

THE UNIVERSITY OF OXFORD



FUNCTIONAL INVESTIGATION OF MICRORNA
PATHWAYS IN HUMAN SPEECH AND LANGUAGE
NETWORKS

JOSES W. HO, SC. B.

MAGDALEN COLLEGE, OXFORD

SUPERVISORS:

PROF. DR. S.E. FISHER

DR. DIANNE NEWBURY

Submitted in partial fulfilment of the requirements
for Doctor of Philosophy in Clinical Medicine

Michaelmas Term 2014

Abstract

The forkhead-box protein P2 (FOXP2) is a transcription factor implicated in an inherited human speech and language disorder. A heterozygous R553H mutation in the gene is associated with a monogenic form of developmental verbal dyspraxia. Reduced functional dosage in mice brains causes altered synaptic plasticity in cortico-striatal circuits, and impairs motor learning. Genome-wide *in vivo* ChIP–chip screens for Foxp2 binding in the developing mouse brain have identified a set of microRNAs (miRNAs) whose promoter regions were occupied by Foxp2. Thus, I employed microarray expression profiling to identify microRNAs regulated by Foxp2 in the embryonic mouse brain. Interestingly, the microarray results were divergent from *in vitro* effects of Foxp2/FOXP2 on these microRNAs, suggesting that regulation of these microRNAs is context-specific. In light of the regulation of *MIR137* by FOXP2, I went on to study a putative genetic network involving FOXP2, *TCF4*, *CNTNAP2* and *MIR137*. While *TCF4*, *CNTNAP2* and *MIR137* each are regulated *in vitro* by FOXP2, the regions bound by FOXP2 were insufficient to generate the anticipated transcriptional regulation when cloned upstream of a reporter gene. To test if FOXP2-regulated microRNAs are risk factors for specific language impairment (SLI), I performed candidate association analyses on a large family-based cohort of SLI probands. I also identified 5 common variants and 3 rare variants in predicted microRNA binding sites in the 3' UTR from the exomes of 48 of these probands. I tested them for association with SLI in the larger cohort. rs72727021, lying in the miR-192/215 binding site of the *C9ORF100* 3' UTR, was significantly associated with nonword repetition in the SLI cohort (p-value = 2.75×10^{-4}). Testing of rs72727021 and the 3 rare variants with a luciferase reporter assay revealed that 2 rare variants affected gene expression levels independently of the microRNA, suggesting that these 2 microRNA binding sites might be risk factors contributing to the SLI phenotype. FOXP2/Foxp2 is likely to act through microRNAs to generate and fine-tune neural gene expression

patterns that underlie neuronal function and neural circuits important for vocal learning, and for human speech and language.

I hereby declare that the work contained herein is entirely my own, unless explicitly and clearly stated.

Joses W. Ho

Approximate Word Count of Thesis: 40, 751 words

Acknowledgements

Throughout the entire course of my DPhil, I received funding from the Agency for Science, Technology, and Research, Singapore (A*STAR). Their generous sponsorship allowed me to travel to the 2012 Society for Neuroscience in New Orleans, USA.

Dr. Dianne Newbury's prompt supervision regarding several aspects of the association analysis was invaluable. I am deeply appreciative of her availability and patience.

Dr. Simon Fisher took on an inexperienced student from Singapore, and his supervision steered me through dead-ends and mistakes. I am grateful for the opportunities he gave me to present my work, and to pursue an area of research I found fascinating.

Several portions of this thesis have relied heavily upon the close instruction of Dr. Sonja Vernes. I have learnt much regarding the planning and execution of all manners of laboratory work and data analysis from her. I am indebted to her tireless bench-side supervision of difficult procedures (such as dissections), her patience at revising and rehearsing scientific presentations with me, and her constant and keen interest in my progress.

Whilst in Oxford, UK, I learnt much about animal husbandry from Rathi Puliyadi. Midway through this DPhil, I moved to Nijmegen, the Netherlands, when the Fisher lab relocated to the Max Planck Institute for Psycholinguistics. I will always be grateful to members of the Language and Genetics Group at for their collegiality, friendship, and food – in no particular order: Elliot, Jasper, Kai, Martin, Paolo, Pedro, Pela, Sara, Sarah, and the rest of the folks happily working on the “dysfunctional” side of the research group.

I was blessed to cross the paths of Joel, Jill, Rob, Rachael, and the Fisher-Hoyrems during my time at Oxford. Their friendship continues to leave a mark on the way I order my time and my thoughts.

And my family: Mom and Dad, Huiran and Lucas, and Matthias – their love and support throughout these years were steadfast and more than a blessing. I thank God for their presence and encouragement, which were more than crucial in the final months.

Each of us finds his way, his place; we rattle around the universe until everything fits; this is life; this is science, or something better than science.

The Fifth Head of Cerberus

Gene Wolfe

Contents

List of Figures	14
List of Tables	16
List of Abbreviations	17
Chapter 1: Introduction	19
1.1 Language: A Unique and Innate Human Trait	19
1.1.1 Communication and Language	19
1.1.2 Speech and Vocal Learning	20
1.2 Neural Basis for Speech and Language	21
1.2.1 Language and the Network Activity of the Brain	21
1.2.2 Cortical Regions and Language: Broca's Area and the Wernicke's Area ..	21
1.2.3 Subcortical Regions and Language: The Basal Ganglia and the Thalamus	23
1.2.4 The Cerebellum and Language	24
1.2.5 Distributed Activity, Functional Integration, and Language: Beyond The Classical Model	25
1.3 Genetic Basis for Speech And Language	26
1.4 The KE Family and <i>FOXP2</i>	27
1.4.1 The KE Family and Their Phenotype	27
1.4.2 Abnormal Activity and Structure of Brain Regions Found In The KE Family	30
1.4.3 The Discovery of <i>FOXP2</i> : From Linkage To A Candidate Gene	32
1.4.4 Mutations in <i>FOXP2</i> Are Associated With Developmental Verbal Dyspraxia	33
1.4.5 The Sequence and Expression of <i>FOXP2</i> : High Inter-species Conservation	34
1.4.6 <i>FOXP2</i> and Other Forkhead Genes	37
1.5 The Function of <i>FOXP2</i> , And Effects of Pathogenic Mutation In The Brain ..	38
1.5.1 The Functions of <i>FOXP2</i> at the Molecular Level: DNA binding and Dimerization With Co-Factors	38

1.5.2 The Functions of <i>FOXP2</i> in Gene Regulation: Genetic Networks Involved in Neuronal Function And Under Positive Selection.....	42
1.5.3 The Functions of <i>FOXP2</i> in the Neuron: Neurogenesis, Differentiation, Morphology, and Synaptogenesis	44
1.5.4 The Functions of <i>FOXP2</i> in Brain Circuits: Roles in Synaptic Activity and Plasticity	47
1.5.5 The Function of <i>FOXP2</i> on Animal Behaviour: Effects on Rodent Motor Skills, Rodent Pup Vocalizations, and Songbird Vocal Learning.....	48
1.5.6 The Functions of <i>FOXP2</i> : From Genes to Neurons to Behaviour, and Back Again	51
1.6 MicroRNAs: A Large Family of Small Noncoding Gene Regulators	51
1.7 Building Bridges Between MicroRNAs and Human Speech and Language ..	54
1.8 Goals of this Thesis	54
Chapter 2: Materials and Methods	57
2.1 <i>In Vitro</i> Cell Culture.....	57
2.2 RNA Extraction	58
2.4 Microarray Processing and Analysis.....	59
2.3 Quantitative Real-time Polymerase Chain Reaction (qRT-PCR)	60
2.4 MicroRNA Target Prediction	63
2.5 Gene Ontology Analysis	63
2.6 Primary Cell Culture.....	63
2.7 Immunofluorescence	64
2.8 Generation of Vectors.....	65
2.10 Luciferase Reporter Assay	67
2.11 Candidate SNP analysis	68
2.12 Exome Analysis	70
Chapter 3: Regulation of MicroRNAs by <i>Foxp2</i> / <i>FOXP2</i> – A Microarray Expression Analysis.....	73
3.1 Introduction	73

3.1.1 MicroRNA Biogenesis In Animals	73
3.1.2 Identifying Promoters of MicroRNA Genes	76
3.1.3 Identifying Transcription Factors That Regulate MicroRNA Genes	78
3.1.4 Transcription Factors and MicroRNA Genes: Feedback Loops In Gene Regulatory Networks	80
3.1.5 Aims of this chapter	83
3.2 Results	84
3.2.1 Normalization, Processing, and Quality Control of Microarrays	84
3.2.2 Eight MicroRNAs are Differentially Expressed in S321X.Foxp2 Embryonic Mutant Mouse Brain	88
3.2.3 Axon Guidance is A Central Theme of Predicted Targets of Top 8 MicroRNAs	92
3.2.4 Validation of Microarray Results <i>in vitro</i>	95
3.3. Discussion	105
3.3.1. Foxp2 regulation of microRNAs: Genetic Networks Underlying Neural Function and Development	105
3.3.2. Foxp2 regulation of microRNAs: Wiring and Building Neuronal Networks	109
3.3.3. A Complex Picture of MicroRNA Regulation by Foxp2: Comparing In Vivo and In Vitro Effects	112
3.3.4. The R553H KE Mutation: Dominant Negative Effects on MicroRNA Regulation?	114
3.3.5 Conclusion	116
Chapter 4: Investigation of A Putative Genetic Network – Beyond FOXP2 and Verbal Dyspraxia?	118
4.1 Introduction	118
4.1.1. FOXP2 Target Genes and Other Disorders: Glimpses of a Common Underlying Genetic Network.....	118
4.1.2. <i>CNTNAP2</i> and Neurodevelopmental Disorders: Autism, Intellectual Disability, Epilepsy, and Language Impairment	118
4.1.3. <i>MIR137</i> and Schizophrenia	121

4.1.4. <i>TCF4</i> and Pitt-Hopkins Syndrome	122
4.1.5. A Mini-network Comprising Four Genes Involved Across A Range of Neurological Disorders	124
4.1.6. Aims of This Chapter	125
4.2 Results	126
4.2.1. Regulation of Direct Foxp2 Murine Targets in Human Neuron-like Cell- lines	126
4.2.2. Expression Patterns Suggest Genetic Interactions of <i>MIR137</i> , <i>CNTNAP2</i> , <i>FOXP2</i> , and <i>TCF4</i>	128
4.2.3. The FOXP2-bound Intronic Region of <i>CNTNAP2</i> Is Insufficient To Produce Transcript Down-regulation	130
4.2.4. The FOXP2-bound Promoter/Intronic Region of <i>TCF4</i> Is Insufficient To Produce Transcript Up-regulation	133
4.2.5. The FOXP2-bound Regions in the <i>MIR137</i> Locus Are Insufficient To Produce Transcript Up-regulation	134
4.3 Discussion	137
4.3.1. Network Motifs in Proposed Mini-network	137
4.3.2. Insufficiency of Binding Sites to Regulate Reporter Expression Suggests a Role of Chromatin Loops in <i>FOXP2</i> Function	138
4.3.3. Conclusion	139
Chapter 5: Are MicroRNAs Risk Factors For Specific Language Impairment?	140
5.1 Introduction	140
5.1.1 Defining Specific Language Impairment	140
5.1.2 The Search for Genetic Factors Underlying SLI	141
5.1.3 Exome Sequencing and Neurodevelopmental Disorders	147
5.1.4 From Variants To Function: Validating Exome-derived Coding Mutations	148
5.1.5 From Variants To Function: Validating Noncoding Mutations.....	149
5.1.6 Genetic Mutations in 3' UTR That Affect MicroRNA Binding Can Cause Human Disease	150

5.1.7 Genetic Mutations in MicroRNAs Can Affect Cellular Function, And Cause Human Disease	151
5.1.8 Aims of This Chapter	152
5.2 Results	154
5.2.1 Candidate Association of SNPs lying in 13 FOXP2-regulated MicroRNAs to SLI	154
5.2.2 Association of SNPs in FOXP2-regulated MicroRNAs to Measures of Language Ability	156
5.2.3 Identifying Variants in MicroRNA Binding Sites From Exome Data	162
5.2.4 Association of Exome-Derived Variants in MicroRNA Binding Sites With SLI	166
5.2.5 Assessing The Conservation and Known Function of Exome Variants in MicroRNA Binding Sites	169
5.2.6 Assessing The Thermodynamic Properties of Exome Variants in MicroRNA Binding Sites	173
5.2.7 Segregation of Rare 3' UTR Variants in MicroRNA Binding Sites	175
5.2.8 Functional Analysis of Exome Variants in MicroRNA Binding Sites With Reporter Gene Assays	177
5.3 Discussion	187
5.3.1 Variation in MicroRNA Primary Transcripts and Association to SLI	187
5.3.2 Functional Investigation of Exome-Derived 3' UTR Binding Site Variants	191
Chapter 6: Conclusion	195
6.1 Summary of Findings of This Thesis	195
6.2 An Expanded Reach of FOXP2 via MicroRNAs	196
6.3 Foxp2/FOXP2 Function, Genetic Background And Cell-type Differences ..	197
6.4 FOXP2 and Genetic Networks Underlying Different Neurodevelopmental Disorders	198
6.5 The Functional Consequences of MicroRNA Variation: Moving Beyond Seed Region and Binding Site Variants.....	200

6.6 Genetic Networks, and Human Speech and Language: The Ways Forward	201
References	203

List of Figures

Figure 1.1. Brain Regions Important For Human Speech And Language.....	22
Figure 1.2. The KE Pedigree.	28
Figure 1.3. Protein-coding Mutations and Disruptions in <i>FOXP2</i> Found In Subjects With Verbal Dyspraxia.	34
Figure 1.4. Alignment of Exon 7 from <i>FOXP2</i> sequences of 13 species.	35
Figure 1.5. Functional Testing of Four Mutations Found in Individuals with Verbal Dyspraxia.	38
Figure 1.6. Three-dimensional Structure of DNA-bound <i>FOXP2</i> Monomers and Homodimers.....	40
Figure 1.7. MicroRNAs are Involved In Several Processes Leading Up to the Formation of Adult Neural Circuits.	53
Figure 3.2. Network motifs of regulatory loops between transcription factors and microRNAs.	81
Figure 3.3. Raw Intensities of Spike-In Controls.	85
Figure 3.4. Boxplot of Probe Intensities Across 12 MicroRNA Microarrays of Mouse Brain RNA.	86
Figure 3.5. Clustered Heatmap of MicroRNA Expression Profiles of Mouse Embryonic Ganglionic Eminences.	87
Figure 3.6. MicroRNAs Differentially Regulated Between the Ganglionic Eminences of Wildtype and Mutant STOP Mice.	89
Figure 3.7. Expression and Clustering of MicroRNAs Previously Identified As <i>Foxp2</i> Targets In Embryonic Mouse Brains.....	90
Figure 3.8. Overlap of Predicted Gene Targets Across Four Prediction Algorithms..	92
Figure 3.9. Predicted MicroRNA Targets of <i>Foxp2</i> Are Involved In Axon Guidance, And Are Localised To The Synapse.	93
Figure 3.10. Overlap of Predicted Targets of <i>Foxp2</i> -regulated microRNAs with Genes Showing <i>Foxp2</i> -dependent Expression in Mouse Ganglionic Eminences.	95
Figure 3.11. qPCR Validation of MicroRNA Targets In Murine Neuro2A Cells.....	98
Figure 3.12. Differentiation Protocol Applied to Stably Transfected SH-SY5Y Cells.	100
Figure 3.13. Selection of Stable MicroRNA Normalisers For qRT-PCR in SH-SY5Y cells.....	101
Figure 3.14. qPCR Validation of MicroRNA Targets In Human Neuron-like SH-SY5Y Cells.	102
Figure 4.1. A Putative Mini-network of Genes Implicated In A Range of Neurodevelopmental Disorders.	125
Figure 4.2. <i>FOXP2</i> Directly Regulates <i>CNTNAP2</i> , <i>TCF4</i> , and miR-137.....	127
Figure 4.3. <i>Tcf4</i> is Expressed in Embryonic Mouse Brain.	128
Figure 4.4. Correlation Heatmap of Expression Patterns of <i>MIR137</i> , <i>CNTNAP2</i> , <i>FOXP2</i> , and <i>TCF4</i> in Adult Human Striatum.	129
Figure 4.6. <i>FOXP2</i> Regulation of <i>TCF4</i> Binding Site in HEK Cells.....	133
Figure 4.7. <i>FOXP2</i> Regulation of <i>MIR137</i> Binding Site in HEK Cells.	135

Figure 4.8. The FOXP2 ChIP-seq Peak Overlaps with a REST Binding Peak.	136
Figure 5.1. Distribution of Language Scores for Children.	154
Figure 5.2. Regional Association Plots for MicroRNA Transcripts Associated With Expressive Language Scores.	157
Figure 5.3. Regional Association Plots for MicroRNA Transcripts Associated With Nonword Repetition Scores.	160
Figure 5.4. Variant Identification in 3' UTRs from Exome Data of 48 SLI Probands: A Summary Workflow.	163
Figure 5.5. Coverage and miRNA Binding Site Density in 3' UTR Regions of 48 SLI Exomes.	164
Figure 5.6. SNPs in LD with the 3' UTR Exome Variant 1 Are Not Significantly Associated with Language Ability.	169
Figure 5.7. The 3 Rare Variants are Highly Conserved Nucleotides.	172
Figure 5.8. Exome-derived 3' UTR Variants are Predicted to Affect Thermodynamic Stability of microRNA-mRNA Secondary Structures.	174
Figure 5.9. The 3 Rare Variants are Inherited Mutations.	176
Figure 5.11. Exome-derived 3' UTR Seed-region Variants Are Unlikely To Be Functional.	181
Figure 5.12. The SNP rs72727021 is in Linkage Disequilibrium With Other Highly Associated SNPs in the C9orf100 Locus.	183
Figure 5.13. Effect of 3' UTR Variants on RNA Folding.	184
Figure 5.14 The pLKO.mir-433 Plasmid Used is Insufficient to Drive Artificial Overexpression of Mature hsa-miR-433.	186
Figure 5.15. Are MicroRNAs Risk Factors For SLI? Integrating Exome Data With Association Analysis.	188

List of Tables

Table 1.1. The FOXP2 DNA binding motif.	39
Table 1.2. Genes Functionally Validated as Direct FOXP2 / Foxp2 Targets.	42
Table 2.1 List of primers used for qRT-PCR experiments.	62
Table 2.2. Tagged Primers Used to Amplify Regions of Interest for the Generation of Luciferase Reporter Constructs.	65
Table 2.3. Tagged Primers Used to Amplify Primary MicroRNA Transcripts for the Generation of MicroRNA Overexpression Constructs.....	66
Table 2.4. Oligos Used to Generate Duplexes With MicroRNA Binding Sites.	67
Table 2.5. Oligos Used to Amplify and Sequence 3' UTRs of Interest.....	71
Table 3.1. Foxp2 Occupies the Promoters of 24 MicroRNAs Precursors in Developing Murine Brains.....	80
Table 3.2. MicroRNAs Differentially Regulated Between the Ganglionic Eminences of Wildtype and Mutant Foxp2.S321X Mice.	88
Table 3.3. Expression and Clustering of MicroRNAs Previously Identified As Foxp2 Targets In Embryonic Mouse Brains.	91
Table 3.4. Gene Ontology (GO) Analysis of Predicted MicroRNA Targets Suggest Foxp2-regulated microRNAs Fine-tune Gene Regulation.....	94
Table 3.5. Overview of qPCR Validation Candidates.....	97
Table 3.6. Summary of Significant Expression Fold Changes of MicroRNA Targets In Human Neuron-like SH-SY5Y Cells.....	104
Table 3.7. Known Neuronal-related Functions of the Top 8 MicroRNAs.	107
Table 3.8. Families of Axon Guidance Ligands and Receptors are Targeted by Foxp2-regulated MicroRNAs.....	110
Figure 4.5. FOXP2 Regulation of CNTNAP2 Binding Site in HEK Cells.....	131
Table 4.1. Putative Transcription Factor Binding in the <i>CNTNAP2</i> Binding Site.	132
Table 4.2. Putative Transcription Factor Binding in <i>MIR137HG</i> Peak A Binding Site.	136
Table 4.3. Summary of Luciferase Results.	137
Table 5.1. Genome-wide Studies Have Uncovered 5 Loci and Several Candidate Genes for SLI.	143
Table 5.2. Genomic Coordinates of 13 Primary MicroRNA Transcripts Selected As Candidates for Association with SLI.	156
Table 5.3. Association Results of SNPs in Primary Transcripts of Top 12 MicroRNAs for SLI.	156
Table 5.4. Five Common Variants Lying in Predicted 3' UTR MicroRNA Binding Sites are Present in the Exomes of 48 SLI Probands.	165
Table 5.5. Three Rare Variants Lying in Predicted 3' UTR MicroRNA Binding Sites are Found in the Exomes of 48 SLI Probands.	165
Table 5.6. The 3' UTR SNP rs72727021 Is Significantly Associated With Nonword Repetition.....	167
Table 5.7. SNPs in LD with the 3' UTR Exome Variant rs1054528 Are Not Significantly Associated with Language Ability.	168

List of Abbreviations

Ago2: Argonaute-2 protein

ASD: Autism Spectrum Disorders

BDNF: Brain-derived neurotrophic factor

bp: basepair

BSA: Bovine serum albumin

CELF-R: Clinical Evaluation of Language Fundamentals, Revised Edition

ceRNA: Competitive endogenous ribonucleic acid

ChIP: Chromatin immunoprecipitation

DGCR8: DiGeorge Syndrome Critical Region-8 protein

DNA: Deoxyribonucleic acid

E16: Embryonic Day 16

EDTA: Ethylenediaminetetraacetic acid

ELS: Expressive language score

EMSA: Electro-mobility shift assay.

