cortex, Str: striatum, OB: olfactory bulb. Scale bar: 1 mm.

Enhancer 37 activity changes during embryonic and adult development

The most promising overlap of known *FOXP2* expression patterns and enhancer 37 activity was in the OB. Therefore, I established a stable mouse line of founder mouse #2, which showed staining in the adult ventricles and OB. I harvested mice at embryonic stages (E13.5, E18.5), neo-natal (P5, P7), juvenile (P18), late adolescent (P39), adult (P129) and late adult age (P218). In the early embryo (E13.5) I did not detect LacZ staining (Figure 13A). At E18.5 I observed staining in the OB (Figure 13B), which was also observed in adolescent mice (P39) (Figure 13C). The late adult individual (P281) did not show OB staining (Figure 13D).

Ventricular staining was first visible at post-natal age P7 (Figure 14A). In late adolescent (P39) and late adult (P281) mice, I observed staining in the lateral and third ventricles (Figure 14B, C). Similar to the adult, the staining is observable in a thin lining of the ventricular walls. The regulatory control of enhancer 37 in the ventricular system seems to start after birth and remains throughout adulthood. In the OB, the enhancer becomes active during late embryonic development and declines in late adult stages. To determine the overlap of this enhancer with *Foxp2* expression it will be crucial to determine the *Foxp2* mRNA and protein levels at these sites.

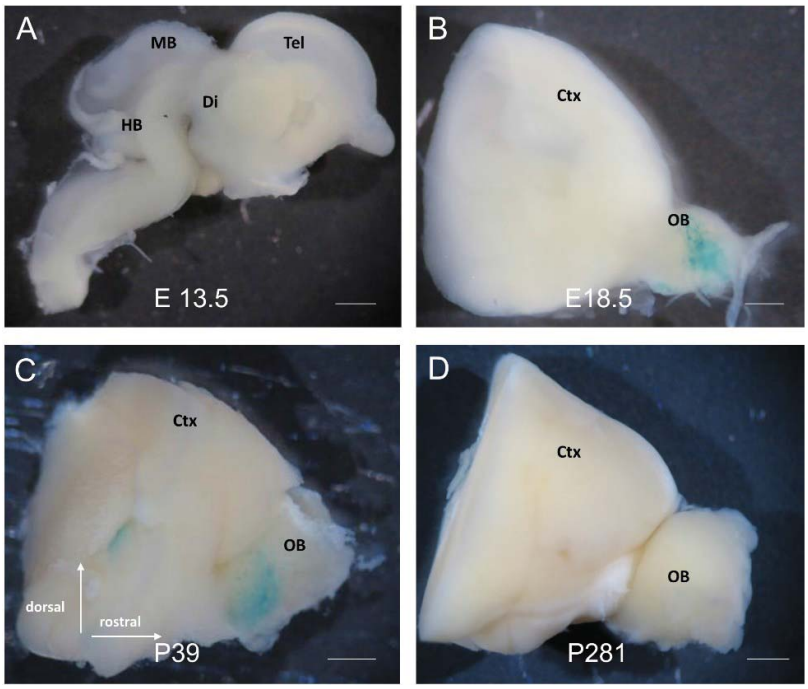

Figure 13: Development of olfactory bulb staining of enhancer 37 in mouse line 2. Stereomicroscopic pictures of mouse brains stained for *LacZ* expression. All mice are offspring of the same founder mouse. (A) *LacZ* staining shown in a sagittal section along the midbrain line at embryonic day 13.5 is not present; (B). OB staining is first visible at age E18.5; (C) OB staining continues to be present in late adolescent mice at P39; (D) No OB staining is not detectable in mature adult mice. A-D) are lateral views of the olfactory bulb of a tissue block that was cut along the frontal cortex. HB: hindbrain, MB: midbrain, Di: diencephalon, Tel: Telencephalon, Ctx: cortex, OB: olfactory bulb. Scale bars: 1 mm

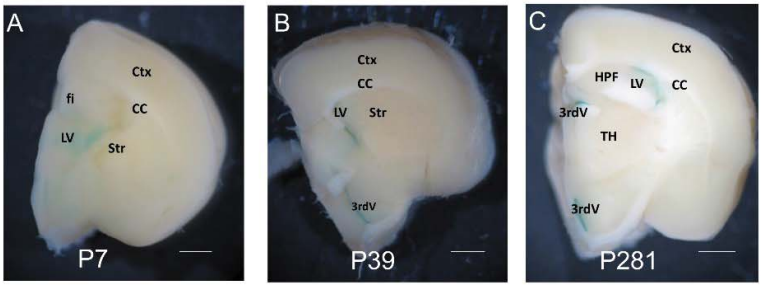

Figure 14: Development of ventricular staining of enhancer 37 in mouse line 2. Stereomicroscopic pictures of mouse brains stained for *LacZ* expression. All mice are offspring of the same founder animal. The staining of the lateral ventricle is first visible at age P7 (A) and is present in late adolescent mice (B). The staining remains intense in mature adult mice (C). fi: fimbria, LV: lateral ventricle, CC: corpus callosum, Ctx: cortex, Str: striatum, 3rdV: third ventricle, HPF: hippocampal formation, TH: thalamus. Scale bar: 1 mm.

Discussion

In this chapter I analysed the activity patterns of two putative *FOXP2* enhancers in the developing and adult central nervous system. I used PNI to insert enhancer-reporter gene constructs into fertilized mouse oocytes. With this method the transgene inserts randomly into the genome. It was therefore crucial to use criteria that distinguish endogenous enhancer activity from spurious activity caused by positional effects (i.e. the impact of neighbouring regulatory regions at the points of genomic insertion). The VISTA enhancer browser collects the results for thousands of enhancers tested with the same PNI-based method (Visel et al., 2007). In the VISTA method, at least five transgenic mice were generated for each enhancer, carrying the sequence at different genomic positions. The staining pattern was analysed in every animal and endogenous activity was defined for staining that is observed in the same tissue in at least three animals (Visel et al., 2007). Both enhancers drive expression in two or more distinct brain regions. The finding that enhancers are active in multiple tissues is in line with data from previously reported enhancers (Visel et al., 2007; Bonn et al., 2012). As both enhancers have been shown to physically contact the transcriptionally active *FOXP2* promoter in human cell lines (Chapter 3) and drive gene expression in reporter assays, they could contribute to *FOXP2* expression in the detected brain regions.

Enhancer 37 activity is present in the OB, which has been reported to express *FOXP2*. The enhancer activity was detected in the outer layers of the OB, but did not overlap with a defined anatomical region. *FOXP2* has been previously found in glomerular and granular cells of the OB (Campbell et al., 2009; Zegary, 2009), of which the glomerular cells form the external neuronal layer. Therefore, enhancer 37 activity and *FOXP2* expression likely overlap in this layer. In the glomerular cells, FOXP2 expression marks one of three defined subtypes (Zegary, 2009). Thus, enhancer 37 may contribute to the expression of *FOXP2* in the OB. Co-localization of enhancer 37 activity with *Foxp2* expression in the transgenic mice is necessary to determine the exact overlap of targeted cell types.

Enhancer 37 was also active in the hippocampus and ventricles, which do not show *FOXP2* expression according to prior research. Enhancers may regulate more than one gene. Thus, enhancer 37 could regulate another gene that is expressed in these structures (Bonn et al., 2012). Enhancer contacts are dynamic between different cell types, but limited to genes within a topological domain (Dixon et al., 2012). Therefore, genes of the same topological domain as *FOXP2* may be targeted by enhancer 37. Next to *FOXP2,* the surrounding topological domain contains the *MDFIC*, *TFEC* and *PPP1R3A* genes (Dixon et al., 2012). In neuronal-like cell lines the enhancer interacted with the *FOXP2* promoter (Chapter3). The promoter was in contact with additional genomic regions, including the promoter of the *MDFIC* gene. It is therefore possible that enhancer 37 interacts with the *MDFIC* promoter and regulates the expression of this gene. *MDFIC* encodes the MyoD family inhibitor domain-containing protein, which regulates the WNT and JNK pathways (Kusano and Raab-Traub, 2002). The Allen Brain Atlas suggests strong expression of MDFIC in the CA1 and dentate gyrus, which matches the observed enhancer activity in the hippocampus (Lein et al., 2007). To further assess the hypothesis that enhancer 37 regulates *MDFIC*, it would be required to determine the direct interaction of enhancer 37 to the *MDFIC* promoter. In addition, co-localization of enhancer 37 with *Mdfic* and *Foxp2* in transgenic mice could further indicate the target gene of this enhancer in the hippocampus.

The tissue, in which enhancer 37 is active has been linked to adult neurogenesis in prior studies (Eriksson et al., 1998; Jessberger and Gage, 2014). Adult neurogenesis occurs in the hippocampus and is involved in memory formation in mammals and humans (Eriksson et al., 1998; Jessberger and Gage, 2014; Ernst and Frisen, 2015). A second site of adult neurogenesis is the subventricular zone (SVZ) of the ventricles. In rodents, the new-born neurons migrate from the SVZ towards the OB, where they differentiate into interneurons and integrate in the existing cellular architecture (Lois et al., 1996; Bergmann et al., 2015). To demonstrate a link between enhancer 37 and neurogenesis it would be necessary to co-label enhancer 37 reporter staining with markers of neurogenesis (Young et al., 2007). Mouse *Foxp2*

has been implicated in embryonic neurogenesis and contributes to the transformation from neural progenitor to differentiated neurons (Tsui et al., 2013). Targets of *FOXP2* in human neuronal-like cells and mouse embryonic brain are enriched for genes involved in neurogenesis (Spiteri et al., 2007; Vernes et al., 2007; Vernes et al., 2011). In adult zebra finch, *FoxP2* expression has been detected in new-born neurons, originating from the SVZ (Rochefort et al., 2007). In zebra finches the SVZ born neurons migrated to and integrated into the striatal nucleus area X. *FOXP2* expression was not observed in the SVZ itself, suggesting that *FOXP2* expression is activated in post-migratory neurons (Rochefort et al., 2007). However, it remains to be shown whether or not *FOXP2* is involved in adult neurogenesis in mammals. Research shows that, in humans, SVZ born neurons migrate and integrate into the striatum (Ernst et al., 2014; Ernst and Frisen, 2015), suggesting that the neurons from the rostral migratory stream could migrate towards the striatum. Enhancer 37 shows low evolutionary conservation across primates and is absent from non-primate species (Figure 3A), suggesting that the function of this enhancer was obtained during more recent evolutionary events. Species-specific genetic variation in enhancers has been shown to contribute to human-specific neuronal changes (Boyd et al., 2015). However, from the presented data it remains speculative whether or not this enhancer could be linked to the human specific traits, such as a changed migratory pathway of new-born neurons. Migration of SVZ neurons to the striatum can be induced in mice after blocking the Notch-signalling pathway (Magnusson et al., 2014). Inducing this pathway in enhancer transgenic mice could be used to assess whether enhancer 37 is active in SVZ-to-striatal migratory neurons.

The other investigated enhancer, enhancer 330, was previously shown to drive forebrain expression at embryonic day E11.5 (Visel et al., 2007). In other studies, *Foxp2* mRNA expression in the ventricular zone and cortical plate has been detected at E14.5 (Ferland et al., 2003). Indeed, after knock-out of the *Dicer* gene, which encodes the enzyme responsible for processing immature micro RNAs (miRNA), cortical expression of *Foxp2* can be detected at E13.5 (Clovis et al., 2012). In the early post-natal cortex, I detected activity of enhancer 330

in all cortical layers, at a time when *Foxp2* is expression is mostly restricted to the deeper cortical layer VI (Ferland et al., 2003). In the post-natal brain the activity of enhancer 330 overlaps *FOXP2* expression in the deeper cortical layers. Thus, enhancer330 activity and *FOXP2* expression could overlap in the embryonic and post-natal cortex suggesting that this enhancer activates *FOXP2* expression during cortical development. Enhancer 330 activity was specific to the six-layered neocortex and absent in the three-layered paleocortex. This is in agreement with the absence of *FOXP2* expression in the paleocortex (Ferland et al., 2003).

In adult brain tissue, enhancer 330 is active in the upper cortical layers, whereas *FOXP2* is expressed specifically in the deeper cortical layer VI (Ferland et al., 2003; Hisaoka et al., 2010). The difference in enhancer 330 activity and *FOXP2* expression may be explained by post-transcriptional and/or post-translational regulation. At the post-transcriptional level *FOXP2* mRNA has been shown to be a target of miRNAs (Clovis et al., 2012; Fu et al., 2014). The miRNAs miR-9 and miR-132 repress *FOXP2* expression specifically in the upper cortical layers (Clovis et al., 2012). In addition, the FOXP2 protein is post-translationally modified and may be targeted for protein degradation. For example, the post-translational addition of a SUMO protein to *FOXP2* has been shown, but this modification has not been shown to target *FOXP2* for degradation (Meredith et al., 2015; Estruch et al., 2016). The LacZ reporter gene does not contain the FOXP2 amino-acid sequences required for protein degradation or 3'-UTR for miRNA-mediated silencing. Therefore, it is likely that staining positive for the reporter gene is observed in cells which endogenously reduce FOXP2 protein levels via post-transcriptional and post-translational mechanisms.

People with mutations of *FOXP2* show structural and functional abnormalities in some cortical areas, such as the classical language associated Wenicke's and Broca's area, the angular gyrus and the precentral gyrus (Vargha-Khadem et al., 1998; Watkins et al., 2002; Belton et al., 2003; Liegeois et al., 2003; Liegeois et al., 2011). Genetic variation of enhancer 330 could in principle impact on enhancer function and affect the cortical expression of *FOXP2.* To my best knowledge, no case has been described with a selective deletion or mutation of this

element and high-throughput sequencing studies of clinical cases commonly focus on genetic variants within the exonic part of the genome and would not detect variation in the enhancer (Rabbani et al., 2014). The common single nucleotide polymorphism (SNP) rs12533005, which is located within 1kb of the upstream flanking sequence (Figure 3A), showed association to human traits in candidate association studies. One study investigated 12 SNPs and detected an association of this SNP to ADHD in a German cohort, which did not replicate in a Spanish cohort (Ribases et al., 2012). Another study investigated the association of 9 SNPs with dyslexia and detected a nominal association to rs12533005 (Wilcke et al., 2012). Wilcke et al. further conducted functional magnetic resonance imaging in 25 individuals carrying the risk-conferring minor allele and 8 non-carrier individuals. They detected that carriers of the minor allele showed reduced activity in the angular gyrus and supramarginal gyri during phonological processing tasks. Thus, enhancer 330 activity and genetic effects of rs12533005 coincide at cortical areas that show anomalies in patients with *FOXP2* mutations. This could suggest that rs12533005 influences the activity of enhancer 330 or tags another functional variant within this enhancer. A third study detected an association of rs12533005 (out of 6 tested variants) to increased left-hemispheric speech perception in healthy individuals (Ocklenburg et al., 2013), further indicating that this SNP may affect cortical functioning. However, the link between this SNP and enhancer 330 requires additional evidence. For example, genotyping of additional SNPs within this enhancer may show that rs12533005 is a proxy of an association to a variant within enhancer 330. Alternatively, a longer version of enhancer 330 could be investigated in reporter gene studies to determine if the sequence at rs12533005 contributes to enhancer function. In addition, the screening of enhancer 330 for mutations in cohorts of people with speech/language disorders is a promising approach to detect pathological variation.

FOXP2 has been recognized as a marker gene for cortical differentiation of cultured neuronal stem cells (Bickenbach et al., 2013; Espuny-Camacho et al., 2013; Belinsky et al., 2014; Raitano et al., 2015). During the *in vitro* differentiation of induced pluripotent stem cells (iPSC)

to cortical neurons, endogenous *FOXP2* expression is switched on (Espuny-Camacho et al., 2013; Belinsky et al., 2014). The activity of the *FOXP2* enhancer 330 in developing cortex could suggest that this enhancer would be activated in the cultured neurons. Chromatin conformation (Chapter 3) and reporter gene experiments (Chapter 4) in cultured stem cells would help to test this hypothesis. Human iPSCs and derived neurons are promising systems to study the neurodevelopmental function of enhancer 330 in human genetic background.

I also detected enhancer 330 activity in the adult cerebellum. The cerebellum develops after E11.5 and enhancer activity at this time may be undetected by the analysis performed for the VISTA enhancer browser. The cerebellar anlage is formed at E12.5 and the cerebellar hemisphere can be distinguished from other anatomical regions at embryonic day E14 (Millen et al., 1994; Fujita and Sugihara, 2012). *FOXP2* is expressed in the majority of neuronal nuclei of the cerebellar hemisphere at E14.5 (Fujita and Sugihara, 2012). The founder mouse and subsequent mice of the established mouse line, which were used to study the longitudinal activity of enhancer 330 did not show cerebellar staining. Therefore, it remains to be determined at what developmental stage *FOXP2* expression and enhancer 330 activity start in the cerebellar anlage/cerebellar hemisphere.

Several *FOXP2* positive brain structures did not show enhancer activity for either enhancer 37 or 330. I therefore expect that additional enhancers regulate the expression of *FOXP2*. In the previous chapters I identified enhancers that were not tested in this chapter and could be promising candidates for recapitulating further parts of the *FOXP2* expression pattern.

In summary, in this chapter I characterized the pattern of activity of two *FOXP2* enhancers in the brain and followed this activity during development. The enhancers recapitulated some parts of the complex *FOXP2* expression pattern. Enhancer 37 is active in the OB, a known region of *FOXP2* expression, and coincides with sites of adult neurogenesis. Enhancer 330 may be involved in modulating *FOXP2* expression in the cortex and PCs. In addition, the

proximity of enhancer 330 to rs12533005 may explain the observed associations of this SNP to language-related traits.

References

Becker M, Devanna P, Fisher SE, Vernes SC (2015) A chromosomal rearrangement in a child with severe speech and language disorder separates FOXP2 from a functional enhancer. Molecular cytogenetics 8:69.

Belinsky GS, Rich MT, Sirois CL, Short SM, Pedrosa E, Lachman HM, Antic SD (2014) Patch-clamp recordings and calcium imaging followed by single-cell PCR reveal the developmental profile of 13 genes in iPSC-derived human neurons. Stem cell research 12:101-118.

Belton E, Salmond CH, Watkins KE, Vargha-Khadem F, Gadian DG (2003) Bilateral brain abnormalities associated with dominantly inherited verbal and orofacial dyspraxia. Human brain mapping 18:194-200.

Bergmann O, Spalding KL, Frisen J (2015) Adult Neurogenesis in Humans. Cold Spring Harbor perspectives in biology 7.

Bickenbach JR, Tomanek-Chalkley A, Wiechert S, Winter MC (2013) Human skin keratinocytes can be reprogrammed to express neuronal genes and proteins after a single treatment with decitabine. BioResearch open access 2:217-221.

Bonn S, Zinzen RP, Girardot C, Gustafson EH, Perez-Gonzalez A, Delhomme N, Ghavi-Helm Y, Wilczynski B, Riddell A, Furlong EE (2012) Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat Genet 44:148-156.

Boyd JL, Skove SL, Rouanet JP, Pilaz LJ, Bepler T, Gordan R, Wray GA, Silver DL (2015) Human-Chimpanzee Differences in a FZD8 Enhancer Alter Cell-Cycle Dynamics in the Developing Neocortex. Current biology : CB.

Campbell P, Reep RL, Stoll ML, Ophir AG, Phelps SM (2009) Conservation and diversity of Foxp2 expression in muroid rodents: functional implications. The Journal of comparative neurology 512:84-100.

Clovis YM, Enard W, Marinaro F, Huttner WB, De Pietri Tonelli D (2012) Convergent repression of Foxp2 3'UTR by miR-9 and miR-132 in embryonic mouse neocortex: implications for radial migration of neurons. Development 139:3332-3342.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485:376-380.

Domene S, Bumaschny VF, de Souza FS, Franchini LF, Nasif S, Low MJ, Rubinstein M (2013) Enhancer turnover and conserved regulatory function in vertebrate evolution. Philosophical transactions of the Royal Society of London Series B, Biological sciences 368:20130027.

Eriksson PS, Perfilieva E, Bjork-Eriksson T, Alborn AM, Nordborg C, Peterson DA, Gage FH (1998) Neurogenesis in the adult human hippocampus. Nature medicine 4:1313-1317.

Ernst A, Frisen J (2015) Adult neurogenesis in humans- common and unique traits in mammals. PLoS biology 13:e1002045.

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

Espuny-Camacho I, Michelsen KA, Gall D, Linaro D, Hasche A, Bonnefont J, Bali C, Orduz D, Bilheu A, Herpoel A, Lambert N, Gaspard N, Peron S, Schiffmann SN, Giugliano M, Gaillard A, Vanderhaeghen P (2013) Pyramidal neurons derived from human pluripotent stem cells integrate efficiently into mouse brain circuits in vivo. Neuron 77:440-456.

Estruch SB, Graham SA, Deriziotis P, Fisher SE (2016) The language-related transcription factor FOXP2 is post-translationally modified with small ubiquitin-like modifiers. Sci Rep 6.

Ferland RJ, Cherry TJ, Preware PO, Morrisey EE, Walsh CA (2003) Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain. The Journal of comparative neurology 460:266-279.

Fu LJ, Shi ZM, Luo GZ, Tu WH, Wang XJ, Fang ZD, Li XC (2014) Multiple microRNAs regulate human FOXP2 gene expression by targeting sequences in its 3 ' untranslated region. Molecular brain 7.

Fujita E, Tanabe Y, Momoi MY, Momoi T (2012) Cntnap2 expression in the cerebellum of Foxp2(R552H) mice, with a mutation related to speech-language disorder. Neuroscience letters 506:277-280.

Fujita H, Sugihara I (2012) FoxP2 expression in the cerebellum and inferior olive: development of the transverse stripe-shaped expression pattern in the mouse cerebellar cortex. The Journal of comparative neurology 520:656-677.

Hawrylycz MJ et al. (2012) An anatomically comprehensive atlas of the adult human brain transcriptome. Nature 489:391-399.

Hisaoka T, Nakamura Y, Senba E, Morikawa Y (2010) The forkhead transcription factors, Foxp1 and Foxp2, identify different subpopulations of projection neurons in the mouse cerebral cortex. Neuroscience 166:551-563.

Ittner LM, Gotz J (2007) Pronuclear injection for the production of transgenic mice. Nature protocols 2:1206-1215.

Jessberger S, Gage FH (2014) Adult neurogenesis: bridging the gap between mice and humans. Trends Cell Biol 24:558-563.

Juers DH, Matthews BW, Huber RE (2012) LacZ beta-galactosidase: structure and function of an enzyme of historical and molecular biological importance. Protein science : a publication of the Protein Society 21:1792-1807.

Kothary R, Clapoff S, Darling S, Perry MD, Moran LA, Rossant J (1989) Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. Development 105:707-714.

Kusano S, Raab-Traub N (2002) I-mfa domain proteins interact with Axin and affect its regulation of the Wnt and c-Jun N-terminal kinase signaling pathways. Molecular and cellular biology 22:6393-6405.

Lai CS, Gerrelli D, Monaco AP, Fisher SE, Copp AJ (2003) FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. Brain : a journal of neurology 126:2455-2462.

Lein ES et al. (2007) Genome-wide atlas of gene expression in the adult mouse brain. Nature 445:168-176.

Liegeois F, Morgan AT, 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 15:283-288.

Liegeois F, Baldeweg T, Connelly A, Gadian DG, Mishkin M, Vargha-Khadem F (2003) Language fMRI abnormalities associated with FOXP2 gene mutation. Nat Neurosci 6:1230-1237.

Lois C, Garcia-Verdugo JM, Alvarez-Buylla A (1996) Chain migration of neuronal precursors. Science 271:978-981.

Magnusson JP, Goritz C, Tatarishvili J, Dias DO, Smith EM, Lindvall O, Kokaia Z, Frisen J (2014) A latent neurogenic program in astrocytes regulated by Notch signaling in the mouse. Science 346:237-241.

Meredith LJ, Wang CM, Nascimento L, Liu R, Wang L, Yang WH (2015) The Key Regulator for Language and Speech Development, FOXP2, is a Novel Substrate for SUMOylation. Journal of cellular biochemistry.

Millen KJ, Wurst W, Herrup K, Joyner AL (1994) Abnormal embryonic cerebellar development and patterning of postnatal foliation in two mouse Engrailed-2 mutants. Development 120:695-706.

Ocklenburg S, Arning L, Gerding WM, Epplen JT, Gunturkun O, Beste C (2013) FOXP2 variation modulates functional hemispheric asymmetries for speech perception. Brain Lang 126:279-284.

Rabbani B, Tekin M, Mahdieh N (2014) The promise of whole-exome sequencing in medical genetics. J Hum Genet 59:5-15.

Raitano S, Ordovas L, De Muynck L, Guo W, Espuny-Camacho I, Geraerts M, Khurana S, Vanuytsel K, Toth BI, Voets T, Vandenberghe R, Cathomen T, Van Den Bosch L, Vanderhaeghen P, Van Damme P, Verfaillie CM (2015) Restoration of progranulin expression rescues cortical neuron generation in an induced pluripotent stem cell model of frontotemporal dementia. Stem cell reports 4:16-24.

Ribases M, Sanchez-Mora C, Ramos-Quiroga JA, Bosch R, Gomez N, Nogueira M, Corrales M, Palomar G, Jacob CP, Gross-Lesch S, Kreiker S, Reif A, Lesch KP, Cormand B, Casas M, Bayes M (2012) An association study of sequence variants in the forkhead box P2 (FOXP2) gene and adulthood attention-deficit/hyperactivity disorder in two European samples. Psychiatric genetics 22:155-160.

Rochefort C, He X, Scotto-Lomassese S, Scharff C (2007) Recruitment of FoxP2-expressing neurons to area X varies during song development. Developmental neurobiology 67:809-817.

Rubinstein M, de Souza FS (2013) Evolution of transcriptional enhancers and animal diversity. Philosophical transactions of the Royal Society of London Series B, Biological sciences 368:20130017.

Shu W, Yang H, Zhang L, Lu MM, Morrisey EE (2001) Characterization of a new subfamily of winged-helix/forkhead (Fox) genes that are expressed in the lung and act as transcriptional repressors. J Biol Chem 276:27488-27497.

Spiteri E, Konopka G, Coppola G, Bomar J, Oldham M, Ou J, Vernes SC, Fisher SE, Ren B, Geschwind DH (2007) Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. Am J Hum Genet 81:1144-1157.

Tsui D, Vessey JP, Tomita H, Kaplan DR, Miller FD (2013) FoxP2 regulates neurogenesis during embryonic cortical development. The Journal of neuroscience : the official journal of the Society for Neuroscience 33:244-258.

Vargha-Khadem F, Watkins KE, Price CJ, Ashburner J, Alcock KJ, Connelly A, Frackowiak RS, Friston KJ, Pembrey ME, Mishkin M, Gadian DG, Passingham RE (1998) Neural basis of an inherited speech and language disorder. Proc Natl Acad Sci U S A 95:12695-12700.

Vernes SC, Spiteri E, Nicod J, Groszer M, Taylor JM, Davies KE, Geschwind DH, Fisher SE (2007) High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders. Am J Hum Genet 81:1232-1250.

Vernes SC, Oliver PL, Spiteri E, Lockstone HE, Puliyadi R, Taylor JM, Ho J, Mombereau C, Brewer A, Lowy E, Nicod J, Groszer M, Baban D, Sahgal N, Cazier JB, Ragoussis J, Davies KE, Geschwind DH, Fisher SE (2011) Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. PLoS Genet 7:e1002145.

Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, Park TJ, Deaville R, Erichsen JT, Jasinska AJ, Turner JM, Bertelsen MF, Murchison EP, Flicek P, Odom DT (2015) Enhancer evolution across 20 mammalian species. Cell 160:554-566.

Visel A, Minovitsky S, Dubchak I, Pennacchio LA (2007) VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res 35:D88-92.

Visel A, Prabhakar S, Akiyama JA, Shoukry M, Lewis KD, Holt A, Plajzer-Frick I, Afzal V, Rubin EM, Pennacchio LA (2008) Ultraconservation identifies a small subset of extremely constrained developmental enhancers. Nat Genet 40:158-160.

Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457:854-858.

Visel A et al. (2013) A high-resolution enhancer atlas of the developing telencephalon. Cell 152:895-908.

Watkins KE, Vargha-Khadem F, Ashburner J, Passingham RE, Connelly A, Friston KJ, Frackowiak RS, Mishkin M, Gadian DG (2002) MRI analysis of an inherited speech and language disorder: structural brain abnormalities. Brain : a journal of neurology 125:465-478.

Wilcke A, Ligges C, Burkhardt J, Alexander M, Wolf C, Quente E, Ahnert P, Hoffmann P, Becker A, Muller-Myhsok B, Cichon S, Boltze J, Kirsten H (2012) Imaging genetics of FOXP2 in dyslexia. European journal of human genetics : EJHG 20:224-229.

Young JI, Otero V, Cerdan MG, Falzone TL, Chan EC, Low MJ, Rubinstein M (1998) Authentic cell-specific and developmentally regulated expression of pro-opiomelanocortin genomic fragments in hypothalamic and hindbrain neurons of transgenic mice. The Journal of neuroscience : the official journal of the Society for Neuroscience 18:6631-6640.

Young KM, Fogarty M, Kessaris N, Richardson WD (2007) Subventricular zone stem cells are heterogeneous with respect to their embryonic origins and neurogenic fates in the adult olfactory bulb. The Journal of neuroscience : the official journal of the Society for Neuroscience 27:8286-8296.

Zegary JA (2009) Transcription Factor Regulation of Olfactory Bulb Interneuron Heterogeneity. Doctoral dissertation of the College of Medicine at the University of Cincinnati

Chapter 6

Early developmental gene enhancers affect subcortical volumes in the adult human brain

This chapter has been published in

Becker M, Guadalupe T, Franke B, Hibar DP, Renteria ME, Stein JL, Thompson PM, Francks C, Vernes SC, Fisher SE (2016) Early developmental gene enhancers affect subcortical volumes in the adult human brain. *Human Brain Mapping* 37:1788-1800.

The supplementary information is available online on the website of the *Human Brain Mapping* journal.

Abstract

Genome-wide association screens (GWAS) aim to identify common genetic variants contributing to the phenotypic variability of complex traits, such as human height or brain morphology. The identified genetic variants are mostly within non-coding genomic regions and the biology of the genotype-phenotype association typically remains unclear. In this paper, we propose a complementary targeted strategy to reveal the genetic underpinnings of variability in subcortical brain volumes, by specifically selecting genomic loci that are experimentally validated forebrain enhancers, active in early embryonic development. We hypothesized that genetic variation within these enhancers may affect the development and ultimately the structure of subcortical brain regions in adults. We tested whether variants in forebrain enhancer regions showed an overall enrichment of association with volumetric variation in subcortical structures of >13,000 healthy adults. We observed significant enrichment of genomic loci that affect the volume of the hippocampus within forebrain enhancers (empirical $p = 0.0015$), a finding which robustly passed the adjusted threshold for testing of multiple brain phenotypes (cut-off of $p < 0.0083$ at an alpha of 0.05). In analyses of individual SNPs, we identified an association upstream of the *ID2* gene with rs7588305 and variation in hippocampal volume. This SNP-based association survived multiple-testing correction for the number of SNPs analyzed but not for the number of subcortical structures. Targeting known regulatory regions offers a way to understand the underlying biology that connects genotypes to phenotypes, particularly in the context of neuroimaging genetics. This biology-driven approach generates testable hypotheses regarding the functional biology of identified associations.

Introduction

Global research efforts are underway to determine how genomic variants contribute to variation in brain volume in the general population. Altered volumes of brain regions are associated with neurodegenerative (Jack et al., 2011) and neuropsychiatric disorders (Videbech and Ravnkilde, 2004; van Erp et al., 2015). Current GWAS studies connect common genetic variants known as single nucleotide polymorphisms (SNPs) with phenotypic variations in brain imaging data in cohorts of normal individuals (Bis et al., 2012; Stein et al., 2012; Cai et al., 2014; Guadalupe et al., 2014b). However, most SNPs identified by GWAS are within non-coding regions and their biological functions remain unclear. We employed a hypothesis-driven strategy involving pre-selection of SNPs within genomic regions of experimentally defined biological function. We found that the selected SNPs are significantly enriched for association with hippocampal volume. This approach can help to generate testable follow-up hypotheses about the underlying genotype-phenotype correlation.

It is challenging to predict and test the consequences of single nucleotide changes in genomic regions that do not encode protein-coding genes or regulatory RNA molecules (Visel et al., 2009a). Nevertheless, a vast amount of non-coding DNA is thought to play some functional role, for example by regulating gene expression as promoters or enhancers. These regions are bound by regulatory proteins such as transcription factors (TFs) to directly influence gene expression and single nucleotide changes within these regions can affect gene expression (Spitz and Furlong, 2012).