EST: Expressed sequence tags

FMRP: Fragile-X mental retardation protein

GE: Ganglionic eminence

GERP: Genomic Evolutionary Rate Profiling

GWAS: Genome-wide Association Study

ID: Intellectual Disability

ISH: In situ hybridization.

kb: kilobases (of DNA basepairs)

LD: Linkage disequilibrium

miRNA: microRNA

mRNA: Messenger ribonucleic acid

ng: Nanogram

nM: nanomoles per litre

NWR: Nonword repetition

PCR: Polymerase chain reaction

pre-miR: precursor microRNA
pri-miR: primary microRNA
qPCR: Quantitative polymerase chain reaction
RISC: RNA-induced silencing complex
RLS: Receptive language score
RNA-seq: Ribonucleic acid sequencing
RNA: Ribonucleic acid
SD: Standard deviation
SEM: Standard error of the mean
SLI: Specific Language Impairment
SNP: Single nucleotide polymorphism
TF: Transcription factor
TRBP: Trans-activation response element RNA binding protein
TSS: Transcription start site
UTR: Untranslated region

Chapter 1: Introduction

1.1 Language: A Unique and Innate Human Trait

1.1.1 Communication and Language

Communication is the transmission of information between individuals or groups, via the generation and reception of signals. These signals can span different modalities (such as vocal, olfactory, and gestural). Several species use signals in various social contexts to affect the behaviour of other individuals. For example, animals communicate to draw others' attention to danger or food. Vervet monkeys use distinct alarm calls to refer to specific types of predators (Seyfarth and Cheney 1990). Honeybees use a waggle dance to indicate the distance and direction of nectar and water to other members of the hive (Riley *et al.* 2005, Grüter and Farina 2009). Communication is also employed during courtship and mating to indicate reproductive fitness to potential mates, and to drive away competitors. An example is the use of birdsong by several species of birds (Nowicki and Searcy 2004).

Language is a specific type of communication: it is the ability to learn and use a finite repertoire of symbols (written, vocalised, or signed) in a rule-based manner to generate an infinite range of meanings (Bishop and Norbury 2008). At first glance, animal communication bears a resemblance to human language: for example, both arbitrarily map signals to specific meanings. However, animals only use a very limited set of signals (a limited lexicon), in contrast to human using upwards of tens of thousands of words. Also, words or signed gestures in human language, unlike signals in animal communication, can be spontaneously combined according to certain rules (a grammar) to describe novel situations (Hauser *et al.* 2002). Therefore, language ability is unique to *Homo sapiens*, and is qualitatively different from other forms of animal communication.

Language ability is argued to be innate to humans: it is acquired very early in development, and with minimal effort. This is perhaps seen most clearly in individuals who are born deaf. Deaf children who did not have exposure to formal sign language can self-develop gestural systems that possess language-like structure (Goldin-Meadow and Mylander 1998, Senghas and Coppola 2001, Coppola and Newport 2005). In contrast, while some nonhuman primates can comprehend and produce signed communications with humans (Sevcik and Savage-Rumbaugh 1994), these abilities require lengthy and repeated training regimes, and none of these trained primates approach the ease with which young children acquire and use language.

1.1.2 Speech and Vocal Learning

Vocal learning is crucial to the acquisition of normal speech and language in humans. It is the ability to modify one's vocal output to either imitate that of another individual, or to create new sounds. Vocal learning requires the ability to hear one's own vocalizations so as to rapidly alter motor control of the vocal apparatus to match a desired sound (Petkov and Jarvis 2012). Unlike speech, vocal learning is also observed in other animal species, although it is remarkably rare in the animal kingdom. Humans are the only primate species capable of vocal learning. In other mammals, only elephants, bats, and certain cetacean species exhibit vocal learning (Nottebohm and Liu, 2010). Only three avian orders display vocal learning: parrots, hummingbirds, and songbirds (Doupe and Kuhl 1999). The importance of vocal learning to human speech, as well as its presence in non-human species, suggests that the neural and vocal apparatus underlying vocal learning are evolutionary substrates for language ability in humans (Jarvis 2004).

1.2 Neural Basis for Speech and Language

1.2.1 Language and the Network Activity of the Brain

Functional imaging and structural studies in the last few decades have revealed that different regions of the brain operate as circuits to understand, process, and produce speech and language (Lieberman 2002, Démonet *et al.* 2005, Price 2010, Price 2012, Hagoort 2014). Distributed activity between these interconnected regions integrates aspects of cognitive function (such as verbal memory) and motor control, allowing rapid understanding and production of speech. These brain regions include cortical, subcortical, and cerebellar areas, and are discussed in detail below.

1.2.2 Cortical Regions and Language: Broca's Area and the Wernicke's Area

There are two classical cortical regions of the human brain that play a role in language ability (Figure 1.1A). The first is Broca's area, which named after its discoverer, the French physician Paul Broca (Broca 1861). It lies in the posterior inferior frontal gyrus of the dominant brain hemisphere (which is the left hemisphere for ~95% of right-handed individuals, and for ~75% of left-handed individuals, Knecht *et al.* 2000). This region is composed of the pars opercularis and the pars triangularis, which correspond to Brodmann's areas 44 and 45 respectively (Amunts and Zilles 2012). Restricted lesions to this area produce an acute form of expressive aphasia (also known as agrammatical aphasia). Affected individuals are unable to produce fluent intelligible speech, but are able to comprehend simple sentences and words (Mohr *et al.* 1978, Alexander *et al.* 1990). This deficit extends to written language as well as signed language (Hickok *et al.* 1996), suggesting that Broca's area plays a role in coordinating language-related motor movements, regardless of modality.

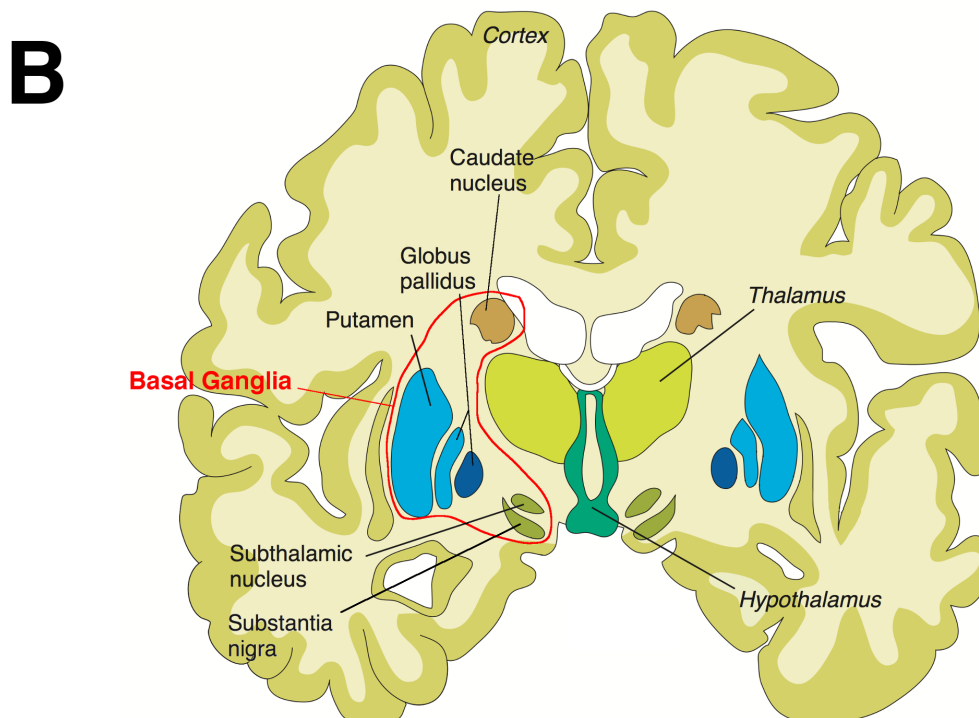
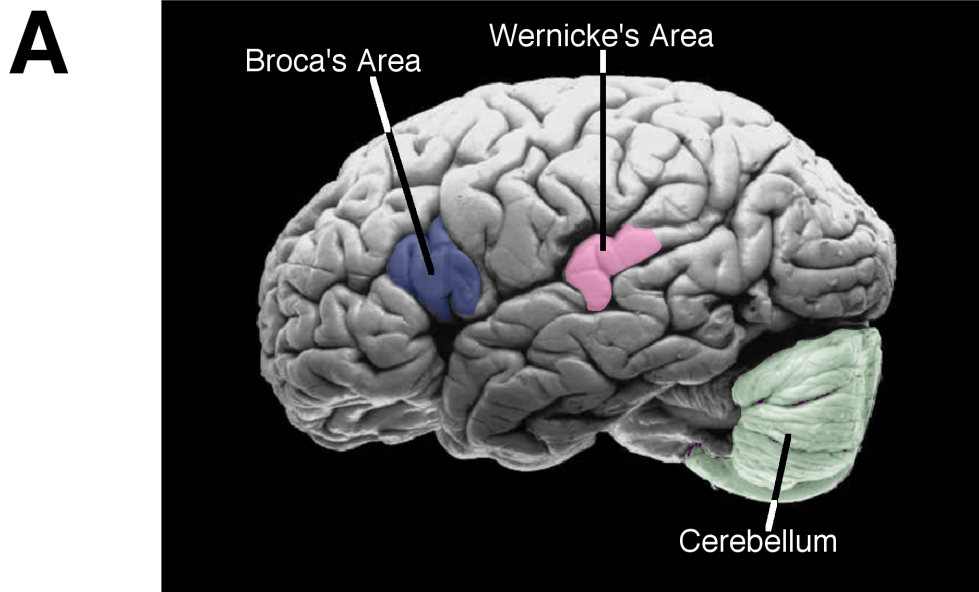


Figure 1.1. Brain Regions Important For Human Speech And Language.

(A) Broca's Area (blue), Wernicke's Area (pink), and the cerebellum (green) in a whole adult human brain. Image modified from the Columbia University Health Sciences Interactive Neuroanatomical Atlas. (Accessible online at <http://www.columbia.edu/itc/hs/medical/neuroanatomy/neuroanat/>) **(B)** A coronal midbrain section. The locations of the basal ganglia (with its components, circled in red) and the thalamus (in olive green) are indicated. Image modified from "Parkinson's Disease: National Clinical Guideline for Diagnosis and Management in Primary and Secondary Care" 2006, published by the UK National Collaborating Centre for Chronic Conditions.

The second classical brain region is Wernicke's area, named after the German scientist Carl Wernicke (Wernicke 1874). It corresponds to the posterior superior temporal gyrus (STG) of the dominant brain hemisphere, and comprises of the posterior section of Brodmann's area 22. Lying adjacent to the left primary auditory cortex, it plays a central role in the auditory processing of human speech (DeWitt and Rauschecker 2013). Wernicke's area is connected as well to the precentral gyrus, where the premotor and motor cortices reside, via a white matter axonal tract known as the arcuate fasciculus (Glasser and Rilling 2008, Bernal and Altman 2009). This connection between the posterior STG and the motor cortices is involved in speech perception, by mapping auditory speech sounds to articulatory motor representations (Wilson *et al.* 2004, Hickok and Poeppel *et al.* 2007). Individuals with injury to Wernicke's area have acute receptive aphasia (also known as fluent aphasia), and are unable to understand written or spoken language. While they are able to speak with relatively normal rhythm and syntax, their sentences are often meaningless and incoherent.

1.2.3 Subcortical Regions and Language: The Basal Ganglia and the Thalamus

The basal ganglia are a collection of bilateral subcortical brain regions in the ventromedial forebrain (Figure 1.1B). It consists of the striatum (subdivided into the caudate nucleus and the putamen), the globus pallidus, the substantia nigra, and the subthalamic nuclei. The basal ganglia form extensive connections with the cortex and the thalamus, and play a central role in motor sequencing (Barbas *et al.* 2013). Motor and language deficits result from dysfunction in cortico-basal ganglia circuits (such as in Parkinson's disease). This strongly supports the hypothesis that sensorimotor sequencing functions of the basal ganglia have been adapted over human evolution to subserve speech and language (Lieberman 2007, Enard 2011).

The thalamus is found in the medial forebrain, lying ventromedial to the basal ganglia (Figure 1.1B). It functions as a relay in language-related brain circuits,

receiving input from the basal ganglia and the cerebellum, and sending recurrent afferent connections to the frontal cortex (Barbas *et al.* 2013). Patients with lesions to the thalamus display lexical-semantic deficits, such as poor performance on picture naming tasks (Raymer *et al.* 1997, Crosson 2013.). The thalamus is activated during processing of syntactic and semantic aspects of aurally presented sentences (Wahl *et al.* 2008). Taken together, this suggests that the thalamus is involved in lexical selection for vocal output and language processing (Crosson *et al.* 2013).

1.2.4 The Cerebellum and Language

The cerebellum is a hemispheric structure found at the base of the brain, where it sits above the spinal cord (Figure 1.1A). It has connections with the cortex and the basal ganglia via the thalamus (Bostan *et al.* 2013). Like the basal ganglia, it is associated with the planning and execution of motor movements. Damage to the cerebellum results in gross movement deficits, arising from an inability to sequence the motor commands necessary for a compound movement. Cerebellar damage can also cause cerebellar cognitive affective syndrome (Schmahmann and Sherman 1998), which includes cognitive dysfunction (such as disturbances in executive function and visual-spatial memory), changes in personality and affect, and deficits in language (such as agrammatism). Behavioural and functional studies have shown that cerebellar function in speech and language is not limited to planning and execution of articulation (Ackermann 2008). Patients with cerebellar degeneration performed poorer on language tasks compared to healthy controls, and this impairment remained significant after controlling for motor deficits such as slower articulation (Stoodley and Schmahmann 2009). The right hemisphere of the cerebellum is activated during a covert verb generation task, indicating that the cerebellum's role in language extends beyond articulatory control, and includes pre-articulatory motor planning (Stoodley *et al.* 2011).

1.2.5 Distributed Activity, Functional Integration, and Language: Beyond The Classical Model

Broca's area and Wernicke's area were previously thought to be self-contained neurological substrates for language ability in humans, (Geschwind 1970), but a distributed network model (Lieberman 2002, Price 2010, Hagoort 2014), involving both cortical and subcortical regions, has supplanted this classical view. Non-invasive imaging techniques have revealed that damage to subcortical structures, in the absence of lesions to Broca's area or Wernicke's area, can also produce the corresponding aphasia (Lieberman 2002).

Indeed, the study of the neurobiology of language has moved beyond the study of single-word or sentence production and comprehension in isolated lesion cases. Functional brain imaging combined with non-invasive techniques (such as trans-cranial magnetic stimulation) have allowed the investigation of brain activity during linguistic tasks which better reflect those underlying natural language comprehension and production. This has revealed that human speech and language depends on distributed patterns of cortical activity that each subserve multiple functions (Price 2010, Price 2012, Hagoort 2014). For example, speech production and comprehension both produce activity in the auditory cortex and the left inferior frontal cortex (Menenti *et al.* 2011). Activity in the premotor cortex has also been implicated in a range of linguistic tasks, such as processing syntactic representation (Segeart *et al.* 2012) and word production (Alario *et al.* 2006). All in all, the roles that individual brain regions play in different aspects of language production and comprehension are thought to be context-dependant, contingent on its functional interactions with other regions in the task-relevant network (Hagoort 2014).

1.3 Genetic Basis for Speech And Language

The heritability of language ability refers to the relative influences of genetic and non-genetic factors on individual differences in language development and performance. Language ability is typically quantified using a range of psychometric tests that measure performance on tasks such as phonological memory, syntactical rule generation, tense marking, and vocabulary. For a given individual, a score on these psychometric tests that falls 1.5 standard deviations below his or her age-appropriate mean score (as measured through large-scale general population studies) is generally thought to be language-impaired.

If genetic factors do affect language ability, then speech and language disorders should show familial aggregation. Indeed, several studies have found that individuals with speech and language problems (probands) tend to have affected biological relatives, compared to non-affected controls (Neils and Aram 1986, Tallal *et al.* 1989, Lewis *et al.* 1993, Felsenfeld and Plomin 1997, Choudhury and Benasich 2003, Lewis *et al.* 2004, Barry *et al.* 2007). Twin studies have also been employed to study the heritability of language development and performance. Classical studies of twins raised together compare phenotypic similarities between monozygotic (MZ) twins versus those of dizygotic (DZ) twins, and allow estimation of the extent to which genetic variation contributes to the phenotype while controlling for the effects of a shared environment. Therefore, if language ability is the result of genetic factors, MZ twins should have a higher concordance for language ability or disability than dizygotic DZ twins. Indeed, non-language impaired MZ twins show a higher correlation in language ability compared to non-impaired DZ twins (Kovas *et al.* 2005). For both written and spoken language impairments, MZ twins have significantly higher concordance rates (35% – 100%) for language disorders than DZ twins (31% – 69%) (Stromswold 2001), providing further evidence for the heritability of language ability.

It must be noted that the diagnostic criteria of speech and/or language impairments used to classify cases affects heritability estimates. Bishop and Hayiou-Thomas (2008) found that children whose language impairments were identified through screening of the general population (primarily on the basis of performance on standardized tests) were likely to represent a different phenotype and etiology from clinically-referred cases. They suggested that a higher heritability might be observed for cases of clinically-referred speech impairment. Indeed, speech and language (and disorders thereof) is a complex trait, with a large heterogeneity in the phenotypes of language disorders (Bishop 1997, Newbury and Monaco 2009). The genetic basis for speech and language disorders is widely agreed to involve multiple genetic elements, interacting with environmental factors, each contributing a small effect (Stromswold 2001, Grigorenko 2009, Newbury and Monaco 2009).

1.4 The KE Family and *FOXP2*

1.4.1 The KE Family and Their Phenotype

Hurst *et al.* were the first to report on a multi-generational family in the UK, termed the KE family (Figure 1.2), where 15 out of 37 family members have a severe articulatory impairment that renders their speech virtually unintelligible (Hurst *et al.* 1990). The impairment manifests early in childhood, and does not fully resolve after speech therapy. Thus, the phenotype presented by affected KE members is described as a severe form of developmental verbal dyspraxia (Hurst *et al.* 1990, Vargha-Khadem *et al.* 1995). Affected members exhibit orofacial dyspraxia, have reduced language ability, with some having cognitive deficits.

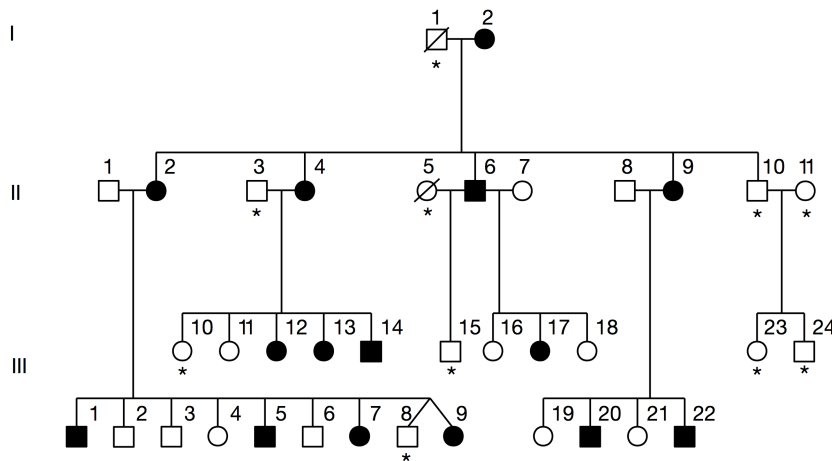


Figure 1.2. The KE Pedigree.

Shown here are three generations of the KE family. Affected individuals are indicated by filled symbols. Asterisks indicate those individuals who were unavailable for genetic analyses. Squares are males, circles are females, and a line through a symbol indicates that the person is deceased. Taken from Lai *et al.* 2001.

1. Orofacial Dyspraxia. Affected family members are unable to perform sequential movements of the tongue and lips, despite being able to execute simple single orofacial movements (Vargha-Khadem *et al.* 1995, Watkins *et al.* 2002a, Alcock *et al.* 2000). This dyspraxia is not due to weakness in or paralysis of orofacial muscles (affected members are thus able to perform basic oral movements necessary for feeding), and all the affected KE members have normal limb praxis (Watkins *et al.* 2002a).

2. Broad and Severe Language Impairment. The language impairment exhibited by affected family members is thought to disturb a wide range of language domains. Affected KE members have reduced grammar abilities (Gopnik 1990, Gopnik and Crago 1991, Vargha-Khadem *et al.* 1995, Watkins *et al.* 2002a), such as an inability to produce changes in word inflection (in both tense and number). Initial studies of the KE family (Gopnik 1990, Gopnik and Crago 1991) suggested that the impairment was specific to morpho-syntactical grammar (namely, an inability to apply or generate syntactic rules, such as those for tense, gender, and number). Further testing, however, uncovered deficits in several other aspects of grammar, such as an inability to

understand relative clauses and complex sentences (Vargha-Khadem and Passingham 1990, Vargha-Khadem *et al.* 1995). Affected KE members also have deficits in receptive language ability (determined by a lexical decision test where nonsense words must be distinguished from actual words) as well as expressive language ability (determined by accuracy of repetition of spoken words) (Vargha-Khadem *et al.* 1995, Watkins *et al.* 2002a). This suggests that a grammatical impairment is not central to the phenotype seen in the KE family. The impairment is not restricted to spoken language: affected individuals also exhibit deficits in written tests (Watkins *et al.* 2002a), strongly suggesting that the phenotypic effects of the deficit are not limited to articulation and orofacial praxis.

3. Cognitive Deficits. Affected family members, as a group, possess lower mean nonverbal intelligence quotient (IQ) scores relative to unaffected members (Vargha-Khadem *et al.* 1995, Watkins *et al.* 2002a). Unaffected members obtained a mean performance IQ score of 98.3 ± 14.6 , while affected members had a significantly lower ($P < 0.003$) mean score of 83.2 ± 10.6 . Gopnik and Goad (1997) note that clinical studies have found a wide variation in performance IQ scores across language-impaired individuals, suggesting the cognitive deficits seen in some affected KE members is not unique. In addition, Watkins and colleagues (2002a) offer three reasons as to why cognitive deficit is unlikely to be the central feature of the impairment. Firstly, reduction in cognitive ability does not co-segregate with the language impairment – some affected members have normal nonverbal IQ, while other unaffected members have low nonverbal IQ scores. Secondly, out of 5 subtests used to determine nonverbal IQ, affected members only performed poorer than unaffected members in one subtest. This specific subtest measures the ability to learn and recall arbitrary associations between numbers and symbols. Lastly, some affected individuals showed a decline in nonverbal IQ scores across time, a phenomenon observed in longitudinal studies of children with specific language impairment (Tallal *et al.* 1991, Stothard *et al.* 1998, Mawhood *et al.* 2000, Conti-Ramsden *et al.* 2001,

Botting 2005). Conti-Ramsden *et al.* (2012) have suggested that the lower nonverbal skills observed across development in language-impaired individuals might in fact be due in part to the language impairment. They conjecture that normal learning processes that rely on understanding and conveying spoken representations are curtailed, because language facilitates much of one's engagement with the external world. Thus, they hypothesise that early language impairment might result in a decreased opportunity to develop nonverbal reasoning skills, leading to a slower growth of these skills over time. All in all, the lower nonverbal IQ of affected KE members is unlikely to be a central feature of their language impairment.

1.4.2 Abnormal Activity and Structure of Brain Regions Found In The KE Family

Functional and structural studies have revealed that affected KE members exhibit aberrant patterns of brain activity during language tasks, as well as grey matter abnormalities in several cortical and subcortical areas known to be important for language production and motor activity.

Broca's Area. Compared to unaffected members, affected KE members had less grey matter in the caudal part of Broca's Area, the pars opercularis of the left inferior frontal gyrus (Vargha-Khadem *et al.* 1998, Watkins *et al.* 2002b). Affected members also had an underactive pars opercularis during both covert and overt verb generation tasks (Liégeois *et al.* 2003).

The caudate nucleus. Affected KE members had less grey matter in the caudate nucleus bilaterally, relative to unaffected family members, as well as aged-matched healthy controls (Vargha-Khadem *et al.* 1998, Watkins *et al.* 2002b). This finding was corroborated by volumetric measurements.

The Putamen. Unlike the pars opercularis and the caudate nucleus, there was more grey matter in the bilateral putamen of affected KE members, compared to age-matched unaffected family members (Vargha-Khadem *et al.* 1998, Watkins *et al.* 2002b). During both covert and overt verb generation tasks, the

activity of the putamen in affected KE members was lower than unaffected members (Liégeois *et al.* 2003).

In addition to the above regions, affected KE members had grey matter abnormalities in other areas such as the supplementary motor area (SMA), the globus pallidus, and the cerebellum (Vargha-Khadem *et al.* 1998, Watkins *et al.* 2002b, Belton *et al.* 2003). Taken together, these studies provide experimental support for the hypothesis, previously suggested by Alcock *et al.* (2002), that orofacial praxis and normal speech ability both rely on common brain regions and neural circuits.

One important result from these functional studies is that regions with reduced grey matter are over-activated during language tasks. It has been suggested (Vargha-Khadem *et al.* 1998, Liégeois *et al.* 2003) that this increased activity might indicate additional reasoning or cognitive effort being used to complete the task. Another key finding is that affected KE members have diffused brain activity in both hemispheres during language tasks that are normally associated with lateralised brain activity in non-language impaired individuals. This is evidence that brain circuits not normally involved in language tasks are being employed in affected KE members to offset the deficit in normal language circuits (Liégeois *et al.* 2003), although it is unclear if this is a conscious employment of alternative strategies, or an unconscious adjustment arising from neural plasticity.

Interestingly, a study that assessed functional and structural differences between 10 individuals with specific language impairment (SLI) with unaffected sibilings and age-matched controls (Badcock *et al.* 2012) observed marked congruence with the abnormalities seen in affected KE members. Individuals with SLI also had reduced grey matter in the caudate nucleus, as well as reduced functional activity in Broca's Area and the putamen during word generation tasks. The similarities between functional and structural abnormalities exhibited by individuals with SLI and by affected members of the KE family are an indication of the parallels in the behavioural deficits and

phenotypic presentation of both disorders.

1.4.3 The Discovery of *FOXP2*: From Linkage To A Candidate Gene

The striking pattern of verbal dyspraxia in the KE family points to a monogenic autosomal dominant form of inheritance. Both males and females are affected in equal proportion, affected members are found in every generation, and all affected individuals have at least one affected parent (Figure 1.2, Fisher *et al.* 1998). Thus, Fisher *et al.* (1998) employed a classical genome-wide parametric linkage technique to identify if any chromosomal regions were shared between affected members (co-segregation), but not with unaffected members, beyond that expected by chance and random assortment. They identified a region on chromosome 7 (7q31) with a maximum logarithm of odds (LOD) score of 6.62, well above the threshold of 3 for significant linkage for a monogenic trait. This locus was termed *SPCH1*, and was the first reported genetic locus linked to a familial form of speech and language impairment.

Lai *et al.* (2000) also analysed an unrelated patient, known as CS, whose language impairment and verbal dyspraxia was very similar to that observed in affected KE members. This patient had a *de novo* balanced reciprocal translocation between chromosome 5 and 7. Crucially, the breakpoint in chromosome 7 lay in the *SPCH1* region of 7q31, and disrupted a gene known as *CAGH44*. As the coding sequence of *CAGH44* was not fully known, Lai and colleagues (2001) performed further work to identify the gene. They found it was a brain-expressed transcript that contained a forkhead box/winged helix (FOX) protein domain. Thus, *CAGH44* was renamed *FOXP2*, in line with the nomenclature for the FOX superfamily of proteins (see Section 1.4.6).

Lai *et al.* (2001) also confirmed that the chromosomal breakpoint in CS disrupted *FOXP2* in the intronic region between exons 3b and 4. They screened the coding sequence of *FOXP2* for mutations in KE family

members, and found a heterozygous G-to-A point mutation in exon 14 that causes an arginine-to-histidine change in amino acid 553 (R553H). This R553H mutation co-segregated perfectly with the verbal dyspraxia in the KE family, and was not found in 364 independent unaffected individuals.

1.4.4 Mutations in *FOXP2* Are Associated With Developmental Verbal Dyspraxia

Other mutations and disruptions that affect the protein sequence of FOXP2 have since been found in unrelated individuals who share the KE phenotype of verbal dyspraxia. These individuals share a primary diagnosis of verbal dyspraxia and impaired speech articulation, in the absence of congenital irregularities and impaired hearing (MacDermot *et al.* 2005). In addition to the KE mutation and the CS breakpoint mutation (Lai *et al.* 2001), 4 point mutations (MacDermot *et al.* 2005, Roll *et al.* 2010, Laffin *et al.* 2012), 1 intragenic deletion (Turner *et al.* 2013), 1 intragenic insertion (MacDermot *et al.* 2005), and 2 breakpoint mutations (Feuk *et al.* 2006, Tomblin *et al.* 2009) have been discovered (Figure 1.3).

22 cases with large interstitial deletions spanning the *FOXP2* locus have also been documented (listed in Palka *et al.* 2012 and Feuk *et al.* 2006, see Zeesman *et al.* 2006, Lennon *et al.* 2007, and Rice *et al.* 2011 as well). As a result of the deletion of several genes, such individuals generally have a complex phenotype, and possess several physical and mental abnormalities in addition to language deficits.

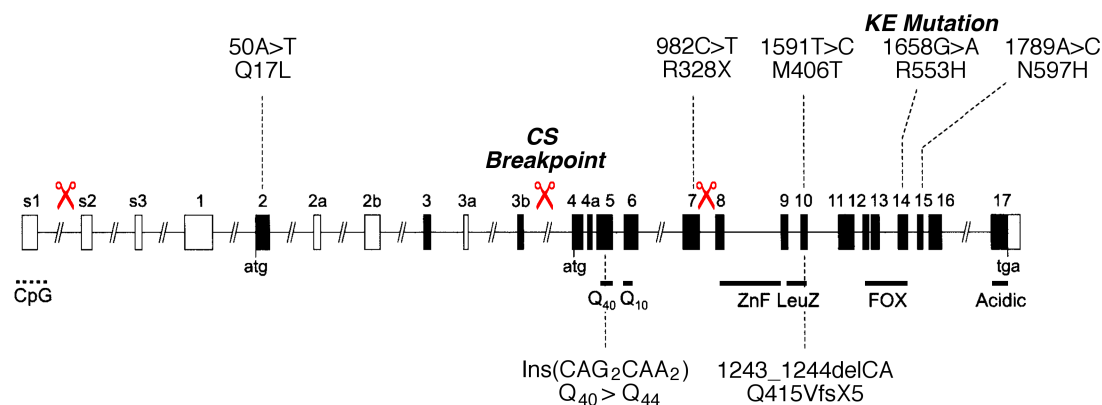


Figure 1.3. Protein-coding Mutations and Disruptions in *FOXP2* Found In Subjects With Verbal Dyspraxia.

This diagram is a schematic of the *FOXP2* locus in humans. Boxes indicate exons; black boxes represent coding exons while white (unshaded) boxes represent noncoding exons. Exons s1–s3 represent 5′ untranslated exons. The codons “atg” and “tga” denote positions of initiation and termination codons respectively. Known functional domains are indicated and include poly-glutamine tracts (Q₄₀ and Q₁₀), a zinc-finger motif (ZnF), a leucine zipper (LeuZ), a forkhead domain (FOX), and an acidic C-terminus (Acidic). Point mutations are shown above the locus, while indels are shown below. Red scissors indicates the location of breakpoints. The locations of the mutations and indels are listed in both genomic and protein-sequence coordinates. Whole-gene deletions of *FOXP2*, and mutations in intronic regions in probands, are not shown here. Modified after MacDermot *et al.* 2005.

It is important to note that all known mutations and disruptions of *FOXP2* are heterozygous. Therefore, verbal dyspraxia results from reduced gene dosage of functional *FOXP2*. Variants in the intronic regions of *FOXP2* have also been identified (Newbury *et al.* 2002, MacDermot *et al.* 2005) but their relevance to protein function remains unknown.

1.4.5 The Sequence and Expression of *FOXP2*: High Inter-species Conservation

1. Sequence Conservation. *FOXP2* is one of the most highly conserved proteins amongst mammals (Enard *et al.* 2002, Zhang *et al.* 2002, Webb and Zhang 2005). Disregarding variations in the poly-glutamine stretches, human *FOXP2* differs from mouse *Foxp2* at only 3 amino acid residues: E80D, T303N, and N325S (Figure 1.4), with the last 2 occurring on the human

lineage after it split from the common ancestor with the chimpanzee. However, some carnivores – such as the dog (*Canis familiaris*) and the cat (*Felis catus*) – and some echolocating bats have the serine residue at position 325 (Zhang *et al.* 2002, Li *et al.* 2007).

[illegible]

Figure 1.4. Alignment of Exon 7 from FOXP2 sequences of 13 species.

Shown here are the protein sequences of FOXP2 exon 7 (residues 259 – 330 of the human protein), from 2 archaic humans, 3 nonhuman primates, 2 carnivores (dog and cat), 2 echolocating bats, the African elephant, the mouse, and a songbird, relative to the human sequence. Residues identical to the human sequence are indicated as a period (.). The N303 and S325 substitutions are indicated by boxes. The increased sequence diversity of exon 7 in bats can be seen in the two representative species shown here.

Despite the high degree of protein sequence conservation, there is an accelerated rate of evolution in the hominid lineage (Enard *et al.* 2002, Zhang *et al.* 2002) as well as echolocating bats (Li *et al.* 2007, Figure 1.5). The diversity of Foxp2 protein sequence in bats is thought to be a sign of increased adaptive pressures arising from the unique demands on auditory-motor feedback required by echolocation (Li *et al.* 2007).

Sequence analysis of modern-day humans supports a selective sweep (rapid fixation of beneficial mutations) occurring in the region surrounding exon 7 of FOXP2, which harbours both the T303N and N325S substitutions (Enard *et al.* 2002, Zhang *et al.* 2002). The fixation of these 2 amino acid changes was initially estimated to occur ~200,000 years ago, around the time when modern *Homo sapiens* emerged (Enard *et al.* 2002). However, further research cast doubt on the locus and the timing of the selective sweep: linkage disequilibrium extended across exon 7 in humans (Ptak *et al.* 2009), and archaic human species – Neanderthals and Denisovans – possessed these

two amino acid changes (Krause *et al.* 2007, Green *et al.* 2010, Reich *et al.* 2010). Maricic *et al.* (2013) showed that a human-specific substitution in intron 8 (falling in the region of the selective sweep) affects POU3F binding and downstream transcription, and thus might be the locus of positive selection instead.

2. Expression. *Foxp2* is expressed in a variety of mammalian tissues during development. It is found in the brain, heart, lung, and intestine, and has been shown to play a role in lung development (Shu *et al.* 2001, Shu *et al.* 2007, Spiteri *et al.* 2007). A similar expression pattern is seen in the adult zebra finch, with FoxP2 detected in the brain, lung, liver, and heart (Haesler *et al.* 2004). The expression of Foxp2 in organs such as the lung and heart is likely to account for the early mortality of mutant mice with either a complete absence of Foxp2, or which only possess non-functional Foxp2 (Shu *et al.* 2005, French *et al.* 2007, Fujita *et al.* 2008, Groszer *et al.* 2008). This early lethality in mice lacking functional Foxp2 likely explains why no homozygous mutations have been reported in humans.

In the developing human brain, *FOXP2* expression is restricted to the cortex, striatum, basal ganglia, and cerebellum (Lai *et al.* 2003, Teramitsu *et al.* 2004). This matches the neural regions in the human brain mentioned earlier (Section 1.2) that play crucial roles in language-related and motor-learning circuits. These regions also overlap with sites of structural and functional abnormalities in affected KE members (Vargha-Khadem *et al.* 1998, Belton *et al.* 2003, Lai *et al.* 2003, Liégeois *et al.* 2003). This expression pattern is highly conserved in mice (Shu *et al.* 2001, Ferland *et al.* 2003, Campbell *et al.* 2009), in zebra finches (Haesler *et al.* 2004, Teramitsu *et al.* 2004), and in zebrafish (Bonkowsky and Chien 2005, Shah *et al.* 2006), indicating a conserved function of FoxP2 in animal brains.

In the zebra finch, *FoxP2* expression is closely related to vocal learning and vocal plasticity. The cortico-striatal anterior forebrain pathway in zebra finches underlies their ability for vocal learning. Within this pathway, a region known

as Area X has received considerable interest, as it is required for vocal learning in juvenile zebra finches (Sohrabji *et al.* 1990). Area X is a region of high *Foxp2* expression, with expression peaking during the song-learning period (Haesler *et al.* 2004). *FoxP2* expression is modulated during developmental and seasonal phases of song learning and vocal plasticity (Haesler *et al.* 2004). *Foxp2* expression also exhibits context-dependent modulation in adult male zebra finches: levels are lower in Area X during undirected singing, but are higher when the male is singing to a female (Teramitsu and White 2006).

Taken together, this suggests that FoxP2 plays a role in the development of corticostriatal structures, which then were then adapted for vocal learning and language (Bonkowsky and Chien 2005, Fisher and Marcus 2006). The high conservation of protein sequence and of neural expression patterns also support the relevance of animal models of FoxP2, even if these animals do not have language ability.

1.4.6 FOXP2 and Other Forkhead Genes

FOXP2 belongs to a superfamily of transcription factors known as forkhead box (FOX) proteins. Members of this superfamily all possess a highly conserved forkhead/winged-helix DNA-binding domain. The name of this superfamily is derived from the first such gene (named *forkhead*) that was discovered in the fruit fly (*Drosophila melanogaster*). Mutations in this gene cause the formation of ectopic head structures, resulting in embryos with a fork-headed appearance (Weigel *et al.* 1989). There are 50 FOX genes in the human genome, which are classified into 19 subfamilies (*FOXA* to *FOXS*) based on phylogenetic analysis of the FOX domain protein sequence (Kaestner *et al.* 2000, Jackson *et al.* 2010). FOXP2 belongs to the FOXP subfamily, which consists of 3 other members: FOXP1, FOXP3, and FOXP4. Besides a forkhead DNA-binding domain, all 4 proteins share a leucine zipper domain and a zinc finger domain.

1.5 The Function of *FOXP2*, And Effects of Pathogenic Mutation In The Brain

FOXP2 has been investigated at several aspects of biological function: at the molecular level, the genetic level, the neuronal level, the level of neuronal circuitry, and the level of animal behaviour. The study of wildtype *FOXP2*/*Foxp2* has proceeded in parallel with functional investigation of *FOXP2* mutations found in individuals with verbal dyspraxia and speech impairments (Section 1.4.4). To date, only four protein-coding mutations in *FOXP2* (Figure 1.6) have undergone functional molecular investigation: the R553H KE mutation, two mutations reported by MacDermot *et al.* 2005 (R328X and Q17L), and a mutation reported by Roll *et al.* 2010 (M406T). The original R553H KE mutation and the R328X mutation have been the focus of much of the work detailed below.

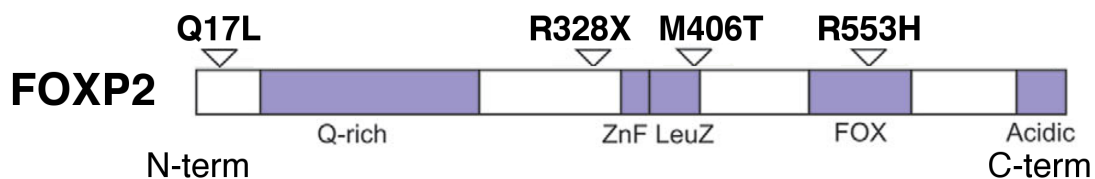


Figure 1.5. Functional Testing of Four Mutations Found in Individuals with Verbal Dyspraxia.

Shown here is a schematic of the *FOXP2* protein (715 amino acids). Four mutations found in probands have undergone functional testing, and are indicated above the schematic. Also indicated are the protein domains. Modified after Vernes *et al.* 2006.

1.5.1 The Functions of *FOXP2* at the Molecular Level: DNA binding and Dimerization With Co-Factors

FOXP2 is found in the nucleus, where it binds to DNA via the FOX domain (Stroud *et al.* 2006, Vernes *et al.* 2006, Nelson *et al.* 2013). This nuclear localization is mediated by a nuclear localization signal (NLS) at the C-terminal, adjacent to the FOX domain (Figure 1.6, Vernes *et al.* 2006). Several studies (Table 1.1) have described the *FOXP2* DNA-binding motif. The considerable divergence amongst the reported motifs can be attributed to

differing experimental designs across the studies. Nevertheless, the FOXP2 motif is likely to be congruent with the 7-basepair FOXP consensus motif (Overdier *et al.* 1994, Wang *et al.* 2003).

Table 1.1. The FOXP2 DNA binding motif.

Experimental System	Method of Fragment Capture	Motif Type	Motif	Reference
SHSY5Y cells stably overexpressing FOXP2	ChIP-chip	FOXP2	TCTTCGT	Vernes et al. 2007
HEK293 cells	EMSA for CNTNAP2 fragment	FOXP2	AATTG	Vernes et al. 2008
Embryonic mice brain	ChIP-chip	Foxp2	ARKTAMYT	Vernes et al. 2011
MITOMI	Microfluidic affinity assay	FOXP2	TRTTTAY	Nelson et al. 2013
<i>In vitro</i>	Cyclic Amplification and Selection of Targets (CAST)	FOXP	TATTTT	Wang et al. 2003
<i>E. coli</i>	Pulldown of GST fusion protein	FOX	TRTTKRY	Overdier et al. 1994

This table summarises the different consensus motifs reported for FOXP2/Foxp2. The canonical FOX motif and the consensus FOXP motif are listed for comparison. International Union of Pure and Applied Chemistry (IUPAC) nucleotide codes for degenerate bases: M = A/C, R = G/A, K = T/G, Y= C/T. Abbreviations: MITOMI stands for “Mechanically Induced Trapping Of Molecular Interactions.” GST stands for Glutathione S-transferase.

FOXP2 forms homo- and heterodimers with FOXP1 and FOXP4 (Li *et al.* 2002, Deriziotis *et al.* 2014a), via interactions at the leucine zipper domain (Wang *et al.* 2003, Li *et al.* 2004) or via domain-swapping at the forkhead domain (Stroud *et al.* 2006). Domain-swapping is a unique feature of the FOXP forkhead domain, and it forces the dimers to bind to non-adjacent DNA sites. This suggests that FOXP2 (and other FOXP members) might be able to regulate higher-order DNA/protein structures (Stroud *et al.* 2006, Jackson *et al.* 2010). While the isolated FOXP2 forkhead box domain is able to bind DNA (Stroud *et al.* 2006, Vernes *et al.* 2006), disruption of dimerization, via deletion or mutation of the leucine zipper domain, abolishes DNA binding and transcriptional activity of FOXP1, FOXP2, and FOXP4 (Li *et al.* 2004). This suggests that dimerization with FOXP members is necessary for FOXP2 function. While the three-dimensional (3D) structure of FOXP2 homodimers bound to DNA has been solved (Stroud *et al.* 2006, Figure 1.7), future elucidation of the full 3D structure of FOXP2 will shed light on the exact

molecular mechanisms by which secondary and tertiary protein structures mediate dimerization and DNA-binding.