A complementary strategy to genome-wide association screening is to look specifically for association within genomic regions that are predicted to be functional; for example by targeting regions known to be bound by a particular TF or marked by a given epigenetic modification. The approach was adopted in a study focusing on the genomic distribution of cell-type specific regulatory regions in hematopoietic cells. These regions were enriched for the presence of genetic variants associated with hematological traits such as platelet count, while no enriched

association was detected for unrelated traits such as body-mass index (BMI) (Paul et al., 2013). Another study applied a similar approach to publicly available functional datasets and found enriched association across a range of human traits and disorders, including Crohn's disease and height (Pickrell, 2014). Thus, *a priori* knowledge of the functional characteristics of genomic regions may help in selecting subsets of genomic variants for association testing. To our knowledge, no prior study has systematically integrated information about functional brain enhancers with genetic data in human individuals. In the current study, we analyzed publicly available data from the ENIGMA (Enhancing Neuroimaging Genetics through Meta-analysis) consortium, a neuroimaging genetics initiative that combines GWAS statistics from 50 cohorts spread world-wide, with a total combined sample size of 30,717 individuals (Thompson et al., 2014). The ENIGMA consortium recently presented GWAS findings on seven subcortical volumes and intracranial volume (Hibar et al., 2015). Five genetic variants were significantly associated with the volumes of the hippocampus and putamen. These SNPs mapped within non-coding regions of the genome and their functional roles are largely unknown. We hypothesized that polymorphisms located within experimentally validated brain enhancers, with established impacts on early neural development may be enriched for association with volumetric changes in subcortical brain structures. A valuable source of enhancers is the VISTA browser, which reports the tissue-specificity of about two thousand potential enhancers. Genomic elements tested in VISTA were chosen for their conservation across species and/or enhancer-specific chromatin marks in brain tissue (Visel et al., 2008; Visel et al., 2009b). Their potential to reproducibly drive expression of a reporter gene was tested at an early developmental timepoint in the mouse brain (Visel et al., 2007). VISTA contains 309 enhancers that drive expression in the embryonic forebrain; the developmental origin of the subcortical regions analyzed in adults by the ENIGMA consortium. Here, we integrate ENIGMA meta-analysis data with information on VISTA forebrain enhancer regions to assess for genetic effects on normal variation in subcortical brain volumes, using gene-set and SNP-based approaches.

Methods and Materials

VISTA enhancer browser of experimentally validated enhancer fragments

The VISTA enhancer browser is a collection of experimentally validated non-coding fragments of the human and mouse genomes that exhibit enhancer activity (Visel et al., 2007). Potential enhancers were predicted based on ChIP-seq experiments or evolutionary conservation. The selected non-coding fragments were cloned in front of a minimal promoter and the LacZ reporter gene. The resulting constructs were injected into the pro-nucleus of mouse embryos and re-implanted into foster mothers. Developing embryos were stained for reporter gene expression at embryonic day 11.5 (E11.5). Non-coding fragments that drive reproducible reporter gene expression in multiple (≥5) embryos are defined as positive enhancers. If no reproducible staining was detected in seven individual embryos, the non-coding fragment was regarded as negative. Thus, the database represents a collection of functional enhancers and their tissue of activity in developing embryos.

ENIGMA meta-analyzed neuroimaging GWAS of subcortical brain volumes

The ENIGMA consortium conducts harmonized analyses that combine neuroimaging genetics data from over 50 cohorts world-wide to search for common gene variants that are associated with brain structure and function. ENIGMA2 describes the second project of this consortium, which tested associations of 8.5 million SNPs with the variance in volumes of the nucleus accumbens, amygdala, caudate nucleus, hippocampus, pallidum, putamen, thalamus, and the intracranial volume, measured from brain MRI (Hibar et al., 2015). The discovery sample for ENIGMA2 consisted of 13,171 subjects of European ancestry. The ENIGMA consortium also had access to a replication set of 17,546 subjects and reported on a total of 30,717 individuals. Subjects were scanned at the individual sites using brain MRI and the resulting images were processed in a standard way across all sites. The volumetric measures of all subcortical structures were corrected for age, sex, genetic homogeneity and intracranial volume. At each

site, genotypes were obtained from commercially available platforms and imputed to the reference panel 1000genomes v1.3 using scripts provided by the ENIGMA working group. GWAS was performed at each site, and the meta-analysis of all 50 cohorts was performed centrally using the software package METAL (Willer et al., 2010). Here, we use the meta-analyzed summary p-values of the 13,171 subjects obtained from the discovery sample. The remaining 17,000 subjects from ENIGMA2 lack meta-analyzed genome-wide statistics and were thus unavailable for use in the present study.

Selecting SNPs within human enhancers

A list of enhancers, including their genomic coordinates (human genome build hg19) and their tissue of activity, was downloaded from the VISTA enhancer browser (Version 19.Sept 2014) (Visel et al., 2007). We filtered all human enhancer elements with a reproducible forebrain activity at embryonic stage E11.5. The forebrain is the developmental origin of the subcortical regions analyzed by ENIGMA. We removed nine enhancer elements located on chromosome X, as this chromosome was not included in the ENIGMA2 data. Four enhancers (hs322, hs1354, hs998, and hs1597) contained within larger enhancers were deleted to obtain a non-redundant list of 296 loci with a mean size of 1936 bp (standard deviation of 941 bp). Enhancer boundaries are not well defined, and flanking regions can be critical for the stability of an enhancers activity (Ludwig et al., 2011). We intersected the genomic enhancer regions, including 500 bp flanking regions with the genomic positions of all markers present in the meta-analyzed GWAS files of ICV-adjusted subcortical volumes (nucleus accumbens, amygdala, caudate nucleus, hippocampus, pallidum, putamen, and thalamus) from ENIGMA2. From this, we derived a list of 2082 SNPs and indels, which represent non-coding genetic variants with a potential impact on enhancer activity. We refer to these variants as *enhancer SNPs*.

Testing enrichment of association with subcortical volumes

ENIGMA2 tested for the association of 8.5 million genome-wide markers against volumes of 7 subcortical brain structures. 2,082 of these markers were located within forebrain enhancer regions. For each of these markers we converted their association p-values into z-scores. Note that we could not specify a priori whether a particular allele of a SNP within an enhancer should lead to an increase or decrease in the activity of the relevant enhancer. Moreover, increases in gene expression may be correlated with either an increase or decrease in volume of a structure, depending on the nature of the gene product; both directions of effect are possible in principle, given what is known about mechanisms by which genes influence brain morphology. Thus, because we did not have prior hypotheses regarding direction of effect, we converted the p-values to right-tailed z-scores. These z-scores were then summed to calculate a "forebrain enhancer" test statistic. This was repeated for each of the tested brain phenotypes.

The probability of observing a similar (or greater) test statistic was determined by sampling 2,082 markers from 296 random genomic regions from the rest of the ENIGMA2 results. For each of the 296 forebrain enhancer regions, a randomly sampled genomic region of similar length was selected (with a difference no greater than 1000 bp), which included at least the same number of markers as the original one. We chose the generated regions to be of similar size to avoid repetitive parts of the genome, such as telomeres or centromeres. We note that since the ENIGMA results are obtained via a large-scale meta-analysis of GWAS statistics from many different populations, we are unable to determine the precise LD structure that would most appropriately match the enrichment analyses. Thus, our permutation approach assumes that the LD structure of the enhancer regions is broadly characteristic of the rest of the genome. From this set of new selected regions, 2,082 markers were used to calculate the new randomly-generated set statistic. This was repeated ten thousand times. The p-value for the enrichment is the result of dividing the number of randomly-generated test statistics that are higher than the observed test statistic by the number of permutations (10,000). The seven

phenotypes are correlated, and we determined that the effective number of independent phenotypes is six, using the Matrix Spectral Decomposition, matSpD software. We accounted for multiple testing by Bonferroni correction for this effective number of independent phenotypes. Thus, at an alpha of 0.05, the p-value threshold for significant enrichment was 0.0083.

Testing enrichment of association with unrelated human traits as a control

As noted above, our permutation approach for assessing enrichment assumes that enhancer regions do not deviate substantially in LD structure, as compared to the rest of the genome. Therefore, we performed additional control experiments to discount the possibility that evidence of enrichment might be an artefact of unusual LD patterns. We downloaded the meta-analyzed GWAS summary statistics for BMI (Speliotes et al., 2010), height (Lango Allen et al., 2010) and waist-to-hip ratio (WHR) (Heid et al., 2010) from the Genetic Investigation of Anthropometric Traits (GIANT) consortium. We used the Galaxy web-based platform (Giardine et al., 2005; Blankenberg et al., 2010; Goecks et al., 2010) to add the SNP's genomic locations (hg19) and carried out enrichment analysis as described above. Based on Bonferroni correction for use of three control phenotypes, at an alpha of 0.05 the p-value threshold for significant enrichment in these analyses was 0.017.

Association of individual enhancer SNPs with subcortical volumes

SNPs within the same enhancer are in close proximity, and often in linkage disequilibrium (LD), so not independent from each other. We therefore wanted to determine the effective number of independent SNPs in the SNP-based association tests. Since the original genotypes for all subjects were not centrally collected by ENIGMA they were unavailable for the present study. Therefore, we calculated the effective number of tested SNPs from a large

subset of ENIGMA for which genotypes were available to us: the Brain Imaging Genetics (BIG) cohort of the Cognomics initiative (Guadalupe et al., 2014a). This cohort contributed approximately 10% of the subjects to the ENIGMA2 discovery sample and consists of healthy subjects of European ancestry. Using the Genetic Type I Error Calculator (Li et al., 2012), we determined that the number of effectively independent tests performed when assessing associations for the 2,082 enhancer SNPs was 770.

We evaluated the evidence for association between enhancer SNPs (spread across 296 enhancers) and variation in the seven subcortical volumes analyzed in ENIGMA2. As noted above, analyses with matSpD (Nyholt, 2004) indicated that the effective number of independent phenotypes is six. We derived an appropriate significance threshold based on Bonferroni correction, taking into consideration both the effective number of tested SNPs and the effective number of independent phenotypes. At an alpha of 0.05, the significance threshold for a SNP was 1.08×10^{-5}.

Analysis of eQTL effects in adult brain samples

We used the BRAINEAC eQTL database (Ramasamy et al., 2014) to investigate if rs7588305 is associated with expression changes in adult human hippocampal samples. Using the webtool we obtained eQTL association p-values for the queried SNP to expression probes of surrounding genes in different brain areas. We corrected the p-values for the number of probes that the rs7588305 SNP was correlated with to determine if the SNP represents an eQTL for one of the surrounding genes.

Predicting transcription factor binding events

To predict the effect of rs7588305 alleles on TF binding events, we performed a motif analysis using the MATCH algorithm, set to minimize false positives (Kel et al., 2003). As the source

for position weight matrices we used the commercially available Transfac2014.4 motif library. We predicted the binding events for a 51-nucleotide long genomic fragment centered on rs7588305, which was downloaded from the UCSC Genome Browser human genome version 19 (hg19). We predicted binding events for both the major and minor alleles.

Results

Developmental brain enhancers are significantly enriched for association with hippocampal volume

The *a priori* selection of functional genomic regions (Figure 1, Supporting table 1) allowed us to test for enriched genetic association among all enhancer elements as a group. We hypothesized that the group of enhancers contains a significant number of SNPs with sub-threshold associations, which we could detect by comparing the enhancer regions to random genomic regions. Using a permutation approach, we detected significantly enriched association of the enhancer region set with hippocampal volume (empirical $p = 0.0015$, passing our predesignated threshold of $p < 0.0083$ based on Bonferroni correction for multiple brain phenotypes). Additionally, we detected nominally significant enrichment in the nucleus accumbens, putamen, and thalamus ($p = 0.016$, $p=0.048$, $p=0.016$, respectively) that did not pass multiple testing correction for the number of subcortical volumes that were analyzed (Table 1). This result supports the hypothesis that experimentally validated forebrain enhancer regions from early developmental timepoints contain a significant number of genetic variants associated with the size of adult subcortical brain structures.

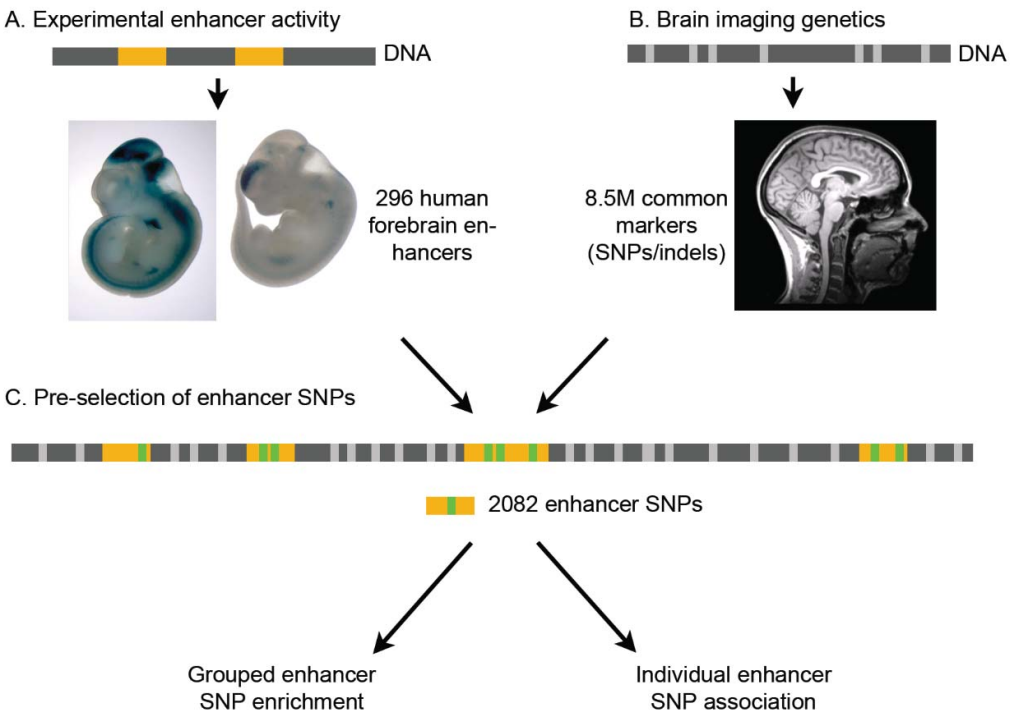

Figure 1: Overview of enhancer SNP association study. In this study we used two independent datasets. (A) Genomic regions of forebrain enhancers as determined in transgenic E11.5 mouse embryos and (B) association of common genetic variants with the volumetric variation of subcortical brain structures, determined by MRI. (C) Preselecting SNPs within enhancer regions allowed us to investigate associations that do not pass genome-wide significance. We used two complementary strategies testing for enriched occurrence of association signals within the group of enhancer regions, and looking for association of individual enhancer SNPs.

Table 1: Enrichment of enhancer SNP association

Brain structure	Enrichment p-value
Accumbens	0.016
Amygdala	0.068
Caudate	0.076
Hippocampus	0.0015
Pallidum	0.075
Putamen	0.048
Thalamus	0.016

Within each permutation the SNPs were randomly generated, but had to be located in genomic regions of similar SNP density to the enhancer regions. Genomic regions with low SNP density may have a different overall LD structure from that of the enhancer regions and may skew the

observed enrichment. Indeed, if we did not limit the genomic distribution of the randomly generated SNPs, we detect a more significant enrichment for the forebrain enhancer regions within the hippocampus ($p < 0.0006$).

Our hypothesis was that specifically forebrain enhancers are enriched for association with subcortical volumes. To assess if the detected enrichment could instead be due to a more general enrichment in enhancer regions, we generated a list of VISTA enhancers that are active in tissues other than the brain. We obtained a comparable number of non-brain enhancers (320 enhancers, Supporting table 2). We calculated the enrichment of association with hippocampal volume using the same approach as for the forebrain enhancer set and found that the control enhancer regions were not significantly enriched for association with hippocampal volume ($p = 0.95$). These findings suggest that the enrichment we detected within forebrain enhancers reflect functional links between these enhancer sequences and their tissue of activity.

Enhancers may function in more than one tissue and many of the forebrain enhancers are also active in other parts of the developing embryo, such as the neural tube (79/296 enhancer regions), limbs (33/296), eye (19/296) or heart (5/296) (Supporting table 3). However, we wanted to determine if the forebrain enhancers are specifically enriched for neuronal traits and exclude that the enrichment is caused by factors other than the biological link. We downloaded GWAS summary statistics for BMI, height and WHR from the GIANT consortium and repeated the enrichment analysis. We detected no enrichment for BMI, height or WHR ($p = 0.06$, $p = 0.99$, $p = 0.97$ respectively, none approaching predesignated threshold of $p < 0.017$ for control phenotypes). This finding supports the view that the significant enrichment of forebrain enhancers with hippocampal volume reflects specific effects, rather than being an artefact e.g. of the particular LD structure of the regions being studied.

Suggestive association of an enhancer SNP upstream of ID2 with variation in hippocampal volume

In parallel we analyzed the individual enhancer SNPs for association with the seven subcortical volumes. After adjusting our significance threshold for the effective number of tested SNPs and the effective number of independent phenotypes, no individual SNP met the strictest significant threshold (Table 2, Supporting table 4). Nonetheless, the strongest evidence of association was between hippocampus volume and SNP rs7588305, which had an unadjusted p-value of 2.9×10^{-5} (Figure 2a, Table 2). This association remained significant after correcting for the effective number of tested SNPs (p = 0.022 after Bonferroni correction), but not after further correction for the number of subcortical brain regions investigated (adjusted p = 0.13). Rs7588305 has a minor allele frequency (MAF) of 0.45. The SNP is located in the enhancer hs1527, which according to the VISTA enhancer browser is active in the forebrain, hindbrain, midbrain, and neural tube (Figure 2b). In the genome this enhancer is approximately 38 kb upstream of the closest gene *ID2* (Figure 2d), which is involved in transcriptional regulation (Ferrer-Vicens et al., 2014). *ID2* expression overlaps hs1527 enhancer activity in the forebrain and midbrain at the same embryonic period (Figure 2c). However, we could not find supportive expression quantitative trait loci (eQTL) association between rs7588305 and *ID2* in a dataset of human adult hippocampal expression data (Ramasamy et al., 2014).

Table 2: Most significant individual SNP associations in each subcortical brain structure

Brain structure	Enhancer ID	Enhancer SNP	Allele 1 [A1]	Allele 2 [A2]	P-value	Effect of A2 [mm^3/allele]	StdErr effect
Accumbens	hs1578	rs139118507	A	C	5.65E-04	-18.16	5.27
Amygdala	hs1300	rs149826134	A	G	9.50E-04	-31.75	9.61
Caudate	hs1362	rs17710617	T	C	8.79E-05	20.66	5.27
Hippocampus	hs1527	rs7588305	C	G	2.88E-05	-21.73	5.20
Pallidum	hs1636	rs36058915	-	T	1.18E-04	7.73	2.01
Putamen	hs1300	rs149826134	A	G	1.33E-04	-88.06	23.04
Thalamus	hs200	rs34415491	A	-	1.29E-04	-26.76	6.99

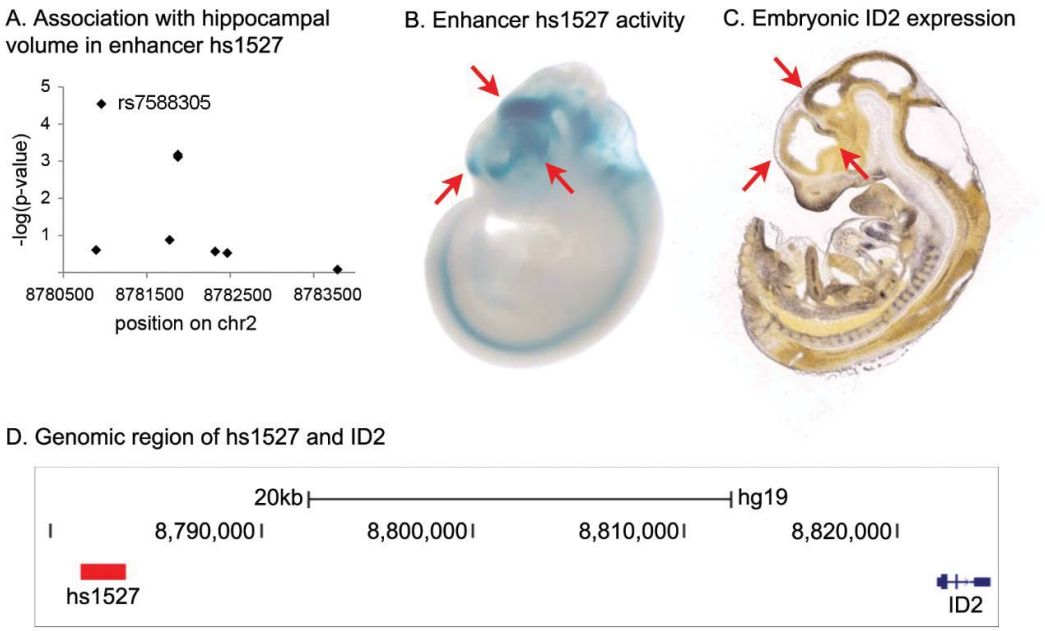

Figure 2: The enhancer SNP rs7588305 is associated with reduced hippocampal volume and is located upstream of the *ID2* gene. Plot (A) shows association *P*-values with hippocampal volume. Plotted are all enhancer SNPs within hs1527 according to their position on human chromosome 2. The rs7588305 SNP is at position 8,780,959 and has a *P*-value of 2.88 3 1025. (B) Representative transgenic mouse embryo injected with the hs1527 construct. Picture downloaded from the VISTA enhancer database (http://enhancer.lbl.gov/cgi-bin/imagedb3.pl?form=presentation&show=1&experiment_id=1527&organism_id=1) (Visel et al., 2007). The hs1527 enhancer activity is detected at embryonic stage E11.5 in parts of the forebrain, midbrain and neural tube as shown by the blue reporter gene staining. (C) *ID2* gene expression in a mouse embryo at stage E11.5, detected with in situ hybridization. Picture downloaded from the Allen Developing Mouse Brain Atlas (http://developing mouse.brain-map.org/experiment/siv?id=100072835&imageId=101217794&initImage=ish, Website: VC 2015 Allen Institute for Brain Science.). *ID2* expression overlaps hs1527 enhancer activity in the forebrain and midbrain (red arrows in b and c). (D) UCSC genome browser (https://genome.ucsc.edu) showing the genomic location of the hs1527 enhancer upstream of the *ID2* gene.

Using the prior knowledge that rs7588305 is located within a forebrain enhancer, we identified a suggestive association of this SNP with hippocampal volume. If this SNP itself (rather than others in LD) is mediating this association, then complementary data should support a functional role. For example, the alternative allele of this SNP might affect the activity of the surrounding enhancer caused by differential binding of TFs. In order to assess such possibilities, we used *in silico* analyses to test if the change from major to minor allele would

affect known TF binding sites. The differential motif analysis of the surrounding genomic fragment predicted a gain of six new motifs and a loss of five motifs for the minor allele (Supporting table 5). Among these predicted differences we found two ubiquitously expressed general TFs that are involved in chromatin structure and looping, CTCF and p300 (Supporting figure 1). The minor allele of rs7588305 gained a binding motif for CTCF - a general TF that stabilizes chromatin loops and is generally found in structural chromatin loops that either block enhancer promoter interactions or are active enhancer loops (Ong and Corces, 2014). The second change for the minor allele was the loss of a p300-binding motif. P300 is a transcriptional coactivator that is found at active enhancer-promoter loops and increases transcription of the target gene (Chan and La Thangue, 2001). The motif analysis of the rs7588305 SNP is thus consistent with the view that the allelic state of this SNP could be associated with changed activity of the enhancer. The variant could potentially decrease the activity of the enhancer, given that the binding site for the co-activator p300 is lost. However, CTCF function is highly dependent on context, and the effect of a gained CTCF binding site is difficult to predict. Thus, to determine the direction of effect of the minor allele would require future experiments, for example using cellular models.

Discussion

The advent of GWAS has allowed for the high-throughput characterization of genetic variants in large cohorts of human subjects. In a comprehensive resource of SNP-trait association data from multiple GWAS studies, collated by the National Institutes of Health (NIH), more than 90% of SNP associations map to non-coding regions (Welter et al., 2014). A substantial number of these SNPs are in genomic regions with regulatory functions, such as enhancers. Thus, pre-selecting functionally tested enhancers may help to detect biological links between genetic variants and phenotypic traits. Here we investigated the association of common variation in well-defined developmental forebrain enhancers with volumetric measures of subcortical regions in a meta-analyzed study of 13,000 adult subjects. For hippocampus volume, we found a significant overall enrichment of association within enhancer regions that was robust to multiple-testing, as well as suggestive individual association of one enhancer SNP. Therefore, this study shows how functional data can help to detect trait-associated genetic variants. Our approach also generates testable hypotheses relating non-coding SNPs to an observed trait of interest.

High-throughput studies have documented the functions of non-coding DNA across a range of tissues and cell lines by characterizing genome-wide distribution of chromatin modifications and/or TF interactions (ENCODE, 2012; Barrett et al., 2013). The resulting public databases allow predictions to be made regarding the regulatory activity of a genomic fragment. They thus provide useful tools to determine the functionality of trait-associated non-coding SNPs, and how they relate to the phenotype of interest (Boyle et al., 2012; Ward and Kellis, 2012). Crucially, in most GWAS efforts these data are only exploited in a post-hoc manner, to assess SNP associations that meet genome-wide significance. Our method is unbiased with respect to the functional regions and our results therefore include sub-threshold associations of SNPs with smaller effect sizes.

Targeted studies of single genomic regions indicate that variants in non-coding DNA are related to variation in normal brain volume and risk for psychiatric illness. For example, a rare variant implicated in autism was found in an enhancer that is active during forebrain development. The variant that may increase risk for autism alters binding of regulatory TFs and reduces enhancer activity in the developing forebrain (Poitras et al., 2010). Another study focused on an enhancer that underwent accelerated evolution on the human lineage, after splitting from that of chimpanzees. In mouse models, the human version of this enhancer prolongs forebrain growth and increases the size of the developing neocortex (Boyd et al., 2015). This supports our finding that human gene enhancers help explain how non-coding SNPs contribute to variation in brain volume.

We identified an enhancer SNP that has a within-phenotype significant association with hippocampal volume. The SNP is contained within the enhancer hs1527 and positioned upstream of the inhibitor of DNA binding 2 (*ID2*) gene. *ID2* is expressed in the embryonic mouse forebrain and in regions of adult rat brain, including the hippocampus (Kitajima et al., 2006). The expression pattern of *ID2* overlaps with the regions of hs1527 enhancer activity (Figure 2b-c). *ID2* enhances cell proliferation (Iavarone et al., 1994) and promotes axon growth in primary rat neurons (Lasorella et al., 2006). In human neuronal cell lines, *ID2* is directly regulated by MECP2. Disruptions of the *MECP2* gene cause Rett syndrome, a neurodevelopmental disorder characterized by severe intellectual disability and microcephaly (Amir et al., 1999). Accordingly, *ID2* expression was higher in post-mortem brain samples from subjects with Rett Syndrome (Peddada et al., 2006). The absence of association between rs7588305 and *ID2* expression in adult human hippocampal samples, as measured by the Braineac eQTL database, does not discount an effect on enhancer activity in embryonic tissue. Differential *ID2* expression during embryonic development may be sufficient to lead to volumetric changes in the adult, which agrees with the role of the gene in cell-proliferation and neurogenesis. Future studies of *ID2* function in neurodevelopment may link this gene to hippocampal development.

A benefit of pre-selecting genetic variants based on functional information is that specific testable hypotheses are generated for follow-up studies. For example, motif analysis using the major and minor alleles of rs7588305 predicted differential binding of CTCF and p300, two ubiquitously expressed proteins involved in chromatin loop formation and transcriptional regulation. The minor allele lacks the motif for p300, a general mediator for enhancer activity, and gains a motif for CTCF, which may affect enhancer loop formation. The shifted binding of these two general chromatin associated factors could conceivably alter the activity of the hs1527 enhancer, a hypothesis that can be tested with future experiments in cellular models, which would also enable the direction of effect to be determined.

Here, we determined if a set of functionally validated enhancers, active during early brain development, was enriched for association with volumetric differences in subcortical brain structures in human adults. We found significant enrichment with hippocampal volume and nominal significance for the nucleus accumbens, pallidum, and thalamus. The selected forebrain enhancers contained an increased number of common genetic variants with sub-threshold associations. The observed enrichment in the hippocampus was found with forebrain enhancers, but was not detectable for non-brain enhancers. In order to exclude that the effects reflected unusual LD structure of the forebrain enhancers, or some other artefact, we performed additional control analyses of other human traits, including BMI, height and WHR, and did not observe enrichment. This finding implies that early developmental enhancer activity has an impact on adult hippocampal volume. During the human lifespan more than 80% of all genes (≥ 20,000) are expressed in the developing and adult human brain (Kang et al., 2011), and each active gene promoter is in contact with several enhancers (Sanyal et al., 2012; Jin et al., 2013). Tens of thousands of enhancers may guide the coordinated expression of genes during development. Indeed, more than 80,000 non-coding genomic regions show histone marks indicative of active enhancers in brain tissue (Vermunt et al., 2014). The VISTA enhancer database likely represents only a fraction of enhancers that act during human brain development, and genetic variants in unsampled enhancers may also contribute to subcortical

volumes. This could likely have limited our ability to detect enrichment beyond the hippocampus. Future testing of all existing brain enhancers may increase the power to identify enrichment of associations.

Another current limitation is that the VISTA database characterizes enhancer activity at E11.5 and anatomically annotates this activity only to the general forebrain region. An annotation with respect to specific adult subcortical areas is not possible at this early embryonic period. It is crucial to look at the most relevant tissue for association studies of enhancers, so more detailed anatomical annotations of enhancer activities at later time points would improve the approach. This may enable mapping of more direct connections from enhancer activity to adult neuroimaging data.

In this study we focused on experimentally validated enhancer elements, because these give a robust set of regions with clear functional links to defined human tissues. The method we have employed may be expanded to study tissue-specific histone modifications as they become more readily available (Vermunt et al., 2014; Roadmap Epigenomics et al., 2015). A benefit of these datasets may be their more fine-grained anatomical description and sampling across several developmental time points.

In summary, we demonstrated genotype-phenotype connections between early neurodevelopmental enhancers and the volumes of adult subcortical brain regions in a dataset of over 13,000 people. As with all genetic association studies, replication of these findings should be sought in independent samples of a similar size where matched genome-wide genotyping and brain measures are available. Comparable analyses of enhancer SNPs in cohorts of patients with neurodevelopmental disorders has the potential to link early embryonic enhancer activity to disease processes. Our study illustrates a novel method for integrating functional genomic and phenotypic data to identify the biological underpinnings of highly heterogeneous traits, including neuropsychological traits. Our method can be applied to other

sets of functional non-coding genomic elements, to connect the function of the regions to studied phenotypes of interest.

References

Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet 23:185-188.

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A (2013) NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res 41:D991-995.

Bis JC et al. (2012) Common variants at 12q14 and 12q24 are associated with hippocampal volume. Nat Genet 44:545-551.

Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J (2010) Galaxy: a web-based genome analysis tool for experimentalists. Current protocols in molecular biology / edited by Frederick M Ausubel [et al] Chapter 19:Unit 19 10 11-21.

Boyd JL, Skove SL, Rouanet JP, Pilaz LJ, Bepler T, Gordan R, Wray GA, Silver DL (2015) Human-Chimpanzee Differences in a FZD8 Enhancer Alter Cell-Cycle Dynamics in the Developing Neocortex. Current biology : CB.

Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M (2012) Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 22:1790-1797.

Cai DC, Fonteijn H, Guadalupe T, Zwiers M, Wittfeld K, Teumer A, Hoogman M, Arias-Vasquez A, Yang Y, Buitelaar J, Fernandez G, Brunner HG, van Bokhoven H, Franke B, Hegenscheid K, Homuth G, Fisher SE, Grabe HJ, Francks C, Hagoort P (2014) A genome-wide search for quantitative trait loci affecting the cortical surface area and thickness of Heschl's gyrus. Genes Brain Behav 13:675-685.

Chan HM, La Thangue NB (2001) p300/CBP proteins: HATs for transcriptional bridges and scaffolds. J Cell Sci 114:2363-2373.

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

Ferrer-Vicens I, Riffo-Campos AL, Zaragoza R, Garcia C, Lopez-Rodas G, Vina JR, Torres L, Garcia-Trevijano ER (2014) In vivo genome-wide binding of Id2 to E2F4 target genes as part of a reversible program in mice liver. Cell Mol Life Sci 71:3583-3597.

Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, Miller W, Kent WJ, Nekrutenko A (2005) Galaxy: a platform for interactive large-scale genome analysis. Genome Res 15:1451-1455.

Goecks J, Nekrutenko A, Taylor J, Galaxy T (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol 11:R86.

Guadalupe T, Zwiers MP, Teumer A, Wittfeld K, Vasquez AA, Hoogman M, Hagoort P, Fernandez G, Buitelaar J, Hegenscheid K, Volzke H, Franke B, Fisher SE, Grabe HJ, Francks C (2014a) Measurement and genetics of human subcortical and hippocampal asymmetries in large datasets. Human brain mapping 35:3277-3289.