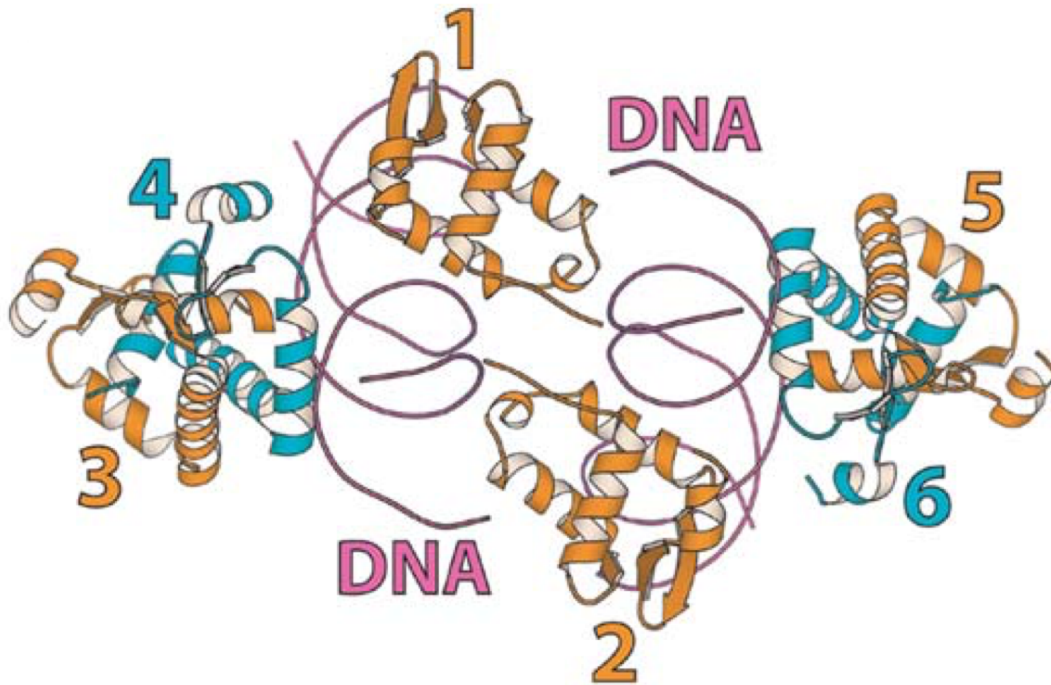


Figure 1.6. Three-dimensional Structure of DNA-bound FOXP2 Monomers and Homodimers.

FOXP2 molecules are shown as ribbon drawings. Molecules within the swapped dimers are both orange (3 and 5) and cyan (4 and 6). Monomers (1 and 2) are orange. The DNA phosphate backbone is shown as a coil in magenta. Taken from Stroud *et al.* 2006.

FOXP2 also interacts with other transcription factors. It binds to CTBP1 via a CTBP1-binding motif located at amino acids 422 and 426, between the leucine zipper and forkhead-box domains (Li *et al.* 2002, Deriziotis *et al.* 2014a). CTBP1 is a transcriptional co-repressor that forms a complex with proteins involved in promoter targeting and chromatin remodeling, such as CoREST and HDAC1 (Shi *et al.* 2003). It has been suggested that CTBP1 might recruit other factors alongside FOXP2 (Li *et al.* 2002), but the exact functional relationship between CTBP1 and FOXP2 remains unknown. Foxp2 also interacts with the transcription factor Nkx2.1 in the lung (Zhou *et al.*

2008), but little is known about the relevance of this interaction in the mammalian brain.

The pathogenic R553H and R328X mutations disrupt DNA binding (Vernes *et al.* 2006, Nelson *et al.* 2013). Crucially, Nelson *et al.* (2013) showed that the R553H mutation completely abolishes the DNA-binding ability of FOXP2, rather than only altering recognition of the consensus motif. This corroborates the finding that the arginine at position 553 directly interacts with DNA base-pairs during DNA-binding, which is likely to be disrupted by the R553H KE mutation (Stroud *et al.* 2006). Interestingly, the Q17L mutation does not disrupt DNA binding (Vernes *et al.* 2006).

The R328X, R553H, and M406T mutants all display aberrant nuclear localization. The lack of the NLS in the R328X accounts for its cytoplasmic localization, but how R553H and M406T mutations disrupt nuclear localization remains unknown. Vernes *et al.* (2006) hypothesised that the R553H mutation might disrupt electrostatic charge across the adjacent NLS, or it might lie in a site with another nuclear localization signal, but these remain to be tested. The M406T is also likely to disrupt interactions with co-factors at the leucine-zipper domain, either by reducing affinity for nuclear co-factors or by increasing affinity for non-nuclear factors, but these mechanisms remain speculative given the paucity of experimental evidence. The Q17L mutation does not disrupt nuclear localization (Vernes *et al.* 2006), and the individual in whom the mutation was first observed also had an affected sibling that did not carry the mutation. Thus, Vernes *et al.* concluded that the Q17L mutation was unlikely to be deleterious to molecular function, but did not discount the possibility that it might produce subtle alterations on protein function that remain to be discovered.

The R328X mutation, but not the R553H mutation, disrupts dimer formation (Deriziotis *et al.* 2014a). This is likely due to the fact that the R328X mutant lacks both the zinc finger and leucine zipper domains, while the R553H has both domains intact. Little is known about the effects of the M406T mutation

on FOXP2 dimerization, but given its location in the leucine zipper domain (Roll *et al.* 2010), it is likely to be deleterious to molecular function.

1.5.2 The Functions of *FOXP2* in Gene Regulation: Genetic Networks Involved in Neuronal Function And Under Positive Selection

The genes whose promoters are bound by FOXP2/Foxp2 were revealed by chromatin immunoprecipitation (ChIP) of cross-linked DNA-protein complexes from foetal human brain (Spiteri *et al.* 2007), human neuron-like cell-lines (Vernes *et al.* 2007), and mouse embryonic brains (Vernes *et al.* 2011). Gene ontology analysis of these genes revealed that neurite outgrowth, axon guidance, and synaptic plasticity pathways were over-represented (Spiteri *et al.* 2007, Vernes *et al.* 2007, Vernes *et al.* 2011), suggesting that FOXP2/Foxp2 plays a role in forming connections between neurons.

Several of these target genes were successfully validated as transcriptional targets *in vitro* using ChIP-PCR techniques (Spiteri *et al.* 2007, Vernes *et al.* 2007, Vernes *et al.* 2011). For most of these target genes, FOXP2/Foxp2 acts as a transcriptional repressor, in line with initial results from viral and artificial promoters (Li *et al.* 2004, Vernes *et al.* 2006). A smaller selection of target genes (namely *CNTNAP2*, *SRPX2*, *uPAR*, *MET*, and *DISC1*) has undergone further functional investigation (Table 1.2).

Table 1.2. Genes Functionally Validated as Direct FOXP2 / Foxp2 Targets.

Gene	Neurological Disorders Implicated In	Experimental Evidence For Regulation by FOXP2 / Foxp2	Effect of Pathogenic FOXP2 / Foxp2 Mutations on Regulation	References
CNTNAP2	Autism; Intellectual disability; SLI	ChIP-PCR (both in vitro and in vivo); ISH	R553H and R328X mutation abolishes FOXP2 / Foxp2 binding and resultant repression.	Vernes et al. 2008; Vernes et al. 2011
SRPX2	Rolandic epilepsy; DVD	EMSA; Luciferase; qPCR	R553H mutation abolishes FOXP2 binding and resultant repression.	Roll et al. 2010
uPAR	None; interacts with SRPX2	EMSA; Luciferase; qPCR	R553H mutation abolishes FOXP2 binding and resultant repression.	Roll et al. 2010
MET	Autism	EMSA; qPCR, Western Blotting; ISH	Unknown.	Mukamel et al. 2011
DISC1	Schizophrenia, bipolar disorder, autism	Luciferase; Western Blotting	R553H and R328X mutation abolishes FOXP2 binding and resultant repression.	Walker et al. 2012

The five genes listed here are direct targets of FOXP2, as borne out by molecular and functional validation. These five genes were of particular interest because they

have each been implicated in neurological disorders such as autism. ISH: *in situ* hybridization. EMSA: electro-mobility shift assay.

The M406T mutant causes a loss of repression of the *SPRX2* promoter (Roll *et al.* 2010). The R553H mutation and the R328X mutation both disrupt the regulation of *CNTNAP2*, *SRPX2*, *uPAR*, and *DISC1*. This is concordant with the observed lack of DNA binding ability of both these pathogenic mutations.

Direct binding by Foxp2, however, does not necessarily lead to a downstream change in gene transcript levels. Vernes and colleagues (2011) found that only 5% of genes showing Foxp2-dependent expression changes in embryonic mouse brain were direct Foxp2 ChIP targets (Vernes *et al.* 2011). They suggested that Foxp2 might be “poised” at the target gene promoter, requiring the presence of other co-factors to bring about a transcriptional effect on the gene. It is important to recognise here the widespread effects of Foxp2 on genetic networks: genes not directly bound by Foxp2 might be regulated via genes directed regulated by Foxp2. The presence of FOXP dimerization factors (namely, FOXP1 and FOXP4) can also have an effect on the gene regulation by FOXP2 the presence of co-factors (Sin *et al.* 2014), and might be responsible for tissue-specific gene expression patterns in brain regions underlying speech and language.

Some of the genes directly bound by FOXP2 are also under positive selection between the human and chimp lineages (Spiteri *et al.* 2007). Differential gene expression has been observed *in vitro* between cells overexpressing human FOXP2 and cells overexpressing chimp FOXP2 (Konopka *et al.* 2009), as well in the striatum of mice carrying a humanised version of *Foxp2* (with the human T320N and N324S substitutions in exon 7), relative to that of wildtype mice (Enard *et al.* 2009). Konopka *et al.* (2009) found significant overlap between genes differentially expressed by human FOXP2, and genes differentially expressed between human and chimp brain regions of high *FOXP2* expression. Taken together, these findings show that the human-specific amino acid changes have a direct effect on genetic networks in the

brain, and support the idea that these genetic networks (and the cellular pathways and brain regions they subserve) underwent evolution to produce human-specific cognitive abilities, such as language.

1.5.3 The Functions of *FOXP2* in the Neuron: Neurogenesis, Differentiation, Morphology, and Synaptogenesis

Investigation of the effect of Foxp2 in neuronal function has exploited murine and songbird models. They have allowed targeted knockdown of Foxp2/FoxP2, as well as the study of non-functional mutant forms of the protein. Of note are 2 mutant mouse models that were generated through N-ethyl-N-nitrosourea (ENU) mutagenesis (Groszer *et al.* 2008). The first such murine model carries a R552H mutation, which is orthologous to the R553H KE mutation (murine Foxp2 has one less glutamine residue in the polyglutamine tract than human FOXP2). The other mutant mouse model is the Foxp2.S321X mutant mouse. This nonsense mutation produces a premature stop codon lying near the human R328X nonsense mutation described by MacDermot *et al.* (2005). S321X homozygotes do not express either full-length or truncated Foxp2 protein, and have very low levels of Foxp2 mRNA, while S321X heterozygotes have intermediate levels of Foxp2 transcript and protein relative to wildtype mice. The Foxp2.S321X transcript is thus likely to undergo nonsense-mediated RNA decay, and the mutation represents a null allele.

R552H and S321X homozygotes both have delayed weight gain and motor impairments, despite normal maternal care and lack of feeding difficulties. Both homozygotes die within 3–4 weeks after birth, even though lung morphology (which is a site of Foxp2 expression) and blood oxygenation levels are normal. Both homozygotes also have smaller cerebellums, with no other detectable abnormalities in gross brain morphology. R552H and S321X heterozygotes, on the other hand, are perfectly viable and develop normally. Heterozygotes have normal gross brain morphology, with no gross or

histological abnormalities in the cerebellum, suggesting that a single copy of functional *Foxp2* is sufficient for normal gross development.

1. Cortical Neurogenesis. In the developing mouse cortex, *Foxp2* promotes neurogenesis by increasing radial precursors at the expense of intermediate progenitors and neurons (Tsui *et al.* 2013). Importantly, while knockdown of *Foxp2* perturbed cortical neurogenesis, overexpression of mouse *Foxp2* did not enhance it, suggesting that *Foxp2* is required, but insufficient, to produce cortical neurons from radial precursors. Tsui *et al.* also found that overexpression of human FOXP2 enhanced neurogenesis, while overexpression of the KE mutant produced the same effect as *Foxp2* knockdown. Thus, with respect to neurogenesis, Tsui and colleagues concluded that human FOXP2 is a gain-of-function relative to the mouse version, and provided evidence that the KE mutant has a dominant loss-of-function effect. *Foxp2* also enhances the differentiation of mouse cortical progenitors into neurons, via the induction of platelet-derived growth factor A (PDGFR α) (Chiu *et al.* 2014).

Foxp2 expression in the cortex is limited to Layer VI, in both the developing and adult brain (Ferland *et al.* 2003). Given that superficial layers of the cortex are derived from the deeper ventral layers, all these findings support a model where *Foxp2* activates neurogenesis and differentiation gene programs in deep-layer cortical precursors during development. Affected members of KE family have reduced cortical thickness in language-related areas such as Broca's area (Vargha-Khadem *et al.* 1998, Watkins *et al.* 2002b), suggesting that the language impairment caused by the KE mutation might be due in part to deficient FOXP2-mediated neurogenesis and neural differentiation. The role of *Foxp2* in the adult cortex, however, remains to be elucidated.

2. Cerebellar Neurogenesis and Development. Mice lacking two functional copies of *Foxp2* have disproportionately reduced cerebellar volume and reduced foliation (Shu *et al.* 2005, Fujita *et al.* 2008, French *et al.* 2007, Groszer *et al.* 2008). The exact cellular mechanisms causing this remain

unclear, as there are inconsistencies in the histologies of these mutant mice. Shu *et al.* (2005) observed an ectopic Purkinje cell layer and impaired migration of granule cell progenitors at P17, but other studies (Fujita *et al.* 2008, French *et al.* 2007, Groszer *et al.* 2008) found normal cytoarchitecture. Such differences might be due to the different genetic background of the mutant mice used by different studies (Fisher and Scharff 2009). The generation of a cerebellar-specific Foxp2 knockout mouse in the future will be invaluable in resolving these issues.

3. Neuronal Differentiation. In the developing mouse striatum, Foxp2 enhances the differentiation of progenitors into medium spiny neurons, but inhibits differentiation into interneurons by inhibiting the sonic hedgehog signaling pathway (Chiu *et al.* 2014). Again, in light of structural and functional abnormalities observed in the striatums of affected KE members (Vargha-Khadem *et al.* 1998, Watkins *et al.* 2002b), these findings suggest that the KE mutation causes a developmental deficit in subcortical regions crucial for vocal learning, resulting in language impairment.

4. Neuronal Morphology. Foxp2 regulates neurite outgrowth, both in murine neuronal cell lines as well as in the developing mouse striatum (Vernes *et al.* 2011, Enard *et al.* 2009, Reimers-Kipping *et al.* 2012). Overexpression of Foxp2 in Neuro2A cells increased neurite length, while primary neurons from the ganglionic eminences of E16 mutant mice homozygous for Foxp2.R552H had reduced branching and shorter neurites (Vernes *et al.* 2011). Mice homozygous for a humanised version of Foxp2 have increased dendrite length at neurons located in cortico-basal ganglia pathways, such as Layer VI cortical neurons and the striatum (Enard *et al.* 2009, Reimers-Kipping *et al.* 2011). These findings suggest that the human-specific changes to FOXP2 contributed to morphological adaptations in cortico-basal ganglia pathways. These results also suggest that the KE mutation adversely affects the function of these brain circuits, resulting in impaired language ability.

5. Synaptogenesis. Foxp2 has been implicated in cortical synaptogenesis: overexpression of Foxp2 in rat cortical cultures reduced the density of excitatory glutamatergic synapses, but did not affect the density of inhibitory GABAergic synapses (Sia *et al.* 2013). This effect is mediated by Foxp2 repression of SRPX2. In zebra finches, knockdown of FoxP2 in both immature and mature spiny neurons of Area X decreased spine density (Schulz *et al.* 2010). While both these findings provide support for a functional role of Foxp2 at the synapse, it remains unknown if these effects are direct (presynaptic) or indirect (postsynaptic) (Schulz *et al.* 2010).

All these findings support a role for FOXP2 in neurogenesis, differentiation, morphology, and synaptic maturation, but the molecular mechanisms by which these effects occur remain unclear. The direct FOXP2 targets responsible for these processes remain to be elucidated. Future investigation can capitalise on the lists of FOXP2/Foxp2 targets (Spiteri *et al.* 2007, Vernes *et al.* 2007, Vernes *et al.* 2011) to investigate molecular links between FOXP2, downstream direct targets, and subsequent effects on neuronal function (an approach used by Sia *et al.* 2013).

1.5.4 The Functions of *FOXP2* in Brain Circuits: Roles in Synaptic Activity and Plasticity

The effects of Foxp2 on neuronal morphology in the striatum have expected downstream consequences on the behaviour of the circuits formed from these very neurons. In mice, R552H heterozygotes have elevated striatal activity during motor skill learning (French *et al.* 2012). In addition, striatal neurons in these mice showed a decrease in firing rate during motor skill learning, as opposed to an increase of firing rate as seen in wildtype mice. Furthermore, R552H heterozygotes have impaired long-term depression (LTD) at glutamatergic inputs to the striatum from the cortex (Groszer *et al.* 2008). In zebra finches, lentiviral knockdown of FoxP2 in Area X of zebra finch brains results in faster signal transmission through the cortico-striatal anterior

forebrain pathway, as well as rendering the circuit insensitive to modulation by dopamine receptors of the D1 subtype (Murugan *et al.* 2013).

On the other hand, mice with humanised *Foxp2* have increased LTD at these corticostriatal synapses (Enard *et al.* 2009, Reimers-Kipping *et al.* 2011), concordant with the trend of humanised *Foxp2* exerting opposite effects on neurite outgrowth (Enard *et al.* 2009, Reimers-Kipping *et al.* 2011) and striatal gene expression (Enard *et al.* 2009), relative to wildtype *Foxp2* and the R552H mutant.

In the cerebellum, at synapses between the parallel fibers and Purkinje cells, R552H heterozygotes had enhanced short-term potentiation at shorter inter-stimulus intervals, and subtly enhanced LTD (Groszer *et al.* 2008). In mice with humanised *Foxp2*, however, no significant differences were observed. These observations suggest that the R552H mutation disrupts synaptic function in the striatum and the hippocampus, but the human-specific changes to exon 7 of *FOXP2* primarily affect the cortico-basal ganglia circuitry (Reimers-Kipping *et al.* 2011).

1.5.5 The Function of *FOXP2* on Animal Behaviour: Effects on Rodent Motor Skills, Rodent Pup Vocalizations, and Songbird Vocal Learning

1. Rodent Motor-skill Learning and Auditory-Motor Association. In mice, R552H heterozygotes exhibit impaired motor-skill learning relative to wildtype mice, as measured by performance on the tilted running wheel (Groszer *et al.* 2008) and the accelerating rotarod (French *et al.* 2012). Related to the motor-skill learning deficit in R552H heterozygotes is their slower rate of learning an auditory-motor association task (Kurt *et al.* 2012).

Importantly, S321X heterozygotes perform even worse on the auditory-motor association task than R552H heterozygotes (Kurt *et al.* 2012). This strongly suggests that the S321X mutant version of *Foxp2* is a null mutant. It will be interesting to see if there are any differences in performance on an auditory-

motor association task between affected KE members (who have the orthologous R553H mutation) and the individual carrying a truncated FOXP2. Any discrepancies in performance will shed light on the different effects of the two mutations in the human brain.

2. Rodent Pup Vocalization. Mouse pups carrying either 1 or 2 copies of non-functional *Foxp2* emit less ultrasonic vocalizations during isolation from the nest (known as isolation calls) compared to wildtypes (Shu *et al.* 2005, Fujita *et al.* 2008, Groszer *et al.* 2008, Gaub *et al.* 2010), with homozygotes producing none at all. However, during a state of higher arousal (elicited by lifting pups above the ground), R552H homozygotes do produce ultrasonic distress calls, albeit fewer in number, relative to both wildtype and heterozygous littermates (Groszer *et al.* 2008, Gaub *et al.* 2010). Both Groszer *et al.* (2008) and Gaub *et al.* (2010) suggested that reduced ultrasonic vocalization by homozygotes is a secondary, rather than primary, consequence of their severe developmental and motor deficits. Taken together, this suggests that while *Foxp2* modulates certain aspects of murine ultrasonic vocalizations, it is unlikely to be necessary for it.

Pups homozygous for a partly humanised version of *Foxp2* (carrying 2 human-specific changes at T303N and N325S) produce as many ultrasonic isolation calls as wildtype pups, but these calls have a lower acoustic frequency (Enard *et al.* 2009). While this suggests that the 2 amino acid changes acquired during human evolution do affect the neural mechanisms underlying vocalization, Enard and colleagues also point out that undetected differences in lung or larynx function might be present in these pups carrying the humanised form of *Foxp2*, suggesting that the altered ultrasonic calls might be secondary effects of humanised *Foxp2* in non-neural tissue.