Guadalupe T, Zwiers MP, Wittfeld K, Teumer A, Vasquez AA, Hoogman M, Hagoort P, Fernandez G, Buitelaar J, van Bokhoven H, Hegenscheid K, Volzke H, Franke B, Fisher SE, Grabe HJ, Francks C (2014b) Asymmetry within and around the human planum temporale is sexually dimorphic and influenced by genes involved in steroid hormone receptor activity. Cortex.

Heid IM et al. (2010) Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat Genet 42:949-960.

Hibar DP et al. (2015) Common genetic variants influence human subcortical brain structures. Nature.

Iavarone A, Garg P, Lasorella A, Hsu J, Israel MA (1994) The helix-loop-helix protein Id-2 enhances cell proliferation and binds to the retinoblastoma protein. Genes Dev 8:1270-1284.

Jack CR, Jr. et al. (2011) Steps to standardization and validation of hippocampal volumetry as a biomarker in clinical trials and diagnostic criterion for Alzheimer's disease. Alzheimer's & dementia : the journal of the Alzheimer's Association 7:474-485 e474.

Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen CA, Schmitt AD, Espinoza CA, Ren B (2013) A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature 503:290-294.

Kang HJ et al. (2011) Spatio-temporal transcriptome of the human brain. Nature 478:483-489.

Kel AE, Gossling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E (2003) MATCH: A tool for searching transcription factor binding sites in DNA sequences. Nucleic Acids Res 31:3576-3579.

Kitajima K, Takahashi R, Yokota Y (2006) Localization of Id2 mRNA in the adult mouse brain. Brain Res 1073-1074:93-102.

Lango Allen H et al. (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467:832-838.

Lasorella A, Stegmuller J, Guardavaccaro D, Liu G, Carro MS, Rothschild G, de la Torre-Ubieta L, Pagano M, Bonni A, Iavarone A (2006) Degradation of Id2 by the anaphase-promoting complex couples cell cycle exit and axonal growth. Nature 442:471-474.

Li MX, Yeung JM, Cherny SS, Sham PC (2012) Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. Human genetics 131:747-756.

Ludwig MZ, Manu, Kittler R, White KP, Kreitman M (2011) Consequences of eukaryotic enhancer architecture for gene expression dynamics, development, and fitness. PLoS Genet 7:e1002364.

Nyholt DR (2004) A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. Am J Hum Genet 74:765-769.

Ong CT, Corces VG (2014) CTCF: an architectural protein bridging genome topology and function. Nat Rev Genet 15:234-246.

Paul DS, Albers CA, Rendon A, Voss K, Stephens J, van der Harst P, Chambers JC, Soranzo N, Ouwehand WH, Deloukas P (2013) Maps of open chromatin highlight cell type-restricted patterns of regulatory sequence variation at hematological trait loci. Genome Res 23:1130-1141.

Peddada S, Yasui DH, LaSalle JM (2006) Inhibitors of differentiation (ID1, ID2, ID3 and ID4) genes are neuronal targets of MeCP2 that are elevated in Rett syndrome. Human molecular genetics 15:2003-2014.

Pickrell JK (2014) Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. Am J Hum Genet 94:559-573.

Poitras L, Yu M, Lesage-Pelletier C, Macdonald RB, Gagne JP, Hatch G, Kelly I, Hamilton SP, Rubenstein JL, Poirier GG, Ekker M (2010) An SNP in an ultraconserved regulatory element affects Dlx5/Dlx6 regulation in the forebrain. Development 137:3089-3097.

Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, De T, Coin L, de Silva R, Cookson MR, Singleton AB, Hardy J, Ryten M, Weale ME (2014) Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat Neurosci 17:1418-1428.

Roadmap Epigenomics C et al. (2015) Integrative analysis of 111 reference human epigenomes. Nature 518:317-330.

Sanyal A, Lajoie BR, Jain G, Dekker J (2012) The long-range interaction landscape of gene promoters. Nature 489:109-113.

Speliotes EK et al. (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet 42:937-948.

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

Stein JL et al. (2012) Identification of common variants associated with human hippocampal and intracranial volumes. Nat Genet 44:552-561.

Thompson PM et al. (2014) The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. Brain Imaging Behav 8:153-182.

van Erp TG et al. (2015) Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. Molecular psychiatry.

Vermunt MW, Reinink P, Korving J, de Bruijn E, Creyghton PM, Basak O, Geeven G, Toonen PW, Lansu N, Meunier C, van Heesch S, Clevers H, de Laat W, Cuppen E, Creyghton MP (2014) Large-scale identification of coregulated enhancer networks in the adult human brain. Cell Rep 9:767-779.

Videbech P, Ravnkilde B (2004) Hippocampal volume and depression: a meta-analysis of MRI studies. Am J Psychiatry 161:1957-1966.

Visel A, Rubin EM, Pennacchio LA (2009a) Genomic views of distant-acting enhancers. Nature 461:199-205.

Visel A, Minovitsky S, Dubchak I, Pennacchio LA (2007) VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res 35:D88-92.

Visel A, Prabhakar S, Akiyama JA, Shoukry M, Lewis KD, Holt A, Plajzer-Frick I, Afzal V, Rubin EM, Pennacchio LA (2008) Ultraconservation identifies a small subset of extremely constrained developmental enhancers. Nat Genet 40:158-160.

Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA (2009b) ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457:854-858.

Ward LD, Kellis M (2012) HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res 40:D930-934.

Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, Parkinson H (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 42:D1001-1006.

Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26:2190-2191.

Chapter 7

Summary and

General discussion

Summary

DNA sequence variants affecting protein-coding genes may cause Mendelian disorders or increase the likelihood to develop a certain phenotype. *FOXP2* mutations cause a condition that is characterized by severe difficulties in performing the oral motor movements required for speech (developmental verbal dyspraxia), as well as reduced receptive and expressive language skills. The expression pattern of *FOXP2* in the brain suggests that this gene supports the development of neuronal circuits involved in sensorimotor integration and learning of motor skills. Thus, the upstream pathways that determine this complex expression pattern are key for understanding its biological effects. In this dissertation, I investigated the regulation of the *FOXP2* gene, from the molecular level, through cultured cell-lines, to the whole organism. I identified regulatory elements of *FOXP2* implicated in disorder (Chapter2), as well as novel enhancers that make direct contact with promoters of the gene (Chapter3). I studied the upstream mechanisms that regulate these elements (Chapter 4) and characterized the tissue activity of the most promising enhancers in the developing mouse brain (Chapter 5). Finally I demonstrated how this type of enhancer information can generate novel insights into the underlying biology of genotype-phenotype associations in relation to human brain development (Chapter 6).

The aim of this thesis was to shed light on the mechanisms of neuronal *FOXP2* regulation in humans and to investigate the contribution of regulatory genetic variants to phenotypic variation in health and disease. In this final chapter, I will summarize the findings of each experimental chapter and highlight the different levels at which I approached the overarching aims of this dissertation. I will combine the results from the single chapters and discuss the findings in a bigger picture. Furthermore, I will discuss follow-up strategies to gain further insight into the genetics and biology of speech and language.

What can we learn from the genome?

Variability in neurodevelopmental enhancers are known to contribute to both normal variation and disease phenotypes (Visel et al., 2013; Ramasamy et al., 2014; Boyd et al., 2015). In chapter 2, I identified an enhancer near the *FOXP2* gene that is displaced in a child with delayed speech development (Becker et al., 2015). The affected individual carries a complex chromosomal rearrangement, including an inversion of chromosome 7. One breakpoint of the inversion is located downstream of *FOXP2* and was mapped to the exact nucleotide position by Moralli et al. (Moralli et al., 2015). The published information, in combination with predicted functional states of the genome, was necessary and sufficient to locate the functional enhancer. The phenotype overlaps that of people with high-penetrance *FOXP2* mutations, which suggested that misregulation of this gene could be involved in the aetiology of the disorder. Thus, the displacement of the identified enhancer could cause aberrant *FOXP2* expression and consequently disrupt speech development.

Moving beyond the linear sequence of the genome, I used the three-dimensional folding of the DNA molecule to help uncover further novel regulatory elements that might regulate *FOXP2* expression. In chapter 3, my approach to identify enhancers was to investigate chromosomal interactions of the *FOXP2* promoter. To achieve this, I identified the transcriptionally active promoter in neuronal-like cells and then measured the chromatin interactions of this promoter with genomic elements located upstream and downstream of the *FOXP2* locus. In total, I identified eleven genomic regions that interacted with the promoter in neuronal-like cells (Figure 1). Promoter-promoter interactions between *FOXP2* and *MDFIC* suggested that these genes could be co-regulated. Indeed, transcriptome sequencing studies of primate forebrain clustered both genes in one group of co-regulated genes (Konopka et al., 2012). However, not much is known about *MDIFC* function and expression. Follow-up studies of this gene with respect to *FOXP2* function may further establish a link between *FOXP2* and the WNT-signalling pathway. I also detected an interaction of the *FOXP2* promoter with the 3'-UTR of the *FOXP2*. Similar interactions have been described for other genes and were reported to

both decrease (Le Cam and Legraverend, 1995; Paul et al., 1998) and increase gene expression (Salerno et al., 2000; Jash et al., 2012). The interaction to the 3'-UTR in *FOXP2* expressing cells was either increased or depleted and thus, the influence of this interaction on *FOXP2* expression remains unclear. Within the eleven interacting regions I detected nine candidate enhancer elements that were characterized by enhancer-specific histone marks in neuronal tissue and increased chromatin interactions in *FOXP2* expressing cell lines.

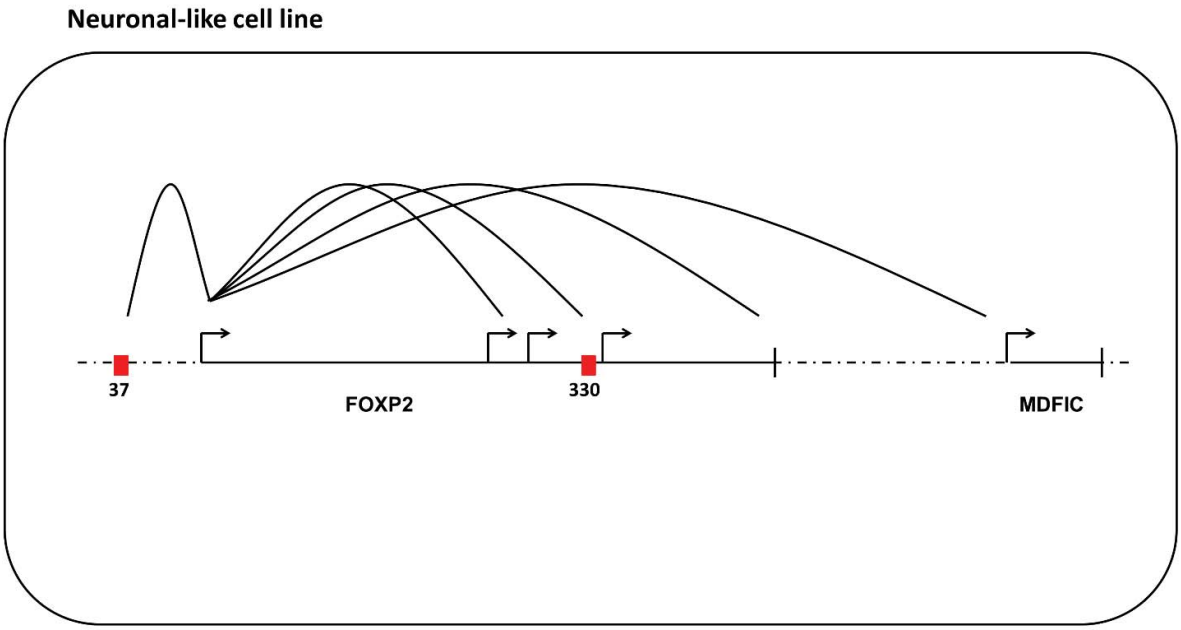

Figure 1: Chromatin interactions at the FOXP2 promoter. I detected chromatin interactions of the active *FOXP2* promoter in human neuronal-like cells (Chapter 3). The distal *FOXP2* promoter interacts with the *MDFIC* promoter, the 3'-UTR and neurodevelopmental enhancers. Interactions to two representative enhancers (red) are depicted. Promoter-promoter interactions, 3'-UTR and enhancer contacts may regulate the expression of *FOXP2* in response to developmental or environmental cues.

What can we learn from enhancers?

The candidate enhancers of *FOXP2* enabled me to study the regulatory mechanism of this gene at different levels. On the molecular level, enhancers are active within genetic networks and respond to specific signaling pathways. On the cellular level, enhancers are active in specific tissues and at developmental periods. On the human population level, variation in enhancer sequences may contribute to the variation in structure and function of the target

tissue. In the following section I summarized what I have learned from studying enhancers on the different levels.

Upstream regulatory networks of FOXP2 identified in molecular studies

In chapter 4 I determined the basal activity of the *FOXP2* promoters and enhancers in human neuronal-like cells. I detected the activity of promoter 1 and demonstrated the *in vitro* enhancer function of three enhancer elements. In combination with the enhancers, I studied the regulatory potential of eight TFs: LEF1, POU3F2, PAX6, TBR1, SOX5, FOXP1, FOXP2 and FOXP4. All TFs except POU3F2 modulated the activity of one or more regulatory element.

I demonstrated an enhancing effect of *FOXP2* on its regulatory elements, which suggested a positive auto-regulatory loop. Auto-regulation is found for TFs that function to maintain the cellular identity of differentiated cells (Bateman, 1998; Meredith et al., 2009) and for some TFs has been suggested as a mechanism to explain haploinsufficiency in disorders (Kleinjan et al., 2004). I detected WNT-signalling TF motifs in all active regulatory elements and consequently determined regulation by the WNT effector LEF1 (Figure 2). The effects of this experiment suggested that *FOXP2* might be actively regulated in response to WNT-signalling. This pathway is important in neurodevelopment and adult neurogenesis (Piccin and Morshead, 2011; Dickins and Salinas, 2013; Oliva et al., 2013). It will be important to determine the developmental period and brain tissue in which *FOXP2* could be regulated by WNT signalling. Additionally, I demonstrated the regulation of the candidate elements by SOX5 and TBR1, which overlap *FOXP2* expression in cortico-fugal interneurons of the cortical layer VI and regulate subtype specification in post-mitotic neurons (Kwan et al., 2008; Lai et al., 2008; Bedogni et al., 2010; Hisaoka et al., 2010; Han et al., 2011; McKenna et al., 2011; Hao et al., 2014). Both TFs increased promoter 1 and enhancer 37 activity, which suggest an activating effect on *FOXP2* expression in cortical layer VI.

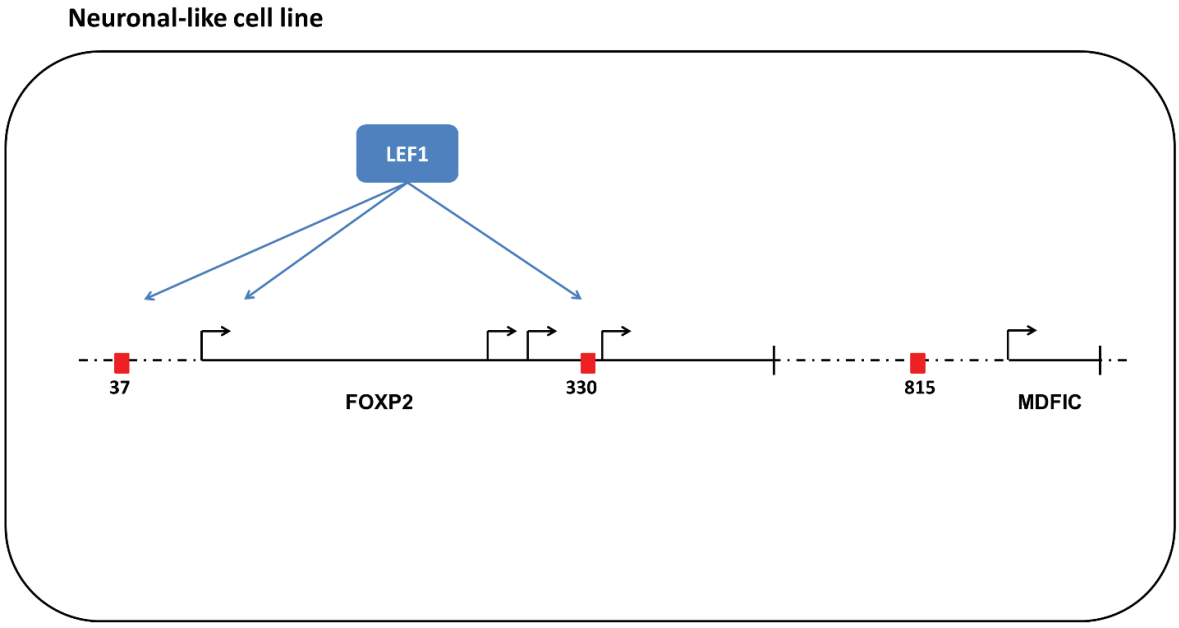

Figure 2: Promoter and enhancers are regulated by transcription factors. Co-transfection luciferase experiments reveal regulatory interactions of the regulatory elements in human neuronal-like cell lines (Chapter 4). The schematic shows effect of LEF1 (blue) on promoter 1, enhancer 37 and enhancer 330. Enhancer 815 activity was not affected by LEF1. In chapter 4 I additionally studied the effects of PAX6, POU3F2, FOXP1, FOXP2, FOXP4, SOX5 and TBR1.

FOXP2 enhancers drive in vivo expression in specific neuronal tissue

In chapter 5, I mapped the activity of two *FOXP2* enhancers in the developing mouse brain. I created transgenic mice that expressed a reporter gene under the control of these enhancers and visualized reporter gene expression in embryonic, postnatal and adult brains. Enhancer 37 activated reporter gene expression in the hippocampus, ventricles and olfactory bulb (OB). The activity of enhancer 37 overlapped known *FOXP2* expression sites only in the OB. The ventricular activity of this enhancer started after birth and remained stable throughout adulthood, whereas the *LacZ* expression in OB was detectable at embryonic day 18.5 and decreased in adult mice. The activity pattern of this enhancer could suggest that it is activated in adult new-born neurons in the ventricles and hippocampus, and remains active in neurons that migrate to the OB (Doetsch and Alvarez-Buylla, 1996; Lois et al., 1996). In humans and songbirds, the ventricular born neurons have been shown to migrate into the striatum (Alvarez-Buylla et al., 1994; Bergmann et al., 2015; Ernst and Frisen, 2015). In songbirds, these neurons

express *FoxP2* and migrate into in the striatal nucleus area X (Rochefort et al., 2007). The increased expression of *FoxP2* in area X was linked to singing behaviour in adult zebra finches (Haesler et al., 2004; Teramitsu and White, 2006; Haesler et al., 2007). It remains to be shown whether or not this gene is active during adult neurogenesis in a mammalian model system. In particular, it is unknown if *FOXP2* is expressed in the migratory stream of adult new-born neurons in humans.

The transgenic mice carrying enhancer 330 showed reporter gene expression in the neocortex and Purkinje Cells (PCs) of the cerebellum (Figure 3). The enhancer was active in the ventricular zone of the embryo and cortex of neonatal brains. Thus, enhancer 330 activity overlaps with known sites of *FOXP2* expression during development. These results suggest that enhancer 330 might be involved in driving early cortical expression of *FOXP2*. In adult mice, the activity of enhancer 330 decreased in deeper cortical layers, but remained active in the upper layers. Interestingly, prior studies have shown that *FOXP2* expression is localized specifically to the deeper cortical layer VI (Ferland et al., 2003). The absence of *FOXP2* mRNA in the upper layers can be explained by micro RNA mediated down-regulation, as the two micro RNAs mir-9 and mir-132 have been shown to repress *FOXP2* expression specifically in the upper cortical layers (Clovis et al., 2012). The lack of enhancer 330 activity in the adult deeper cortical layers indicates that *FOXP2* transcription in this layer is maintained by regulatory mechanisms that only act on the promoter or on other enhancers. In summary, the enhancer 330 activity overlapped *FOXP2* expression in the adult PCs and developing neocortex, which strongly suggests that this enhancer contributes to *FOXP2* expression in these structures.

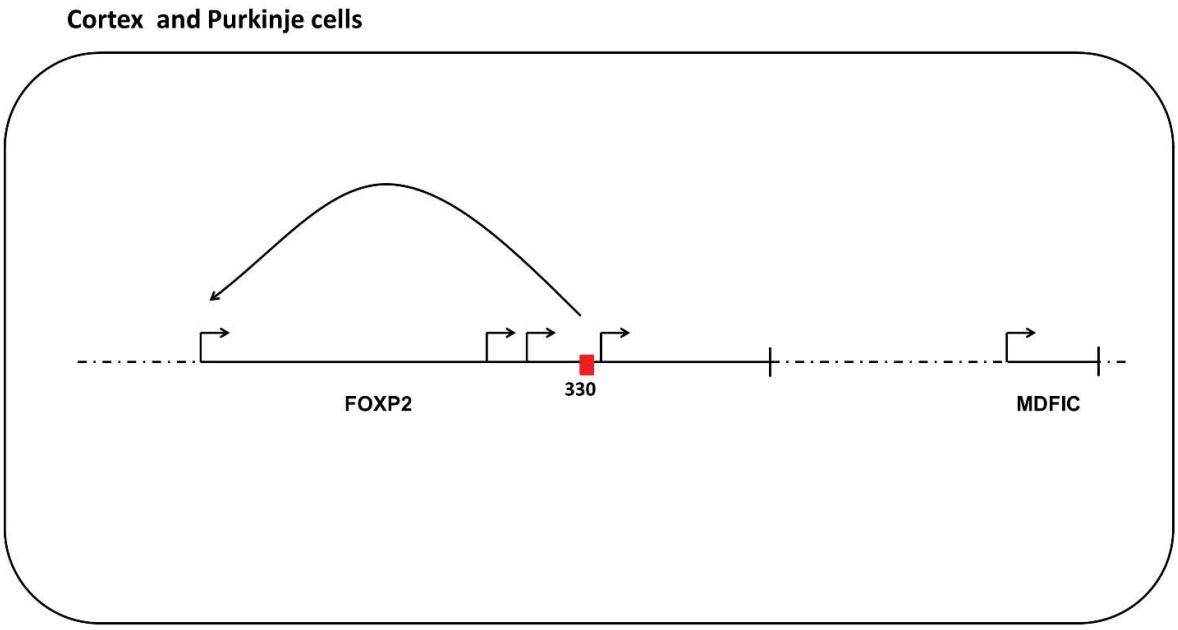

Figure 3: Neurodevelopmental enhancers drive tissue-specific expression *in vivo.* I determined the *in vivo* enhancer activity of two elements in transgenic mice expressing the LacZ reporter gene under control of the enhancer elements. In the schematic, enhancer 330 is depicted as active enhancer in the endogenous genomic loci controlling *FOXP2* in the cortex and PCs.

Genetic variants in neurodevelopmental enhancers explain structural variation in brain tissue

In chapter 6 I studied common genetic variants in a set of neurodevelopmental enhancers. The hypothesis was that genetic variants in enhancers affect the structure of the enhancer's target tissue in the adult brain. To test this hypothesis I selected common genetic variants located within experimentally validated regulatory elements. Because I expected subtle effects for single genetic enhancer variants, I analysed a set of 296 enhancers of the prior VISTA study, publically available through the VISTA enhancer browser (Visel et al., 2007). The results of this chapter are therefore not specific to the *FOXP2* enhancers. I focused on enhancers that had been shown to be active in the embryonic forebrain, which develops into the cortex and most of the subcortical brain regions. I hypothesized that variants in these enhancers influence the enhancer activity and consequently the expression of their target genes. Differential expression of neural genes could potentially alter the function and/or structure of the forebrain. I selected the enhancer variants from the summary data of the ENIGMA2 genome-wide

association study (GWAS) (Hibar et al., 2015). The ENIMGA2 study aimed to identify associations between genetic variants and variation in subcortical brain volumes of 30,000 healthy human adults, of which I analysed the discovery sample (13,000) in chapter 6. Within the group of forebrain enhancer variants I detected significant enrichment for variants associated with hippocampal volume. This enrichment was specific to subcortical phenotypes and not detectable in control phenotypes, such as body-mass index. The results of this chapter demonstrate how targeting experimentally-defined enhancers may help identify the underlying molecular aetiology of genotype-phenotype associations.

Integration of results across the chapters

Point mutations in *FOXP2* cause a developmental speech and language disorder (Lai et al., 2001). People with *FOXP2* mutations show functional and structural differences in brain structures related to motor control and language (Vargha-Khadem et al., 2005; Liegeois et al., 2011). The aim of this thesis was to identify the upstream regulatory networks that control *FOXP2* expression, which might help shed further light on genetic pathways involved in the development of neuronal circuits supporting human speech and language. Each chapter produced independent results and provided insight into components of the regulatory mechanisms. I obtained results on the molecular, cellular, tissue and human population levels. In order to understand the complex regulatory mechanisms and their possible implications for pathogenic and normal variation it is important to discuss the results across the chapters.

Purkinje cells maintenance mediated by WNT-signalling

In this thesis, I demonstrated that enhancer 330 is a neurodevelopmental enhancer and active in the PCs of the cerebellum (Chapter 5). The enhancer activity could be increased by *LEF1* (Chapter 4), which is known to be expressed in the developing cerebellum and in adult PCs

(Wisniewska et al., 2010). LEF1 is the effector TF of the WNT-signalling pathway, which is active in a subset of adult PCs (Selvadurai and Mason, 2011). Thus, my findings raise the possibility that *LEF1* might regulate FOXP2 expression during cerebellar development and in adult PCs (Figure 4). However, I did not study the embryonic activity pattern of enhancer 330 in the cerebellum, which is necessary to determine the overlap with WNT-signalling. In a prior study of zebrafish embryos, *lef1* knock-down was reported to abolish f*oxp2* expression in the hindbrain (Bonkowsky et al., 2008), which is the developmental origin of the cerebellum. It has been established that WNT-signalling in the cerebellum controls axon maintenance and synapse formation (Salinas, 1999; Dickins and Salinas, 2013) and that PCs express different components of the WNT-signalling pathway (Salinas et al., 1994; Patapoutian and Reichardt, 2000). Mouse *Foxp2* has been shown to promote neuronal differentiation, neurite formation and axon guidance in the striatum (Vernes et al., 2011). It can therefore be hypothesized that *LEF1* may potentially control *FOXP2* expression to regulate aspects of axon growth and synapse plasticity in the developing cerebellum (Figure 1), something that could be specifically tested with future functional investigations. Notably, mice homozygous for aetiological *Foxp2* mutations develop a smaller cerebellum (Groszer et al., 2008). The hypothesis that a *Lef1-Foxp2* pathway is important for normal cerebellar development could be followed up by conditional knock-out of *Foxp2* or *Lef1* in PCs. Homozygous knock-out mice of *Lef1* die before weening, whereas heterozygotes appear normal with no effect on cerebellar development (Vangenderen et al., 1994; Galceran et al., 2000). Conditional knock-outs of *Lef1*, where the gene is disrupted specifically in the PCs, may be required to determine subtle effects on cerebellar development, which cannot be detected with conventional knock-out. In conclusion, *Foxp2* is important for normal cerebellar development and my findings suggest that its cerebellar expression may be regulated by enhancer 330. The human enhancer 330 is itself regulated by *LEF1,* which may indicate a *LEF1-FOXP2* pathway in the developing cerebellum.

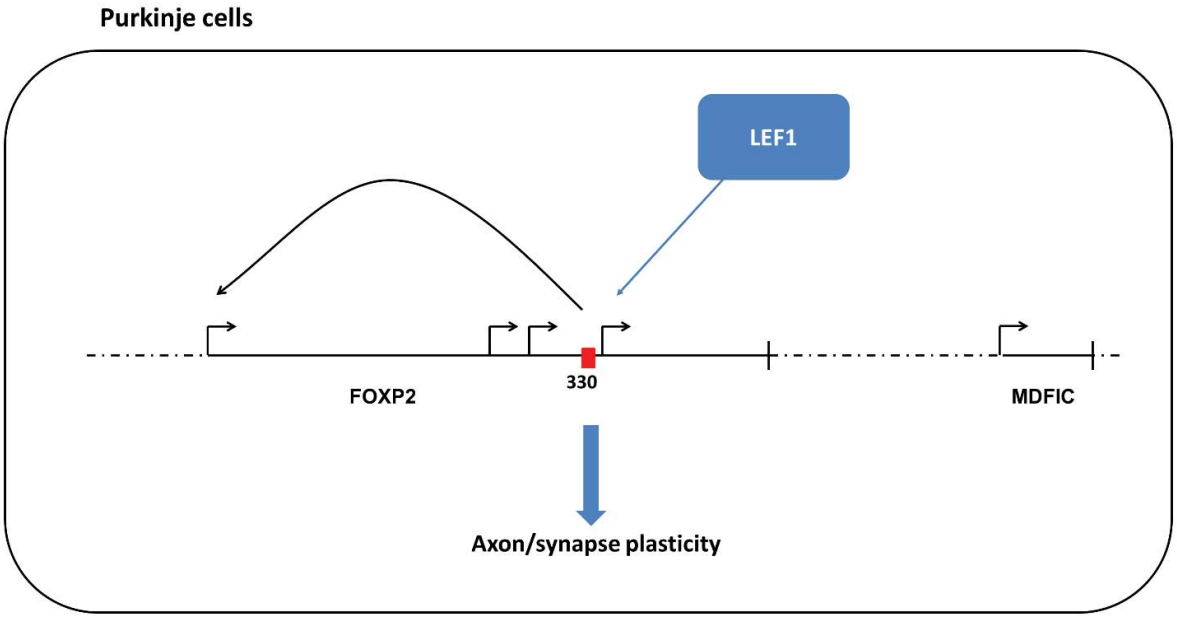

Figure 4: Integration of *FOXP2* in cell-specific signaling pathways. Enhancer 330 (red) is located within the *FOXP2* gene body. This enhancer contacts the *FOXP2* promoter (Chapter 3) and increases gene expression (Chapter 4). The enhancer is, among others, active in developing and adult PCs (Chapter 5). PCs express LEF1 (blue), which is the transcription factor downstream of the WNT-signaling pathway and activates genes responsible for synapse development and axon maintenance. LEF1 increases the activity of enhancer 330 (Chapter 4). The hypothesis is that WNT-signaling via LEF1 regulates *FOXP2* expression in the PCs to control axon and synapse plasticity.

Is cortical neuron subtype specificity mediated by enhancer 330?

I detected reproducible activity of enhancer 330 in the developing and adult neocortex (Chapter 5). This enhancer was previously shown to reproducibly drive cortical expression in mouse embryos at day E11.5 (Visel et al., 2007). In earlier studies, cortical *Foxp2* mRNA expression has been detected as early as E14.5 in the ventricular zone and in the postnatal cortex (Ferland et al., 2003). In the cortex of neonatal mice, enhancer 330 is active in all cortical layers. *FOXP2* expression is known to be primarily limited to the cortical layer VI in the adult cortex, whereas I found that enhancer 330 showed the strongest activity in the higher cortical layers (Chapter 5). Thus, enhancer 330 and *FOXP2* overlap in embryonic and neonatal brains, but not in the adult cortex.

I detected reduced enhancer 330 activity in response to *TBR1* and *SOX5* overexpression in human neuronal-like cell lines (Chapter 4) and enhancer 330 activity in the deeper cortical layers of adult brains (Figure 5). The results from both chapters together suggest that *TBR1* and *SOX5* downregulate the expression of *FOXP2* in the developing cortex. The expression of SOX5 and TBR1 is present in the developing cortical subplate and cortical layer VI (Kwan et al., 2008; McKenna et al., 2011). Enhancer 330 activity, *SOX5* and *TBR1* expression thus likely overlap in embryonic and prenatal neurons. Immunohistochemistry or *in situ* hybridization of *SOX5*/*TBR1* with *FOXP2* is required to determine the exact expression overlap of these TFs. Cortical development of cortico-thalamic projection neurons is regulated by SOX5 and TBR1, which establish the cellular-identity by repressing TFs that are found in other cortical subtypes, such as FEZF2 and CTIP (Kwan et al., 2008; Lai et al., 2008; McKenna et al., 2011; Greig et al., 2013). In the adult cortex, SOX5 and TBR1 expression is limited to layer VI (Lai et al., 2008) and enhancer 330 activity in this layer decreases in adult animals. This suggests that *SOX5* and *TBR1* may repress enhancer 330 in adult cortical layer VI. However, *FOXP2* remains expressed in adult cortical layer VI neurons and all *FOXP2* positive cortical neurons in the postnatal cortex have been reported to be *TBR1* positive (Hisaoka et al., 2010). Thus, the adult activity of enhancer 330 cannot explain adult *FOXP2* expression in the cortical layers.

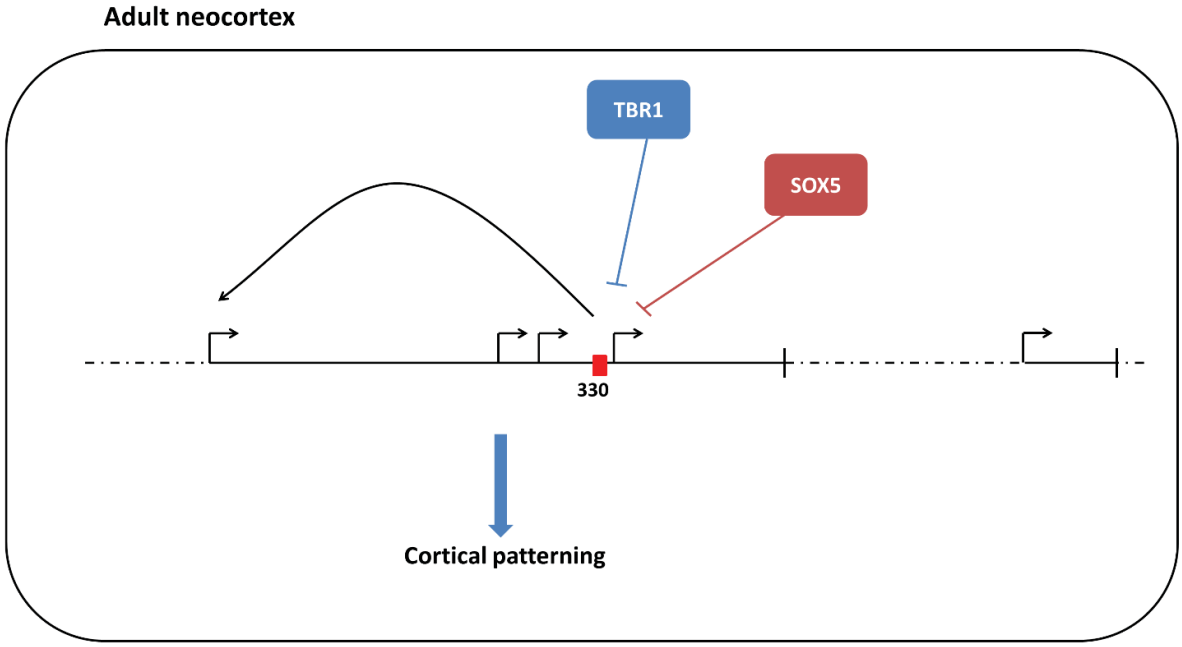

Figure 5: Integration of *FOXP2* in cortical laminar development. Enhancer 330 (red) is located within the *FOXP2* gene body. This enhancer contacts the *FOXP2* promoter (Chapter 3) and increases gene expression (Chapter 4). The enhancer is active in the developing cortex and upper cortical layers in the adult (Chapter 5). FOXP2 expression overlaps enhancer 330 activity in the developing cortex. In the adult cortex *FOXP2* is specifically expressed in the deeper cortical layer VI. TBR1 and SOX5 are expressed in the cortical layer VI and decrease the activity of enhancer 330 (Chapter 4).

Downregulation of *FOXP2* in upper cortical layers by post-transcriptional mechanisms may explain the missing *FOXP2* expression in these layers. Indeed, mir-9 and mir-132 have been shown to specifically repress *FOXP2* mRNA in the upper cortical layers (Clovis et al., 2012). Thus, enhancer 330 may activate expression in all cortical progenitor cells and the layer VI specificity is achieved by the post-transcriptional down-regulation. The expression in the deeper cortical layers could potentially be maintained by *FOXP2* itself, via autoregulation at other regulatory elements (Chapter 4). Thus, enhancer 330 may only be required for embryonic and early postnatal expression of *FOXP2.* In conclusion, my data suggest that *FOXP2* expression in the cortex could be initiated by enhancer 330. *SOX5* and *TBR1* could be involved in the regulation of enhancer 330, but the timing of the regulatory interaction will have to be investigated in model systems of neural development.

Could enhancer 37 regulate the expression of FOXP2 in adult neurogenesis?

I mapped the *in vivo* activity of enhancer 37 to the hippocampus, ventricles and OB. Adult neurogenesis occurs in the hippocampus and ventricles (Eriksson et al., 1998; Ernst et al., 2014; Bergmann et al., 2015). In mice, the ventricular born neurons migrate to the OB in the rostral migratory stream (Doetsch and Alvarez-Buylla, 1996; Wichterle et al., 1997). The ventricular and OB activity of enhancer 37 could suggest that this enhancer is active in neurons that migrate in the rostral migratory stream. I also detected that TF motifs of the WNT-signalling pathway are located in enhancer 37 and that it is upregulated by *LEF1* (Chapter 4). *LEF1* in complex with *TCF-1* regulates the expression of WNT-signalling target genes (Dickins and Salinas, 2013; Oliva et al., 2013). WNT signalling has been shown to regulate neurogenesis in the ventricles (Piccin and Morshead, 2011), the hippocampus (Lie et al., 2005) and OB (Booker-Dwyer et al., 2008). Enhancer 37 could thus be a target regulatory element of WNT-signalling during adult neurodevelopment (Figure 6).

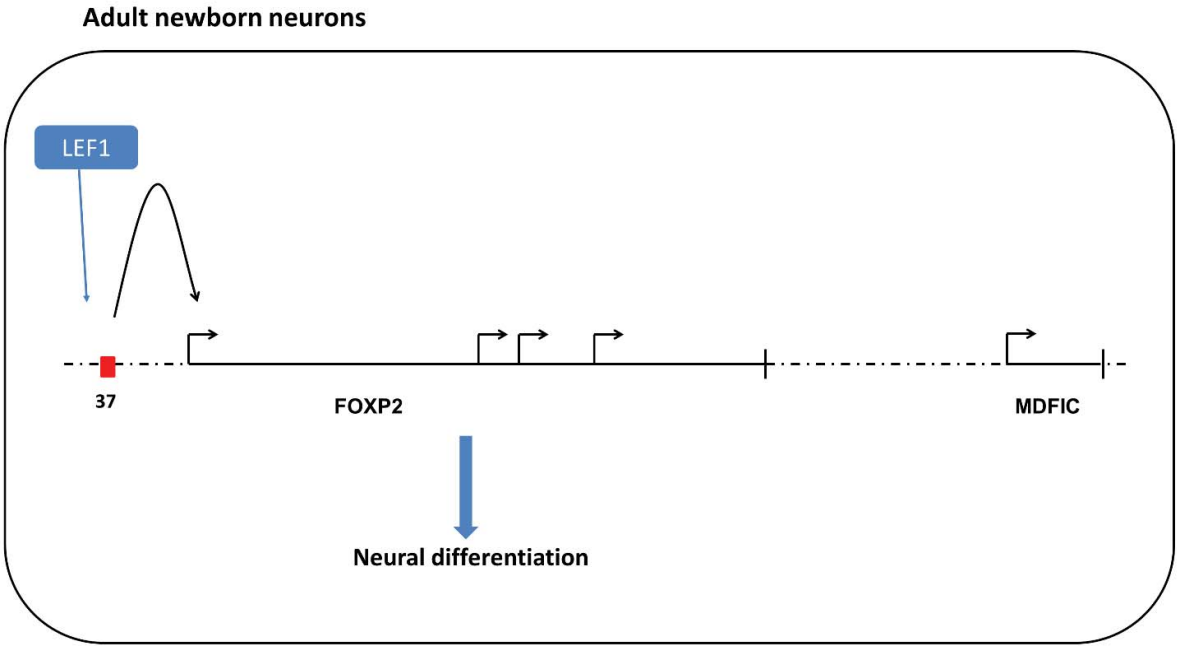

Figure 6: Integration of *FOXP2* in adult neurogenesis. Enhancer 37 (red) is located upstream of the *FOXP2* gene. This enhancer and the promoter region of *MDFIC* contact the *FOXP2* promoter (Chapter 3). Enhancer 37 increases gene expression (Chapter 4). The enhancer is active in the OB, the outer layer of the ventricles, the dentate gyrus and CA1 neurons of the hippocampus (Chapter 5). Ventricular born neurons migrate to OB in mice; and to the striatum in human and songbirds. The canonical WNT-signaling pathway, with the downstream activators LEF1 and TCF-1 regulate neural differentiation in adult new-born neurons. LEF1 increases the activity of enhancer 37. The hypothesis is that WNT-signaling via LEF1 regulates *FOXP2* expression in adult new-born neurons in the ventricles and the neurons migrate to the species-specific target tissue.

In humans, the ventricular born neurons do not migrate to the OB but have been shown to migrate into the striatum, where they differentiate into interneurons (Ernst et al., 2014; Ernst and Frisen, 2015). Similarly, in songbirds ventricular born neurons migrate into the striatal song-learning nucleus called Area X, where they differentiate and integrate into the existing neuronal architecture (Alvarez-Buylla et al., 1994). The regulation of *FoxP2* in Area X has been shown to be linked to singing behaviour in the songbirds (Chen et al., 2013; Mendoza et al., 2015). The increase of *FoxP2* expression is partly due to new-born neurons in adult animals that migrated into the striatum (Rochefort et al., 2007). Thus, adult neurogenesis may be linked to *FoxP2* expression and singing behaviour in songbirds (Schulz et al., 2010). However, in songbirds, the new-born neurons form medium spiny neurons, whereas in humans they form interneurons (Ernst and Frisen, 2015). The new-born human interneurons are DARPP32

negative and express parvalbumin and calbindin. The co-expression of *FOXP2* with these marker genes in striatal interneurons would, therefore, be crucial to determine a possible relation between *FOXP2* and human adult neurogenesis. Molecular studies of adult neurogenesis in humans are not feasible, because of the lack of neuronal tissue for experimental investigations. As new-born neurons in mice migrate into the OB, this animal is not a suitable system to study the link between *FOXP2* and neurogenesis. Possibly, this link could be studied in other mammals, such as bats. Recently bats obtained attention as a model system for mammalian vocal and communication system (Rodenas-Cuadrado et al., 2015). Ventricular born neurons in bats at least partially migrate towards the OB (Amrein et al., 2007), but the migratory route to the striatum has not yet been investigated.

Signalling pathways involved in language-related phenotypes

I identified enhancer 815 based on the information about a child with delayed speech development, described in the literature (Chapter 2). The child carries an inversion that separates *FOXP2* from this enhancer, suggesting that misregulation of the gene contributes to the disorder. My in vitro experiments suggested that activity of enhancer 815 may be regulated by *PAX6*, *FOXP1* and *FOXP2* (Chapter 4), which suggests that, in the affected child, the *FOXP2* gene lacks enhancer 815 mediated regulatory input from these TFs. Similarly, loss-of-function mutations in these TFs could affect the regulation of *FOXP2* in other clinical cases. The study of phenotypes associated with *PAX6* and *FOXP1* mutations and their overlap with phenotypes seen for people with *FOXP2* mutations may shed light on an overlapping molecular aetiology. Genetic disruptions of *FOXP1* can cause autism spectrum disorder, intellectual disability, motor delay and expressive language problems (Pariani et al., 2009; Carr et al., 2010; Hamdan et al., 2010; Horn et al., 2010; O'Roak et al., 2011; Bacon and Rappold, 2012; Talkowski et al., 2012). The phenotypic spectra associated with *FOXP1* and *FOXP2* disruptions were previously described in a review and show overlap for language problems

(Bacon and Rappold, 2012). *PAX6* mutations generally cause aniridia and other malformations of the eye (Hanson, 2003). A splice site mutation in *PAX6* was linked to Gillespie syndrome, which among other problems includes impairments in orofacial motor control and dysarthric speech (Ticho et al., 2006). While dysarthria has been suggested to be a phenotype associated with *FOXP2* disruptions at least in some cases (Shriberg et al., 2006), it was not reported for the child with the chromosomal rearrangement that separates *FOXP2* from enhancer 815 (Moralli et al., 2015). Another patient carrying a *de novo* nonsense mutation in *PAX6* presented with the typical aniridia phenotype, but further included coordination impairment, general dyspraxia, intellectual disability and poor speech (Graziano et al., 2007). Detailed comparisons between people with *PAX6* and *FOXP2* mutations would be required to determine a possible common overlapping phenotype.

Co-expression of *FOXP2* with *FOXP1* has been described in the basal ganglia, thalamus and inferior olive (Teramitsu et al., 2004; Takahashi et al., 2008; Kaoru et al., 2010; Mendoza et al., 2015) and with *PAX6* in the amygdala and striatum (Kaoru et al., 2010). Notably, co-expression of all three TFs has been seen in the striatum of monkeys and rats (Kaoru et al., 2010). Thus, the observed regulation of enhancer 815 by *FOXP1* and *PAX6* could suggest that this enhancer is active in the striatum (Figure 7). Perhaps the overlap in speech/language-related phenotypes between cases of *PAX6, FOXP1* and *FOXP2* mutation can be partially explained by regulatory interactions in the striatum. Overlapping functional or structural aberrations in the striatum of the child that led to the discovery of enhancer 815 would further support this hypothesis. Brain imaging studies of this child have not been reported, and comparison to findings in additional clinical cases would be required to thoroughly evaluate this hypothesis.

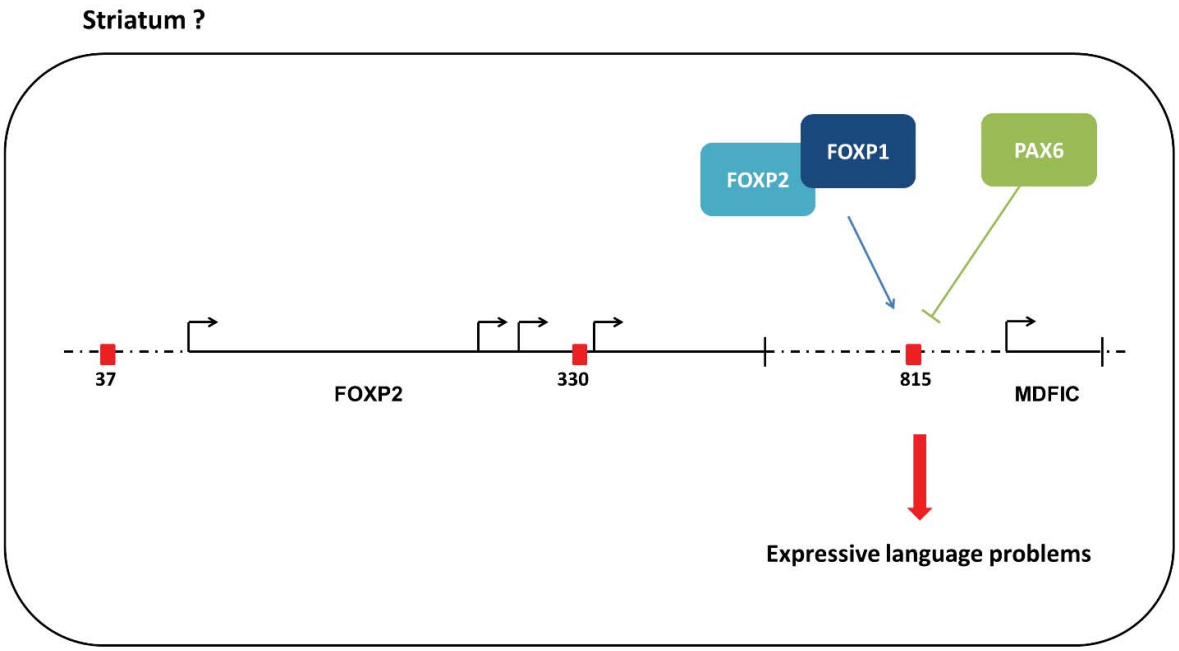

Figure 7: Disruption of enhancer 815 regulation in language development. Enhancer 815 (red) is located downstream of the *FOXP2* gene. This enhancer is separated from *FOXP2* in a patient with language problems (Chapter 2). Enhancer 815 is repressed by PAX6 and activated by FOXP1 and FOXP2 (Chapter 4). Mutations in PAX6 and FOXP2 can cause language problems. The hypothesis is that the language phenotypes caused by mutations of FOXP1, FOXP2 and PAX6 are the result of aberrant *FOXP2* regulation. This regulatory interaction possibly occurs in the developing striatum.

Future direction

In the previous section, I presented models of *FOXP2* regulation that emerged from analysing results across the individual chapters and suggested follow-up strategies to investigate these. In the following section, I address additional future directions that follow from my results.

Enhancers as tools to study FOXP2

The experiments in each chapter investigated a limited amount of the possible search space. Future findings on *FOXP2* regulation may highlight the importance of genomic regions which have not been tested in chapter 3, and TFs which have not been investigated in chapter 4. Clinically relevant copy number variants near *FOXP2* may justify the extension of the chromatin conformation capture design to more distant genomic regions. My in silico analyses detected hundreds of TF motifs in the putative regulatory elements that I identified. Many of the relevant

TFs are expressed in the brain and known to be involved in neuronal development (chapter 4). For example, the overlapping expression of *foxp2* with *emx1* in the zebrafish forebrain could be an interesting link to follow up (Shah et al., 2006). Also retinoic acid binding receptor alpha was detected in the motif analysis and accumulating evidence shows that retinoic acid signalling is linked to *FOXP2* function (Devanna et al., 2014) and language-related neural networks (van Rhijn and Vernes, 2015). Furthermore, I selected the most promising enhancers to map the *in vivo* activity (Chapter 5) and did not investigate enhancer 815 using these techniques. As previously discussed, this enhancer is likely involved in the pathogenesis of a child with delayed speech development. Uncovering the tissue-specific activity of this enhancer will be important to determine the molecular aetiology underlying the phenotype of this case.

Genetic variation underlying language-related phenotypes

The identification of *FOXP2* mutations in speech and language disorders has sparked speculations on the potential contributions of this gene to other pathological and normal human cognitive phenotypes. Consequently, researchers have investigated the *FOXP2* gene locus for genetic association to various traits. Such studies investigated the association of common variants within the *FOXP2* gene to auditory-visual hallucinations (Sanjuan et al., 2006; McCarthy-Jones et al., 2014), schizophrenia (Tolosa et al., 2010; Spaniel et al., 2011), ADHD (Ribases et al., 2012), autism (Park et al., 2014), enhanced language skills (Chandrasekaran et al., 2015) or brain volumes (Hoogman et al., 2014). However, most of these studies report negative findings or have not been replicated. Genetic variants that affect the identified regulatory elements from this thesis could potentially alter *FOXP2* expression and disrupting the development of *FOXP2* positive neuronal circuits. According to this hypothesis, in chapter 6 I demonstrated on a broader scale that neurodevelopmental enhancers are enriched for association to subtle common phenotypes. Thus future candidate gene association studies of *FOXP2* would benefit from targeting genetic variants within the identified regulatory enhancer elements from this thesis. Similarly, next generation sequencing of samples of people with speech- and language-related disorders has been limited to the exome, focusing on coding

sequences (Villanueva et al., 2015). The regulatory regions of *FOXP2* are promising candidate regions to contain rare variants, which would remain undetected with exome-sequencing. As the field of clinical genetics moves rapidly from exome-sequencing to whole-genome sequencing it will be crucial to filter the variants in a justified and meaningful way. Thus, the incorporation of enhancer and other regulatory sequences will be necessary for all genes. The field of language genetics is well equipped to identify genetic variants that may affect the regulation of *FOXP2* and contribute to language-related traits.

Conclusion

In my thesis, I investigated the regulation of *FOXP2* at the molecular, cellular, tissue and phenotypic level. I identified the first human enhancers of the *FOXP2* gene, which led to the identification of upstream signalling pathways of *FOXP2.* I mapped the enhancer activity in the developing and adult brain, which relates the identified upstream pathways to the neuronal tissue in which they regulate *FOXP2*. I linked a set of neurodevelopmental enhancers, unrelated to *FOXP2*, to normal variation in brain structure and pathologic variation of language skills. The integration of results across these levels created new hypotheses regarding the mechanisms underlying the neurobiology of *FOXP2.*

References

Alvarez-Buylla A, Ling CY, Yu WS (1994) Contribution of neurons born during embryonic, juvenile, and adult life to the brain of adult canaries: regional specificity and delayed birth of neurons in the song-control nuclei. The Journal of comparative neurology 347:233-248.

Amrein I, Dechmann DK, Winter Y, Lipp HP (2007) Absent or low rate of adult neurogenesis in the hippocampus of bats (Chiroptera). PloS one 2:e455.

Bacon C, Rappold GA (2012) The distinct and overlapping phenotypic spectra of FOXP1 and FOXP2 in cognitive disorders. Human genetics 131:1687-1698.

Bateman E (1998) Autoregulation of eukaryotic transcription factors. Progress in nucleic acid research and molecular biology 60:133-168.

Becker M, Devanna P, Fisher SE, Vernes SC (2015) A chromosomal rearrangement in a child with severe speech and language disorder separates FOXP2 from a functional enhancer. Molecular cytogenetics 8:69.

Bedogni F, Hodge RD, Elsen GE, Nelson BR, Daza RA, Beyer RP, Bammler TK, Rubenstein JL, Hevner RF (2010) Tbr1 regulates regional and laminar identity of postmitotic neurons in developing neocortex. Proc Natl Acad Sci U S A 107:13129-13134.

Bergmann O, Spalding KL, Frisen J (2015) Adult Neurogenesis in Humans. Cold Spring Harbor perspectives in biology 7.

Bonkowsky JL, Wang X, Fujimoto E, Lee JE, Chien CB, Dorsky RI (2008) Domain-specific regulation of foxP2 CNS expression by lef1. BMC developmental biology 8:103.

Booker-Dwyer T, Hirsh S, Zhao H (2008) A unique cell population in the mouse olfactory bulb displays nuclear beta-catenin signaling during development and olfactory sensory neuron regeneration. Developmental neurobiology 68:859-869.

Boyd JL, Skove SL, Rouanet JP, Pilaz LJ, Bepler T, Gordan R, Wray GA, Silver DL (2015) Human-Chimpanzee Differences in a FZD8 Enhancer Alter Cell-Cycle Dynamics in the Developing Neocortex. Current biology : CB.

Carr CW, Moreno-De-Luca D, Parker C, Zimmerman HH, Ledbetter N, Martin CL, Dobyns WB, Abdul-Rahman OA (2010) Chiari I malformation, delayed gross motor skills, severe speech delay, and epileptiform discharges in a child with FOXP1 haploinsufficiency. European journal of human genetics : EJHG 18:1216-1220.

Chandrasekaran B, Yi HG, Blanco NJ, McGeary JE, Maddox WT (2015) Enhanced procedural learning of speech sound categories in a genetic variant of FOXP2. The Journal of neuroscience : the official journal of the Society for Neuroscience 35:7808-7812.

Chen Q, Heston JB, Burkett ZD, White SA (2013) Expression analysis of the speech-related genes FoxP1 and FoxP2 and their relation to singing behavior in two songbird species. The Journal of experimental biology 216:3682-3692.

Clovis YM, Enard W, Marinaro F, Huttner WB, De Pietri Tonelli D (2012) Convergent repression of Foxp2 3'UTR by miR-9 and miR-132 in embryonic mouse neocortex: implications for radial migration of neurons. Development 139:3332-3342.

Devanna P, Middelbeek J, Vernes SC (2014) FOXP2 drives neuronal differentiation by interacting with retinoic acid signaling pathways. Frontiers in cellular neuroscience 8:305.

Dickins EM, Salinas PC (2013) Wnts in action: from synapse formation to synaptic maintenance. Frontiers in cellular neuroscience 7:162.

Doetsch F, Alvarez-Buylla A (1996) Network of tangential pathways for neuronal migration in adult mammalian brain. Proc Natl Acad Sci U S A 93:14895-14900.

Eriksson PS, Perfilieva E, Bjork-Eriksson T, Alborn AM, Nordborg C, Peterson DA, Gage FH (1998) Neurogenesis in the adult human hippocampus. Nature medicine 4:1313-1317.

Ernst A, Frisen J (2015) Adult neurogenesis in humans- common and unique traits in mammals. PLoS biology 13:e1002045.

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

Ferland RJ, Cherry TJ, Preware PO, Morrisey EE, Walsh CA (2003) Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain. The Journal of comparative neurology 460:266-279.

Galceran J, Miyashita-Lin EM, Devaney E, Rubenstein JLR, Grosschedl R (2000) Hippocampus development and generation of dentate gyrus granule cells is regulated by LEF1. Development 127:469-482.

Graziano C, D'Elia AV, Mazzanti L, Moscano F, Guidelli Guidi S, Scarano E, Turchetti D, Franzoni E, Romeo G, Damante G, Seri M (2007) A de novo nonsense mutation of PAX6 gene in a patient with aniridia, ataxia, and mental retardation. American journal of medical genetics Part A 143A:1802-1805.

Greig LC, Woodworth MB, Galazo MJ, Padmanabhan H, Macklis JD (2013) Molecular logic of neocortical projection neuron specification, development and diversity. Nat Rev Neurosci 14:755-769.

Groszer M et al. (2008) Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits. Current biology : CB 18:354-362.

Haesler S, Rochefort C, Georgi B, Licznerski P, Osten P, Scharff C (2007) Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. PLoS biology 5:e321.

Haesler S, Wada K, Nshdejan A, Morrisey EE, Lints T, Jarvis ED, Scharff C (2004) FoxP2 expression in avian vocal learners and non-learners. The Journal of neuroscience : the official journal of the Society for Neuroscience 24:3164-3175.

Hamdan FF, Daoud H, Rochefort D, Piton A, Gauthier J, Langlois M, Foomani G, Dobrzeniecka S, Krebs MO, Joober R, Lafreniere RG, Lacaille JC, Mottron L, Drapeau P, Beauchamp MH, Phillips MS, Fombonne E, Rouleau GA, Michaud JL (2010) De novo mutations in FOXP1 in cases with intellectual disability, autism, and language impairment. Am J Hum Genet 87:671-678.

Han W, Kwan KY, Shim S, Lam MM, Shin Y, Xu X, Zhu Y, Li M, Sestan N (2011) TBR1 directly represses Fezf2 to control the laminar origin and development of the corticospinal tract. Proc Natl Acad Sci U S A 108:3041-3046.

Hanson IM (2003) PAX6 and congenital eye malformations. Pediatr Res 54:791-796.

Hao H, Li Y, Tzatzalos E, Gilbert J, Zala D, Bhaumik M, Cai L (2014) Identification of a transient Sox5 expressing progenitor population in the neonatal ventral forebrain by a novel cis-regulatory element. Developmental biology 393:183-193.

Hibar DP et al. (2015) Common genetic variants influence human subcortical brain structures. Nature 520:224-229.

Hisaoka T, Nakamura Y, Senba E, Morikawa Y (2010) The forkhead transcription factors, Foxp1 and Foxp2, identify different subpopulations of projection neurons in the mouse cerebral cortex. Neuroscience 166:551-563.

Hoogman M, Guadalupe T, Zwiers MP, Klarenbeek P, Francks C, Fisher SE (2014) Assessing the effects of common variation in the FOXP2 gene on human brain structure. Frontiers in human neuroscience 8:473.

Horn D et al. (2010) Identification of FOXP1 deletions in three unrelated patients with mental retardation and significant speech and language deficits. Human mutation 31:E1851-1860.

Jash A, Yun K, Sahoo A, So JS, Im SH (2012) Looping mediated interaction between the promoter and 3' UTR regulates type II collagen expression in chondrocytes. PloS one 7:e40828.

Kaoru T, Liu FC, Ishida M, Oishi T, Hayashi M, Kitagawa M, Shimoda K, Takahashi H (2010) Molecular characterization of the intercalated cell masses of the amygdala: implications for the relationship with the striatum. Neuroscience 166:220-230.

Kleinjan DA, Seawright A, Childs AJ, van Heyningen V (2004) Conserved elements in Pax6 intron 7 involved in (auto)regulation and alternative transcription. Developmental biology 265:462-477.

Konopka G, Friedrich T, Davis-Turak J, Winden K, Oldham MC, Gao F, Chen L, Wang GZ, Luo R, Preuss TM, Geschwind DH (2012) Human-specific transcriptional networks in the brain. Neuron 75:601-617.

Kwan KY, Lam MM, Krsnik Z, Kawasawa YI, Lefebvre V, Sestan N (2008) SOX5 postmitotically regulates migration, postmigratory differentiation, and projections of subplate and deep-layer neocortical neurons. Proc Natl Acad Sci U S A 105:16021-16026.

Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. Nature 413:519-523.

Lai T, Jabaudon D, Molyneaux BJ, Azim E, Arlotta P, Menezes JR, Macklis JD (2008) SOX5 controls the sequential generation of distinct corticofugal neuron subtypes. Neuron 57:232-247.

Le Cam A, Legraverend C (1995) Transcriptional repression, a novel function for 3' untranslated regions. European journal of biochemistry / FEBS 231:620-627.

Lie DC, Colamarino SA, Song HJ, Desire L, Mira H, Consiglio A, Lein ES, Jessberger S, Lansford H, Dearie AR, Gage FH (2005) Wnt signalling regulates adult hippocampal neurogenesis. Nature 437:1370-1375.

Liegeois F, Morgan AT, 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 15:283-288.
Lois C, Garcia-Verdugo JM, Alvarez-Buylla A (1996) Chain migration of neuronal precursors. Science 271:978-981.
McCarthy-Jones S, Green MJ, Scott RJ, Tooney PA, Cairns MJ, Wu JQ, Oldmeadow C, Carr V, Australian Schizophrenia Research B (2014) Preliminary evidence of an interaction between the FOXP2 gene and childhood emotional abuse predicting likelihood of auditory verbal hallucinations in schizophrenia. Journal of psychiatric research 50:66-72.
McKenna WL, Betancourt J, Larkin KA, Abrams B, Guo C, Rubenstein JL, Chen B (2011) Tbr1 and Fezf2 regulate alternate corticofugal neuronal identities during neocortical development. The Journal of neuroscience : the official journal of the Society for Neuroscience 31:549-564.
Mendoza E, Tokarev K, During DN, Retamosa EC, Weiss M, Arpenik N, Scharff C (2015) Differential coexpression of FoxP1, FoxP2, and FoxP4 in the Zebra Finch (Taeniopygia guttata) song system. The Journal of comparative neurology 523:1318-1340.
Meredith DM, Masui T, Swift GH, MacDonald RJ, Johnson JE (2009) Multiple transcriptional mechanisms control Ptf1a levels during neural development including autoregulation by the PTF1-J complex. The Journal of neuroscience : the official journal of the Society for Neuroscience 29:11139-11148.
Moralli D, Nudel R, Chan MT, Green CM, Volpi EV, Benitez-Burraco A, Newbury DF, Garcia-Bellido P (2015) Language impairment in a case of a complex chromosomal rearrangement with a breakpoint downstream of FOXP2. Molecular cytogenetics 8:36.
O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, Karakoc E, Mackenzie AP, Ng SB, Baker C, Rieder MJ, Nickerson DA, Bernier R, Fisher SE, Shendure J, Eichler EE (2011) Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet 43:585-589.
Oliva CA, Vargas JY, Inestrosa NC (2013) Wnts in adult brain: from synaptic plasticity to cognitive deficiencies. Frontiers in cellular neuroscience 7:224.
Pariani MJ, Spencer A, Graham JM, Jr., Rimoin DL (2009) A 785kb deletion of 3p14.1p13, including the FOXP1 gene, associated with speech delay, contractures, hypertonia and blepharophimosis. European journal of medical genetics 52:123-127.
Park Y, Won S, Nam M, Chung JH, Kwack K (2014) Interaction between MAOA and FOXP2 in association with autism and verbal communication in a Korean population. Journal of child neurology 29:NP207-211.
Patapoutian A, Reichardt LF (2000) Roles of Wnt proteins in neural development and maintenance. Current opinion in neurobiology 10:392-399.
Paul C, Simar-Blanchet AE, Ro HS, Le Cam A (1998) Characterization of three transcriptional repressor sites within the 3' untranslated region of the rat serine protease inhibitor 2.3 gene. European journal of biochemistry / FEBS 254:538-546.
Piccin D, Morshead CM (2011) Wnt signaling regulates symmetry of division of neural stem cells in the adult brain and in response to injury. Stem cells 29:528-538.
Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, De T, Coin L, de Silva R, Cookson MR, Singleton AB, Hardy J, Ryten M, Weale ME (2014) Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat Neurosci 17:1418-1428.
Ribases M, Sanchez-Mora C, Ramos-Quiroga JA, Bosch R, Gomez N, Nogueira M, Corrales M, Palomar G, Jacob CP, Gross-Lesch S, Kreiker S, Reif A, Lesch KP, Cormand B, Casas M, Bayes M (2012) An association study of sequence variants in the forkhead box P2 (FOXP2) gene and adulthood attention-deficit/hyperactivity disorder in two European samples. Psychiatric genetics 22:155-160.
Rochefort C, He X, Scotto-Lomassese S, Scharff C (2007) Recruitment of FoxP2-expressing neurons to area X varies during song development. Developmental neurobiology 67:809-817.
Rodenas-Cuadrado P, Chen XS, Wiegrebe L, Firzlaff U, Vernes SC (2015) A novel approach identifies the first transcriptome networks in bats: a new genetic model for vocal communication. BMC genomics 16:836.
Salerno MS, Mordvinov VA, Sanderson CJ (2000) Binding of octamer factors to a novel 3'-positive regulatory element in the mouse interleukin-5 gene. J Biol Chem 275:4525-4531.
Salinas PC (1999) Wnt factors in axonal remodelling and synaptogenesis. Biochemical Society symposium 65:101-109.