A recent study (Bowers *et al.* 2013) found sex differences in rat ultrasonic isolation calls, with male pups producing more isolation calls compared to females and eliciting preferential maternal retrieval. The sex-specific patterns in ultrasonic isolation calls and preferential retrieval were reversed following

transient knockdown of Foxp2. While this study provided additional evidence for a role of Foxp2 in rodent ultrasonic vocalization, the relevance of sex-specific differences in Foxp2 expression to vocalization remain to be clarified.

It is important to note that pup ultrasonic vocalizations are believed to be innate rather than learned (Fischer and Hammerschmidt, 2010), as they occur when the pups are still deaf (Ehret 2005). Also ultrasonic vocalizations are produced via whistle-like constriction through which air is forced, rather than through vibrations of vocal cords (Fischer and Hammerschmidt, 2010). Thus, establishing their relevance to non-innate learned vocalization (as seen in human speech and language) requires more clarity regarding the adult neural circuitry underlying it.

3. Songbird Vocal Learning. In zebra finches, FoxP2 is required in Area X for complete and accurate song imitation by young birds ('pupils') undergoing song learning from an adult 'tutor' (Haesler *et al.* 2007). Pupils who underwent acute lentiviral knockdown of FoxP2 in Area X omitted syllables during song imitation, and copied syllables were inaccurately imitated. This impaired song imitation manifested early during song learning, and persisted into adulthood (Haesler *et al.* 2007, Murugan *et al.* 2013).

FoxP2 also has acute effects on zebra finch song during adulthood. Knockdown of Foxp2 in Area X in adult male zebra finches reduced pitch variability normally observed between directed and undirected singing (Murugan *et al.* 2013). This reveals the functional relevance of higher FoxP2 expression in Area X during directed singing (Teramitsu and White 2006).

Vocal learning in songbirds is believed to rely on rapid reinforcement-based motor-learning, where motor patterns controlling vocal output are dynamically adjusted in response to auditory feedback (Haesler *et al.* 2007). Thus, impairments in zebra finch vocal learning caused by lentiviral FoxP2 knockdown are congruous with the motor-learning deficits observed in mice carrying pathogenic mutations in Foxp2. Given that a central feature of the KE

phenotype is an inability to properly execute sequential orofacial movements required for speech, these behavioural studies highlight the centrality of motor learning deficits in *FOXP2* dysfunction. Taken together, these behavioural studies reveal that *FOXP2*, beyond having a developmental function in the brain, has an online role in aspects of motor-learning and vocalization. They also provide support for an evolutionarily conserved neural substrate for human speech and language, involving orthologous brain circuits in mice and songbirds that are sites of *Foxp2* function.

1.5.6 The Functions of *FOXP2*: From Genes to Neurons to Behaviour, and Back Again

All of the studies reviewed above show that *FOXP2* can produce behavioural changes in animals by exerting effects at the molecular and genetic levels of brain function. Pathogenic mutations affect the immediate DNA-binding function of *FOXP2*, which has ramifications at all levels of brain function. The differences in gene regulation caused by the two amino acid changes in exon 7 of human *FOXP2* also account for the observed changes in neuronal morphology and neuronal plasticity (Enard *et al.* 2009, Reimers-Kipping *et al.* 2011). Insight into the developmental and behavioural roles of *FOXP2* in the brain, as well as its contribution to the evolution of human speech and language, can thus be gained from investigation at the genetic and neuronal levels.

1.6 MicroRNAs: A Large Family of Small Noncoding Gene Regulators

Amongst the set of genes bound by *Foxp2* in the murine brain (Vernes *et al.* 2011), there are a number of non-coding genes known as microRNAs. MicroRNAs are small (18 – 22 nucleotides in length) RNA molecules that regulate mRNA stability and turnover – and thus gene expression – by sequence-specific base-pairing at the 3' untranslated region (3' UTR) of mRNA transcripts. One microRNA is able to regulate several thousand

transcripts on the basis of complementary base-pairing (Lewis *et al.* 2003). MicroRNAs are thought to fine-tune gene expression, rather than fully repressing genes in an on-off fashion (Bartel and Chen 2004, Kerres *et al.* 2007).

Several microRNAs are widely expressed in the brain, with a number of these restricted to the central nervous system (Landgraf *et al.* 2007). They regulate the maturation of neurons, and have roles in neurogenesis (Visvanathan *et al.* 2007) and differentiation (Conaco *et al.* 2006). MicroRNAs play a role in neuronal morphology, such as modulating dendrite development (Schratt *et al.* 2006) and axonal outgrowth (Hengst *et al.* 2006). MicroRNAs have also been shown to be involved in activity-dependent synaptic plasticity (Vo *et al.* 2005; Wayman *et al.* 2008; Edbauer *et al.* 2010). In line with these findings, several microRNAs are localised to axonal and dendritic regions (Schratt 2009, Natera-Naranjo *et al.* 2010), and contribute to rapid and dynamic modulation of local protein synthesis. Thus, microRNAs regulate several steps of leading to the formation of adult neural circuits (Figure 1.7).

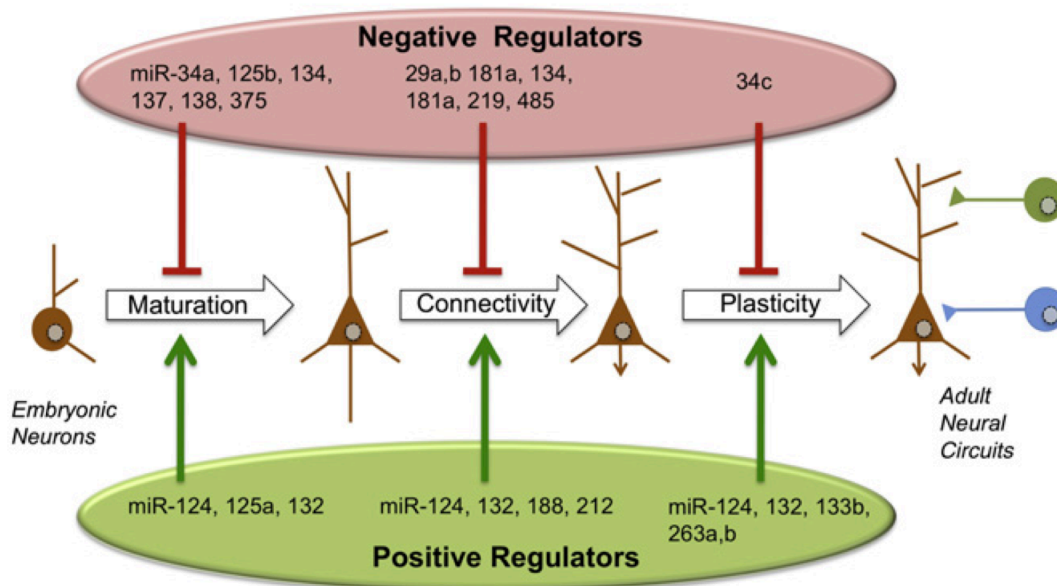


Figure 1.7. MicroRNAs are Involved In Several Processes Leading Up to the Formation of Adult Neural Circuits.

MicroRNAs both promote and dampen the processes that produce brain circuits from individual embryonic neurons, highlighting their pervasive influence on the brain at several varied levels of function. Taken from McNeill and Van Vactor 2010.

MicroRNAs have also been implicated in neuro-psychological disorders like schizophrenia (Perkins *et al.* 2007), bipolar disease (Moreau *et al.* 2011), and autism (Abu-Elneel 2008). These studies, and several others, have identified aberrant postmortem levels of several microRNAs in the brain tissues of affected individuals, as well as in murine models (Loohuis *et al.* 2011, Beveridge and Cairns 2012). This suggests that dysregulation of microRNA levels can result in aberrant gene expression, leading to neuro-psychological disorders. A recent large-scale genome-wide association study (GWAS), which compares genomic variants between individuals with a trait (cases) and individuals without the trait (controls) to identify if any genomic variant or region is associated with the trait, uncovered a highly significant association signal between schizophrenia and a single nucleotide polymorphism (SNP) lying near miR-137 (Ripke *et al.* 2011). Furthermore, miR-137 has been found to regulate other genes implicated in schizophrenia (such as *TCF4*, Kwon *et*

al. 2011) and autism (such as *RORa*, Devanna and Vernes 2014). All this underscores the importance of microRNAs in molecular functions underlying cognitive aspects of the human brain.

1.7 Building Bridges Between MicroRNAs and Human Speech and Language

The microRNAs miR-9 and miR-140-5p are upregulated in Area X of adult male zebra finches during undirected singing (Shi *et al.* 2013). Also, microRNA levels in the zebra finch auditory forebrain are regulated in response to song exposure (Gunaratne *et al.* 2010, Warren *et al.* 2010). Taken together, this suggests that microRNAs play a role in auditory processing in songbirds. Given that auditory feedback is necessary for vocal learning in male zebra finches (Nordeen and Nordeen 1992), and that pathogenic *Foxp2* mutations impair auditory-motor association learning in mice (Kurt *et al.* 2012), it is also possible to hypothesise that microRNAs play a role in vocal learning.

Amongst the validated targets of murine *Foxp2*, there are several microRNAs important to neuronal function, such as miR-9 and miR-124 (Vernes *et al.* 2011). *FoxP2* itself undergoes regulation by miR-9, miR-132, and miR-140-5p in the mouse and zebra finch (Clovis *et al.* 2012, Shi *et al.* 2013). MicroRNAs and *FOXP2*/*Foxp2* might thus participate in gene networks that act to bring about gene expression patterns. These expression patterns might underpin neuronal function in circuits and regions important for human speech and language.

1.8 Goals of this Thesis

In light of tentative links between microRNAs and human speech and language, I will investigate the genetic connections between microRNAs, and networks regulated by *FOXP2*.

I will begin by identifying microRNAs regulated by murine Foxp2 in embryonic brain using microarrays. With a list of significantly regulated microRNAs, I will identify biological themes associated with the targets of these microRNAs. To both validate the microarray results as well as identify conserved microRNA regulation by human FOXP2, I will investigate microRNA regulation in established *in vitro* models for FOXP2 function. The results of this work are detailed in Chapter 3.

In Chapter 4, I will investigate a genetic network involving FOXP2 and 3 of its downstream targets – *TCF4*, *CNTNAP2*, and *MIR137*. These genes have each been implicated in a range of neuropsychological disorders, and also have been found to interact with each other. Given the regulatory cross-talk between these 4 genes, as well as the striking similarities between the cognitive phenotypes produced by mutations in these individual genes, it is possible that these 4 genes are members of a common genetic network that is disrupted in neurological disorders such as language impairment and schizophrenia. Elucidation of this putative genetic network might refine the understanding of molecular mechanisms involved in language disorders, as well as those in neurological deficits. To test if FOXP2 acts as a hub in this network, I will use reporter gene assays to see if direct binding of FOXP2 to previously identified binding sites is sufficient to produce the observed transcriptional regulation.

In Chapter 5, I will investigate if variants in microRNAs represent risk factors for specific language impairment (SLI). FOXP2 targets such as *CNTNAP2* have been shown to be associated with language deficits in families with SLI (Vernes *et al.* 2008). If microRNAs play an important role in the function of FOXP2 in the brain, SNPs lying in FOXP2-regulated primary microRNA transcripts might be associated with SLI as well. I will test this hypothesis with a candidate association study in a family-based SLI cohort. Noncoding variants in the 3' UTR that lie in predicted microRNA binding sites can also affect microRNA function. Given the relative ease of validating these

noncoding mutations *in vitro*, I will describe and apply a novel workflow to identify and functionally test the effects 3' UTR variants in exome data from SLI probands.

Chapter 2: Materials and Methods

2.1 *In Vitro* Cell Culture

1. SH-SY5Y cell-lines. Stably-transfected human SH-SY5Y neuroblastoma cell-lines (Vernes *et al.* 2007) were cultured in Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (DMEM:F-12) (Sigma, USA), supplemented with 10% Foetal Calf Serum (FCS) (PAA Laboratories GmbH), 2 mM L-glutamate (Sigma, USA), and 2 mM penicillin/streptomycin (Sigma, USA). Three separate stable cell-lines were used: one expressing FOXP2 (SH.FOX), another expressing a mutant version of FOXP2 with the R553H KE mutation (SH.KE), and a third line transfected with the empty vector, pcDNA3.1 (SH.EMPTY) (Invitrogen, USA). Cells were grown at 37°C in 5% CO₂.

Neuronal differentiation of SH-SY5Y cells was achieved with a two-stage differentiation protocol (Encinas *et al.* 2000, performed by P. Devanna and P. Rodenas-Cuadrado). Cells are seeded at 10⁴ cells/cm² per well in 6-well plates, and allowed to grow for 48 hours. They are then exposed to 10µM all-trans retinoic acid in low-serum (2% FCS) media for 5 days, followed by exposure to 50ng/mL brain-derived neurotrophic factor (BDNF, from Alomone Laboratories, Israel) in serum-free media for 10 days, with replenishment every 5 days.

2. Neuro2A cell-lines. Stably-transfected murine Neuro2a (N2A) neuroblastoma cell-lines (Vernes *et al.* 2011) were cultured in Modified Eagle's Medium (MEM) (Sigma, USA), supplemented with 10% Foetal Calf Serum (PAA Laboratories GmbH), 2 mM L-glutamate (Sigma, USA), and 2 mM penicillin/streptomycin (Sigma, USA). Stable Neuro2a cell-lines overexpressing Foxp2 protein (N2A.FOX), the mutant R552H form (N2A.KE), or the empty vector pcDNA3.1 (N2A.EMPTY) (Invitrogen, USA) were used. Cells were grown at 37°C in 5% CO₂.

3. HEK392 cell-lines. HEK293 cells (derived from human embryonic kidney) were grown in Dulbecco's Modified Eagles Medium (DMEM) (Invitrogen, UK) supplemented with 10% foetal calf serum (Sigma, USA), 2 mM L-glutamine (Sigma, USA) and 2 mM penicillin/streptomycin (Sigma, USA), and were grown at 37°C in 5% CO₂.

2.2 RNA Extraction

1. RNA extraction from mouse embryos. Whole brains from embryonic day 16 (E16) mouse embryos were transferred into ice-cold Dissection Buffer (15 mM HEPES, 0.01% NaHCO₃, and 25 mM glucose in calcium and magnesium-free Hank's Buffered Salt Solution). Each brain was kept on ice while the ganglionic eminences from both hemispheres were removed, followed by homogenization in TRIzol® (Invitrogen, UK). (Performed by S. Vernes.) Total RNA was then harvested according to manufacturer's instructions, and yield and purity was assessed using a NanoDrop ND-1000 spectrophotometer. (NanoDrop Technologies, USA). Purified samples were stored at nuclease-free water (Ambion, USA) at -80°C.

2. *In vitro* RNA extraction. SH-SY5Y cell-lines were harvested following 5 days of culture in regular growth media (undifferentiated) or following the differentiation protocol (Day 5, 10 or 15). N2A cell-lines were harvested at 70 – 80% confluency. Cells were washed twice with sterile PBS (Gibco, UK), and then harvested by scraping, followed by pelleting at 300 *g* for 3 minutes. The pellet was then resuspended and homogenised in TRIzol® (Invitrogen, UK).

For each sample, separate mRNA and microRNA fractions were obtained using the RNeasy kit (QIAGEN, Netherlands) in combination with the Purelink miRNA isolation kit (Invitrogen), according to manufacturers' instructions. The yield and purity was assessed via a NanoDrop. Purified samples were stored at nuclease-free water (Ambion, USA) at -80°C.

2.4 Microarray Processing and Analysis

Each sample consisted of 800 ng of total RNA (from 6 wildtype C57BL/6J mice embryos and 6 homozygous *Foxp2.S321X* mutant littermates, Vernes *et al.* 2011), and was poly(A)-tailed and labeled with biotin via ligation using the FlashTag® Biotin HSR RNA labelling kit (Genisphere, USA). This was followed by hybridisation to GeneChip® miRNA 1.0 arrays (Affymetrix, USA). These GeneChip arrays comprised of 46,228 probes that form 7,815 probe sets and represent over 6703 microRNA sequences found across 71 organisms from the Sanger miRNA database v11 (<ftp://mirbase.org/pub/mirbase/11.0/>) and an additional 922 sequences of Human snoRNA and scaRNA from the Ensembl database and snoRNABase. Four copies of each miRNA probe were included on the array. Each probe was a 25-mer oligonucleotide, unless the miRNA was less than 25 nucleotides in length. All probe sets are perfect match only (perfectly complementary to the target miRNA sequence), and were designed by the manufacturer. (RNA labeling and microarray hybridization performed by E. Ballabio.)

Data from the Affymetrix GeneChip arrays was imported into miRNAQCtool v1.0.33.0 (Affymetrix, USA). The Robust Multichip Average method (Irizarry *et al.* 2003), involving quantile normalization and median polish summarization, was used to correct for background fluorescence and to normalise probe intensity. MicroRNA probe sets are assayed for detection based on a Wilcoxon Rank-sum test of the miRNA probe signal, compared to the distribution of signals from GC-content matched anti-genomic probes. If the obtained p-value for a probe is < 0.05 , that probe was declared “Detected” in that sample.

The data was then imported into Partek (version 6.11.0701), and the Batch RemoverTM function was applied to remove effects of litter and array processing date. An analysis of variance (ANOVA) was performed to identify differentially expressed genes. Heatmaps were generated in R (<http://www.R->

project.org/), using agglomerative clustering based on the Euclidean dissimilarity of gene expression values, which were normalised to a mean of zero and a standard deviation of one.

2.3 Quantitative Real-time Polymerase Chain Reaction (qRT-PCR)

1. Reverse transcription of mRNA transcripts and pri-miR transcripts. Reverse transcription of mRNA fractions was performed using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, USA), according to manufacturer's instructions. Briefly, 800 ng of RNA was added to 4 mM deoxynucleotide triphosphates (dNTPs), 50 units of MultiScribe™ Reverse Transcriptase, 20 units of Superase-In RNase Inhibitor (Ambion, USA), and random primers on ice. The mixture was then incubated in a thermal cycler with the following protocol: 10 min at 25°C, 120 min at 37°C, 5 min at 85°C, hold at 4°C. A negative control reaction lacking reverse transcriptase was performed in parallel, to ensure that any amplification observed during subsequent PCR experiments is solely due to cDNA, and not genomic DNA.

2. Reverse transcription of microRNAs from small molecular weight fractions. RNA fractions of low-molecular weight were polyadenylated prior to reverse transcription, using the NCode miRNA First-strand cDNA synthesis kit (Invitrogen, USA), as per the manufacturers' protocol. Briefly, 250 ng of low-molecular weight RNA was mixed with 2.5 mM MnCl₂, 5 nM adenosine triphosphate (ATP), and 1 unit of Poly-A polymerase, in miRNA Reaction Buffer (Invitrogen, USA), on ice. The mixture was incubated at 37°C for 15 minutes. Reverse transcription was performed with Superscript III Reverse Transcriptase kit (Invitrogen, USA), in accordance with the manufacturer's instructions. Briefly, the polyA-tailed sample was diluted with Universal RT Primer (final concentration 9.375 μM) in Annealing Buffer (Invitrogen, USA), and incubated at 65°C for 5 minutes, followed by a 1-minute incubation on ice. 200 units of SuperScript III reverse transcriptase (Invitrogen, USA), and 40 units of RNaseOUT™ Recombinant Ribonuclease Inhibitor (Invitrogen, USA)

were then added in the presence of First-strand Buffer (Invitrogen, USA), Reverse transcription was achieved by incubating the final mixture in a thermal cycler with the following parameters: 50 minutes at 50°C, 5 minutes at 85°C, hold at 4°C. A negative control reaction lacking reverse transcriptase was included as well to ensure the preparation was free from genomic contamination.

3. Quantitative Polymerase Chain Reactions (qPCRs). qPCRs were run in technical duplicates with SYBR Green Supermix (Biorad, USA), and performed on the CFX96 Real-Time PCR Detection System thermal cycler (BioRad, USA), according to manufacturers' instructions. Briefly, forward and reverse primers (each with a final concentration of 0.5 μ M) were added to 1.5 μ l of cDNA, and mixed with SYBR Green Supermix. The thermal cycling protocol used was as follows:

1. 95°C for 3 minutes, followed by 45 cycles of:
2. 95°C for 10 seconds
3. 60°C for 30 seconds (Annealing/Extension step)

Fluorescence data was acquired as the end of each annealing/extension step. Melting curve analysis was performed using the CFX Manager Software (Biorad, USA), to evaluate the specificity of the primers used.

Quantification was done via the $2^{-\Delta\Delta C_T}$ method (Livak and Schmittgen 2001). For mRNA transcripts and pri-miR transcripts, the raw threshold cycle data was normalised to the geometric mean of *GAPDH* and *POLR2* expression (Hoerndli *et al.* 2004). Normalisers for microRNA quantification were selected using the GENorm algorithm (Vandesompele *et al.* 2002, Mestdagh *et al.* 2009). Fold changes are relative to the cell-lines expressing the empty vector, and are expressed as mean of values from 3 biological replicates from each cell-line, along with the standard error of the mean (SEM). Statistical significance was determined using ANOVA followed by Tukey's post-hoc test.

The primers used for quantification of mRNA and pri-miR transcripts are listed in Table 2.1.

Table 2.1 List of primers used for qRT-PCR experiments.

Gene	Forward primer (5'→3')	Reverse primer (5'→3')
<i>GAPDH</i>	CAGTCCATGCCATCACTGC	TTCGTTGTCATACCGGAAATG
<i>POLR2F</i>	CCCGAAAGATCCCCATCAT	CACCCCCCAGTCTTCATAGC
<i>DICER</i>	TGATCCCCCTGTGAATTGGCTTC	TCGCCAGCATGCAGTCTTTTGTC
<i>TCF4</i>	GAGATGGAAATGGACATTGCCACG	GGACCCTGAGCTACTTCTGTCTTC
<i>CNTNAP2</i>	TCCCTCCACGTCCCAAAAATG	TCTTGGCATAGCCGGGAGAA
pri-mir-137 (<i>MIR137</i>)		
pri-mir-221 (<i>MIR221</i>)	AACATCCAGGTCTGGGGCATGAAC	GGAGAACATGTTTCCAGGTAGCCT
pri-mir-340 (<i>MIR340</i>)	CCCTACTCCTTTTCCAGCTTGAGTC	CGGGCCATGATCTGTGGTATAAG
pri-mir-941 (<i>DNAJC5</i>)	GGAGTCATTGTACCACGTCCTTGG	GGATGGCGTGC GCGTTGTTGATCT
let-7d-5p [^]	AGAGGTAGTAGGTTGCATAGTT	Universal RT Primer ^b
miR-9-5p [^]	TCTTTGGTTATCTAGCTGTATGA	
miR-20a-5p [^]	TAAAGTGCTTATAGTGCAGGTAG	
miR-25-3p [^]	CATTGCACTTGTCTCGGTCTGA	
miR-30c-5p [^]	TGTAAACATCCTACACTCTCAGC	
mmu-miR-124-3p	CGTGTTCCACAGCGGACCTTG	
mmu-miR-125b2-3p	ACAAGTCAGGTTCTTGGGACCT	
miR-137-3p [^]	TTATTGCTTAAGAATACGCGTAG	
mmu-miR-181b-5p	AACATTCATTGCTGTCTGGTGGGT	
mmu-miR-320-3p	AAAAGCTGGGTTGAGAGGGCG	
miR-615-5p [^]	GGGGGTCCCCGGTGCTC	
mmu-miR-1195	TGAGTTCGAGGCCAGCCTG	
hsa-let-7b-5p	TGAGGTAGTAGGTTGTGTGGTT	
hsa-let-7e-5p	TGAGGTAGGAGGTTGTATAGTT	
hsa-mir-17-5p	CAAAGTGCTTACAGTGCAGGTA	
hsa-miR-23b-3p	ATCACATTGCCAGGGATTACC	
hsa-miR-24-3p	TGGCTCAGTTCAGCAGGAACAG	
hsa-miR-92a-3p	TATTGCACTTGTCCCGGCCTGT	
hsa-miR-93-5p	CAAAGTGCTGTTTCGTGCAGGTAG	
hsa-miR-103a-3p	AGCAGCATTGTACAGGGCTATGA	
hsa-miR-191-5p	CAACGGAATCCCAAAAGCAGCTG	
hsa-miR-342-3p	TCTCACACAGAAATCGCACCCG	
hsa-miR-346	TGTCTGCCCCGCATGCCTGCCT	
hsa-miR-433-3p	ATCATGATGGGCTCCTCGGTG	
hsa-miR-638	AGGGATCGCGGGCGGGTG	
mouse U6	CTGCGCAAGGATGACACG	
human U6	CTGACAACACGCATACGGTTAAG	

[^] – Primers that were used for both human and mouse experiments. ^b – The reverse primer for miRNA and U6 reactions was the Universal qPCR reverse primer, supplied with the NCode miRNA First-strand cDNA synthesis kit (Invitrogen, USA).

2.4 MicroRNA Target Prediction

Target predictions for the mouse genome were downloaded from four databases: TargetScan (Lewis *et al.* 2003), miRanda (Betel *et al.* 2008), PicTar (Krek *et al.* 2005, Anders *et al.* 2012), and Diana-microT (Reczko *et al.* 2012). Each database employs a different algorithm to predict targets, and each one assigns a score on the prediction quality and likelihood. To ensure a high confidence of target predictions, only TargetScan predictions with a “context+” score of ≤ 0.1 (Grimson *et al.* 2007), miRanda predictions with a miRSVR score of < -0.8 (Betel *et al.* 2010), and Diana-microT predictions with a miTG score > 0.8 were used (Maragkakis *et al.* 2010). Overlap between the different target predictions was determined in R (<http://www.R-project.org>).

2.5 Gene Ontology Analysis

The Webgestalt program (<http://bioinfo.vanderbilt.edu/webgestalt/>) was used to perform gene ontology analysis (Wang *et al.* 2013). Significantly enriched gene ontology categories were determined with hypergeometric testing, with multiple-testing correction performed via the Benjamini and Hochberg method (Benjamini and Hochberg 1995).

2.6 Primary Cell Culture

The cortices and ganglionic eminences were dissected out from E16 mouse embryos in ice-cold Dissection Buffer, and immediately chopped with a sterile razor on ice. The chopped tissues were then pelleted at 150 *g* for 5 minutes at 4°C. The pellet was then incubated with 500 μ L of cold papain solution (20 units/ml papain, 1 mM L-cysteine, 0.5 mM EDTA, 100 units/ml DNaseI, in ice-cold Dissection Buffer), first on ice for 5 minutes, then at 37°C for 10 minutes, with agitation every 2 minutes. Two mL of 3.7-BSA solution (10 mg/ml ovomucoid, 10 mg/ml BSA in Dissection Buffer, filtered sterilised) was added with gentle mixed to stop the enzymatic digest. Cells were pelleted at 233 *g* at 4°C for 5 min and then washed in complete medium (Neurobasal medium (Gibco, UK) with 2 mM Glutamax (Gibco, UK), 2 mM Penicillin/Streptomycin

(Gibco, UK), and 1 X B27 supplement (Gibco, UK)).

The resultant cell suspension then underwent 2 successive titrations: 10 times with a plastic pipette, followed by 10 times with a fire-polished Pastuer pipette. Finally, the suspension was passed through a 70 μ m cell strainer. Cells were plated onto laminin and poly-D-lysine coated coverslips (BD Biosciences, USA) at a density of 6×10^4 cells per well into 24 well plates and grown at 37°C, in the presence of 5% CO₂, in complete medium.

2.7 Immunofluorescence

Twenty-one days after plating onto coverslips, primary mouse neurons from the ganglionic eminences and the cortices were fixed using 4% paraformaldehyde solution (Electron Microscopy Sciences, USA) for 10 minutes at room temperature, and then washed with phosphate-buffered solution (LEM, Netherlands). After 14 days in culture, cells were fixed in 4% paraformaldehyde solution (Electron Microscopy Sciences, USA) for 15 minutes at room temperature, and then washed with 0.1% Triton X-100 in Tris-buffered saline. Antibodies were diluted in a blocking solution consisting of 1% fish gelatin (Sigma, USA), 0.1% Triton X-100 (Sigma, USA) and 5% bovine serum albumin (Sigma, USA) dissolved in phosphate-buffered saline (LEM, Netherlands). Cells were incubated with two primary antibodies overnight at 4°C in the dark, with gentle shaking: a rabbit anti-Tcf4 monoclonal antibody (1:500, Merck Millipore #04-1080) and a chicken anti-Map2 polyclonal antibody (1:800, Merck Millipore #AB5622).

The Tcf4 signal was amplified via four rounds of incubation with secondary and tertiary antibodies. Cells were incubated in the dark at room temperature with a biotinylated goat anti-rabbit antibody (1:100, Vector Laboratories #BA-1000) and a goat anti-chicken TRITC antibody (1:500, Molecular Probes #A-11039). After 4 washes with blocking solution, the cells were further incubated with a goat anti-biotin FITC antibody (1:500, Vector Laboratories #SP-3040) in the same conditions. Nuclei were visualised using VECTASHIELD® Mounting

Media with a DAPI counterstain (Vector Laboratories, USA). Images were acquired using an Axiovert A-1 fluorescent microscope, with a Zeiss AxioCam MRm camera and ZEN Image Software (Zeiss, Germany).

2.8 Generation of Vectors

1. Luciferase Reporter Constructs. 3' UTR regions of interest and FOXP2 binding sites of interest were amplified from human genomic DNA, using the primers from Table 2.2. The Expand High Fidelity PCR System (Roche, Germany) was used according to manufacturer's instructions, with final primer concentrations of 300 nM. The amplified 3' UTR fragments were recovered with the Wizard® SV Gel and PCR Clean-Up System (Promega, USA) according to manufacturer's instructions. The amplicons were then cloned into the pGL4.23 expression vector (Promega, USA), with XbaI (New England Biolabs, USA) and FseI (New England Biolabs, USA), downstream of a firefly luciferase reporter gene that is driven by the SV40 promoter. The rs72727021 minor allele construct for the *C9ORF100* 3' UTR were generated using the QuickChange Site-Directed Mutagenesis kit (Stratagene, USA), according to the manufacturer's protocol, with the primers indicated in Table 2.2.

Table 2.2. Tagged Primers Used to Amplify Regions of Interest for the Generation of Luciferase Reporter Constructs.

3' UTR	Forward primer (5'→ 3')	Reverse primer (5'→ 3')
<i>BTN2A1</i>	TCTAGAACAGCCATGTAGACA AGCCC	GGCCGGCCTCTCCTCTCTGG TGGTCCTG
<i>C9ORF100</i>	TCTAGACCACAGCAGCCAGA AAAAC	GGCCGGCCCACTCACTGTGT TGCCCAGG
<i>CENPJ</i>	TCTAGAACACGGAGCTGTGA CGATCCTCAT	GGCCGGCCACAAGTCCAGTG CTCTACGGCT
<i>MTMR3</i>	TCTAGACCAACTGAAGCTCAG TGACCTGG	GGCCGGCCTTTCTTCCCTGC TCACGGTCTCA
rs72727021 minor allele site-directed mutagenesis for <i>C9ORF100</i>	GAACTCAGGATACCTCAGGG AGAGGTCACAGCC	GGCTGTGACCTCTCCCTGAG GTATCCTGAGTTC

The FOXP2 binding sites were cloned upstream of the SV40 promoter instead, with KpnI (Fermentas, USA) and XhoI (Fermentas, USA). All restriction digests were performed at 37°C for 1 hour, and recovery of digested fragments was achieved with the Wizard® SV Gel and PCR Clean-Up System according to manufacturer's instructions (Promega, USA). All ligations were performed by incubating digested fragments and vector with T4 DNA Ligase (New England Biolabs, USA) in Ligase Reaction Buffer (New England Biolabs, USA) for 1 hour at room temperature.

2. MicroRNA overexpression constructs. To generate overexpression constructs for microRNAs of interest, regions containing the primary transcripts were amplified from human genomic DNA with the primers listed in Table 2.3. The PCR reagents and thermal cycling programs used were the same as above. These amplicons were then cloned into the pLKO.1 vector (Addgene, USA) using AgeI (Fermentas, USA) and EcoRI (Fermentas, USA) restriction enzymes, downstream of the human U6 promoter.

Table 2.3. Tagged Primers Used to Amplify Primary MicroRNA Transcripts for the Generation of MicroRNA Overexpression Constructs.

microRNA	Forward primer (5'→ 3')	Reverse primer (5'→ 3')
<i>MIR215</i>	ACCGGTGGACTCTCATTTGATT CCAGCAG	GAATTCCTGCCATTGCGTTGGTT GCGT
<i>MIR342</i>	ACCGGTGGCACTTCCTGCCATT GCATCC	GAATTCAGTATGCACACCTCGG GGC
<i>MIR346</i>	ACCGGTTTGGCTGCAGGTTGG AGGGCTT	GAATTCGCTGACTGTGGAGGTAG GTTGG
<i>MIR433</i>	ACCGGTCCGGAGGTTTCTCAC AGGTGG	GAATTCCTCCCGCAACATCTCCC CTATC

3. Positive Control Reporter Constructs. Positive control reporter constructs (also referred to as “sponge” constructs) were generated according to previously reported protocols (Kluiver *et al.* 2012). Briefly, for each of the four microRNAs of interest, complementary oligonucleotide sequences (listed in

Table 2.4) were designed, containing 2 binding sites for the microRNA separated by a 6-nucleotide spacer, and with overhangs at the 5' end compatible with KflI (Fermentas, USA). Oligos were annealed to produce a duplex, phosphorylated with T4 polynucleotide kinase (New England Biolabs, USA), and then ligated into a modified pGL4.23 expression vector (Devanna and Vernes, 2014) carrying a KflI site downstream of a firefly luciferase reporter. This produces a construct with the sponge duplex in the 3' UTR of the luciferase reporter. Because of the palindromic KflI recognition sequence, multiple copies of the duplex will be directionally inserted into the vector. PCR screening of positive clones was thus performed, and sponge constructs carrying 2 duplexes (or 4 binding sites) were selected for use as positive controls in luciferase reporter assays.

Table 2.4. Oligos Used to Generate Duplexes With MicroRNA Binding Sites.

Sponge for	Sense Oligo (5'→ 3')	Antisense Oligo (5'→ 3')
hsa-miR-215-5p	GACCCGTCTGTCAACCATAG GTCATATATTCGTCTGTCAAC CATAGGTCATGG	GTCCCATGACCTATGGTTGACAG ACGAATATATGACCTATGGTTGAC AGACGG
miR-342-3p	GACCCACGGGTGCGATTTAG TGTGAGAATATTCACGGGTG CGATTTAGTGTGAGAGG	GTCCCTGTCTGCCCCGTGGCCTGC CTCTGAATATTGTCTGCCCGTGGC CTGCCTCTGG
hsa-miR-346	GACCCAGAGGCAGGCCACGG GCAGACAATATTCAGAGGCA GGCCACGGGCAGACAGG	GTCCCATCATGATGTTACCTCGGT GTGAATATATCATGATGTTACCTC GGTGTGG
hsa-miR-433-3p	GACCCACACCGAGGTAACAT CATGATATATTCACACCGAGG TAACATCATGATGG	GTCCCTCTCACAATAAATCGCACC CGTGAATATTCTCACAATAAATCG CACCCGTGG

2.10 Luciferase Reporter Assay

Twenty-four hours prior to transfection, 50,000 HEK293 cells were plated per well in 24-well plates. For each well, 150 ng of firefly luciferase construct, 50 ng of *renilla* luciferase construct, and 100 ng of pLKO microRNA expression construct (or pcDNA4 construct) were mixed with 0.6 μ l of GeneJuice (Merck, Germany). OptiMEM (Gibco, UK) was then added to obtain the desired

volume. For each combination of constructs that was tested, transfection was performed in biological triplicate.

Transfected cells were grown for 48 hours, and then harvested for the luciferase assay with the Promega Dual-Luciferase® Reporter Assay System (Promega Corp., USA), according to manufacturer's instructions. Luminescence readings for firefly and *renilla* luciferase (integrated over 10 seconds) were taken with a Tecan F200PRO microplate reader (Tecan Group Ltd., Switzerland). Each sample was measured in technical duplicate, and the average of both readings was used in further data analysis. Background luminescence was determined by measuring luminescence produced by mock-transfected wells, and was subtracted from raw readings.

2.11 Candidate SNP analysis

1. Study Cohort. The cohort was previously reported in the following studies: The SLI Consortium 2002 and 2004, Monaco and the SLI Consortium 2007, Falcaro *et al.* 2008, Newbury *et al.* 2009, Nudel *et al.* 2014a and 2014b. The SLI Consortium (SLIC) recruited 278 nuclear families (which included 297 cases) from five centers across the UK—the Evelina Children's Hospital in London (previously known as the Newcomen Centre at Guy's Hospital), the Cambridge Language and Speech Project (CLASP), the Child Life and Health Department at the University of Edinburgh, the Department of Child Health at the University of Aberdeen, and the Manchester Language Study. All individuals self-reported as white British, and furnished informed consent for their participation in the studies. In addition, local ethics committees for each site gave ethical approval for the studies.

The proband from each family was originally identified on the basis on deficient language scores but normal nonverbal test performance. Probands had expressive and/or receptive language scores at least 1.5 standard deviations (SDs) below their age-appropriate mean (normalized population mean = 100, SD = 15), and with non-verbal intelligence scores (assessed with

the Wechsler Intelligence Scale for Children Perceptual Organization) of more than 77.5, which is within 1.5 SD of their age-appropriate score. Language scores were obtained with the revised version of the Clinical Evaluation of Language Fundamentals (CELF-R) (Semel *et al.* 1992), as was done in previous studies.

Probands did not have a diagnosis of autism or hearing impairment. Given the high degree of co-occurrence of SLI and ADHD or dyslexia, probands were not excluded on the basis of a diagnosis of ADHD or dyslexia alone. Data regarding the presence of hyperactivity, coordination, and reading problems was available for some of the samples, allowing Nudel *et al.* (2014a and 2014b) to estimate that approximately one third of the probands exhibited some evidence of ADHD or developmental coordination disorder, and about half of the probands had reading problems.

Any sibling with an expressive language score (ELS) or a receptive language score (RLS) ≤ 1.5 standard deviations (SD) below the age-appropriate population mean was also classified as a case. The expressive language score measures performance on linguistic tasks related to production and output of spoken and written language, while the receptive language score measures performance on comprehension of spoken and written language (Semel *et al.* 1992). None of the cases were diagnosed with autism or hearing impairment. DNA and phenotypic data were collected from all children regardless of SLI diagnosis, while DNA only was collected from all available parents. DNA was collected via whole blood samples (London and Edinburgh) or via buccal swab samples (Cambridge, Aberdeen and Manchester).

2. Genotyping and Imputation. Samples underwent genotyping, quality control, and imputation as described in Nudel *et al.* 2014b. Briefly, samples were genotyped using Illumina Omni-Express v12.1 genome-wide arrays (Illumina Inc., USA), which covered a total of 733,202 SNPs. Two-stage quality control was performed as described in Nudel *et al.* 2014b. The genotypes of SNPs not directly captured by the arrays were imputed against

1000 Genomes data, resulting in an additional 5,276,564 SNPs. (Performed by D. Newbury and R. Nudel.)

3. Gene-dropping Simulations and Regional Association Plots. Gene-dropping simulations were performed in MERLIN (Abecasis *et al.*, 2002) using the "--simulate" flag, with 5000 reruns across 10 random seeds (500 runs per random seed), to determine the empirical false-discovery rate of the observed association p-value for association of rs72727201, which was significantly associated with nonword repetition ability in the cohort. Regional association plots for candidate SNPs were generated using LocusZoom (Pruim *et al.* 2010).

2.12 Exome Analysis

1. Variant Identification From Exome Data. Forty-eight SLI probands (from a larger family-based cohort consisting of 297 probands from 278 nuclear families, see Nudel *et al.* 2014a, 2014b) were selected for whole-exome capture with the Agilent SureSelect v2 platform, and then underwent high-throughput sequencing on the SOLiD 5500xl System (Chen *et al.*, in prep.). Variant calling and quality control was performed with Genome Analysis Tool Kit (GATK version 2.0, DePristo *et al.* 2011) by S. Chen, with default parameters.

The genomic coordinates of predicted microRNA binding sites were obtained from TargetScan 2.0 (Lewis *et al.* 2003). Exome-derived variants found within these binding sites were extracted with BEDtools-intersectBed. The by-base coverage of exome sequencing was obtained using SAMtools-mpileup. A number of custom Perl scripts were written to calculate coverage information of various kinds. This analysis was conceived by J. Ho and S. Chen, and was performed by S. Chen. The final list of variants lying in seed regions was further filtered by J. Ho on the following criteria: average read depth of ≥ 10 , and with at least 20% of total reads per individual called as variant (Meynert *et*

al. 2014). These criteria were empirically observed by S. Chen to produce a read quality

2. Variant Validation. Primers were designed to amplify 500bp of genomic DNA flanking the variant of interest. Fragments of interest were amplified from proband and family member DNA samples with the Expand High Fidelity PCR System (Roche, Germany), according to manufacturer's instructions, using primers listed in Table 2.5 at a final concentration of 300 nM. The fragments were recovered with the Wizard® SV Gel and PCR Clean-Up System (Promega, USA) according to manufacturer's instructions. The presence of variants in the fragments was then determined with bidirectional Sanger sequencing.

Table 2.5. Oligos Used to Amplify and Sequence 3' UTRs of Interest.

3' UTR	Forward primer (5'→ 3')	Reverse primer (5'→ 3')
<i>BTN2A1</i>	ACAGCCATGTAGACAAGCCC	TGGATGTGCGCTCTGGAAAT
<i>CENPJ</i>	GCCATTTCTAACCAATAAACA CACCC	GTGAAAGAAGGGTCCCTTTTA
<i>MTMR</i>	TCTGCTAGGTAGGGACACCC	TTTCTTCCCTGCTCACGGTC

3. Conservation Analysis. The evolutionary conservation of 3' UTR variants of interest were evaluated with rejected substitution scores produced by the Genomic Evolutionary Rate Profiling (GERP) algorithm (Davydov *et al.* 2010, Goode *et al.* 2010). Briefly, the difference between the expected number of substitutions at a nucleotide of interest, and the actual observed number of substitutions at that basepair is computed, under an assumption of neutral evolution in 36 mammalian genomes, giving rise to a "GERP score". GERP scores were downloaded for the hg19 version of the human genome from the UCSC Genome Browser (<http://genome.ucsc.edu/>, Karolchick *et al.* 2014).