Salinas PC, Fletcher C, Copeland NG, Jenkins NA, Nusse R (1994) Maintenance of Wnt-3 expression in Purkinje cells of the mouse cerebellum depends on interactions with granule cells. Development 120:1277-1286.
Sanjuan J, Tolosa A, Gonzalez JC, Aguilar EJ, Perez-Tur J, Najera C, Molto MD, de Frutos R (2006) Association between FOXP2 polymorphisms and schizophrenia with auditory hallucinations. Psychiatric genetics 16:67-72.
Schulz SB, Haesler S, Scharff C, Rochefort C (2010) Knockdown of FoxP2 alters spine density in Area X of the zebra finch. Genes Brain Behav 9:732-740.
Selvadurai HJ, Mason JO (2011) Wnt/beta-catenin signalling is active in a highly dynamic pattern during development of the mouse cerebellum. PloS one 6:e23012.
Shah R, Medina-Martinez O, Chu LF, Samaco RC, Jamrich M (2006) Expression of FoxP2 during zebrafish development and in the adult brain. The International journal of developmental biology 50:435-438.
Shriberg LD, Ballard KJ, Tomblin JB, Duffy JR, Odell KH, Williams CA (2006) Speech, prosody, and voice characteristics of a mother and daughter with a 7;13 translocation affecting FOXP2. Journal of speech, language, and hearing research : JSLHR 49:500-525.
Spaniel F, Horacek J, Tintera J, Ibrahim I, Novak T, Cermak J, Klirova M, Hoschl C (2011) Genetic variation in FOXP2 alters grey matter concentrations in schizophrenia patients. Neuroscience letters 493:131-135.
Takahashi K, Liu FC, Hirokawa K, Takahashi H (2008) Expression of Foxp4 in the developing and adult rat forebrain. Journal of neuroscience research 86:3106-3116.
Talkowski ME et al. (2012) Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. Cell 149:525-537.
Teramitsu I, White SA (2006) FoxP2 regulation during undirected singing in adult songbirds. The Journal of neuroscience : the official journal of the Society for Neuroscience 26:7390-7394.
Teramitsu I, Kudo LC, London SE, Geschwind DH, White SA (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 24:3152-3163.
Ticho BH, Hilchie-Schmidt C, Egel RT, Traboulsi EI, Howarth RJ, Robinson D (2006) Ocular findings in Gillespie-like syndrome: association with a new PAX6 mutation. Ophthalmic genetics 27:145-149.
Tolosa A, Sanjuan J, Dagnall AM, Molto MD, Herrero N, de Frutos R (2010) FOXP2 gene and language impairment in schizophrenia: association and epigenetic studies. BMC medical genetics 11:114.
van Rhijn JR, Vernes SC (2015) Retinoic Acid Signaling: A New Piece in the Spoken Language Puzzle. Frontiers in psychology 6:1816.
Vangenderen C, Okamura RM, Farinas I, Quo RG, Parslow TG, Bruhn L, Grosschedl R (1994) Development of Several Organs That Require Inductive Epithelial-Mesenchymal Interactions Is Impaired in Lef-1-Deficient Mice. Genes Dev 8:2691-2703.
Vargha-Khadem F, Gadian DG, Copp A, Mishkin M (2005) FOXP2 and the neuroanatomy of speech and language. Nat Rev Neurosci 6:131-138.
Vernes SC, Oliver PL, Spiteri E, Lockstone HE, Puliyadi R, Taylor JM, Ho J, Mombereau C, Brewer A, Lowy E, Nicod J, Groszer M, Baban D, Sahgal N, Cazier JB, Ragoussis J, Davies KE, Geschwind DH, Fisher SE (2011) Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. PLoS Genet 7:e1002145.
Villanueva P et al. (2015) Exome sequencing in an admixed isolated population indicates NFXL1 variants confer a risk for specific language impairment. PLoS Genet 11:e1004925.
Visel A, Minovitsky S, Dubchak I, Pennacchio LA (2007) VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res 35:D88-92.
Visel A et al. (2013) A high-resolution enhancer atlas of the developing telencephalon. Cell 152:895-908.
Wichterle H, Garcia-Verdugo JM, Alvarez-Buylla A (1997) Direct evidence for homotypic, glia-independent neuronal migration. Neuron 18:779-791.
Wisniewska MB, Misztal K, Michowski W, Szczot M, Purta E, Lesniak W, Klejman ME, Dabrowski M, Filipkowski RK, Nagalski A, Mozrzymas JW, Kuznicki J (2010) LEF1/beta-catenin complex regulates transcription of the Cav3.1 calcium channel gene (Cacna1g) in thalamic neurons of the adult brain. The Journal of neuroscience : the official journal of the Society for Neuroscience 30:4957-4969.

Appendix 1

Appendix table 1: Expression primer

Gene	Exon	Direction	Sequence
FOXP2	5	Forward	ACAGCATCCTGGAAAGCAAG
FOXP2	6	Reverse	ATGGAGATGAGTCCCTGACG
GAPDH	2	Forward	AAGGTGAAGGTCGGAGTCAAC
GAPDH	3	Reverse	GGGGTCATTGATGGCAACAATA

Appendix table 2: ChIP-qPCR Primer

		Forward Primer			Reverse Primer			
Chromosome	**ID**	**Start [hg19]**	**End [hg19]**	**Sequence**	**Start [hg19]**	**End [hg19]**	**Sequence**	**Amplicon Size**
chr7	1	113724817	113724837	GTTCAAGTTGGCCGTTCTTTC	113724901	113724923	TGAAACTTTCCTCCAAATAGCTG	106
chr7	2	113726213	113726234	GCGAGCAGTGTGTGTACAGTTC	113726290	113726306	GTGTGTCTGTGTGTGTGAATGTG	93
chr7	3	113727436	113727458	TGGCCTGGTTTTACTTTTATTCC	113727572	113727593	AGGTAACAGCCAAGTGAGAAGC	157
chr7	4	114053697	114053719	AAGGTGAAATGTGAAATTTGCTG	114053820	114053842	ACATCTGCCACTACTTTTGTTGC	145
chr7	5	114054910	114054932	TGGTGCTTTTGTCTCTCTCTCTC	114055046	114055068	GGTGACAAAGGTTCAAGACTGAG	158
chr7	6	114056133	114056155	AGCTTGGCTGGGCTATAAATTAC	114056230	114056252	GTCATCAGGCAATTACTGTAGCC	119
chr7	7	114057557	114057579	TTGCAAGTCCTGTAATGATGATG	114057648	114057670	GTGATATTGAAACAGCCCAATTC	113
chr7	8	114065128	114065151	TGTATTTAGTGTTTGTGTGCATCG	114065193	114065215	AATGCAAGCTTCCTATCTCTTCC	87
chr7	9	114066378	114066400	GATTAGCATCTGGATGAAAGCTG	114066505	114066527	ACACGTTACACCTTCACACTTCC	149
chr7	10	114067773	114067794	AAGTTAAAGCAAATGCCCACAC	114067874	114067900	GATTATTCTTGAAAACAATGCCTAAAG	127
chr12	GAPDH promoter	6643535	6643554	CGGCTACTAGCGGTTTTACG	6643653	6643672	GCTGCGGGCTCAATTTATAG	137
chr22	Myoglobin Exon2	36013261	36013280	CAGACGTTCAGCACCAACTG	36013332	36013351	ACCCAGTGAGCCCATACTTG	90

Appendix table 3: 3C target and anchor primer

Chromosome	Start [hg19]	End [hg19]	ID [distance to TSS1]	Sequence
chr7	113619659	113619682	-106682	TTCTTTCAGTACCACTGCTAGGTG
chr7	113652257	113652276	-74088	TTCTCTCCCCTCCCCATATC
chr7	113658218	113658234	-68130	CCACATGGGAGGAGCAAG
chr7	113667023	113667041	-59323	CCCCATTAAATGCTTGCAG
chr7	113686885	113686904	-39460	GTTCCTTTTCTGGCAGGATG
chr7	113688356	113688375	-37989	CAATGCCTTGGTCATTGTTG
chr7	113689702	113689721	-36643	CACACTGATGGATGGCTCAC
chr7	113706877	113706896	-19468	GCAAGCCAAAATTCCAAATC
chr7	113708164	113708186	-18178	TTGCTCATAGACACAAATCAAGG
chr7	113710846	113710867	-15497	AAAACCAGGAAACTGACATTGG
chr7	113715771	113715790	-10574	GGGTCTGCTTCCAAGTTCAG
chr7	113721054	113721074	-5290	TCTTGCATCAGGAAATTCCAC
chr7	113732140	113732161	Anchor 5797	TGAAAAGCACATTGCTGTGAAG
chr7	113739754	113739773	13409	CTCCCTTGGACACCAGAATC
chr7	113743011	113743032	16668	TTGTTCAGTGCTTGACTTTTGG
chr7	113755630	113755649	29285	GCCAGGAATATTGGCTGTTG
chr7	113764551	113764570	38206	TCCAATGGTGCTCACAAAAG
chr7	113772692	113772711	46347	TAGATGGAATGGCCAGGAAG
chr7	113779379	113779407	53043	GTTGTTTTCGGGGAGGTTG
chr7	113792245	113792265	65901	TGTGTGGCACAATGATGAAAC
chr7	113796607	113796626	70262	CCACCTTTCAAGACCAGACC
chr7	113803849	113803870	77506	CTGGAATGTTACCTAGGGCATC
chr7	113815841	113815863	89499	CGATTCTCTTCCCTGTAGAATCC
chr7	113978968	113978990	252626	TTCCATGCTCTGGAGATTAAAAC
chr7	113988151	113988176	261812	GAGAATTACGTGGTAAATCACAGTTG
chr7	113998725	113998744	272380	AGAGACTGGAGGCATTTTGC
chr7	114034128	114034148	307784	TCCAAAAGCAATTAGCACACC
chr7	114055582	114055601	329237	GGGAGACCAGACACAGGAAC
chr7	114056404	114056423	330059	TTCCCAGTGACATGAGCAAG
chr7	114061389	114061413	335049	TTTCATTGGCTTTACTTATGTTTCC
chr7	114068822	114068845	342481	TGATGCAGTTATTGCCTTTAATTC
chr7	114072519	114072541	346177	AACTAAGGTTTTGCAGCTGACTG
chr7	114079808	114079827	353463	ATGTGGAGACAGGGTTTTGG
chr7	114329855	114329874	603510	GGAAGAGCCAGTGATTGCAG
chr7	114329855	114329874	603510	TTGGAAAGGGCTACAGTGACA
chr7	114329855	114329874	603510	CACCCATTTCCTCTTTTTCCA
chr7	114408672	114408691	682327	GTCGTCTTGAGAGGCAGCAC
chr7	114432476	114432495	706131	GAAGTTGACCCACCACTCCA
chr7	114449271	114449290	722926	ACAGCACAGTCAAGGCCACT
chr7	114463228	114463247	736883	TGGGCCTCTGTTGTCTCAGT
chr7	114478670	114478689	752325	GCAGAACAAGGGTGGACAAA
chr7	114498528	114498547	772183	AGCTTTTCTGGGGAGCTGAC
chr7	114508327	114508347	781983	TAAACCTGGGAGGTGGAGGT
chr7	114535036	114535055	808691	TCCTTTAGCGCTGCTCTGAT
chr7	114549286	114549307	822943	CAAGCATTTGTCCTTTCTTTGA
chr7	114559396	114559415	833051	CAGTGGCCCACATAGCTCAC
chr7	114568978	114568997	842633	CAGCTGTGAGAACCCCTGAG
chr7	114590707	114590726	864362	CAGGGGAGGTAAAGGCTCAA
chr7	115112801	115112820	1386456	CACAAGGGCCCAAGTTACAA
chr7	115117519	115117538	1391174	GGAGATCCCGAGAAAAGCTC

Appendix 2

Table S1: Cloning Primer

Reg. Element	Orientation	Start [hg19]	End [hg19]	Sequence	Size
Element 37	Forward	113688009	113688029	ATCAGGTACCCACAGGGGAACCTCTGCAAAACTGC	20
Element 37	Reverse	113475999	113476018	ATCACTCGAGCACAATTTGTGGTGGGTGGGTGGG	19
Element 330	Forward	114056847	114056867	ATCAGGTACCGGTTTTCACGGAAAGCAGAGT	20
Element 330	Reverse	114058626	114058647	ATCAGAGCTCGACGTCTGTTCTGTGAGTTTCC	21
Element 815	Forward	114541369	114541389	ATCAGGTACCAGTATTGGCACACCCAGCAA	20
Element 815	Reverse	114542182	114542201	ATCAGAGCTCTGTTCCTAGGCGTGTCTGTG	19
Element 700	Forward	114424203	114424224	ATCAGAGCTCACCAGAGGAAACCCTTCTTACC	21
Element 700	Reverse	114427828	114427847	ATCACTCGAGAAGCATGTCCCAGAAAGC	19
Element 704	Forward	114427887	114427907	ATCAGGTACCTCAGACACAACCGACATTCC	20
Element 704	Reverse	114431955	114431974	ATCACTCGAGACCTCGAGCACATGTTTTCC	19
Element 843	Forward	114568454	114568473	ATCAGGTACCAGTGTGATGGCAGAACATGG	19
Element 843	Reverse	114572392	114572411	ATCACTCGAGACAACTCCCCAATTTCTCGC	19
Promoter 1	Forward	113724817	113724838	ATCAGAGCTCGTTCAAGTTGGCCGTTCTTTCC	21
Promoter 1	Reverse	113726588	113726609	ATCACTCGAGAAGCCCCCAAACAACACACC	21
Promoter 2	Forward	114051220	114051242	ATCAGAGCTCCTCACATCGCCAAAGGTATACC	22
Promoter 2	Reverse	114055298	114055324	ATCACTCGAGGCACAGAAAACAGTATAAAAAGAAGGG	26
Promoter 3	Forward	114055454	114055475	ATCAGAGCTCTAGAGAGGGGTGGGATTTTACG	21
Promoter 3	Reverse	114056438	114056459	ATCACTCGAGGGTCAGAGGAAAAAGCTGTTGG	21

Table S2: Predicted TF motifs

Enhancer 37, 330 and 700		
Reg. Element	**Transcription Factor**	**# of motifs/ element**
Enhancer 37	AML	1
Enhancer 37	AML1	7
Enhancer 37	AP1	24
Enhancer 37	AP1FJ	3
Enhancer 37	AP2A	1
Enhancer 37	AP2C	1
Enhancer 37	AP3	1
Enhancer 37	AP4	3
Enhancer 37	ARP1	1
Enhancer 37	ATF3	1
Enhancer 37	BACH1	1
Enhancer 37	BACH2	1
Enhancer 37	BARX1	3
Enhancer 37	BARX2	4
Enhancer 37	BATF	2
Enhancer 37	BATF::JUN	2
Enhancer 37	BCL6	9
Enhancer 37	BLIMP1	2
Enhancer 37	CART1	4
Enhancer 37	CEBP	4
Enhancer 37	CEBPA	3
Enhancer 37	CEBPB	5
Enhancer 37	CEBPD	1
Enhancer 37	CEBPDELTA	2
Enhancer 37	CEBPE	1
Enhancer 37	CEBPG	1
Enhancer 37	CEBPGAMMA	2
Enhancer 37	CMYB	1
Enhancer 37	COUPTF	2
Enhancer 37	CP2	2
Enhancer 37	CPEB1	5
Enhancer 37	DBP	2
Enhancer 37	EBOX	1
Enhancer 37	EGR2	4
Enhancer 37	EHF	2
Enhancer 37	ELF1	6
Enhancer 37	ELF3	2
Enhancer 37	Elf5	3
Enhancer 37	ELK3	2
Enhancer 37	EPAS1	1
Enhancer 37	ERR3	1
Enhancer 37	ESE1	2
Enhancer 37	ESR1	2
Enhancer 37	ESR2	4
Enhancer 37	ESRRA	1
Enhancer 37	ESRRG	1
Enhancer 37	ETS	2
Enhancer 37	ETS1	3
Enhancer 37	ETS2	2
Enhancer 37	ETV4	1
Enhancer 37	ETV5	1
Enhancer 37	FEV	5
Enhancer 37	FOS	5
Enhancer 37	FOSB	2
Enhancer 37	FOSL1	5
Enhancer 37	FOSL2	5
Enhancer 37	FOXH1	1
Enhancer 37	Foxj3	7
Enhancer 37	FOXO4	1
Enhancer 37	FRA1	3
Enhancer 37	GABP1+GABP2	1
Enhancer 37	GATA3	6
Enhancer 37	GCM1	5
Enhancer 37	GLI1	1
Enhancer 37	GLI3	2
Enhancer 37	HBP1	3
Enhancer 37	HELIOSA	2
Enhancer 37	HES1	2
Enhancer 37	HLTF	1
Enhancer 37	HMGA2	1
Enhancer 37	HMGIY	2
Enhancer 37	HNF1	3
Enhancer 37	HNF4	2
Enhancer 37	HNF4G	1
Enhancer 37	HOXA4	1
Enhancer 37	HOXA5	1
Enhancer 37	HOXD8	1

Enhancer 704, 815 and 843		
Reg. Element	**Transcription Factor**	**# of motifs/ element**
Enhancer 704	ALX3	4
Enhancer 704	AML1	4
Enhancer 704	AP1	18
Enhancer 704	AP1FJ	2
Enhancer 704	ARI3A	1
Enhancer 704	ARX	6
Enhancer 704	Atoh1	2
Enhancer 704	BARHL1	1
Enhancer 704	BARX1	1
Enhancer 704	BARX2	2
Enhancer 704	BATF	2
Enhancer 704	BCL6	5
Enhancer 704	BHLHA15	2
Enhancer 704	BHLHE23	1
Enhancer 704	BPTF	1
Enhancer 704	BRCA1	1
Enhancer 704	CART1	4
Enhancer 704	CDP	2
Enhancer 704	CDX2	3
Enhancer 704	CEBPGAMMA	1
Enhancer 704	CEBPZ	1
Enhancer 704	COT2	2
Enhancer 704	COUP	1
Enhancer 704	COUPTF	1
Enhancer 704	CPEB1	2
Enhancer 704	CTCF	1
Enhancer 704	DEC	1
Enhancer 704	DLX1	2
Enhancer 704	DLX2	2
Enhancer 704	DLX3	2
Enhancer 704	DLX7	1
Enhancer 704	DR1	1
Enhancer 704	DR3	1
Enhancer 704	EAR2	1
Enhancer 704	EFC	1
Enhancer 704	EHF	1
Enhancer 704	ELF1	1
Enhancer 704	EMX2	1
Enhancer 704	EN1	2
Enhancer 704	EN2	2
Enhancer 704	ERR1	2
Enhancer 704	ERR2	1
Enhancer 704	ESX1	3
Enhancer 704	EVI1	3
Enhancer 704	EVX1	1
Enhancer 704	FEV	1
Enhancer 704	FOS	3
Enhancer 704	FOSB	3
Enhancer 704	FOSL1	2
Enhancer 704	FOSL2	6
Enhancer 704	FOXB1	7
Enhancer 704	FOXC1	5
Enhancer 704	FOXC2	3
Enhancer 704	FOXI1	3
Enhancer 704	FOXJ2	8
Enhancer 704	Foxj3	6
Enhancer 704	FOXL1	6
Enhancer 704	FOXM1	4
Enhancer 704	FOXO1	5
Enhancer 704	FOXO3	4
Enhancer 704	FRA1	2
Enhancer 704	GATA1	2
Enhancer 704	GATA3	5
Enhancer 704	GATA4	2
Enhancer 704	GATA5	2
Enhancer 704	GATA6	2
Enhancer 704	GBX1	1
Enhancer 704	GBX2	4
Enhancer 704	GCR	1
Enhancer 704	GFI1	2
Enhancer 704	GFI1B	1
Enhancer 704	GKLF	2
Enhancer 704	GLI	1
Enhancer 704	GLI3	6
Enhancer 704	GR	1
Enhancer 704	GSX1	1
Enhancer 704	HAND1	2

Promoter 1, 2 and 3		
Reg. Element	**Transcription Factor**	**# of motifs/ element**
Promoter 1	AHR	2
Promoter 1	AHRHIF	1
Promoter 1	AML	1
Promoter 1	AP2	5
Promoter 1	AP2ALPHA	4
Promoter 1	AP2B	1
Promoter 1	AP2D	1
Promoter 1	AP2GAMMA	1
Promoter 1	AREB6	1
Promoter 1	ARNT2	1
Promoter 1	Ascl2	1
Promoter 1	ATF	2
Promoter 1	ATF1	3
Promoter 1	ATF2+ATF4	2
Promoter 1	ATF3	1
Promoter 1	ATF4	2
Promoter 1	BHLHE22	1
Promoter 1	BHLHE23	2
Promoter 1	BPTF	3
Promoter 1	BRCA	1
Promoter 1	BRCA1	1
Promoter 1	CEBPZ	1
Promoter 1	CETS1P54	1
Promoter 1	CHCH	1
Promoter 1	CKROX	2
Promoter 1	COE1	1
Promoter 1	CP2	5
Promoter 1	CREB	10
Promoter 1	CREB1	4
Promoter 1	CREB3	1
Promoter 1	CREBATF	2
Promoter 1	CREBP1	2
Promoter 1	CREBP1CJUN	2
Promoter 1	CREM	2
Promoter 1	CXXC1	1
Promoter 1	E12	1
Promoter 1	E2A	1
Promoter 1	E2F	4
Promoter 1	E2F1	6
Promoter 1	E2F6	1
Promoter 1	E47	1
Promoter 1	EBF	1
Promoter 1	EBF1	1
Promoter 1	EGR	1
Promoter 1	EGR1	1
Promoter 1	EGR2	2
Promoter 1	EGR3	2
Promoter 1	EGR4	2
Promoter 1	EHF	1
Promoter 1	ELF2	1
Promoter 1	ELF5	2
Promoter 1	ELK1	3
Promoter 1	ELK3	1
Promoter 1	ELK4	1
Promoter 1	EPAS1	2
Promoter 1	ER	1
Promoter 1	ERF	1
Promoter 1	ERG	2
Promoter 1	ESR2	1
Promoter 1	ETF	2
Promoter 1	ETS	1
Promoter 1	ETS1	2
Promoter 1	ETV1	1
Promoter 1	ETV2	1
Promoter 1	ETV4	1
Promoter 1	ETV5	2
Promoter 1	EVI1	4
Promoter 1	FAC1	2
Promoter 1	FEV	1
Promoter 1	FIGLA	2
Promoter 1	FLI1	2
Promoter 1	FOS	3
Promoter 1	FOXA1	7
Promoter 1	FOXA2	4
Promoter 1	FOXA3	4
Promoter 1	FOXB1	6
Promoter 1	FOXC1	12

Enhancer 37	HXD4	1
Enhancer 37	HXD9	3
Enhancer 37	ICSBP	2
Enhancer 37	IRF	3
Enhancer 37	IRF1	5
Enhancer 37	IRF2	2
Enhancer 37	IRF3	5
Enhancer 37	IRF4	2
Enhancer 37	IRF5	2
Enhancer 37	IRF7	3
Enhancer 37	IRF8	5
Enhancer 37	IRF9	1
Enhancer 37	ISL1	1
Enhancer 37	Jdp2	3
Enhancer 37	JUN	3
Enhancer 37	JUN(var.2)	2
Enhancer 37	JUN::FOS	2
Enhancer 37	JUNB	4
Enhancer 37	JUND	3
Enhancer 37	KROX	1
Enhancer 37	LBX2	1
Enhancer 37	LEF1	2
Enhancer 37	LHX2	1
Enhancer 37	LHX4	1
Enhancer 37	LMO2COM	2
Enhancer 37	MAF	1
Enhancer 37	Mafb	2
Enhancer 37	MAFG	3
Enhancer 37	MAFK	5
Enhancer 37	MATH1	1
Enhancer 37	MAZ	1
Enhancer 37	MCR	1
Enhancer 37	MEF2	1
Enhancer 37	MEF2C	2
Enhancer 37	MITF	1
Enhancer 37	MSX1	1
Enhancer 37	MSX2	1
Enhancer 37	Msx3	1
Enhancer 37	MYB	1
Enhancer 37	MYBL1	1
Enhancer 37	MYCMAX	1
Enhancer 37	MYOD	1
Enhancer 37	MYOG	1
Enhancer 37	MZF1	2
Enhancer 37	NEUROD	1
Enhancer 37	NF2L1	1
Enhancer 37	NF2L2	1
Enhancer 37	NFAC2	1
Enhancer 37	NFAC4	1
Enhancer 37	NFAT	3
Enhancer 37	NFAT1	1
Enhancer 37	NFAT2	2
Enhancer 37	NFATC2	1
Enhancer 37	NFE2::MAF	1
Enhancer 37	NFIA+NFIB+NFIC+NFIX	2
Enhancer 37	NFIX	1
Enhancer 37	NKX2-3	4
Enhancer 37	NKX25	3
Enhancer 37	NKX28	2
Enhancer 37	NKX2-8	4
Enhancer 37	NKX3-1	2
Enhancer 37	NKX32	3
Enhancer 37	NOTO	1
Enhancer 37	Nr2e1	2
Enhancer 37	NR2F6	1
Enhancer 37	NR4A1	1
Enhancer 37	NR4A2	4
Enhancer 37	NUR77	1
Enhancer 37	NURR1	1
Enhancer 37	OCT	5
Enhancer 37	OCT1	14
Enhancer 37	OCT2	2
Enhancer 37	OCTAMER	1
Enhancer 37	P53	4
Enhancer 37	PARP	2
Enhancer 37	PAX	2
Enhancer 37	PAX2	3
Enhancer 37	PAX8	5
Enhancer 37	PEA3	1
Enhancer 37	PEBP	1
Enhancer 37	PITX2	2
Enhancer 37	PO2F1	3
Enhancer 37	PO2F2	2
Enhancer 37	PO3F1	2
Enhancer 37	PO3F2	2

Enhancer 704	HAND1E47	2
Enhancer 704	HB9	1
Enhancer 704	HESX1	3
Enhancer 704	Hic1	3
Enhancer 704	HINFP	1
Enhancer 704	HLF	2
Enhancer 704	HLTF	1
Enhancer 704	HMBOX1	2
Enhancer 704	HMGA1	2
Enhancer 704	HMGIY	2
Enhancer 704	HMX2	1
Enhancer 704	HMX3	1
Enhancer 704	HNF1	3
Enhancer 704	HNF1A	2
Enhancer 704	HNF1B	4
Enhancer 704	HNF3	5
Enhancer 704	HNF4	3
Enhancer 704	HNF4ALPHA	1
Enhancer 704	HOX13	1
Enhancer 704	HOXA10	1
Enhancer 704	Hoxa11	1
Enhancer 704	HOXA13	7
Enhancer 704	HOXA2	2
Enhancer 704	HOXA4	1
Enhancer 704	HOXA5	3
Enhancer 704	HOXA6	1
Enhancer 704	HOXA7	2
Enhancer 704	HOXB13	2
Enhancer 704	HOXB5	1
Enhancer 704	HOXC10	2
Enhancer 704	HOXC12	1
Enhancer 704	HOXC13	4
Enhancer 704	HOXC5	1
Enhancer 704	HOXC8	1
Enhancer 704	HOXD12	3
Enhancer 704	HOXD13	7
Enhancer 704	HOXD3	1
Enhancer 704	HOXD8	1
Enhancer 704	Hoxd9	2
Enhancer 704	HSF1	2
Enhancer 704	HTF4	1
Enhancer 704	HXA5	1
Enhancer 704	HXA7	1
Enhancer 704	HXD10	2
Enhancer 704	IK	1
Enhancer 704	IKZF1	1
Enhancer 704	INSM1	1
Enhancer 704	IPF	1
Enhancer 704	IPF1	4
Enhancer 704	IRF	1
Enhancer 704	IRF1	4
Enhancer 704	IRF3	1
Enhancer 704	IRF4	1
Enhancer 704	IRF5	1
Enhancer 704	IRF8	2
Enhancer 704	IRX2	2
Enhancer 704	Irx3	2
Enhancer 704	ISL2	1
Enhancer 704	ITF2	1
Enhancer 704	JUN	4
Enhancer 704	JUN(var.2)	4
Enhancer 704	JUN::FOS	2
Enhancer 704	JUNB	5
Enhancer 704	JUND	5
Enhancer 704	K2B	1
Enhancer 704	KLF1	2
Enhancer 704	Klf12	1
Enhancer 704	KLF3	1
Enhancer 704	KLF8	1
Enhancer 704	LBX2	1
Enhancer 704	LEF1	4
Enhancer 704	LEF1TCF1	1
Enhancer 704	LH2	1
Enhancer 704	LHX2	2
Enhancer 704	LHX3	1
Enhancer 704	LHX4	1
Enhancer 704	LHX6	1
Enhancer 704	LHX61	2
Enhancer 704	LMO2COM	1
Enhancer 704	LTF	1
Enhancer 704	LYF1	1
Enhancer 704	MAX	1
Enhancer 704	MEF2	4
Enhancer 704	MEF2A	5
Enhancer 704	MEF2C	3

Promoter 1	FOXC2	6
Promoter 1	FOXD1	6
Promoter 1	FOXD3	9
Promoter 1	FOXF1	4
Promoter 1	FOXF2	4
Promoter 1	Foxg1	3
Promoter 1	FOXI1	6
Promoter 1	FOXJ2	12
Promoter 1	Foxj3	12
Promoter 1	FOXL1	4
Promoter 1	FOXM1	3
Promoter 1	FOXO1	11
Promoter 1	FOXO3	11
Promoter 1	FOXO4	9
Promoter 1	FOXO6	3
Promoter 1	FOXP2	6
Promoter 1	FOXP3	7
Promoter 1	FPM315	2
Promoter 1	FREAC2	3
Promoter 1	FREAC4	3
Promoter 1	FREAC7	2
Promoter 1	FUBP1	2
Promoter 1	GABP1+GABP2	1
Promoter 1	GABPALPHA	1
Promoter 1	GATA	2
Promoter 1	GATA3	2
Promoter 1	GFI1	3
Promoter 1	GKLF	1
Promoter 1	GLI2	1
Promoter 1	GLI3	1
Promoter 1	HEB	2
Promoter 1	HES1	2
Promoter 1	HEY2	1
Promoter 1	HFH3	3
Promoter 1	HFH8	3
Promoter 1	HIC1	1
Promoter 1	HIF1	1
Promoter 1	HIF1A	1
Promoter 1	HINFP	2
Promoter 1	HINFP1	1
Promoter 1	HNF3	5
Promoter 1	HNF3A	3
Promoter 1	HNF3ALPHA	4
Promoter 1	HNF3B	2
Promoter 1	HOXA13	5
Promoter 1	HSF1	1
Promoter 1	HSFY2	4
Promoter 1	HTF4	1
Promoter 1	HXA10	1
Promoter 1	IRF1	5
Promoter 1	IRF4	3
Promoter 1	KLF1	1
Promoter 1	KLF15	2
Promoter 1	KLF16	2
Promoter 1	KLF4	2
Promoter 1	KLF5	2
Promoter 1	KLF6	1
Promoter 1	KLF8	2
Promoter 1	KROX	1
Promoter 1	LBP1	2
Promoter 1	LHX61	1
Promoter 1	MATH1	2
Promoter 1	MAZ	4
Promoter 1	MAZR	1
Promoter 1	MBD2	1
Promoter 1	MCR	1
Promoter 1	MECP2	1
Promoter 1	MEF2	4
Promoter 1	MEIS3	3
Promoter 1	MGA	1
Promoter 1	MLXPL	1
Promoter 1	MYCMAX	2
Promoter 1	MYF6	2
Promoter 1	MYOD	1
Promoter 1	MYOD1	1
Promoter 1	MYOG	1
Promoter 1	MYOGENIN	1
Promoter 1	MZF1	10
Promoter 1	NANOG	3
Promoter 1	NDF1	2
Promoter 1	NEUROD	1
Promoter 1	NEUROD2	2
Promoter 1	NFAT1	1
Promoter 1	NFATC1	1
Promoter 1	NFIA	1