4. RNA Secondary Structure Determination. To visualise the impact of the variants identified on the secondary structure of their 3' UTRs, the RNAfold algorithm with default settings (Hofacker and Stadler 2006) was run in Geneious 7.1 (Biomatters Ltd., New Zealand). To visualise the secondary structures of microRNA-3'UTR hybrids, as well as to obtain predicted ΔG values of duplex formation, the STarMiR algorithm (Liu *et al.* 2003, available online at <http://sfold.wadsworth.org/cgi-bin/starmirtest2.pl>) was used.

Chapter 3: Regulation of MicroRNAs by Foxp2/FOXP2 – A Microarray Expression Analysis

3.1 Introduction

3.1.1 MicroRNA Biogenesis In Animals

A single microRNA is able to regulate several hundred gene transcripts. Thus, microRNA biogenesis is a tightly controlled process in the cell, involving several different proteins and biochemical steps, and is summarised below and in Figure 3.1.

MicroRNA biogenesis begins in the nucleus, where microRNA genes are transcribed by RNA polymerase II as primary microRNA (pri-miR) transcripts, in the same fashion as protein-coding genes (Lee *et al.* 2004). As such, they have a 5' methyl-guanylate cap and a 3' poly(A) tail (Cai *et al.* 2004). However, microRNA genes that lie near Alu sequences are instead transcribed by RNA polymerase III (Borchert *et al.* 2006). A pri-miR transcript can vary in length from a few hundred nucleotides to several kilobases (Saini *et al.* 2008), and can be composed of one or more precursor microRNA (pre-miR) hairpin structures (Lee *et al.* 2002), which are about 65 nucleotides in length. An example of a pri-miR transcript with multiple pre-miR hairpins is the pri-miR 17/18a/19a transcript (O'Donnell *et al.* 2005).

The pri-miR transcript is then cleaved into pre-miR hairpins by a protein complex, which consists of the enzymes Drosha and DiGeorge Syndrome Critical Region 8, or DGCR8. (Lee *et al.* 2003, Gregory *et al.* 2004, Han *et al.* 2006). Both enzymes are necessary for the production of pri-miR hairpins (Gregory *et al.* 2004), as only DGCR8, but not Drosha, is able to bind to pri-miRs (Yeom *et al.* 2005, Han *et al.* 2006). It is thought that DGCR8 acts as a

“molecular anchor” (Han *et al.* 2006) to properly orient Drosha with the appropriate cleavage site in the pri-miR, allowing release of the pre-miR.

Some microRNAs that occupy an entire intron of a protein-coding host gene are termed mirtrons (Okamura *et al.* 2006), and are transcribed along with their host gene. Release of the pre-miR hairpin from the host gene transcript is achieved by the activity of either Drosha (Kim and Kim 2007) or by splicing and de-branching proteins hairpins (Okamura *et al.* 2006, Ruby *et al.* 2007), highlighting the complexity of microRNA biogenesis.

Exportin-5 (a member of the karyopherin protein family that is involved in transporting molecules between the cytoplasm and the nucleus) then exports the pre-microRNA hairpins out of the nucleus and into the cytoplasm (Yi *et al.* 2003). This process requires the energy-releasing GTPase activity of Ran to shuttle the hairpins across the nuclear envelope (Bohnsack *et al.* 2004, Lund *et al.* 2004). Once in the cytoplasm, the pre-miR is cleaved by the RNase III enzyme Dicer (Hutvagner *et al.* 2001, Ketting *et al.* 2001). The hairpin loop is removed, producing a microRNA duplex, often with imperfect Watson-Crick base-pairing, that is about 22 nucleotides in length (Lund and Dhalberg 2005).

While either strand of the microRNA duplex can theoretically function as a microRNA, only one strand is preferentially loaded into the RNA-induced silencing complex (RISC) (Khvorova *et al.* 2002, Schwarz *et al.* 2003, Matranga *et al.* 2005). This strand is known as the mature guide strand. The other strand comprising the duplex is the passenger strand. The RISC is composed of Dicer, the trans-activation response element RNA binding protein (TRBP), Argonaut 2 (Ago2), and Fragile-X Mental Retardation protein (FMRP) (Chendrimada *et al.* 2005, Gregory *et al.* 2005). TRBP is thought to stabilise the Dicer-microRNA complex (Haase *et al.* 2005), and recruits Ago2 to the complex (Chendrimada *et al.* 2005).

The “catalytic centre” of the RISC is Ago2 (Gregory *et al.* 2005) as it performs the key functions involved in microRNA-mediated gene repression: strand

selection and target mRNA identification. Target mRNA identification occurs via complementary base-pairing between the target 3' UTR and a 7-nucleotide region in the mature microRNA, known as the seed region. It has been hypothesised that the RISC-microRNA complex may act in parallel with RNA processing machinery during mRNA transcription to identify target sites in the 3' UTR (Brown *et al.* 2005), but exact details remain to be elucidated.

When a microRNA binds to a site in the 3' UTR of a gene transcript, repression by the RISC can occur in three ways (Djuranovic *et al.* 2012): inhibition of translation (Olsen and Ambros 1999, Petersen *et al.* 2005), deadenylation of the mRNA transcript (Wu *et al.* 2006), and degradation of the mRNA target via Ago2-mediated cleavage (Liu *et al.* 2004, Meister *et al.* 2004, Bagga *et al.* 2005).

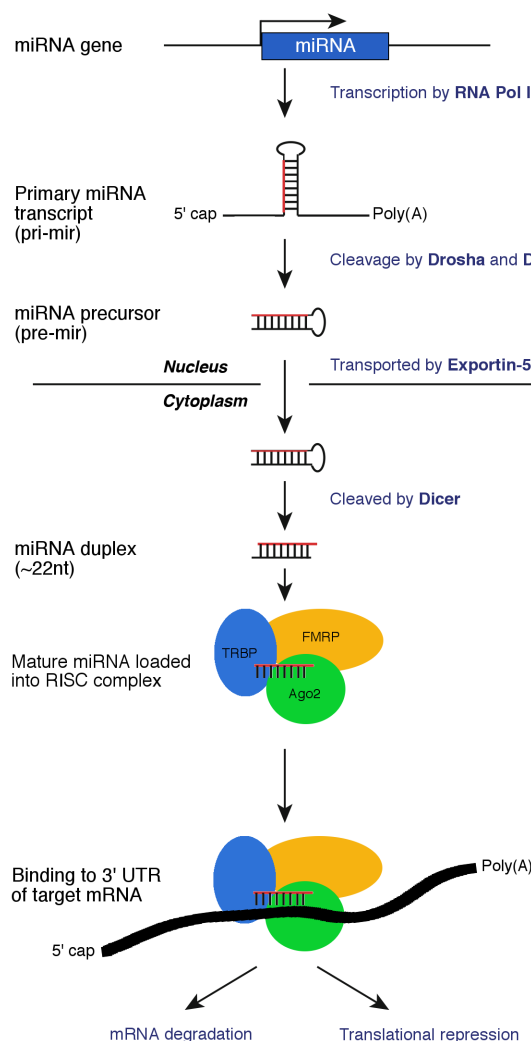


Figure 3.1. MicroRNA Biogenesis.

Shown here are the steps involved in producing a mature microRNA in a cell. Details are found in the text. Figure modified after Li and Zhang 2013.

3.1.2 Identifying Promoters of MicroRNA Genes

Intragenic microRNAs lie within a host gene (either as an intron or an exon) and in the same orientation. Intergenic microRNA genes do not overlap with annotated mRNA transcripts, and operate as independent “transcription units” (Rodriguez *et al.* 2004). Much like canonical gene transcripts generated by RNA Pol II, they possess a transcription start site (TSS) and a poly(A) tail. It is thought that intragenic microRNAs generally undergo the same transcriptional regulation as their host genes, while intergenic microRNAs have their own promoters and transcriptional start sites.

Intergenic microRNAs are generally co-transcribed with the host gene, and the pri-miR transcript is identical to the host gene mRNA transcript. Early studies on intragenic microRNA expression found they often had similar expression patterns with their host gene transcripts (Rodriguez *et al.* 2004, Baskerville and Bartel, 2005). Further work, however, benefitting from a growing body of annotated microRNA transcripts in the mammalian genome, found that intragenic microRNAs could have different expression patterns from their host genes. The host gene and miR might have independent promoters (Hayashita *et al.* 2005, Woods *et al.* 2007, Ozsolak *et al.* 2008), or the host gene transcript might undergo alternative splicing (Bell *et al.* 2010).

An obvious starting point for studying the promoter of a microRNA gene is to first identify the sequence of its pri-miR transcript. However, pri-miR transcripts are difficult to detect via traditional methods because of rapid co-transcriptional processing, leading to a short half-life (Lee *et al.* 2002, Cai *et al.* 2004). Therefore, identification of pri-miR transcripts and their TSS has relied upon genome-wide signatures of histone modification, nucleosome occupancy, and RNA Polymerase II/III occupancy (Marson *et al.* 2008, Ozsolak *et al.* 2008, Corcoran *et al.* 2009). These studies were motivated by the fact that for protein-coding genes, specific modifications (such as histone modification and nucleosome depletion) are indicators of transcriptionally active promoters (Guenther *et al.* 2007, Ozsolak *et al.* 2007).

The integration of these chromatin signatures with other genome-wide annotations (such as EST sequences and CpG islands) have enabled the identification of TSS, transcript coordinates, and putative promoter regions for a range of human and mouse microRNAs. The ability of chromatin signatures that designate promoter regions of protein-coding genes to also identify microRNA promoters indicates that microRNAs and protein-coding genes are regulated by similar transcriptional mechanisms. One noteworthy finding emerging from these studies of microRNA promoters was that a sizable proportion (30% – 35%) of intronic microRNAs have promoters independent of their host gene promoters (Ozsalak *et al.* 2008, Monteys *et al.* 2010), and might account for the discordant expression patterns between some intronic microRNAs and their host genes (Rodriguez *et al.* 2004, Baskerville and Bartel, 2005, Liang *et al.* 2007).

More recent work has begun to study the transcriptional regulation of microRNAs in the context of three-dimensional chromatin structure (Chen *et al.* 2014). By integrating genome-wide signatures of methylation and histone modifications with long-range chromatin interactions, Chen and colleagues found that microRNA genes are organised (along with protein-coding genes) into distinct topological domains with correlated expression patterns. This showed that higher-order genomic and chromatin structure can give insight into the regulation and function of microRNA genes.

An important caveat here is that these studies can only identify promoters that are active in the cell-line or tissue being studied. The promoters of microRNA genes that are not expressed are unlikely to possess the genomic signatures of transcriptional activity, and will escape identification. Thus, Monteys and colleagues used a computational approach to predict RNA Pol II binding sites within 5kb of intronic microRNAs (Monteys *et al.* 2010). They found and validated 6 novel microRNA promoters that previous studies ChIP-based studies did not identify. Thus, they highlighted the need for promoter identification in different cell-types and at different developmental stages.

3.1.3 Identifying Transcription Factors That Regulate MicroRNA Genes

Given that microRNA genes share the same features as protein-coding genes, it is no surprise that transcription factors (TFs) regulate the levels of both these genes. For example, the oncogenic TF c-Myc binds to and up-regulates the miR-17-92 cluster (O'Donnell *et al.* 2005), while repressing several other microRNAs such as the miR-30d/miR30b cluster, miR-34a, and miR-146a (Chang *et al.* 2008). Another oncogenic factor, SNAIL2, has been shown to repress miR-203 (Ding *et al.* 2013). The tumour suppressor p53 has also been shown to up-regulate the miR-34 family (Feng *et al.* 2011), miR-145 (Sachdeva *et al.* 2009), and miR-192/215 (Braun *et al.* 2008, Georges *et al.* 2008). The tissue-specific epithelial TF HNF-1 α was found to up-regulate miR-194 intestinal epithelial cells by binding to its promoter element (Hino *et al.* 2008), showing that tissue-specific microRNA expression patterns can be the result of transcription factor regulation.

In a search for genes regulated by Foxp2, Vernes and colleagues (Vernes *et al.* 2011) performed ChIP-chip on 3 sets of pooled wildtype and litter-matched mutant S321X brains (a total of 6 chromatin samples) from E16 mice. As previously mentioned in Chapter 1, Foxp2.S321X brains lack functional Foxp2 (Groszer *et al.* 2008) and represent a null mutant. Each ChIP sample was pooled from 5-6 brains of the same genotype, and were hybridised to a microarray platform (Agilent Mouse Promoter ChIP-on-Chip Microarray Set 244K, G4490A) with probes for the promoter regions of approximately 17,000 mouse genes, including those for 348 microRNAs. Each promoter region was covered by about 25 probes, spaced about 200bp apart, spanning 5.5kb upstream and 2.5kb downstream of the TSS, for a total of ~8kb of tiled coverage.

A probe score for each probe was computed as the mean normalized log₂ fold enrichment score (ChIP sample versus input sample) of the three replicate samples of each genotype. This probe score was then used to compute a

window-adjusted score for each probe, which was defined as the relative median value of a probe and both its nearest upstream and downstream neighbours within a ± 500 bp window. Analysis of the window-adjusted scores from the S321X null samples allowed the empirical derivation of a significance threshold that excluded 99% of the probes from the S321X data (window-adjusted score ≥ 0.974). By applying this empirical threshold the wildtype data, as well as excluding genes with probes that exceeded the empirical threshold score in the S321X null samples, a list of 1,164 genes that were consistently enriched by Foxp2-ChIP across 3 wildtype replicates was obtained. This list included 24 microRNAs, such as mmu-mir-9 and mmu-let-7d (Table 3.1). Vernes *et al.* reported a total of 25 microRNAs, including mmu-mir-469, but this microRNA is no longer considered a canonical annotated microRNA, as several deep sequencing reads from the mir-469 locus reads mapped to the loop of the putative hairpin, a pattern that is inconsistent with Dicer processing (Chiang *et al.* 2010).

To obtain a high-confidence short-list with stricter thresholds, Vernes *et al.* retained only promoters that had at least one probe with a window-adjusted score ≥ 1.5 . Two hundred and sixty-four (264) genes met this criterion, including 3 microRNAs (mmu-let-7a-1, mmu-mir-137, and mmu-mir-19b-1). Vernes *et al.* were able to validate a selection of these 24 microRNAs in murine Neuro2A neural cell-lines that stably expressed Foxp2. Several of these targeted microRNAs are important to neural development and function, such as mmu-mir-9, mmu-mir-137, and mmu-mir-124. This was the first evidence that Foxp2 regulates the expression of microRNAs, and implicated microRNA expression and function in speech and language development.

Table 3.1. Foxp2 Occupies the Promoters of 24 MicroRNAs Precursors in Developing Murine Brains.

MicroRNA	Chr	Start	End	Strand	Shortlist (Window Score ≥ 1.5)
mmu-mir-30a	1	23272269	23272339	+	
mmu-mir-205	1	193507463	193507530	-	
mmu-mir-124-3	2	180894040	180894107	+	
mmu-mir-137	3	118433857	118433929	+	✓
mmu-mir-29a	6	31062660	31062747	-	
mmu-mir-34b	9	51103562	51103645	-	
mmu-let-7g	9	106178840	106178927	+	
mmu-mir-216	11	28757012	28757083	+	
mmu-mir-217	11	28763728	28763835	+	
mmu-mir-22	11	75463716	75463810	+	
mmu-mir-10a	11	96317165	96317274	+	
mmu-mir-342	12	108658620	108658718	+	
mmu-mir-382	12	109733771	109733846	+	
mmu-let-7a-1	13	48538179	48538272	-	✓
mmu-mir-9-2	13	83738814	83738885	+	
mmu-mir-320	14	70443510	70443591	+	
mmu-mir-17	14	115043671	115043754	+	
mmu-mir-19b-1	14	115044305	115044391	+	✓
mmu-mir-30d	15	68341208	68341289	-	
mmu-mir-125b-2	16	77646273	77646343	+	
mmu-mir-201	X	67988096	67988161	-	
mmu-mir-325	X	105379082	105379179	-	
mmu-mir-448	X	147158210	147158321	+	
mmu-let-7f-2	X	151912346	151912428	+	

This table shows the 24 microRNA precursors whose promoter regions were significantly enriched (window score ≥ 0.974 , see text for details) in Foxp2-ChIP samples from wildtype brains, relative to those from S321X null mutant brains (Vernes *et al.* 2011). Three microRNA precursors (mir-137, let-7a-1, and mir-19b) survived a stricter cut-off (window score ≥ 1.5), and were included in a high-confidence shortlist. The genome coordinates correspond to the Genome Reference Consortium Mouse Build 38 (GRCm38/mm10) assembly.

3.1.4 Transcription Factors and MicroRNA Genes: Feedback Loops In Gene Regulatory Networks

Besides regulating the levels of microRNAs, TFs can themselves be regulated by microRNAs. A microRNA and a TF might regulate each other (feedback loop), or both the microRNA and the TF might regulate the same target gene

(feedforward loop). The types of motifs, along with relevant examples, are summarised in Figure 3.2.

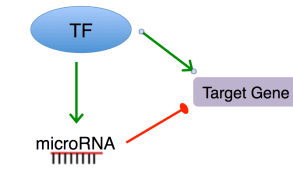
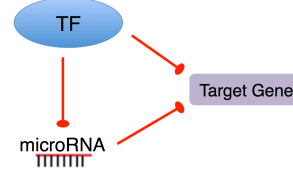
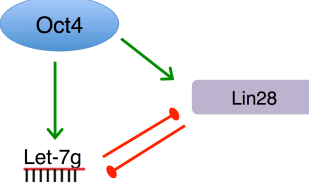
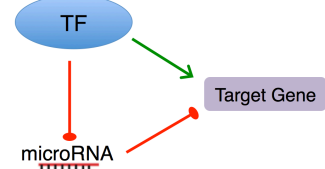
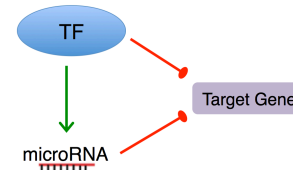
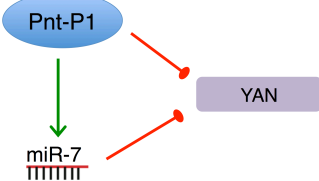
Type of Motif		Scheme	Example	Ref.
Negative Feedback Loop	Unilateral			Kim <i>et al.</i> 2007
	Reciprocal			Burk <i>et al.</i> 2008
Feedforward Loop	Incoherent (Type I)	 		Marson <i>et al.</i> 2008
	Coherent (Type II)	 		Li <i>et al.</i> 2008

Figure 3.2. Network motifs of regulatory loops between transcription factors and microRNAs.

Shown here are four types of loops that transcription factors (TFs) and microRNAs can participate in. The sharp green arrow indicates up-regulation, while the blunt red arrow indicates down-regulation.

A negative feedback loop occurs when a TF is repressed by a microRNA that it regulates. Such feedback loops can either be unilateral, where the TF up-regulates the microRNA, such as the feedback loop between PITX3 and miR-

133b in midbrain dopaminergic neurons (Kim *et al.* 2007). Such an interaction is thought to fine-tune the levels of the TF to a steady state, leading to robustness of TF expression and function (Krol *et al.* 2010). Negative feedback loops can also be reciprocal, where the microRNA represses a TF that negatively regulates it. An example is ZEB1 and miR-200 in cancer cells (Burk *et al.* 2008). Such reciprocal negative feedback loops act to maintain expression of the TF.

A feedforward loop (FFL) occurs when a TF and microRNA both regulate a common target gene, and when the microRNA is regulated by the TF as well. FFLs can either be incoherent (also known as a Type I FFL) or coherent (Type II FFL) (Mangan and Alon 2003).

An incoherent FFL occurs when the TF regulates both the microRNA and the target gene in the same fashion (See Figure 3.2). Since the microRNA down-regulates the target gene, there is discordance between the direct effect of the TF on the target gene and the indirect microRNA-mediated effect. An example is observed in embryonic stem cells, where Oct4 up-regulates let-7g and its target Lin28 (Marson *et al.* 2008). Incoherent TF-microRNA FFLs are thought to result in a rapid buffering of the target gene against (potentially destabilising) fluctuations in TF levels (Tsang *et al.* 2007). An incoherent FFL can also produce a ON-OFF pulsed expression of the target gene following sustained activation of the TF (Vera *et al.* 2012), where the target gene level rises following increase in TF levels, but falls as the miR levels rises (in response to the TF). This enables high levels of the TF to be sustained even after the expression spike of the target gene has been generated (Herranz and Cohen 2010).

A coherent FFL is formed when the TF regulates the microRNA in an opposite fashion from the target gene, resulting in concordant effects on the target gene. An example is found in the developing *Drosophila* eye: Pnt-P1 up-regulates miR-7, while also repressing its target Yan (Li *et al.* 2009). Coherent FFLs normally result in mutually exclusive expression patterns

between the microRNA and the target gene, and acts to prevent unwanted, “leaky” expression of the target gene (Hornstein and Shomron 2006).

Another way that TFs can regulate microRNA levels is by acting as competitive endogenous RNAs (abbreviated as ceRNAs, see Salmena *et al.* 2011). The mRNA transcript of a given TF can act as a natural microRNA sponge, limiting the availability of a microRNA in the cytoplasm. This results in less microRNA binding to other target mRNAs, resulting in an indirect *trans* regulation of these target mRNAs by the mRNA transcript. In other words, protein-coding mRNA transcripts can co-regulate each other by competing for microRNA binding. Experimental data on the interplay between ceRNAs and microRNAs is growing (Tay *et al.* 2014), and represents an additional avenue by which microRNA levels are regulated.