Enhancer 37	PO4F2	1	Enhancer 704	MEF2D	2	Promoter 1	NFIA+NFIB+NFIC+NFIX	1
Enhancer 37	POU1F1	5	Enhancer 704	MEIS1	1	Promoter 1	NFIC	1
Enhancer 37	POU2F1	6	Enhancer 704	MEIS1BHOXA9	1	Promoter 1	NFIX	2
Enhancer 37	POU2F2	8	Enhancer 704	MEIS3	1	Promoter 1	NFKB	1
Enhancer 37	POU2F3	9	Enhancer 704	MEOX1	1	Promoter 1	NFY	1
Enhancer 37	POU3F1	6	Enhancer 704	MEOX2	4	Promoter 1	NFYA	2
Enhancer 37	POU3F2	5	Enhancer 704	MESP1	1	Promoter 1	NFYB	2
Enhancer 37	POU3F3	9	Enhancer 704	MGA	1	Promoter 1	NFYC	1
Enhancer 37	POU3F4	3	Enhancer 704	MITF	1	Promoter 1	NHLH1	2
Enhancer 37	POU4F1	2	Enhancer 704	MMEF2	1	Promoter 1	NKX2-3	2
Enhancer 37	POU5F1	3	Enhancer 704	MNX1	1	Promoter 1	NKX25	6
Enhancer 37	POU5F1P1	6	Enhancer 704	MSX1	2	Promoter 1	NKX2-8	2
Enhancer 37	PRDM1	5	Enhancer 704	MSX2	2	Promoter 1	NR0B1	2
Enhancer 37	PRRX1	2	Enhancer 704	MYBB	1	Promoter 1	NR2F6	1
Enhancer 37	PU1	2	Enhancer 704	NANOG	1	Promoter 1	NRF1	2
Enhancer 37	RAX	1	Enhancer 704	NCX	2	Promoter 1	NUR77	2
Enhancer 37	RFX5	1	Enhancer 704	NEUROD2	1	Promoter 1	OCT1	14
Enhancer 37	Rhox11	1	Enhancer 704	NFAC1	3	Promoter 1	OLIG2	4
Enhancer 37	RREB1	2	Enhancer 704	NFAC2	2	Promoter 1	OLIG3	2
Enhancer 37	RUNX2	4	Enhancer 704	NFAT2	1	Promoter 1	P50P50	1
Enhancer 37	RUNX3	1	Enhancer 704	NFAT5	2	Promoter 1	PAX3	1
Enhancer 37	SF1	1	Enhancer 704	NFIA	1	Promoter 1	PAX4	2
Enhancer 37	SIX4	1	Enhancer 704	NFKB1	1	Promoter 1	PAX5	3
Enhancer 37	SMAD1	2	Enhancer 704	NFY	1	Promoter 1	PAX8	4
Enhancer 37	SMAD2	1	Enhancer 704	NFYA	1	Promoter 1	PBX3	1
Enhancer 37	SMAD3	2	Enhancer 704	NFYB	1	Promoter 1	PLAG1	3
Enhancer 37	SMAD4	1	Enhancer 704	NKX22	1	Promoter 1	PPARD	1
Enhancer 37	SMRC1	1	Enhancer 704	NKX2-3	2	Promoter 1	PURA	2
Enhancer 37	SOX10	8	Enhancer 704	NKX25	2	Promoter 1	RARA	1
Enhancer 37	SOX13	1	Enhancer 704	NKX2-8	2	Promoter 1	RARB	1
Enhancer 37	SOX15	2	Enhancer 704	NKX31	2	Promoter 1	RELB	1
Enhancer 37	SOX18	2	Enhancer 704	NKX3-1	4	Promoter 1	RFX1	1
Enhancer 37	SOX4	2	Enhancer 704	NKX32	3	Promoter 1	RREB1	4
Enhancer 37	SOX5	5	Enhancer 704	NKX3-2	1	Promoter 1	SMAD	1
Enhancer 37	SOX9	3	Enhancer 704	NKX3A	4	Promoter 1	SMAD2	2
Enhancer 37	SPIB	1	Enhancer 704	NKX61	2	Promoter 1	SNAI1	1
Enhancer 37	SRY	7	Enhancer 704	NKX62	2	Promoter 1	SNAI2	1
Enhancer 37	STAT1	2	Enhancer 704	NOBOX	1	Promoter 1	SOX13	1
Enhancer 37	STAT2	1	Enhancer 704	NR1D1	1	Promoter 1	SOX15	1
Enhancer 37	STAT2::STAT1	1	Enhancer 704	NR2C2	1	Promoter 1	SOX4	2
Enhancer 37	STAT3	3	Enhancer 704	Nr2e1	3	Promoter 1	SOX5	2
Enhancer 37	STAT4	1	Enhancer 704	NR2E3	2	Promoter 1	SOX9	1
Enhancer 37	STAT6	1	Enhancer 704	NR2F1	1	Promoter 1	SP1	16
Enhancer 37	STF1	1	Enhancer 704	NR4A1	1	Promoter 1	SP1SP3	1
Enhancer 37	TF7L2	1	Enhancer 704	NR4A2	1	Promoter 1	SP2	4
Enhancer 37	TFAP2A	4	Enhancer 704	NURR1	2	Promoter 1	SP3	4
Enhancer 37	TFAP2C	4	Enhancer 704	OCT	3	Promoter 1	SP4	3
Enhancer 37	TFAP4	1	Enhancer 704	OCT1	11	Promoter 1	SPDEF	2
Enhancer 37	TFCP2	2	Enhancer 704	OCT2	1	Promoter 1	SPIB	1
Enhancer 37	TFE3	3	Enhancer 704	OCTAMER	2	Promoter 1	SRBP2	2
Enhancer 37	TFEB	2	Enhancer 704	OLIG2	4	Promoter 1	STAT1	2
Enhancer 37	TFEC	2	Enhancer 704	OLIG3	2	Promoter 1	STAT3	1
Enhancer 37	THAP1	1	Enhancer 704	OSF2	1	Promoter 1	SZF11	1
Enhancer 37	TLX1	1	Enhancer 704	OTX2	5	Promoter 1	TAL1	3
Enhancer 37	UBIP1	2	Enhancer 704	OVOL1	1	Promoter 1	TAL1::TCF3	2
Enhancer 37	USF1	2	Enhancer 704	PAX	2	Promoter 1	TBX1	1
Enhancer 37	USF2	1	Enhancer 704	PAX4	1	Promoter 1	TBX15	1
Enhancer 330	AIRE	3	Enhancer 704	PAX5	4	Promoter 1	TBX5	2
Enhancer 330	Alx1	4	Enhancer 704	PAX8	1	Promoter 1	TCF4	1
Enhancer 330	ALX3	8	Enhancer 704	PBX1	2	Promoter 1	TCF7	1
Enhancer 330	ALX4	4	Enhancer 704	PIT1	2	Promoter 1	TCF7L2	1
Enhancer 330	AML	2	Enhancer 704	PITX2	3	Promoter 1	Tcfap2a	2
Enhancer 330	AML1	4	Enhancer 704	PMX2A	1	Promoter 1	TEF	1
Enhancer 330	AREB6	1	Enhancer 704	PO2F1	3	Promoter 1	TF7L2	1
Enhancer 330	ARI3A	3	Enhancer 704	PO2F2	1	Promoter 1	TFAP2A	2
Enhancer 330	ARI5B	1	Enhancer 704	PO3F1	1	Promoter 1	TFAP2B	1
Enhancer 330	ARX	2	Enhancer 704	PO3F2	2	Promoter 1	TFAP4	1
Enhancer 330	ATF4	1	Enhancer 704	PO4F2	1	Promoter 1	TFDP1	1
Enhancer 330	ATF5	1	Enhancer 704	PO5F1	2	Promoter 1	TFE2	1
Enhancer 330	Atoh1	1	Enhancer 704	POU1F1	4	Promoter 1	TFIII	1
Enhancer 330	BARHL1	5	Enhancer 704	POU2F1	5	Promoter 1	THA	1
Enhancer 330	BARHL2	10	Enhancer 704	POU2F2	6	Promoter 1	THAP1	1
Enhancer 330	BARX1	4	Enhancer 704	POU2F3	5	Promoter 1	THB	2
Enhancer 330	BARX2	2	Enhancer 704	POU3F1	5	Promoter 1	TLX1	1
Enhancer 330	BCL6	2	Enhancer 704	POU3F2	4	Promoter 1	TWST1	1
Enhancer 330	BHLHE22	1	Enhancer 704	POU3F3	5	Promoter 1	VDR	2
Enhancer 330	BHLHE23	1	Enhancer 704	POU3F4	3	Promoter 1	WHN	1
Enhancer 330	BRN2	2	Enhancer 704	POU4F1	2	Promoter 1	WT1	3
Enhancer 330	BRN4	1	Enhancer 704	POU4F3	2	Promoter 1	XBP1	2
Enhancer 330	BSX	2	Enhancer 704	POU5F1	3	Promoter 1	YBOX1	1
Enhancer 330	CART1	5	Enhancer 704	POU5F1P1	2	Promoter 1	ZBT7B	2
Enhancer 330	CDC5L	1	Enhancer 704	POU6F1	1	Promoter 1	ZBTB33	1
Enhancer 330	CDP	2	Enhancer 704	PPAR	1	Promoter 1	Zfp740	1
Enhancer 330	CDX	2	Enhancer 704	PPARA	1	Promoter 1	ZFX	3
Enhancer 330	CDX1	4	Enhancer 704	PPARD	1	Promoter 1	ZN148	2

Enhancer 330	CDX2	7
Enhancer 330	CEBP	8
Enhancer 330	CEBPA	3
Enhancer 330	CEBPB	1
Enhancer 330	CEBPD	1
Enhancer 330	CEBPE	1
Enhancer 330	CEBPG	1
Enhancer 330	CEBPZ	1
Enhancer 330	CHX10	1
Enhancer 330	CREB	1
Enhancer 330	CREBP1	1
Enhancer 330	CRX	3
Enhancer 330	CUX1	1
Enhancer 330	DBP	2
Enhancer 330	DLX1	5
Enhancer 330	DLX2	7
Enhancer 330	DLX3	4
Enhancer 330	DLX4	2
Enhancer 330	DLX5	4
Enhancer 330	DLX6	2
Enhancer 330	DMRT2	1
Enhancer 330	DMRT3	1
Enhancer 330	DRGX	2
Enhancer 330	DRI1	1
Enhancer 330	DUX4	1
Enhancer 330	E2F1	1
Enhancer 330	E4BP4	1
Enhancer 330	EGR1	1
Enhancer 330	EMX1	3
Enhancer 330	EMX2	6
Enhancer 330	EN1	8
Enhancer 330	EN2	7
Enhancer 330	ER	1
Enhancer 330	ESX1	4
Enhancer 330	EVX1	4
Enhancer 330	EVX2	2
Enhancer 330	FOS	1
Enhancer 330	FOXA1	3
Enhancer 330	FOXA2	1
Enhancer 330	FOXA3	2
Enhancer 330	FOXB1	2
Enhancer 330	FOXC1	6
Enhancer 330	FOXC2	3
Enhancer 330	FOXD1	2
Enhancer 330	FOXD2	2
Enhancer 330	FOXD3	4
Enhancer 330	FOXF1	1
Enhancer 330	FOXF2	2
Enhancer 330	Foxg1	1
Enhancer 330	FOXI1	3
Enhancer 330	FOXJ2	3
Enhancer 330	FOXJ3	2
Enhancer 330	Foxk1	1
Enhancer 330	FOXL1	4
Enhancer 330	FOXM1	1
Enhancer 330	FOXO1	6
Enhancer 330	FOXO3	5
Enhancer 330	FOXO3A	1
Enhancer 330	FOXO4	5
Enhancer 330	FOXO6	1
Enhancer 330	FOXP1	1
Enhancer 330	FOXP2	2
Enhancer 330	FOXP3	3
Enhancer 330	FREAC7	1
Enhancer 330	FUBP1	1
Enhancer 330	GATA1	2
Enhancer 330	GATA2	2
Enhancer 330	GATA3	2
Enhancer 330	GATA4	1
Enhancer 330	GATA5	2
Enhancer 330	GATA6	2
Enhancer 330	Gbx1	4
Enhancer 330	GBX2	9
Enhancer 330	GFI1	2
Enhancer 330	GR	2
Enhancer 330	GSC	1
Enhancer 330	GSX1	2
Enhancer 330	GSX2	2
Enhancer 330	HAND1	1
Enhancer 330	HAND1E47	1
Enhancer 330	HB9	1
Enhancer 330	HBP1	1
Enhancer 330	HELIOSA	1
Enhancer 330	HESX1	2
Enhancer 330	HFH3	1
Enhancer 330	HFH4	1
Enhancer 330	HLF	6

Enhancer 704	PR	2
Enhancer 704	PRGR	1
Enhancer 704	PRRX1	2
Enhancer 704	PRRX2	1
Enhancer 704	PU1	1
Enhancer 704	PXRRXR	2
Enhancer 704	RAX	1
Enhancer 704	RBPJK	1
Enhancer 704	RFX1	1
Enhancer 704	RORA	4
Enhancer 704	RORA1	1
Enhancer 704	RORA2	1
Enhancer 704	RORBETA	1
Enhancer 704	RORG	1
Enhancer 704	RUNX2	1
Enhancer 704	RXRB	1
Enhancer 704	RXRG	1
Enhancer 704	SATB1	1
Enhancer 704	SF1	3
Enhancer 704	SMAD1	3
Enhancer 704	SMAD3	2
Enhancer 704	SMAD4	2
Enhancer 704	SMRC1	3
Enhancer 704	SOX	1
Enhancer 704	SOX10	6
Enhancer 704	SOX13	1
Enhancer 704	SOX15	2
Enhancer 704	SOX2	1
Enhancer 704	SOX4	1
Enhancer 704	SOX5	4
Enhancer 704	SOX9	4
Enhancer 704	SOX91	1
Enhancer 704	SRF	2
Enhancer 704	SRY	6
Enhancer 704	STA5A	3
Enhancer 704	STA5B	2
Enhancer 704	STAT	1
Enhancer 704	STAT1	6
Enhancer 704	STAT1STAT1	1
Enhancer 704	STAT2	2
Enhancer 704	STAT2::STAT1	1
Enhancer 704	STAT3	4
Enhancer 704	STAT4	3
Enhancer 704	STAT5A	3
Enhancer 704	STAT5B	2
Enhancer 704	SUH	1
Enhancer 704	TAL1::TCF3	1
Enhancer 704	TBP	3
Enhancer 704	TBX1	1
Enhancer 704	TBX15	1
Enhancer 704	TBX20	1
Enhancer 704	TBX4	1
Enhancer 704	TBX5	1
Enhancer 704	TCF11	1
Enhancer 704	TCF3	2
Enhancer 704	TCF4	1
Enhancer 704	TEAD1	11
Enhancer 704	TEAD3	5
Enhancer 704	TEAD4	4
Enhancer 704	TEF	8
Enhancer 704	TEF1	2
Enhancer 704	TFE	1
Enhancer 704	TFE2	2
Enhancer 704	TFEB	1
Enhancer 704	TFEC	1
Enhancer 704	TGIF1	1
Enhancer 704	THA	2
Enhancer 704	TR4	1
Enhancer 704	TWST1	1
Enhancer 704	UBIP1	1
Enhancer 704	Uncx	3
Enhancer 704	USF	1
Enhancer 704	USF1	2
Enhancer 704	USF2	1
Enhancer 704	VAX1	1
Enhancer 704	VBP	2
Enhancer 704	VSX1	1
Enhancer 704	VSX2	1
Enhancer 704	ZABC1	2
Enhancer 704	ZEB1	1
Enhancer 704	ZFHX3	1
Enhancer 704	ZNF333	2
Enhancer 704	ZNF354C	1
Enhancer 815	AFP1	2
Enhancer 815	Alx1	5
Enhancer 815	ALX3	3
Enhancer 815	ALX4	6

Promoter 1	ZN219	2
Promoter 1	ZNF219	2
Promoter 1	ZNF263	2
Promoter 1	ZNF354C	1
Promoter 1	ZNF740	3
Promoter 2	AIRE	2
Promoter 2	ALX1	2
Promoter 2	ALX3	3
Promoter 2	ALX4	2
Promoter 2	AP2ALPHA	3
Promoter 2	AP3	1
Promoter 2	ARI3A	4
Promoter 2	ARI5B	1
Promoter 2	ATF5	2
Promoter 2	Barhl1	4
Promoter 2	BARHL2	7
Promoter 2	BARX1	5
Promoter 2	BARX2	2
Promoter 2	BCL6	3
Promoter 2	BPTF	1
Promoter 2	BRN4	2
Promoter 2	BSX	1
Promoter 2	CDC5	1
Promoter 2	CDC5L	1
Promoter 2	CDX1	2
Promoter 2	CDX2	3
Promoter 2	CEBP	2
Promoter 2	CEBPA	2
Promoter 2	CEBPB	1
Promoter 2	CEBPGAMMA	1
Promoter 2	CHOP	1
Promoter 2	COT2	1
Promoter 2	CP2	1
Promoter 2	CPEB1	1
Promoter 2	CRX	3
Promoter 2	CUX1	1
Promoter 2	DBP	1
Promoter 2	DLX1	3
Promoter 2	DLX2	6
Promoter 2	DLX3	3
Promoter 2	DLX4	1
Promoter 2	DLX5	2
Promoter 2	DLX6	1
Promoter 2	DLX7	1
Promoter 2	DMRT1	2
Promoter 2	DMRT2	1
Promoter 2	DMRT3	1
Promoter 2	DPRX	1
Promoter 2	DRI1	1
Promoter 2	EGR3	1
Promoter 2	EMX1	2
Promoter 2	EMX2	2
Promoter 2	EN1	6
Promoter 2	EN2	3
Promoter 2	EPAS1	1
Promoter 2	ERR1	1
Promoter 2	ERR3	1
Promoter 2	ESE1	1
Promoter 2	ESX1	2
Promoter 2	ETV7	1
Promoter 2	EVI1	3
Promoter 2	FAC1	1
Promoter 2	FOS	1
Promoter 2	FOX	2
Promoter 2	FOXA1	4
Promoter 2	FOXA2	2
Promoter 2	FOXA3	2
Promoter 2	FOXB1	4
Promoter 2	FOXC1	8
Promoter 2	FOXC2	4
Promoter 2	FOXD1	4
Promoter 2	FOXD2	4
Promoter 2	FOXD3	7
Promoter 2	FOXF1	2
Promoter 2	FOXF2	3
Promoter 2	Foxg1	2
Promoter 2	FOXI1	3
Promoter 2	FOXJ2	7
Promoter 2	Foxj3	6
Promoter 2	Foxk1	2
Promoter 2	FOXL1	3
Promoter 2	FOXM1	2
Promoter 2	FOXO1	6
Promoter 2	FOXO3	6
Promoter 2	FOXO3A	1
Promoter 2	FOXO4	4
Promoter 2	FOXO6	2

Enhancer 330	HMBOX1	1
Enhancer 330	HMGA1	2
Enhancer 330	HMGIY	4
Enhancer 330	HMX1	1
Enhancer 330	HMX2	1
Enhancer 330	HMX3	2
Enhancer 330	HNF1	4
Enhancer 330	HNF1A	1
Enhancer 330	HNF1B	3
Enhancer 330	HNF3	3
Enhancer 330	HNF3A	1
Enhancer 330	HNF3ALPHA	2
Enhancer 330	HNF3B	2
Enhancer 330	HNF4	2
Enhancer 330	HNF6	3
Enhancer 330	HOX13	2
Enhancer 330	HOXA1	3
Enhancer 330	HOXA10	5
Enhancer 330	Hoxa11	3
Enhancer 330	HOXA13	2
Enhancer 330	HOXA2	6
Enhancer 330	HOXA3	2
Enhancer 330	HOXA4	4
Enhancer 330	HOXA6	2
Enhancer 330	HOXA9	2
Enhancer 330	HOXB2	2
Enhancer 330	HOXB3	2
Enhancer 330	HOXB5	4
Enhancer 330	HOXB6	2
Enhancer 330	HOXB9	1
Enhancer 330	HOXC10	8
Enhancer 330	HOXC11	4
Enhancer 330	HOXC12	2
Enhancer 330	HOXC4	1
Enhancer 330	HOXC5	1
Enhancer 330	HOXC8	2
Enhancer 330	HOXC9	1
Enhancer 330	HOXD11	2
Enhancer 330	HOXD12	3
Enhancer 330	HOXD13	3
Enhancer 330	HOXD3	4
Enhancer 330	HOXD8	3
Enhancer 330	Hoxd9	3
Enhancer 330	HSF1	2
Enhancer 330	HSF2	2
Enhancer 330	HXA10	1
Enhancer 330	HXA13	1
Enhancer 330	HXA5	1
Enhancer 330	HXD10	1
Enhancer 330	HXD4	3
Enhancer 330	HXD9	3
Enhancer 330	INSM1	1
Enhancer 330	IPF	1
Enhancer 330	IPF1	14
Enhancer 330	IRF	4
Enhancer 330	IRF1	4
Enhancer 330	IRF3	4
Enhancer 330	IRF4	4
Enhancer 330	IRF7	4
Enhancer 330	IRX2	2
Enhancer 330	IRX4	2
Enhancer 330	IRX5	2
Enhancer 330	IRXB3	1
Enhancer 330	ISL1	2
Enhancer 330	ISL2	1
Enhancer 330	ISRE	1
Enhancer 330	ISX	4
Enhancer 330	JUN	3
Enhancer 330	JUND(var.2)	1
Enhancer 330	K2B	1
Enhancer 330	LBX2	4
Enhancer 330	LEF1	6
Enhancer 330	LH2	2
Enhancer 330	LHX2	6
Enhancer 330	LHX3	1
Enhancer 330	LHX4	5
Enhancer 330	LHX5	1
Enhancer 330	LHX6	2
Enhancer 330	LHX61	3
Enhancer 330	Lhx8	2
Enhancer 330	LHX9	2
Enhancer 330	LMX1	1
Enhancer 330	LMX1A	2
Enhancer 330	LMX1B	4
Enhancer 330	MAFG	4
Enhancer 330	MEF2A	1
Enhancer 330	MEF2C	2

Enhancer 815	AP1	13
Enhancer 815	AP1FJ	1
Enhancer 815	AP4	1
Enhancer 815	ARI5B	1
Enhancer 815	ARX	4
Enhancer 815	ATF3	1
Enhancer 815	ATF5	3
Enhancer 815	BACH2	1
Enhancer 815	BARHL2	1
Enhancer 815	BATF	2
Enhancer 815	BATF::JUN	1
Enhancer 815	BCL6	4
Enhancer 815	CART1	2
Enhancer 815	CDX	2
Enhancer 815	CDX1	1
Enhancer 815	CDX2	3
Enhancer 815	CEBP	8
Enhancer 815	CEBPA	6
Enhancer 815	CEBPB	6
Enhancer 815	CEBPD	2
Enhancer 815	CEBPDELTA	1
Enhancer 815	COT2	3
Enhancer 815	CRX	3
Enhancer 815	DBP	2
Enhancer 815	DMBX1	1
Enhancer 815	DMRT3	1
Enhancer 815	DPRX	1
Enhancer 815	DRGX	3
Enhancer 815	EPAS1	1
Enhancer 815	EVI1	7
Enhancer 815	FOS	2
Enhancer 815	FOSB	3
Enhancer 815	FOSL1	1
Enhancer 815	FOSL2	4
Enhancer 815	FOX	3
Enhancer 815	FOXA1	5
Enhancer 815	FOXA2	3
Enhancer 815	FOXA3	3
Enhancer 815	FOXB1	6
Enhancer 815	FOXC1	7
Enhancer 815	FOXC2	5
Enhancer 815	FOXD1	5
Enhancer 815	FOXD2	4
Enhancer 815	FOXD3	8
Enhancer 815	FOXF1	3
Enhancer 815	FOXF2	5
Enhancer 815	Foxg1	2
Enhancer 815	FOXI1	4
Enhancer 815	FOXJ2	7
Enhancer 815	FOXJ3	5
Enhancer 815	Foxk1	2
Enhancer 815	FOXL1	7
Enhancer 815	FOXM1	5
Enhancer 815	FOXO1	5
Enhancer 815	FOXO3	5
Enhancer 815	FOXO3A	1
Enhancer 815	FOXO4	5
Enhancer 815	FOXO6	2
Enhancer 815	FOXP1	1
Enhancer 815	FOXP2	5
Enhancer 815	FOXP3	4
Enhancer 815	FRA1	3
Enhancer 815	FREAC2	2
Enhancer 815	FREAC3	2
Enhancer 815	FUBP1	2
Enhancer 815	GAF	1
Enhancer 815	GATA1	3
Enhancer 815	GATA3	7
Enhancer 815	GATA5	3
Enhancer 815	GATA6	4
Enhancer 815	GFI1	4
Enhancer 815	GKLF	1
Enhancer 815	GLI2	1
Enhancer 815	GLI3	3
Enhancer 815	GSC	1
Enhancer 815	GSC2	1
Enhancer 815	HESX1	2
Enhancer 815	HFH3	4
Enhancer 815	HFH8	2
Enhancer 815	Hic1	3
Enhancer 815	HIC2	2
Enhancer 815	HLF	1
Enhancer 815	HMBOX1	1
Enhancer 815	HMGA1	2
Enhancer 815	HMGIY	4
Enhancer 815	HNF3	6
Enhancer 815	HNF3A	3

Promoter 2	FOXP1	2
Promoter 2	FOXP2	4
Promoter 2	FOXP3	5
Promoter 2	FREAC2	2
Promoter 2	FREAC4	2
Promoter 2	FREAC7	1
Promoter 2	FUBP1	1
Promoter 2	GATA	1
Promoter 2	GATA1	4
Promoter 2	GATA2	1
Promoter 2	GATA3	1
Promoter 2	GATA4	1
Promoter 2	GATA5	2
Promoter 2	GATA6	1
Promoter 2	Gbx1	2
Promoter 2	GBX2	3
Promoter 2	GFI1	2
Promoter 2	GMEB2	1
Promoter 2	GR	1
Promoter 2	GSC	1
Promoter 2	GSC2	1
Promoter 2	GSX1	1
Promoter 2	GSX2	1
Promoter 2	HB9	1
Promoter 2	HELIOSA	1
Promoter 2	HESX1	2
Promoter 2	HFH3	1
Promoter 2	HFH4	1
Promoter 2	HFH8	2
Promoter 2	HIF2A	2
Promoter 2	HMGA1	1
Promoter 2	HMGIY	4
Promoter 2	HMX1	1
Promoter 2	HMX2	1
Promoter 2	HMX3	2
Promoter 2	HNF1	3
Promoter 2	HNF1A	2
Promoter 2	HNF1B	4
Promoter 2	HNF3	3
Promoter 2	HNF3A	2
Promoter 2	HNF3ALPHA	2
Promoter 2	HNF3B	1
Promoter 2	HOX13	1
Promoter 2	HOXA1	1
Promoter 2	HOXA10	4
Promoter 2	Hoxa11	2
Promoter 2	HOXA13	4
Promoter 2	HOXA2	2
Promoter 2	HOXA3	2
Promoter 2	HOXA4	1
Promoter 2	HOXA6	2
Promoter 2	HOXA7	1
Promoter 2	HOXA9	1
Promoter 2	HOXB13	1
Promoter 2	HOXB2	1
Promoter 2	HOXB3	1
Promoter 2	HOXB5	2
Promoter 2	HOXB6	2
Promoter 2	HOXB9	1
Promoter 2	HOXC10	5
Promoter 2	HOXC11	4
Promoter 2	HOXC12	1
Promoter 2	HOXC13	1
Promoter 2	HOXC4	1
Promoter 2	HOXC5	2
Promoter 2	HOXC8	2
Promoter 2	HOXC9	1
Promoter 2	HOXD11	2
Promoter 2	HOXD12	2
Promoter 2	HOXD13	3
Promoter 2	HOXD3	3
Promoter 2	Hoxd9	2
Promoter 2	HSF2	1
Promoter 2	HXA13	1
Promoter 2	HXA5	1
Promoter 2	HXC8	1
Promoter 2	HXD10	2
Promoter 2	HXD13	1
Promoter 2	HXD4	2
Promoter 2	HXD9	2
Promoter 2	ID4	1
Promoter 2	IPF	1
Promoter 2	IPF1	4
Promoter 2	IRF	1
Promoter 2	IRF1	4
Promoter 2	IRF2	1
Promoter 2	IRF4	2

Enhancer 330	MEIS1	1
Enhancer 330	MEOX1	2
Enhancer 330	MEOX2	8
Enhancer 330	MIXL1	2
Enhancer 330	MMEF2	1
Enhancer 330	MNX1	2
Enhancer 330	MOX1	1
Enhancer 330	MSX1	9
Enhancer 330	MSX2	9
Enhancer 330	Msx3	5
Enhancer 330	MYB	2
Enhancer 330	NANOG	1
Enhancer 330	NCX	2
Enhancer 330	NF2L1	2
Enhancer 330	NFAC2	2
Enhancer 330	NFAC3	1
Enhancer 330	NFAT	5
Enhancer 330	NFIL3	5
Enhancer 330	NFKB2	1
Enhancer 330	NFY	4
Enhancer 330	NFYA	2
Enhancer 330	NFYB	3
Enhancer 330	NFYC	2
Enhancer 330	NKX25	3
Enhancer 330	NKX31	1
Enhancer 330	NKX3A	2
Enhancer 330	NKX61	4
Enhancer 330	Nkx6-1	6
Enhancer 330	NKX62	2
Enhancer 330	NKX6-2	4
Enhancer 330	NOBOX	2
Enhancer 330	NOTO	3
Enhancer 330	NR1I3	1
Enhancer 330	Nr2e1	5
Enhancer 330	NR2E3	1
Enhancer 330	NRL	1
Enhancer 330	OCT	6
Enhancer 330	OCT1	23
Enhancer 330	OCT2	3
Enhancer 330	OCTAMER	4
Enhancer 330	OLIG2	2
Enhancer 330	OLIG3	1
Enhancer 330	OTX	1
Enhancer 330	OVOL1	1
Enhancer 330	PARP	2
Enhancer 330	PAX2	3
Enhancer 330	PAX3	1
Enhancer 330	PAX4	8
Enhancer 330	PAX6	3
Enhancer 330	PAX7	2
Enhancer 330	PBX1	1
Enhancer 330	PBX2	2
Enhancer 330	PDX1	5
Enhancer 330	PEBB	1
Enhancer 330	PEBP	2
Enhancer 330	PHOX2A	2
Enhancer 330	PHOX2B	4
Enhancer 330	PIT1	2
Enhancer 330	PITX1	1
Enhancer 330	PITX2	3
Enhancer 330	PITX3	1
Enhancer 330	PLZF	1
Enhancer 330	PMX2A	2
Enhancer 330	PMX2B	2
Enhancer 330	PO2F1	3
Enhancer 330	PO2F2	2
Enhancer 330	PO3F1	3
Enhancer 330	PO3F2	3
Enhancer 330	PO5F1	2
Enhancer 330	POU1F1	7
Enhancer 330	POU2F1	6
Enhancer 330	POU2F2	11
Enhancer 330	POU2F3	10
Enhancer 330	POU3F1	6
Enhancer 330	POU3F2	5
Enhancer 330	POU3F3	8
Enhancer 330	POU3F4	4
Enhancer 330	POU4F1	2
Enhancer 330	POU4F2	2
Enhancer 330	POU4F3	1
Enhancer 330	POU5F1	2
Enhancer 330	POU5F1P1	6
Enhancer 330	POU6F1	4
Enhancer 330	POU6F2	4
Enhancer 330	PR	2
Enhancer 330	PROP1	7

Enhancer 815	HNF3ALPHA	4
Enhancer 815	HNF3B	3
Enhancer 815	HNF4	2
Enhancer 815	HOXA5	1
Enhancer 815	HOXA7	1
Enhancer 815	HOXD11	1
Enhancer 815	HOXD13	5
Enhancer 815	Hoxd9	1
Enhancer 815	HSF1	1
Enhancer 815	HXD10	2
Enhancer 815	HXD9	2
Enhancer 815	IRF1	2
Enhancer 815	IRX2	1
Enhancer 815	Irx3	1
Enhancer 815	IRX5	2
Enhancer 815	IRXB3	1
Enhancer 815	ISL1	1
Enhancer 815	ISX	1
Enhancer 815	JDP2	2
Enhancer 815	JUN	2
Enhancer 815	JUN(var.2)	3
Enhancer 815	JUN::FOS	1
Enhancer 815	JUNB	5
Enhancer 815	JUND	4
Enhancer 815	KLF4	1
Enhancer 815	LFA1	1
Enhancer 815	Lhx8	1
Enhancer 815	LRH1	1
Enhancer 815	MAF	1
Enhancer 815	MEF2	4
Enhancer 815	MEF2A	4
Enhancer 815	MEF2C	1
Enhancer 815	MRF2	1
Enhancer 815	MYB	5
Enhancer 815	MYF6	1
Enhancer 815	MYOGENIN	1
Enhancer 815	NDF1	1
Enhancer 815	NFAC2	2
Enhancer 815	NFAC3	1
Enhancer 815	NFAT3	2
Enhancer 815	NFAT5	1
Enhancer 815	NFATC2	2
Enhancer 815	NFE2	2
Enhancer 815	NFIA+NFIB+NFIC+NFIX	3
Enhancer 815	NFIC	1
Enhancer 815	NFIL3	2
Enhancer 815	NFKAPPAB	1
Enhancer 815	NFKB	1
Enhancer 815	NKX31	2
Enhancer 815	NKX3-1	4
Enhancer 815	NKX3A	2
Enhancer 815	NKX61	1
Enhancer 815	NKX62	1
Enhancer 815	NR2E3	1
Enhancer 815	NRF2	1
Enhancer 815	OCT	4
Enhancer 815	OCT1	13
Enhancer 815	OCTAMER	2
Enhancer 815	OTX	2
Enhancer 815	OTX1	3
Enhancer 815	OTX2	6
Enhancer 815	P50P50	1
Enhancer 815	P53	1
Enhancer 815	P63	1
Enhancer 815	PAX	1
Enhancer 815	PAX5	3
Enhancer 815	PBX	3
Enhancer 815	PBX1	5
Enhancer 815	PHOX2A	2
Enhancer 815	PHOX2B	4
Enhancer 815	PIT1	4
Enhancer 815	PITX1	4
Enhancer 815	PITX2	4
Enhancer 815	PITX3	2
Enhancer 815	PO2F1	3
Enhancer 815	PO3F2	3
Enhancer 815	PO5F1	1
Enhancer 815	POU1F1	6
Enhancer 815	POU2F1	6
Enhancer 815	POU2F2	9
Enhancer 815	POU2F3	7
Enhancer 815	POU3F1	5
Enhancer 815	POU3F2	5
Enhancer 815	POU3F3	7
Enhancer 815	POU3F4	4
Enhancer 815	POU4F1	1

Promoter 2	IRF7	2
Promoter 2	IRF8	3
Promoter 2	IRF9	1
Promoter 2	IRX2	1
Promoter 2	Irx3	1
Promoter 2	IRX5	1
Promoter 2	ISL1	1
Promoter 2	ISL2	1
Promoter 2	ISX	2
Promoter 2	ITF2	1
Promoter 2	JUND(var.2)	1
Promoter 2	K2B	1
Promoter 2	KLF3	1
Promoter 2	LEF1	4
Promoter 2	LEF1TCF1	1
Promoter 2	LHX2	2
Promoter 2	LHX3	4
Promoter 2	Lhx4	2
Promoter 2	LHX5	1
Promoter 2	LHX6	1
Promoter 2	Lhx8	3
Promoter 2	LHX9	2
Promoter 2	LMX1A	2
Promoter 2	LMX1B	3
Promoter 2	Mafb	1
Promoter 2	MAFF	1
Promoter 2	MAFK	5
Promoter 2	MEF2	3
Promoter 2	MEF2A	3
Promoter 2	MEF2C	1
Promoter 2	MEIS1	2
Promoter 2	MEIS1AHOXA9	1
Promoter 2	MEIS2	3
Promoter 2	MEIS3	2
Promoter 2	MEOX1	1
Promoter 2	MEOX2	4
Promoter 2	MIXL1	1
Promoter 2	MNX1	1
Promoter 2	MOX1	2
Promoter 2	MSX1	5
Promoter 2	MSX2	3
Promoter 2	Msx3	2
Promoter 2	MYBL1	1
Promoter 2	MYBL2	1
Promoter 2	MYF6	1
Promoter 2	NANOG	2
Promoter 2	NCX	2
Promoter 2	NDF1	1
Promoter 2	NFAC1	1
Promoter 2	NFAC2	1
Promoter 2	NFAT3	1
Promoter 2	NFATC2	1
Promoter 2	NFIL3	1
Promoter 2	NKX22	3
Promoter 2	NKX25	2
Promoter 2	NKX31	1
Promoter 2	Nkx3-1	2
Promoter 2	NKX3A	3
Promoter 2	NKX61	3
Promoter 2	Nkx6-1	3
Promoter 2	NKX62	2
Promoter 2	NKX6-2	2
Promoter 2	NOBOX	1
Promoter 2	NR0B1	1
Promoter 2	NR1I2	1
Promoter 2	NR1I3	2
Promoter 2	NR2C1	1
Promoter 2	NR2E3	1
Promoter 2	NR2F1	1
Promoter 2	NR4A1	1
Promoter 2	NR4A2	3
Promoter 2	NRL	2
Promoter 2	NUR77	1
Promoter 2	NURR1	1
Promoter 2	OCT	2
Promoter 2	OCT1	13
Promoter 2	OTX	1
Promoter 2	Otx1	2
Promoter 2	OTX2	1
Promoter 2	PAX	1
Promoter 2	PAX4	1
Promoter 2	PAX6	1
Promoter 2	PAX8	2
Promoter 2	PBX1	2
Promoter 2	PBX2	1
Promoter 2	PDX1	3

Enhancer 330	PRRX1	7
Enhancer 330	PRRX2	6
Enhancer 330	RAX	5
Enhancer 330	RAXL1	2
Enhancer 330	RFX2	1
Enhancer 330	Rhox11	2
Enhancer 330	RORA	1
Enhancer 330	RUNX1	1
Enhancer 330	RUNX2	5
Enhancer 330	SHOX	2
Enhancer 330	Shox2	4
Enhancer 330	SMAD1	1
Enhancer 330	SOX10	4
Enhancer 330	SOX13	1
Enhancer 330	SOX15	2
Enhancer 330	SOX2	3
Enhancer 330	SOX5	1
Enhancer 330	SOX9	5
Enhancer 330	SOX91	1
Enhancer 330	SRF	2
Enhancer 330	SRY	5
Enhancer 330	STAT1STAT1	1
Enhancer 330	STAT2::STAT1	1
Enhancer 330	STAT3	1
Enhancer 330	STAT4	1
Enhancer 330	STAT5A	1
Enhancer 330	STAT6	2
Enhancer 330	TAL1	1
Enhancer 330	TATA	2
Enhancer 330	TBP	1
Enhancer 330	TCF11	1
Enhancer 330	TCF3	2
Enhancer 330	Tcf7	1
Enhancer 330	TCF7L2	1
Enhancer 330	TEAD1	2
Enhancer 330	TEAD3	1
Enhancer 330	TEAD4	1
Enhancer 330	TEF	3
Enhancer 330	TEF1	1
Enhancer 330	TGIF1	1
Enhancer 330	TITF1	1
Enhancer 330	TLX1	1
Enhancer 330	UBIP1	2
Enhancer 330	UNCX	8
Enhancer 330	VAX1	2
Enhancer 330	VAX2	4
Enhancer 330	VBP	2
Enhancer 330	VENTX	1
Enhancer 330	VSX1	6
Enhancer 330	VSX2	4
Enhancer 330	YY1	1
Enhancer 330	ZFHX3	1
Enhancer 330	ZNF333	2
Enhancer 330	ZNF354C	1
Enhancer 700	AFP1	1
Enhancer 700	AIRE	2
Enhancer 700	AP1	10
Enhancer 700	AP1FJ	2
Enhancer 700	AP4	1
Enhancer 700	ARI3A	4
Enhancer 700	ATF5	1
Enhancer 700	Atoh1	4
Enhancer 700	BARX2	3
Enhancer 700	BCL6	8
Enhancer 700	BEN	1
Enhancer 700	BHLHA15	4
Enhancer 700	BHLHE22	2
Enhancer 700	BHLHE23	2
Enhancer 700	BLIMP1	1
Enhancer 700	BRN2	1
Enhancer 700	CDX	2
Enhancer 700	CDX2	6
Enhancer 700	CEBPDELTA	1
Enhancer 700	CEBPZ	2
Enhancer 700	CETS1P54	1
Enhancer 700	COT1	2
Enhancer 700	COT2	2
Enhancer 700	CP2	1
Enhancer 700	CPEB1	4
Enhancer 700	CUX1	2
Enhancer 700	CUX2	1
Enhancer 700	DEC	2
Enhancer 700	DLX2	3
Enhancer 700	DLX3	3
Enhancer 700	DMRT3	1
Enhancer 815	POU4F3	1
Enhancer 815	POU5F1	3
Enhancer 815	POU5F1P1	5
Enhancer 815	PROP1	7
Enhancer 815	PRRX1	3
Enhancer 815	RHOXF1	2
Enhancer 815	SMAD1	3
Enhancer 815	SMAD2	1
Enhancer 815	SMAD3	1
Enhancer 815	SMAD4	1
Enhancer 815	SMRC1	3
Enhancer 815	SOX10	2
Enhancer 815	SOX17	2
Enhancer 815	SOX5	1
Enhancer 815	SP1	1
Enhancer 815	SRF	1
Enhancer 815	SRY	4
Enhancer 815	STA5A	1
Enhancer 815	STAT1	1
Enhancer 815	STAT3	2
Enhancer 815	STAT4	1
Enhancer 815	STAT5A	2
Enhancer 815	STAT6	2
Enhancer 815	TAL1::GATA1	2
Enhancer 815	TAL1BETAITF2	1
Enhancer 815	TBP	1
Enhancer 815	TCF3	1
Enhancer 815	TEAD1	7
Enhancer 815	TEAD3	3
Enhancer 815	TEAD4	4
Enhancer 815	TEF	7
Enhancer 815	TFE2	1
Enhancer 815	TFIII	3
Enhancer 815	UNCX	2
Enhancer 815	VBP	1
Enhancer 815	YBOX1	1
Enhancer 815	ZBTB4	2
Enhancer 815	ZN589	1
Enhancer 815	ZNF333	1
Enhancer 843	AFP1	1
Enhancer 843	Alx1	3
Enhancer 843	ALX3	1
Enhancer 843	ALX4	2
Enhancer 843	AML	1
Enhancer 843	AML1	3
Enhancer 843	AP1	6
Enhancer 843	AR	3
Enhancer 843	AREB6	3
Enhancer 843	ARI3A	1
Enhancer 843	ATF5	2
Enhancer 843	BATF	2
Enhancer 843	BCL6	4
Enhancer 843	BPTF	1
Enhancer 843	CART1	1
Enhancer 843	CDC5	2
Enhancer 843	CDP	3
Enhancer 843	CDX	1
Enhancer 843	CDX2	4
Enhancer 843	CEBP	7
Enhancer 843	CEBPA	7
Enhancer 843	CEBPB	5
Enhancer 843	CEBPD	3
Enhancer 843	CEBPG	3
Enhancer 843	CEBPGAMMA	1
Enhancer 843	COT1	1
Enhancer 843	COT2	2
Enhancer 843	CPEB1	2
Enhancer 843	CTF1	1
Enhancer 843	CUX1	2
Enhancer 843	CUX2	1
Enhancer 843	DBP	3
Enhancer 843	DELTAEF1	1
Enhancer 843	DLX3	3
Enhancer 843	DMRT2	1
Enhancer 843	DRGX	1
Enhancer 843	DRI1	1
Enhancer 843	E12	1
Enhancer 843	E2F6	2
Enhancer 843	EFC	1
Enhancer 843	EHF	2
Enhancer 843	ELF1	3
Enhancer 843	ELF3	3
Enhancer 843	ELF4	1
Enhancer 843	Elf5	3
Enhancer 843	ELK1	5
Promoter 2	PIT1	4
Promoter 2	PITX1	3
Promoter 2	PITX2	2
Promoter 2	PITX3	2
Promoter 2	PKNX1	1
Promoter 2	PO2F1	2
Promoter 2	PO2F2	1
Promoter 2	PO3F1	2
Promoter 2	PO3F2	1
Promoter 2	POU1F1	4
Promoter 2	POU2F1	4
Promoter 2	POU2F2	8
Promoter 2	POU2F3	3
Promoter 2	POU3F1	4
Promoter 2	POU3F2	5
Promoter 2	POU3F3	5
Promoter 2	POU3F4	4
Promoter 2	POU4F1	2
Promoter 2	POU4F3	1
Promoter 2	POU5F1	1
Promoter 2	POU5F1P1	3
Promoter 2	POU6F2	1
Promoter 2	PPARA	1
Promoter 2	PPARG	1
Promoter 2	PROP1	1
Promoter 2	PRRX1	3
Promoter 2	PRRX2	2
Promoter 2	PXR	1
Promoter 2	PXRRXR	2
Promoter 2	RARA	1
Promoter 2	RARG	1
Promoter 2	RAX	1
Promoter 2	RAXL1	1
Promoter 2	RFX1	2
Promoter 2	RFX2	6
Promoter 2	Rfx3	4
Promoter 2	RFX4	2
Promoter 2	RFX5	7
Promoter 2	Rhox11	1
Promoter 2	RHOXF1	3
Promoter 2	RORA	1
Promoter 2	RORA1	1
Promoter 2	RORBETA	1
Promoter 2	RSRFC4	2
Promoter 2	RUSH1A	1
Promoter 2	SHOX	1
Promoter 2	SHOX2	2
Promoter 2	SIX1	1
Promoter 2	SIX4	2
Promoter 2	SIX6	3
Promoter 2	SMAD	1
Promoter 2	SMAD1	1
Promoter 2	SMAD2::SMAD3::SMAD4	1
Promoter 2	SNAI2	1
Promoter 2	SOX10	3
Promoter 2	SOX13	1
Promoter 2	SOX15	1
Promoter 2	SOX17	1
Promoter 2	SOX18	1
Promoter 2	SOX21	1
Promoter 2	SOX4	2
Promoter 2	SOX5	2
Promoter 2	SOX9	3
Promoter 2	SOX91	1
Promoter 2	SRF	7
Promoter 2	SRY	5
Promoter 2	STAT1STAT1	1
Promoter 2	STAT2	1
Promoter 2	STAT2::STAT1	1
Promoter 2	STAT4	1
Promoter 2	TAL1	3
Promoter 2	TATA	1
Promoter 2	TBP	4
Promoter 2	TBX20	1
Promoter 2	TBX5	1
Promoter 2	TCF3	2
Promoter 2	TCF4	2
Promoter 2	TCF7	1
Promoter 2	TCF7L2	2
Promoter 2	Tcfap2a	2
Promoter 2	TEAD1	1
Promoter 2	TEAD3	2
Promoter 2	TEAD4	1
Promoter 2	TEF	4
Promoter 2	TF7L2	2

Enhancer 700	DR1	2
Enhancer 700	E2F6	1
Enhancer 700	EAR2	2
Enhancer 700	EGR1	2
Enhancer 700	EGR2	3
Enhancer 700	EGR3	2
Enhancer 700	EGR4	1
Enhancer 700	ELF1	4
Enhancer 700	ELF3	1
Enhancer 700	ELK3	1
Enhancer 700	EMX2	2
Enhancer 700	EN1	5
Enhancer 700	EOMES	4
Enhancer 700	ER	1
Enhancer 700	ESE1	1
Enhancer 700	ESR2	1
Enhancer 700	ETS1	1
Enhancer 700	EVI1	3
Enhancer 700	FEV	3
Enhancer 700	FOXA1	2
Enhancer 700	FOXA3	1
Enhancer 700	FOXB1	6
Enhancer 700	FOXC1	5
Enhancer 700	FOXC2	4
Enhancer 700	FOXD1	2
Enhancer 700	FOXD2	1
Enhancer 700	FOXD3	2
Enhancer 700	FOXF1	1
Enhancer 700	FOXI1	3
Enhancer 700	FOXJ2	6
Enhancer 700	FOXJ3	6
Enhancer 700	FOXM1	3
Enhancer 700	FOXO1	2
Enhancer 700	FOXO3A	1
Enhancer 700	FOXP2	1
Enhancer 700	FOXP3	2
Enhancer 700	FUBP1	1
Enhancer 700	FXR	1
Enhancer 700	GATA	2
Enhancer 700	GATA1	2
Enhancer 700	GATA2	2
Enhancer 700	GATA3	4
Enhancer 700	GBX2	5
Enhancer 700	GCM1	2
Enhancer 700	GFI1B	2
Enhancer 700	GLI3	4
Enhancer 700	HELIOSA	1
Enhancer 700	HFH3	2
Enhancer 700	HFH8	1
Enhancer 700	Hic1	3
Enhancer 700	HMGA1	3
Enhancer 700	HMX3	2
Enhancer 700	HNF1	5
Enhancer 700	HNF1A	4
Enhancer 700	HNF1B	7
Enhancer 700	HNF3	3
Enhancer 700	HNF3ALPHA	2
Enhancer 700	HNF3B	2
Enhancer 700	HNF4	2
Enhancer 700	HOXA10	2
Enhancer 700	Hoxa11	2
Enhancer 700	HOXA13	13
Enhancer 700	HOXB13	3
Enhancer 700	HOXC10	5
Enhancer 700	HOXC11	3
Enhancer 700	HOXC12	3
Enhancer 700	HOXC13	7
Enhancer 700	HOXD11	1
Enhancer 700	HOXD12	5
Enhancer 700	Hoxd13	8
Enhancer 700	Hoxd9	2
Enhancer 700	HSF1	6
Enhancer 700	HSF2	3
Enhancer 700	HSF4	1
Enhancer 700	HXD10	2
Enhancer 700	HXD9	2
Enhancer 700	ICSBP	1
Enhancer 700	IPF1	7
Enhancer 700	IRF	2
Enhancer 700	IRF1	5
Enhancer 700	IRF3	3
Enhancer 700	IRF4	2
Enhancer 700	IRF5	2
Enhancer 700	IRF8	3
Enhancer 700	IRX2	2
Enhancer 700	Irx3	2
Enhancer 700	IRX5	1

Enhancer 843	ELK3	2
Enhancer 843	ELK4	1
Enhancer 843	ERR1	2
Enhancer 843	ERR2	3
Enhancer 843	ERR3	1
Enhancer 843	ESE1	1
Enhancer 843	Esrra	4
Enhancer 843	ESRRG	1
Enhancer 843	ETS	2
Enhancer 843	ETS2	2
Enhancer 843	ETV6	1
Enhancer 843	EVI1	6
Enhancer 843	EVX1	1
Enhancer 843	FEV	1
Enhancer 843	FLI1	2
Enhancer 843	FOS	1
Enhancer 843	FOSB	2
Enhancer 843	FOX	2
Enhancer 843	FOXA1	4
Enhancer 843	FOXA2	3
Enhancer 843	FOXA3	3
Enhancer 843	FOXB1	5
Enhancer 843	FOXC1	9
Enhancer 843	FOXC2	4
Enhancer 843	FOXD1	3
Enhancer 843	FOXD2	4
Enhancer 843	FOXD3	5
Enhancer 843	FOXF1	4
Enhancer 843	FOXF2	5
Enhancer 843	FOXH1	2
Enhancer 843	FOXI1	3
Enhancer 843	FOXJ2	8
Enhancer 843	Foxj3	8
Enhancer 843	FOXL1	4
Enhancer 843	FOXM1	4
Enhancer 843	FOXO1	7
Enhancer 843	FOXO3	4
Enhancer 843	FOXO4	5
Enhancer 843	FOXP2	3
Enhancer 843	FPM315	1
Enhancer 843	FRA1	3
Enhancer 843	FREAC2	3
Enhancer 843	FREAC7	3
Enhancer 843	FUBP1	2
Enhancer 843	GABP1+GABP2	1
Enhancer 843	GABPA	2
Enhancer 843	GABPALPHA	1
Enhancer 843	GADP	1
Enhancer 843	GATA	5
Enhancer 843	GATA1	2
Enhancer 843	GATA3	7
Enhancer 843	GATA4	4
Enhancer 843	GATA5	4
Enhancer 843	GATA6	5
Enhancer 843	GCM	1
Enhancer 843	GFI1	5
Enhancer 843	GFI1B	5
Enhancer 843	GRHL1	4
Enhancer 843	HEB	1
Enhancer 843	HELIOSA	2
Enhancer 843	HESX1	2
Enhancer 843	HFH3	3
Enhancer 843	HFH4	1
Enhancer 843	HFH8	3
Enhancer 843	Hic1	2
Enhancer 843	HIC2	1
Enhancer 843	HMGA1	1
Enhancer 843	HMGIY	6
Enhancer 843	HMX2	1
Enhancer 843	HNF1	4
Enhancer 843	HNF1B	1
Enhancer 843	HNF3	5
Enhancer 843	HNF3A	1
Enhancer 843	HNF3ALPHA	4
Enhancer 843	HNF3B	2
Enhancer 843	HNF4ALPHA	2
Enhancer 843	HNF6	1
Enhancer 843	HOXA10	1
Enhancer 843	HOXA13	3
Enhancer 843	HOXB13	1
Enhancer 843	HOXC13	1
Enhancer 843	HOXD13	4
Enhancer 843	HXA7	1
Enhancer 843	HXA9	3
Enhancer 843	HXD10	3
Enhancer 843	HXD13	3
Enhancer 843	HXD4	2

Promoter 2	TFAP2A	4
Promoter 2	TFAP2B	2
Promoter 2	TFAP2C	4
Promoter 2	TFE2	1
Promoter 2	TFIIA	1
Promoter 2	TGIF1	2
Promoter 2	TITF1	1
Promoter 2	UNCX	3
Promoter 2	VAX1	1
Promoter 2	VAX2	1
Promoter 2	VDR	1
Promoter 2	VENTX	1
Promoter 2	Vsx1	2
Promoter 2	YY1	1
Promoter 2	ZBTB7B	1
Promoter 2	ZFHX3	2
Promoter 2	ZN384	1
Promoter 3	ALX1	1
Promoter 3	AML1	1
Promoter 3	AP1	5
Promoter 3	AP1FJ	1
Promoter 3	AP3	1
Promoter 3	AR	1
Promoter 3	AREB6	2
Promoter 3	ATF3	2
Promoter 3	Atf4	4
Promoter 3	Barhl1	1
Promoter 3	BARHL2	2
Promoter 3	BARX1	5
Promoter 3	BARX2	3
Promoter 3	BATF	1
Promoter 3	BATF::JUN	1
Promoter 3	BCL6	1
Promoter 3	BSX	1
Promoter 3	CDP	1
Promoter 3	CDX1	1
Promoter 3	CDX2	2
Promoter 3	CEBP	4
Promoter 3	CEBPA	4
Promoter 3	CEBPB	3
Promoter 3	CEBPD	2
Promoter 3	CEBPDELTA	2
Promoter 3	CEBPG	2
Promoter 3	COUPTF	1
Promoter 3	CREM	3
Promoter 3	DBP	2
Promoter 3	DLX1	2
Promoter 3	DLX2	5
Promoter 3	DLX3	2
Promoter 3	DLX4	1
Promoter 3	DLX5	3
Promoter 3	DLX6	1
Promoter 3	DLX7	1
Promoter 3	ELF3	1
Promoter 3	EMX2	1
Promoter 3	EN1	5
Promoter 3	EOMES	1
Promoter 3	ERR2	1
Promoter 3	ETS2	1
Promoter 3	EVI1	2
Promoter 3	FOSB	1
Promoter 3	FOSL1	1
Promoter 3	FOSL2	2
Promoter 3	FOX	1
Promoter 3	FOXA1	6
Promoter 3	FOXA2	4
Promoter 3	FOXA3	4
Promoter 3	FOXB1	4
Promoter 3	FOXC1	9
Promoter 3	FOXC2	3
Promoter 3	FOXD1	3
Promoter 3	FOXD2	2
Promoter 3	FOXD3	4
Promoter 3	FOXF1	4
Promoter 3	FOXF2	3
Promoter 3	FOXG1	3
Promoter 3	FOXH1	1
Promoter 3	FOXI1	4
Promoter 3	FOXJ2	10
Promoter 3	Foxj3	12
Promoter 3	Foxk1	2
Promoter 3	FOXL1	3
Promoter 3	FOXM1	2
Promoter 3	FOXO1	11
Promoter 3	FOXO3	7
Promoter 3	FOXO4	9
Promoter 3	FOXP1	1

Enhancer 700	JUN	3
Enhancer 700	JUNB	3
Enhancer 700	JUND(var.2)	1
Enhancer 700	KLF3	2
Enhancer 700	KLF8	2
Enhancer 700	LHX3	2
Enhancer 700	LHX6	2
Enhancer 700	Mafb	1
Enhancer 700	MAFK	3
Enhancer 700	MAX	2
Enhancer 700	MEF2A	5
Enhancer 700	MEF2C	3
Enhancer 700	MEIS1	4
Enhancer 700	Meis2	1
Enhancer 700	Meis3	2
Enhancer 700	MGA	2
Enhancer 700	MRF2	1
Enhancer 700	MYBB	2
Enhancer 700	NANOG	2
Enhancer 700	NCX	4
Enhancer 700	NDF1	1
Enhancer 700	NEUROD2	2
Enhancer 700	NEUROG2	4
Enhancer 700	NF1	1
Enhancer 700	NFAC1	4
Enhancer 700	NFAT	1
Enhancer 700	NFAT2	2
Enhancer 700	NFIA+NFIB+NFIC+NFIX	1
Enhancer 700	NGFIC	1
Enhancer 700	NKX21	1
Enhancer 700	NKX2-3	4
Enhancer 700	NKX25	4
Enhancer 700	NKX28	1
Enhancer 700	NKX2-8	4
Enhancer 700	NKX31	3
Enhancer 700	NKX3-1	4
Enhancer 700	NKX32	5
Enhancer 700	NKX3-2	2
Enhancer 700	NKX3A	4
Enhancer 700	NKX61	3
Enhancer 700	NR2F1	2
Enhancer 700	NR4A2	2
Enhancer 700	NR4A3	1
Enhancer 700	NRF2	1
Enhancer 700	OCT	5
Enhancer 700	OCT1	17
Enhancer 700	OCT2	2
Enhancer 700	OLIG1	2
Enhancer 700	OLIG2	8
Enhancer 700	OLIG3	4
Enhancer 700	P53	3
Enhancer 700	P63	1
Enhancer 700	P73	2
Enhancer 700	PARP	1
Enhancer 700	PAX2	1
Enhancer 700	PAX7	1
Enhancer 700	PIT1	3
Enhancer 700	PLZF	1
Enhancer 700	PO2F1	4
Enhancer 700	PO2F2	2
Enhancer 700	PO3F1	2
Enhancer 700	PO3F2	2
Enhancer 700	PO6F1	1
Enhancer 700	POU1F1	4
Enhancer 700	POU2F1	7
Enhancer 700	POU2F2	9
Enhancer 700	POU2F3	9
Enhancer 700	POU3F1	7
Enhancer 700	POU3F2	5
Enhancer 700	POU3F3	9
Enhancer 700	POU3F4	3
Enhancer 700	POU4F1	2
Enhancer 700	POU4F2	1
Enhancer 700	POU4F3	2
Enhancer 700	POU5F1	4
Enhancer 700	POU5F1P1	4
Enhancer 700	POU6F2	4
Enhancer 700	PPAR	2
Enhancer 700	PPARA	3
Enhancer 700	PPARD	2
Enhancer 700	PPARG	1
Enhancer 700	PRDM1	2
Enhancer 700	PRRX1	2
Enhancer 700	PU1	2
Enhancer 700	RARA	1

Enhancer 843	HXD9	2
Enhancer 843	ICSBP	1
Enhancer 843	IK	1
Enhancer 843	IKZF1	1
Enhancer 843	IPF1	5
Enhancer 843	IRF	2
Enhancer 843	IRF1	2
Enhancer 843	IRF7	1
Enhancer 843	IRF8	3
Enhancer 843	ISRE	1
Enhancer 843	JUN(var.2)	2
Enhancer 843	JUNB	4
Enhancer 843	JUND	3
Enhancer 843	KAISO	1
Enhancer 843	KLF1	1
Enhancer 843	KLF8	1
Enhancer 843	LEF1	7
Enhancer 843	LHX61	1
Enhancer 843	LYF1	1
Enhancer 843	MAX	1
Enhancer 843	MAZ	1
Enhancer 843	MEF2	1
Enhancer 843	MEF2A	2
Enhancer 843	MEF2D	1
Enhancer 843	MEIS1	8
Enhancer 843	MEIS1AHOXA9	1
Enhancer 843	MEIS2	6
Enhancer 843	Meis3	7
Enhancer 843	MITF	1
Enhancer 843	MSX1	4
Enhancer 843	Msx3	2
Enhancer 843	MYB	2
Enhancer 843	MYBB	2
Enhancer 843	NCX	1
Enhancer 843	NEUROD2	1
Enhancer 843	NF1	3
Enhancer 843	NFAC2	1
Enhancer 843	NFAC4	1
Enhancer 843	NFAT1	1
Enhancer 843	NFAT2	1
Enhancer 843	NFAT3	1
Enhancer 843	NFATC2	1
Enhancer 843	NFE2	1
Enhancer 843	NFIA	3
Enhancer 843	NFIA+NFIB+NFIC+NFIX	3
Enhancer 843	NFIB	1
Enhancer 843	NFIX	6
Enhancer 843	NKX22	1
Enhancer 843	NKX2-3	2
Enhancer 843	NKX25	1
Enhancer 843	NKX31	2
Enhancer 843	NKX3-1	4
Enhancer 843	NKX3-2	1
Enhancer 843	NKX3A	2
Enhancer 843	NR0B1	1
Enhancer 843	NR2F1	2
Enhancer 843	NR2F6	2
Enhancer 843	NR4A1	1
Enhancer 843	NR4A2	3
Enhancer 843	NUR77	1
Enhancer 843	OCT	4
Enhancer 843	OCT1	16
Enhancer 843	OCT2	2
Enhancer 843	OCTAMER	3
Enhancer 843	ONEC2	1
Enhancer 843	OTX	2
Enhancer 843	OTX2	2
Enhancer 843	PAX4	4
Enhancer 843	PAX5	1
Enhancer 843	PAX6	2
Enhancer 843	PAX7	1
Enhancer 843	PAX8	2
Enhancer 843	PBX	2
Enhancer 843	PBX1	6
Enhancer 843	PBX2	3
Enhancer 843	PBX3	1
Enhancer 843	PEA3	1
Enhancer 843	PEBB	2
Enhancer 843	PEBP	1
Enhancer 843	PHOX2A	1
Enhancer 843	PHOX2B	3
Enhancer 843	PIT1	6
Enhancer 843	PITX2	2
Enhancer 843	PITX3	1
Enhancer 843	PO2F1	3

Promoter 3	FOXP2	4
Promoter 3	FOXP3	6
Promoter 3	FOXQ1	1
Promoter 3	FRA1	1
Promoter 3	FREAC2	3
Promoter 3	FREAC4	2
Promoter 3	FREAC7	2
Promoter 3	FUBP1	2
Promoter 3	GATA	2
Promoter 3	GATA2	2
Promoter 3	GATA3	3
Promoter 3	GATA4	2
Promoter 3	GATA5	2
Promoter 3	GATA6	2
Promoter 3	GBX2	1
Promoter 3	GFI1	3
Promoter 3	GFI1B	3
Promoter 3	GMEB2	1
Promoter 3	HBP1	2
Promoter 3	HESX1	1
Promoter 3	HFH3	3
Promoter 3	HFH8	3
Promoter 3	HLF	1
Promoter 3	HMGIY	3
Promoter 3	HNF1	2
Promoter 3	HNF3	4
Promoter 3	HNF3ALPHA	4
Promoter 3	HNF3B	2
Promoter 3	HNF4ALPHA	1
Promoter 3	HOX13	2
Promoter 3	HOXA1	2
Promoter 3	HOXA13	2
Promoter 3	HOXA2	1
Promoter 3	HOXA5	1
Promoter 3	HOXA6	2
Promoter 3	HOXB5	1
Promoter 3	HOXB6	1
Promoter 3	HOXC10	1
Promoter 3	HOXC8	2
Promoter 3	HOXD13	1
Promoter 3	Hoxd9	1
Promoter 3	HXA1	1
Promoter 3	HXA10	2
Promoter 3	HXA9	2
Promoter 3	HXD10	2
Promoter 3	HXD13	1
Promoter 3	HXD4	1
Promoter 3	HXD9	1
Promoter 3	IPF1	3
Promoter 3	IRF	1
Promoter 3	IRF1	2
Promoter 3	IRX2	1
Promoter 3	IRX5	1
Promoter 3	JUN	1
Promoter 3	JUN(var.2)	1
Promoter 3	JUN::FOS	2
Promoter 3	JUNB	2
Promoter 3	JUND	2
Promoter 3	LEF1	3
Promoter 3	LEF1TCF1	1
Promoter 3	LFA1	1
Promoter 3	LHX2	2
Promoter 3	LMX1B	1
Promoter 3	MAFK	1
Promoter 3	MEIS1	6
Promoter 3	Meis2	2
Promoter 3	MEIS3	4
Promoter 3	MEOX2	2
Promoter 3	MGA	2
Promoter 3	MOX1	2
Promoter 3	MSX1	3
Promoter 3	MSX2	2
Promoter 3	Msx3	1
Promoter 3	MYB	1
Promoter 3	MYBB	1
Promoter 3	MYBL1	1
Promoter 3	NANOG	2
Promoter 3	NF2L1	1
Promoter 3	NFIA	2
Promoter 3	NFIA+NFIB+NFIC+NFIX	2
Promoter 3	NFIX	4
Promoter 3	NKX31	1
Promoter 3	NKX3-1	1
Promoter 3	NKX3A	1
Promoter 3	NKX61	1

Enhancer 700	RARG	1
Enhancer 700	RBPJK	2
Enhancer 700	RFX	2
Enhancer 700	RFX1	2
Enhancer 700	RORA	5
Enhancer 700	RORA1	2
Enhancer 700	RREB1	1
Enhancer 700	RSRFC4	1
Enhancer 700	RUSH1A	1
Enhancer 700	RXRB	3
Enhancer 700	SMAD	1
Enhancer 700	SMAD1	2
Enhancer 700	SMAD3	2
Enhancer 700	SNAI1	1
Enhancer 700	SOX	2
Enhancer 700	SOX10	7
Enhancer 700	SOX15	3
Enhancer 700	SOX18	1
Enhancer 700	SOX2	2
Enhancer 700	SOX21	3
Enhancer 700	SOX4	2
Enhancer 700	SOX5	7
Enhancer 700	SOX9	6
Enhancer 700	SOX91	2
Enhancer 700	SRBP2	1
Enhancer 700	SREBF2	1
Enhancer 700	SREBP	1
Enhancer 700	SRF	3
Enhancer 700	SRY	11
Enhancer 700	T3R	2
Enhancer 700	TAL1	1
Enhancer 700	TBP	4
Enhancer 700	TBR1	1
Enhancer 700	TBX1	2
Enhancer 700	TBX2	2
Enhancer 700	TBX20	3
Enhancer 700	TBX21	2
Enhancer 700	TBX3	1
Enhancer 700	TBX4	4
Enhancer 700	TBX5	4
Enhancer 700	TCF11	2
Enhancer 700	TCF11MAFG	1
Enhancer 700	TCF7L1	1
Enhancer 700	Tcfap2a	1
Enhancer 700	TEAD1	5
Enhancer 700	TERALPHA	1
Enhancer 700	TFAP2A	2
Enhancer 700	TFAP2B	1
Enhancer 700	TFAP2C	2
Enhancer 700	TFE	2
Enhancer 700	TFIII	1
Enhancer 700	TGIF1	2
Enhancer 700	THA	3
Enhancer 700	TP63	1
Enhancer 700	Tp73	2
Enhancer 700	TWST1	3
Enhancer 700	UBIP1	2
Enhancer 700	VAX1	2
Enhancer 700	VAX2	1
Enhancer 700	VSX2	2
Enhancer 700	ZABC1	3
Enhancer 700	ZBTB49	1
Enhancer 700	ZBTB6	1
Enhancer 700	ZFHX3	2
Enhancer 700	ZN333	1
Enhancer 700	ZN384	1
Enhancer 700	ZNF263	1
Enhancer 700	ZSCAN4	1

Enhancer 843	PO2F2	2
Enhancer 843	PO3F1	2
Enhancer 843	PO3F2	3
Enhancer 843	POU1F1	7
Enhancer 843	POU2F1	6
Enhancer 843	POU2F2	12
Enhancer 843	POU2F3	8
Enhancer 843	POU3F1	6
Enhancer 843	POU3F2	4
Enhancer 843	POU3F3	9
Enhancer 843	POU3F4	3
Enhancer 843	POU5F1	4
Enhancer 843	POU5F1P1	6
Enhancer 843	PPARA	2
Enhancer 843	PR	1
Enhancer 843	PROP1	2
Enhancer 843	PRRX1	2
Enhancer 843	PTF1A	1
Enhancer 843	PU1	1
Enhancer 843	Rarb	1
Enhancer 843	RARG	2
Enhancer 843	RBPJK	3
Enhancer 843	RORA1	1
Enhancer 843	RUNX1	2
Enhancer 843	RUNX2	2
Enhancer 843	RUNX3	1
Enhancer 843	RUSH1A	1
Enhancer 843	SF1	2
Enhancer 843	SIX1	1
Enhancer 843	SIX4	1
Enhancer 843	SMAD1	3
Enhancer 843	SMRC1	1
Enhancer 843	SOX	1
Enhancer 843	SOX10	4
Enhancer 843	SOX15	2
Enhancer 843	SOX17	1
Enhancer 843	SOX18	1
Enhancer 843	SOX2	1
Enhancer 843	SOX4	1
Enhancer 843	SOX9	3
Enhancer 843	Spic	1
Enhancer 843	SRY	7
Enhancer 843	STAT	2
Enhancer 843	STAT1STAT1	1
Enhancer 843	STAT2	1
Enhancer 843	STAT3	5
Enhancer 843	STAT5A	3
Enhancer 843	SUH	1
Enhancer 843	TAL1	2
Enhancer 843	TAL1::GATA1	1
Enhancer 843	TATA	1
Enhancer 843	TBX5	5
Enhancer 843	TCF4	3
Enhancer 843	Tcf7	2
Enhancer 843	TCF7L2	4
Enhancer 843	TEAD4	1
Enhancer 843	TEF	3
Enhancer 843	TEF1	1
Enhancer 843	TF7L2	3
Enhancer 843	TFAP2A	1
Enhancer 843	TFAP2C	1
Enhancer 843	TFE	1
Enhancer 843	TFEB	1
Enhancer 843	TFIII	2
Enhancer 843	TGIF1	4
Enhancer 843	TLX1::NFIC	1
Enhancer 843	TR4	1
Enhancer 843	TWST1	1
Enhancer 843	USF	2
Enhancer 843	USF1	2
Enhancer 843	USF2	1
Enhancer 843	VSX2	1
Enhancer 843	YY1	2
Enhancer 843	ZNF263	2
Enhancer 843	ZNF524	2

Promoter 3	Nkx6-1	1
Promoter 3	NKX62	1
Promoter 3	NKX6-2	1
Promoter 3	NOBOX	1
Promoter 3	NR1I2	1
Promoter 3	NR1I3	1
Promoter 3	OCT	1
Promoter 3	OCT1	7
Promoter 3	OCTAMER	2
Promoter 3	OTX	1
Promoter 3	OTX1	1
Promoter 3	P50RELAP65	1
Promoter 3	P53	2
Promoter 3	P63	1
Promoter 3	PAX2	3
Promoter 3	PAX4	4
Promoter 3	PAX6	1
Promoter 3	PBX1	2
Promoter 3	PBX2	1
Promoter 3	PEBB	1
Promoter 3	PHOX2B	1
Promoter 3	PIT1	3
Promoter 3	PKNX1	1
Promoter 3	PO2F1	1
Promoter 3	PO2F2	1
Promoter 3	PO3F1	1
Promoter 3	PO3F2	1
Promoter 3	PO5F1	1
Promoter 3	POU1F1	4
Promoter 3	POU2F1	2
Promoter 3	POU2F2	5
Promoter 3	POU2F3	2
Promoter 3	POU3F1	3
Promoter 3	POU3F2	2
Promoter 3	POU3F3	4
Promoter 3	POU3F4	1
Promoter 3	POU4F1	1
Promoter 3	POU4F2	1
Promoter 3	POU4F3	1
Promoter 3	POU5F1	2
Promoter 3	POU5F1P1	2
Promoter 3	PRRX2	1
Promoter 3	PXRRXR	1
Promoter 3	RUNX1	1
Promoter 3	SMAD1	2
Promoter 3	SMAD3	1
Promoter 3	SMAD4	1
Promoter 3	SOX10	3
Promoter 3	SOX15	2
Promoter 3	SOX9	2
Promoter 3	SPZ1	1
Promoter 3	SREBP1	1
Promoter 3	SRY	4
Promoter 3	STA5A	1
Promoter 3	STA5B	2
Promoter 3	STAT	1
Promoter 3	STAT1	7
Promoter 3	STAT3	5
Promoter 3	STAT4	1
Promoter 3	STAT5A	2
Promoter 3	STAT5B	2
Promoter 3	TAL1	4
Promoter 3	TAL1::TCF3	3
Promoter 3	TAL1ALPHAE47	1
Promoter 3	TBP	3
Promoter 3	TBR1	2
Promoter 3	TBX1	2
Promoter 3	TBX15	2
Promoter 3	TBX2	1
Promoter 3	TBX20	2
Promoter 3	TBX21	2
Promoter 3	TBX4	1
Promoter 3	TBX5	6
Promoter 3	TCF3	1
Promoter 3	TCF4	3
Promoter 3	Tcf7	2
Promoter 3	TCF7L2	3
Promoter 3	TEF	2
Promoter 3	TF7L2	3
Promoter 3	TFE2	2
Promoter 3	TFE3	1
Promoter 3	TGIF	3
Promoter 3	TGIF1	2
Promoter 3	TP53	1
Promoter 3	TP63	2
Promoter 3	UBIP1	1
Promoter 3	UNCX	1

Promoter 3	WHN	2
Promoter 3	ZN384	1
Promoter 3	ZNF333	1
Promoter 3	ZNF75A	1

Table S3: Predicted Transcription factors expressed in the brain

Gene	Gene Name
ALX1	ALX homeobox 1
ALX3	ALX homeobox 3
ALX4	ALX homeobox 4
BACH2	BTB and CNC homology 1, basic leucine zipper transcription factor 2
CEBPE	CCAAT/enhancer binding protein (C/EBP), epsilon
CTCF	CCCTC-binding factor (zinc finger protein)
CDC5L	CDC5 cell division cycle 5-like (S. pombe)
DBP	D site of albumin promoter (albumin D-box) binding protein
E2F1	E2F transcription factor 1
ELK1	ELK1, member of ETS oncogene family
ELK3	ELK3, ETS-domain protein (SRF accessory protein 2)
ESX1	ESX homeobox 1
ERF	Ets2 repressor factor
FOSB	FBJ murine osteosarcoma viral oncogene homolog B
FEV	FEV (ETS oncogene family)
GABPA	GA binding protein transcription factor, alpha subunit 60kDa
GATA2	GATA binding protein 2
GLI2	GLI family zinc finger 2
GSX2	GS homeobox 2
HBP1	HMG-box transcription factor 1
Klf12	Kruppel-like factor 12
KLF5	Kruppel-like factor 5 (intestinal)
KLF8	Kruppel-like factor 8
LHX2	LIM homeobox 2
MGA	MAX gene associated
MAX	MYC associated factor X
MAZ	MYC-associated zinc finger protein (purine-binding transcription factor)
Meis2	Meis homeobox 2
Meis3	Meis homeobox 3; Meis homeobox 3 pseudogene 2
MIXL1	Mix1 homeobox-like 1 (Xenopus laevis)
NKX2-8	NK2 homeobox 8
POU2F3	POU class 2 homeobox 3
POU3F2	POU class 3 homeobox 2
POU3F3	POU class 3 homeobox 3
POU3F4	POU class 3 homeobox 4
POU6F1	POU class 6 homeobox 1
SMAD3	SMAD family member 3
SOX15	SRY (sex determining region Y)-box 15
SOX18	SRY (sex determining region Y)-box 18
SOX5	SRY (sex determining region Y)-box 5
SP1	Sp1 transcription factor
SP2	Sp2 transcription factor
SP3	Sp3 transcription factor
SP4	Sp4 transcription factor
TBX2	T-box 2
TBX3	T-box 3
TBR1	T-box, brain, 1
TGIF1	TGFB-induced factor homeobox 1
THAP1	THAP domain containing, apoptosis associated protein 1
WT1	Wilms tumor 1
YY1	YY1 transcription factor
ATF3	activating transcription factor 3
ATF5	activating transcription factor 5
AR	androgen receptor
ARNT2	aryl-hydrocarbon receptor nuclear translocator 2
CREB3	cAMP responsive element binding protein 3
CREM	cAMP responsive element modulator

CTF1	cardiotrophin 1
CPEB1	cytoplasmic polyadenylation element binding protein 1
DLX1	distal-less homeobox 1
DLX2	distal-less homeobox 2
DMRT1	doublesex and mab-3 related transcription factor 1
EGR3	early growth response 3
EGR4	early growth response 4
EMX1	empty spiracles homeobox 1
EMX2	empty spiracles homeobox 2
EN2	engrailed homeobox 2
EOMES	eomesodermin homolog (Xenopus laevis)
ESR1	estrogen receptor 1
ESRRA	estrogen-related receptor alpha
ESRRG	estrogen-related receptor gamma
ETV1	ets variant 1
ETV7	ets variant 7
FUBP1	far upstream element (FUSE) binding protein 1
FIGLA	folliculogenesis specific basic helix-loop-helix
Foxg1	forkhead box G1
Foxj3	forkhead box J3
Foxk1	forkhead box K1
FOXP2	forkhead box P2
GBX2	gastrulation brain homeobox 2
GMEB2	glucocorticoid modulatory element binding protein 2
GFI1B	growth factor independent 1B transcription repressor
HAND1	heart and neural crest derivatives expressed 1
HSF1	heat shock transcription factor 1
HSF2	heat shock transcription factor 2
HLTF	helicase-like transcription factor
HLF	hepatic leukemia factor
Hoxd3	homeobox D3
HMBOX1	homeobox containing 1
HIC2	hypermethylated in cancer 2
IRF3	interferon regulatory factor 3
IRF8	interferon regulatory factor 8
MECP2	methyl CpG binding protein 2 (Rett syndrome)
MZF1	myeloid zinc finger 1
MEF2A	myocyte enhancer factor 2A
MEF2D	myocyte enhancer factor 2D
NHLH1	nescient helix loop helix 1
NF1	neurofibromin 1
NEUROD2	neurogenic differentiation 2
NEUROG2	neurogenin 2
NFIA	nuclear factor I/A
NFIX	nuclear factor I/X (CCAAT-binding transcription factor)
NFAT5	nuclear factor of activated T-cells 5, tonicity-responsive
NFATC1	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 1
NR2E1	nuclear receptor subfamily 2, group E, member 1
NR4A2	nuclear receptor subfamily 4, group A, member 2
NFYC	nuclear transcription factor Y, gamma
OLIG2	oligodendrocyte lineage transcription factor 2
OLIG1	oligodendrocyte transcription factor 1
OTX1	orthodenticle homeobox 1
PAX6	paired box 6
PHOX2A	paired-like homeobox 2a
PHOX2B	paired-like homeobox 2b
PLAG1	pleiomorphic adenoma gene 1
PBX1	pre-B-cell leukemia homeobox 1
PURA	purine-rich element binding protein A

RREB1	ras responsive element binding protein 1
RFX4	regulatory factor X, 4 (influences HLA class II expression)
RARA	retinoic acid receptor, alpha
RXRB	retinoid X receptor, beta
SRF	serum response factor (c-fos serum response element-binding transcription factor)
STAT1	signal transducer and activator of transcription 1, 91kDa
STAT4	signal transducer and activator of transcription 4
STAT5A	signal transducer and activator of transcription 5A
STAT5B	signal transducer and activator of transcription 5B
IK	similar to CG18005; IK cytokine, down-regulator of HLA II
SPZ1	spermatogenic leucine zipper 1
SF1	splicing factor 1
TEF	thyrotrophic embryonic factor
TFAP4	transcription factor AP-4 (activating enhancer binding protein 4)
TFEB	transcription factor EB
TFEC	transcription factor EC
TP53	tumor protein p53
ETS2	v-ets erythroblastosis virus E26 oncogene homolog 2 (avian)
Mafb	v-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian)
MAFK	v-maf musculoaponeurotic fibrosarcoma oncogene homolog K (avian)
MYBL2	v-myb myeloblastosis viral oncogene homolog (avian)-like 2
VAX1	ventral anterior homeobox 1
ZEB1	zinc finger E-box binding homeobox 1
ZBTB4	zinc finger and BTB domain containing 4
ZNF219	zinc finger protein 219
ZNF333	zinc finger protein 333
ZBTB49	zinc finger protein 509
ZNF524	zinc finger protein 524

Table S4: Predicted Transcription factors in the "Neuron differentiation" category

Gene	Gene Name
BARHL2	BarH-like homeobox 2
CEBPB	CCAAT/enhancer binding protein (C/EBP), beta
GATA2	GATA binding protein 2
GLI2	GLI family zinc finger 2
GLI3	GLI family zinc finger 3
GSX1	GS homeobox 1
GSX2	GS homeobox 2
ISL1	ISL LIM homeobox 1
ISL2	ISL LIM homeobox 2
LHX2	LIM homeobox 2
LHX4	LIM homeobox 4
LHX5	LIM homeobox 5
LHX6	LIM homeobox 6
Lhx8	LIM homeobox 8
LMX1A	LIM homeobox transcription factor 1, alpha
LMX1B	LIM homeobox transcription factor 1, beta
NKX2-8	NK2 homeobox 8
NKX6-1	NK6 homeobox 1
NKX6-2	NK6 homeobox 2
POU3F2	POU class 3 homeobox 2
POU3F4	POU class 3 homeobox 4
POU4F1	POU class 4 homeobox 1
POU4F2	POU class 4 homeobox 2
POU4F3	POU class 4 homeobox 3
RORA	RAR-related orphan receptor A
SMAD4	SMAD family member 4
SOX2	SRY (sex determining region Y)-box 2
SOX5	SRY (sex determining region Y)-box 5
TBR1	T-box, brain, 1
TLX1	T-cell leukemia homeobox 1
ARX	aristaless related homeobox
Atoh1	atonal homolog 1 (Drosophila)
BHLHE22	basic helix-loop-helix family, member e22
CREB1	cAMP responsive element binding protein 1
CTF1	cardiotrophin 1
CUX1	cut-like homeobox 1
DLX1	distal-less homeobox 1
DLX2	distal-less homeobox 2
DLX5	distal-less homeobox 5
EGR2	early growth response 2
EMX1	empty spiracles homeobox 1
EMX2	empty spiracles homeobox 2
EN1	engrailed homeobox 1
EN2	engrailed homeobox 2
ETV4	ets variant 4
EVX1	even-skipped homeobox 1
FOXA1	forkhead box A1
FOXA2	forkhead box A2
Foxg1	forkhead box G1
GBX2	gastrulation brain homeobox 2
GFI1	growth factor independent 1 transcription repressor
HES1	hairy and enhancer of split 1, (Drosophila)
HOXA1	homeobox A1
HOXA2	homeobox A2
HOXC10	homeobox C10
HOXC8	homeobox C8
Hoxd9	homeobox D9

ID4	inhibitor of DNA binding 4, dominant negative helix-loop-helix protein
IRX5	iroquois homeobox 5
MNX1	motor neuron and pancreas homeobox 1
NRL	neural retina leucine zipper
NEUROD2	neurogenic differentiation 2
NEUROG2	neurogenin 2
NR2E1	nuclear receptor subfamily 2, group E, member 1
NR2E3	nuclear receptor subfamily 2, group E, member 3
NR2F6	nuclear receptor subfamily 2, group F, member 6
NR4A2	nuclear receptor subfamily 4, group A, member 2
OLIG2	oligodendrocyte lineage transcription factor 2
OLIG1	oligodendrocyte transcription factor 1
OTX2	orthodenticle homeobox 2
PAX2	paired box 2
PAX3	paired box 3
PAX6	paired box 6
PAX7	paired box 7
PHOX2A	paired-like homeobox 2a
PITX3	paired-like homeodomain 3
PTF1A	pancreas specific transcription factor, 1a
PBX3	pre-B-cell leukemia homeobox 3
RUNX3	runt-related transcription factor 3
SRF	serum response factor (c-fos serum response element-binding transcription factor)
STAT3	signal transducer and activator of transcription 3 (acute-phase response factor)
Tp73	tumor protein p73
VAX1	ventral anterior homeobox 1
VAX2	ventral anterior homeobox 2
VSX1	visual system homeobox 1
VSX2	visual system homeobox 2

Nederlandse samenvatting

Mutaties in het FOXP2-gen veroorzaken problemen bij de ontwikkeling van spraak en taal. Het gen codeert voor een eiwit uit de transcriptiefactor (TF)-familie, dat met het DNA interageert en daardoor honderden andere genen aan en uit kan zetten. De mutaties in FOXP2 zorgen ervoor dat het eiwit niet aan het DNA kan binden of niet goed geproduceerd kan worden. Onder normale omstandigheden wordt FOXP2 geproduceerd in de longen, het hart, de darm en de hersenen. In de hersenen wordt het gen aangezet in bepaalde structuren, zoals de hersenenschors, de kleine hersenen en de basale gangliën. De taalproblemen bij mensen met FOXP2-mutaties worden vooral veroorzaakt door het ontbreken van het normale FOXP2-eiwit in de hersenen.

De effecten van de mutatie op moleculair en cellulair niveau zijn intensief bestudeerd en zij leveren veel kennis op over de rol van het gen in de ontwikkeling van neuronen en de formatie van neurale netwerken die de menselijke taalvaardigheden beïnvloeden. In hoofdstuk 1 van deze PhD-thesis beschrijf ik hoe de eerste mutaties in FOXP2 ontdekt zijn, wat voor aandoeningen mensen met FOXP2-mutaties precies hebben en wat er al is ontdekt over de functie van FOXP2 in menselijke en dierlijke hersencellen. Hoewel al veel bekend is over de processen die het FOXP2-eiwit reguleert, weten wij vrij weinig over de processen die de productie van het FOXP2-gen zelf reguleren. Daarom was het doel van dit PhD-project om te onderzoeken hoe het FOXP2-gen wordt aangezet, wat voor processen daar een rol bij spelen en wat er gebeurt als de regulatie misgaat.

Mijn onderzoek uit hoofdstuk 2 is gebaseerd op een eerdere beschrijving van een kind met een vertraagde taalontwikkeling. In het genoom van het kind werd aangetoond dat sommige delen van het DNA in volgorde zijn omgedraaid. Eén van de breukpunten, waar delen van het genoom zijn omgedraaid, ligt in de buurt van het FOXP2-gen en het gen zelf werd niet veranderd. Het idee is dus, dat door de omgedraaide volgorde een regulerend DNA-element verplaatst werd en het FOXP2-eiwit niet goed wordt aangemaakt. Ik beschrijf in hoofdstuk 2

hoe ik een regulerend element, enhancer genoemd, met behulp van het bovengenoemde breukpunt kon identificeren. Bij het bovengenoemde kind is de enhancer verplaatst en zal het FOXP2-gen niet kunnen reguleren, met als gevolg dat FOXP2 niet goed wordt aangemaakt.

Enhancers reguleren de productie in bepaalde weefsels en cellen. Genen met een complexe expressie, zoals FOXP2, hebben meestal meerdere enhancers. In hoofdstuk 3 was het mijn doel om verdere enhancers van het gen op te sporen. Een eigenschap van enhancers is dat ze de productie van genen reguleren door contact te maken met het beginpunt van het gen, de zogenaamde promoter. Om FOXP2-enhancers op te sporen heb ik de contacten van de FOXP2-promoter met andere delen van het genoom bepaald. Ik verwachtte in cellen die FOXP2 produceren contacten te ontdekken die afwezig -of minder sterk- zijn in cellen zonder FOXP2-productie. Inderdaad kon ik een aantal van deze contacten detecteren en daardoor potentiële enhancers ontdekken.

De activiteit van enhancers en promoters wordt mede bepaald door TFs – waarvan FOXP2 er éé is-, die aan specifieke DNA-motieven binden. In hoofdstuk 4 heb ik de activiteit van enhancers in menselijke cellen getest en vervolgens bepaald of de activiteit afhankelijk is van bepaalde TFs. In totaal heb ik de invloed van 8 TFs bestudeerd op basis van vermoede associaties uit de literatuur. Inderdaad zag ik voor 7 TFs dat ze de activiteit van de enhancers of promoters verhogen of verlagen. Ik kon bijvoorbeeld zien dat FOXP2 zelf de activiteit van zijn promoter en een enhancer verhoogt. Het gen zou dus zijn eigen productie positief kunnen beïnvloeden. Deze positieve autoregulatie zou een bijdrage kunnen leveren aan het feit dat mutaties in één van de twee FOXP2-kopieën voldoende is om de bekende aandoening in taalontwikkeling te veroorzaken.

Zoals eerder genoemd, zijn enhancers slechts in bepaalde weefsels en cellen actief. Zij dragen dus bij aan de productie van het gen in een beperkt aantal cellen. Om te begrijpen waar een enhancer door TFs wordt aangestuurd is het dus noodzakelijk om de weefsel-

specifieke activiteit van de enhancers te onderzoeken. In het vijfde hoofdstuk heb ik twee enhancers in transgene muizen bestudeerd. Deze muizen produceren een reportergen onder de controle van de ontdekte enhancer. Het reportergen is dan eenvoudig te visualiseren in de weefsels en cellen waar de enhancer actief is. Ik zag dat een enhancer actief was in de cortex, en juist rond de geboorte activiteit vertoonde in het specifieke weefsel waar ook FOXP2 wordt geproduceerd. Deze enhancer zal dus FOXP2 tijdens de ontwikkeling aan kunnen sturen. Verdere overeenstemming tussen enhanceractiviteit en FOXP2-productie was te zien in de cerebellum. De enhancer stuurt dus de productie van FOXP2 tijdens een belangrijke fase van de hersenontwikkeling. Genetische variatie in deze enhancer zou kunnen veroorzaken dat FOXP2 in deze weefsels minder goed wordt geproduceerd en zou dus problemen in taal en spraak kunnen veroorzaken. Verder onderzoek naar mutaties in deze enhancer bij mensen met specifieke taalproblemen zou dus een genetische verklaring van deze specifieke aandoeningen kunnen opleveren.

Voor de zesde hoofdstuk was mijn doel om aan te tonen dat normale genetische veranderingen in enhancers de structuur van de hersenen kan beïnvloeden. Normale genetische veranderingen zijn variaties van enkele DNA-bouwstenen die bij iedereen aanwezig kunnen zijn. Er zijn miljoenen posities in het genoom bekend waar mensen van elkaar verschillen. De effecten van enkele variaties zullen dus heel klein zijn. Om toch een effect te ontdekken, heb ik zo'n 300 enhancers uit de wetenschappelijke literatuur onderzocht, waarvan bekend is dat ze in de hersenen actief zijn. Ik wilde weten of normale genetische verandering in één van de enhancers invloed op de hersenenstructuur van gezonde mensen heeft. Een vergelijkbare vraag werd eerder onderzocht door een wereldwijd consortium, Enhancing Neuro Imaging Genetics by Meta-Analysis genaamd, kortweg ENIGMA. Dit consortium heeft de normale genetische variatie en hersenenstructuur van 30,000 mensen bestudeerd en voor meer dan 8 miljoen variaties statistisch berekend wat de invloed op de structuren in het midden van de hersenen zou kunnen zijn. Bij 8 miljoen bekeken variaties is het heel moeilijk om statistisch aan te tonen dat het effect van één

variant echt de hersenenstructuur beïnvloedt, of dat de berekende invloed door toeval is ontstaan. Omdat ik niet naar het hele genoom keek, maar naar een beperkt aantal interessante posities, de 300 enhancers, was het mogelijk om statistische effecten in deze enhancers aan te tonen. In hoofdstuk zes laat ik zien dat normale genetische variatie in enhancers, die in de hersenen actief zijn, de normale variatie in de hersenstructuur beïnvloeden. Verder onderzoek naar genetische variatie in enhancers van FOXP2 zou mogelijk effecten op kunnen leveren in de hersendelen waar FOXP2 geproduceerd wordt of in de neuronale functies van deze hersengebieden.

Samenvattend, ik heb de basiselementen van genregulatie op moleculair en cellulair niveau onderzocht en de mogelijke effecten van mis-regulatie bestudeerd. Mijn onderzoek was speciaal gericht op het FOXP2-gen, dat bekend is als veroorzaker van mogelijke ontwikkelingsstoornissen van taal en spraak. De resultaten van mijn project tonen genetische elementen aan die de productie van FOXP2 sturen en factoren die de productie verder kunnen versterken of verminderen. Bovendien heb ik voor twee enhancers de activiteit in de hersenen zien laten zien. De afwezigheid van deze enhancers zou een taalontwikkelingsstoornis kunnen veroorzaken. Normale genetische variatie in deze enhancers zou ook een invloed kunnen hebben op de normale variatie in de hersenstructuur. Met behulp van mijn resultaten kunnen toekomstige studies onderzoeken hoe de productie van het FOXP2-gen de hersenontwikkeling verandert.

Curriculum Vitae

Martin Becker was born on the 21st of March, 1985 in Berlin, Germany and obtained his Abitur from the Alfred-Krupp-Schule in Essen, Germany in 2004. The next two years he visited the Berufskolleg Hilden and graduated in 2006 as biological research assistant (Biologisch-technischer Assistant) with a specialisation in molecular biological techniques obtained from the chamber of industry and commerce (Industrie und Handelskammer) Düsseldorf.

In 2009 he finished his Bachelor of Science in the Molecular Lifescience program of the Radboud University Nijmegen, The Netherlands. In 2011, Martin obtained the Master of Science degree from the medical faculty of the Radboud University after completing the Molecular Mechanism of Disease program. During that time, he studied gene enhancer variants associated with colorectal cancer at the Wellcome Trust Centre for Human Genetics, Oxford University, UK and chromosomal rearrangements in patients with intellectual disability at the Human Genetics department of the Radboud University Medical Centre, Nijmegen.

In September 2011, Martin started his PhD project at the Max Planck Institute for Psycholinguistics, Nijmegen under the supervision of Prof. Dr. Simon Fisher and Dr. Sonja Vernes. Martin investigated the transcriptional regulation of the *FOXP2* gene and neurodevelopmental gene enhancers in health and disease and obtained his doctoral degree in October 2016 from the Radboud Univserity Nijmegen.

Since March 2016, Martin is working in Dr. Kristiina Tammimies group at the Karolinska Institute Centre for Neurodevelopmental Disorders in Stockholm, Sweden and investigates the interaction of genetic variants and environmental exposure in the development of autism.

List of publications

Becker, M., Guadalupe, T., Franke B., Hibar, D.P., Renteria, M.E., Stein, J.L., Thompson, P.M., Francks, C., Vernes, S.C., Fisher, S.E. (2016) Early developmental gene enhancers affect subcortical volumes in the adult human brain. *Human Brain Mapping, 37:1788-1800.*

Becker, M., Devanna P., Fisher, S.E. & Vernes, S.C. (2015). A chromosomal rearrangement in a child with severe speech and language disorder separates FOXP2 from a functional enhancer. *Molecular Cytogenetics, 8:69*.

Lewis, A., Freeman-Mills, L., de la Calle-Mustienes, E., Giráldez-Pérez, R.M., Davis, H., Jaeger, E., **Becker, M.**, Hubner, N.C., Nguyen, L.N., Zeron-Medina, J., Bond, G., Stunnenberg, H.G., Carvajal, J.J., Gomez-Skarmeta, J.L., Leedham, S., Tomlinson, I. (2014). A polymorphic enhancer near GREM1 influences bowel cancer risk through differential CDX2 and TCF7L2 binding. *Cell Reports, 8, 983-990.*

Jaeger, E., Leedham, S., Lewis, A., Segditsas, S., **Becker, M**., Rodenas-Cuadrado, P., Davis, H., Kaur, K., Heinimann, K., Howarth, K., East, J., Taylor, J., Thomas, H., & Tomlinson, I. (2012). Hereditary mixed polyposis syndrome is caused by a 40-kb upstream duplication that leads to increased and ectopic expression of the BMP antagonist GREM1. *Nature Genetics, 44, 699-703*.

Acknowledgments

First of all, I would like to thank my promoter Simon Fisher for giving me the opportunity to pursue my PhD project in his department and for taking his time in giving me so much valuable feedback. I want to thank my co-promoter Sonja Vernes for challenging me and my ideas to constantly improve my critical thinking and experimental approaches.

Further I want to thank everyone in the lab! All colleagues who created a great lab atmosphere and have all helped me a lot to get around in the lab. Pela for being my favourite Greek. Sarah for showing me what really good organization skills are. Pedro for sharing the sunset over Elafonisos with me. Paolo for keeping me awake with amazing coffee and teaching me some of his lab magician skills. Jasper for his magnificent lab management skills that made lab work so much smoother. Joses for showing me how to hug people. Sara for being hilarious company as a car passenger. Kai for inspiring my T-shirt selection. Elliot for understanding my jokes. Arianna for keeping Paolo in check and Moritz for helping me out with the mouse brains.

Many thanks go to the computational side of the department: Clyde for his initial support to pursue the idea underlying chapter 6. Tulio for the great exchange of ideas and explaining concepts in statistical genetics. Rick for teaching me in Python. Dan for always asking smart questions and Amaia for fun talks about science.

I further want to thank all my friends in Nijmegen, who have significantly improved my life in the Netherlands. On a similar note, I want to thank my friends from Essen, Germany, who are the reason that this city will keep on being the anchor point of my life.

My biggest thanks go to Antje for having my back during the last year, helping me to stay sane, on track and to move forward in life. Finally, I want to thank my sister Maria for being a leading example throughout my life and my parents for raising a little scientist in me.

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-theoretical perspective. *Victor Kuperman*
51. A grammar of Savosavo, a Papuan language of the Solomon Islands. *Claudia Wegener*
52. Prosodic structure in speech production and perception. *Claudia Kuzla*
53. The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production and comprehension. *Sarah Schimke*
54. Studies on intonation and information structure in child and adult German. *Laura de Ruiter*
55. Processing the fine temporal structure of spoken words. *Eva Reinisch*
56. Semantics and (ir)regular inflection in morphological processing. *Wieke Tabak*
57. Processing strongly reduced forms in casual speech. *Susanne Brouwer*
58. Ambiguous pronoun resolution in L1 and L2 German and Dutch. *Miriam Ellert*
59. Lexical interactions in non-native speech comprehension: Evidence from electro-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. Incrementality and Flexibility in Sentence Production. *Maartje van de Velde*
98. Social learning dynamics in chimpanzees: Reflections on (nonhuman) animal culture. *Edwin van Leeuwen*

99. The request system in Italian interaction. *Giovanni Rossi*
100. Timing turns in conversation: A temporal preparation account. *Lilla Magyari*
101. Assessing birth language memory in young adoptees**.** *Wencui Zhou*
102. A social and neurobiological approach to pointing in speech and gesture. *David Peeters*
103. Investigating the genetic basis of reading and language skills. *Alessandro Gialluisi*
104. Conversation Electrified: The Electrophysiology of Spoken Speech Act Recognition**.** *Rósa Signý 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*