Given that Foxp2 regulates both protein-coding genes and microRNAs in the mouse brain (Vernes *et al.* 2011), such regulatory loops are likely to occur in genetic networks regulated by Foxp2. Identifying microRNAs regulated by FOXP2/Foxp2 can reveal novel genetic factors underlying motor learning and speech and language, and shed light on how FOXP2/Foxp2 exerts additional transcriptional control on its protein-coding targets. However, the full repertoire of microRNAs regulated by Foxp2/FOXP2, as well as the secondary downstream effects of microRNAs with Foxp2-dependent expression, remains poorly understood.

3.1.5 Aims of this chapter

I will identify microRNAs regulated by Foxp2 in embryonic mouse brains using whole-genome microRNA expression microarrays. I will determine putative targets of significantly regulated microRNAs, and I will use gene ontology analysis and pathway analysis to discern biological themes associated with these putative targets. I will then perform validation of the microarray findings in well-studied cell-lines that overexpress FOXP2/Foxp2 and the KE mutant.

3.2 Results

3.2.1 Normalization, Processing, and Quality Control of Microarrays

To assess the regulation of microRNAs by *Foxp2*, bilateral ganglionic eminences (GE) were dissected from 6 wildtype and 6 mutant (*Foxp2.S321X*) litter-matched embryonic mice 16 days after fertilization (E16) (Vernes *et al.* 2011). The ganglionic eminences are regions of high *Foxp2* expression (Lai *et al.* 2003) and thus were selected for study of *Foxp2* function. Total RNA was extracted and hybridised onto GeneChip microRNA microarrays (version 1.0; Affymetrix). To enable detection of hybridisation on the array, the RNA samples are first poly(A)-tailed and then labeled with biotin via a ligation reaction. The presence of hybridised probes is then detected with streptavidin-phycoerythrin conjugate.

These microarrays have probes for 613 mouse microRNAs that are annotated in miRBase v11.0. MicroRNAs that were not detected in any of the 12 samples were excluded from further analysis, resulting in a smaller list of 432 microRNAs. This list of microRNAs consisted of microRNAs whose expression domains included brain regions (e.g. *mmu-let-7a-5p* and *mmu-let-7d-5p*), as well as microRNAs exclusively expressed in the brain (e.g. *mmu-miR-124-3p* and *mmu-miR-125b-5p*).

As a quality control step, the raw intensities of the spike-in control probes (added to each RNA sample before poly(A)-tailing) were obtained (Figure 3.3). The raw $\log_2$ intensities for all 5 spike-in controls in all 12 arrays were greater than 9.8, which is the threshold set by the manufacturer for successful RNA labeling and detection (Kaddis *et al.* 2012).

Spike-in controls 2, 23, and 29 are RNA oligos, and high raw intensities of these controls after microarray hybridisation confirm that poly(A) tailing and ligation-mediated biotin labelling was successful. Spike-in control 31 is a poly(A)-tailed RNA oligo, and also serves to confirm that ligation worked. Lastly, spike-in control 36 is a poly(A)-tailed DNA oligo, and confirms ligation and lack of RNases in the RNA sample.

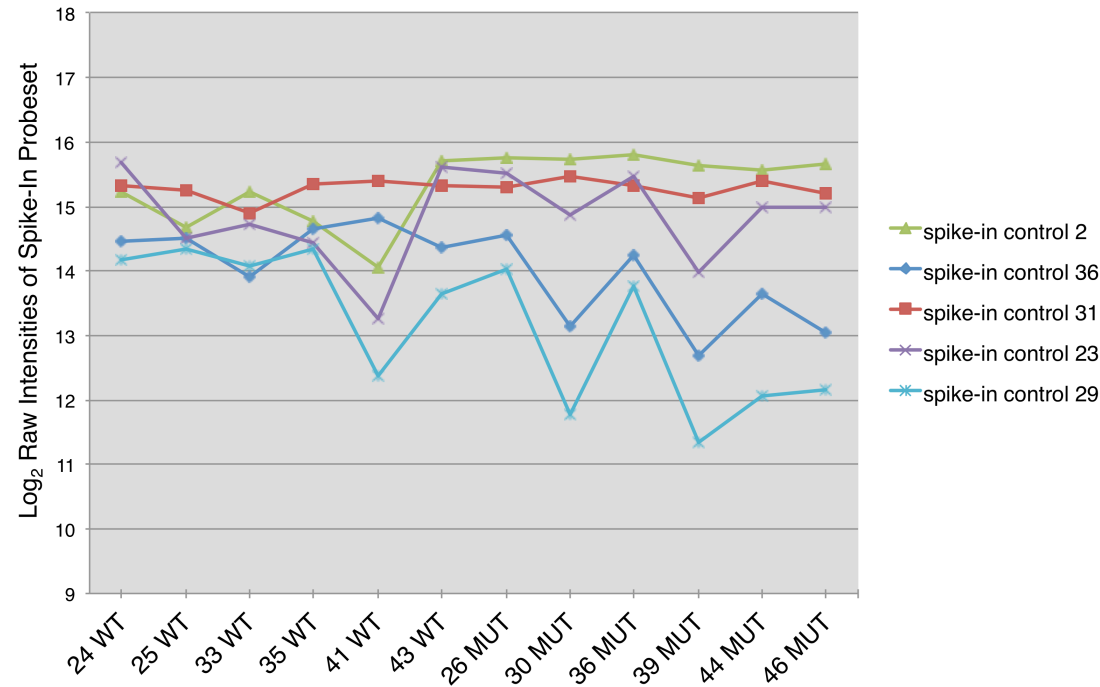


Figure 3.3. Raw Intensities of Spike-In Controls.

Shown here are mean log-transformed intensities of probesets designed against 5 proprietary synthetic spike-in controls. These spike-in controls are added as part of the sample processing workflow as a positive control. Their $\log_2$ intensities beyond the manufacturer-recommended threshold of 9.8, and robust detection across the microarrays, indicates that the sample processing steps are successful.

Next, the probe intensities underwent normalization and median polish summarization according to the Robust Multichip Average method (Irizarry *et al.* 2003). The boxplots of the probe intensity distributions across all 12 arrays before and after normalization were obtained as well (Figure 3.4). All the arrays had similar array-wide probe intensities, lending confidence to the quality of the normalized microarray data.

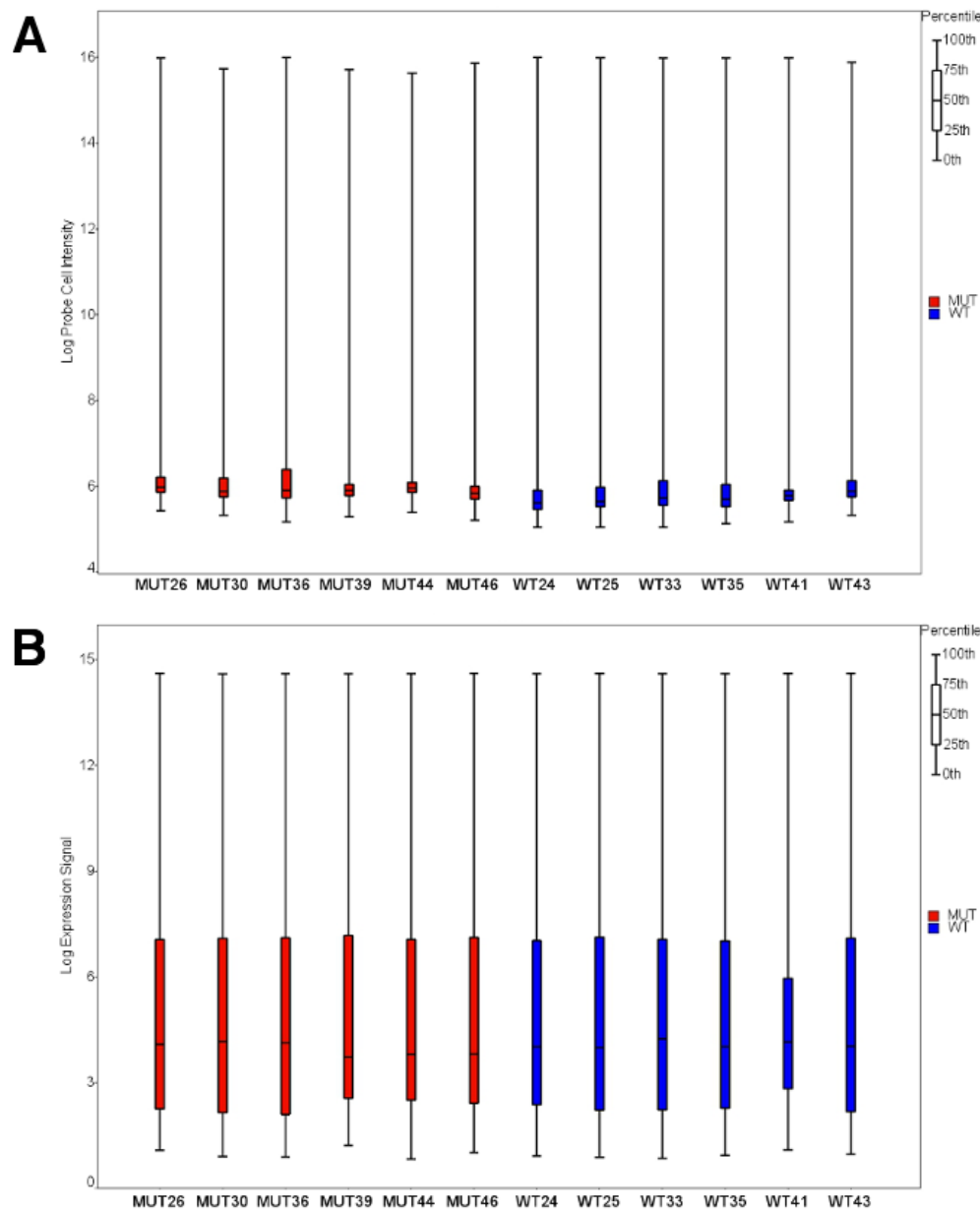


Figure 3.4. Boxplot of Probe Intensities Across 12 MicroRNA Microarrays of Mouse Brain RNA.

Shown here are boxplots of log-transformed probe intensities of the twelve microarrays studied here (A) before and (B) after normalization.

If *Foxp2* status is the main difference between the litter-matched mouse samples, the microRNA expression profiles of these 12 mouse brains should cluster according to the presence or absence of *Foxp2* in the samples. Unsupervised hierarchical clustering of the 12 mouse arrays produced two clusters according to *Foxp2* mutant status (Figure 3.5), supporting this hypothesis.

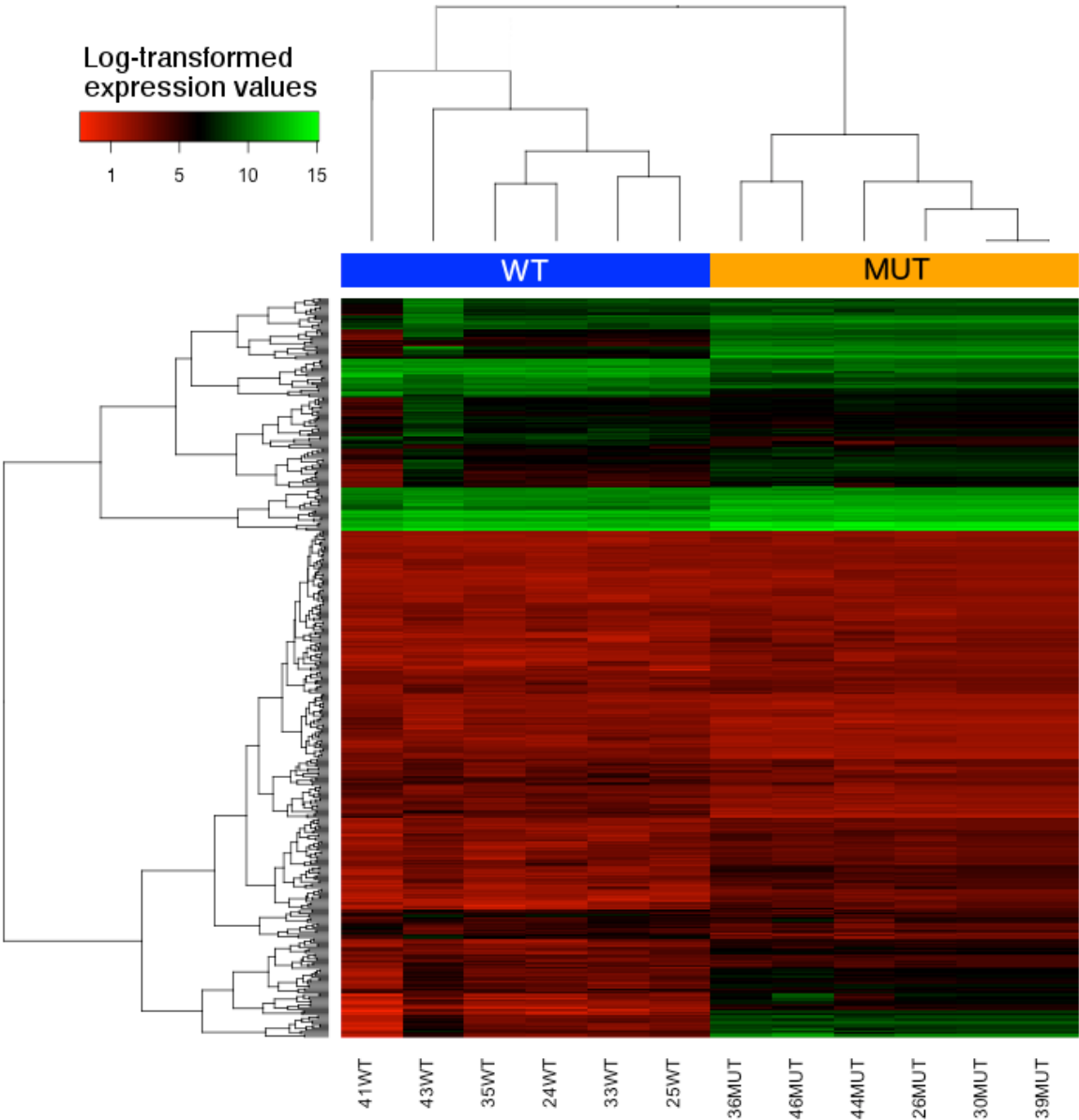


Figure 3.5. Clustered Heatmap of MicroRNA Expression Profiles of Mouse Embryonic Ganglionic Eminences.

Shown here are the microRNA expression profiles of ganglionic eminences dissected from E16 mouse brains. The heatmap displays the log-transformed expression values for 432 probes expressed in either wildtype brains, or litter-matched *Foxp2*.S321X mutant brains that do not express *Foxp2*.

3.2.2 Eight MicroRNAs are Differentially Expressed in S321X.Foxp2 Embryonic Mutant Mouse Brain

To identify differentially expressed genes, an analysis of variance (ANOVA) was performed. Eight microRNAs were differentially expressed (unadjusted p-value < 0.01) in the GEs of mutant mice versus that of wildtype mice (Table 3.2). A heat map showing expression changes across the samples (Figure 3.6) was generated through unsupervised cluster analysis. Five microRNAs were up-regulated in mutant brains relative to wildtype brains, while three microRNAs were down-regulated.

Table 3.2. MicroRNAs Differentially Regulated Between the Ganglionic Eminences of Wildtype and Mutant Foxp2.S321X Mice.

miR	log ₂ Fold Change	Average Normalized Expression	Unadjusted p-value	Promoter Region captured on ChIP chip platform?	ChIP-chip target of Foxp2?
mmu-miR-30c-5p	4.35	7.85	2.24E-04	✓	
mmu-miR-25-3p	5.27	7.68	3.04E-03	✓	
mmu-miR-1195	-3.34	10.64	4.26E-03		
mmu-miR-20a-5p	6.21	9.36	4.84E-03		
mmu-miR-615-5p	-2.95	1.98	5.26E-03	✓	
mmu-miR-125b-2-3p	4.45	3.24	7.88E-03	✓	✓
mmu-miR-181b-5p	2.64	11.01	8.07E-03	✓	
mmu-miR-320-3p	-2.17	11.05	8.53E-03	✓	✓

This table shows the log₂ fold changes of the microRNAs in mutant Foxp2.S321X mouse GEs, relative to wildtype GEs. Positive fold changes (where the microRNA is more highly expressed in the mutant) are in green, while negative fold changes are in red. Listed alongside these are the average log₂ expression values of the microRNA in the samples, as well as the associated unadjusted p-values.

To identify which of these 8 microRNAs are direct targets of Foxp2, I looked at which of these microRNA promoter regions were interrogated in the ChIP-chip study performed in Vernes *et al.* 2011. The ChIP-chip platform used interrogated the promoters of all these microRNAs except those of mmu-miR-1195 and mmu-miR-20a-5p. Thus, it is uncertain if Foxp2 regulates these 2 microRNAs by directly binding to their promoters. Only the promoters of mmu-miR-125b-2-3p and mmu-miR-320-3p, however, were bound by Foxp2 in the

embryonic mouse brain. (Vernes *et al.* 2011). Therefore, miR-30c-5p, miR-25-3p, miR-615-5p, and miR-181b-5p are either indirectly regulated by Foxp2, or directly regulated by another binding site not captured through ChIP-chip.

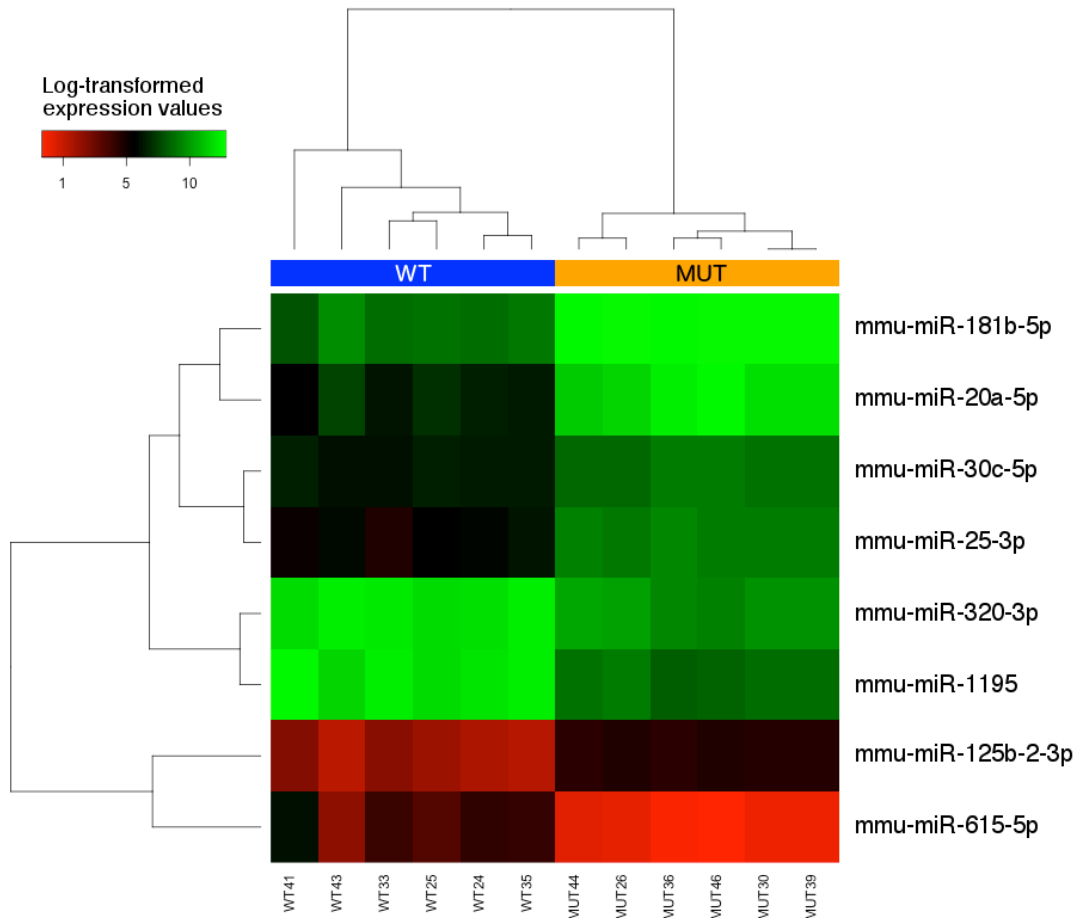


Figure 3.6. MicroRNAs Differentially Regulated Between the Ganglionic Eminences of Wildtype and Mutant STOP Mice.

Shown here are all microRNAs with significant differences (unadjusted p-value < 0.01) in levels between wildtype and mutant Foxp2.S321X mouse ganglionic eminences (GEs). The heatmap displays the log-transformed expression values of the microRNAs, along with hierarchal clustering of the samples.

In addition to mir-125b-2 and mmu-miR-320, the Vernes *et al.* 2011 ChIP-chip study identified 22 other microRNA precursors whose promoters were directly bound by Foxp2, for a total of 24 microRNA precursors. They give rise to a total of 33 mature microRNAs. If these 24 precursors are indeed direct targets of Foxp2, the expression of the 33 mature microRNAs should be disrupted in the brains of Foxp2.S321X mutant mice. To test this, a heatmap of the relative expression levels of these 33 mature microRNAs was generated using unsupervised clustering (Figure 3.7). Samples from wildtype brains clustered

together and were distinct from the mutant samples, providing evidence for functional relevance of Foxp2 binding to the promoter regions of these microRNA precursors to regulate their expression. Of the 33 mature microRNAs, 26 were up-regulated in mutant mouse brains (Table 3.3), which is in line with previous observations that FoxP2 generally acts as a transcriptional repressor (Vernes *et al.* 2007).

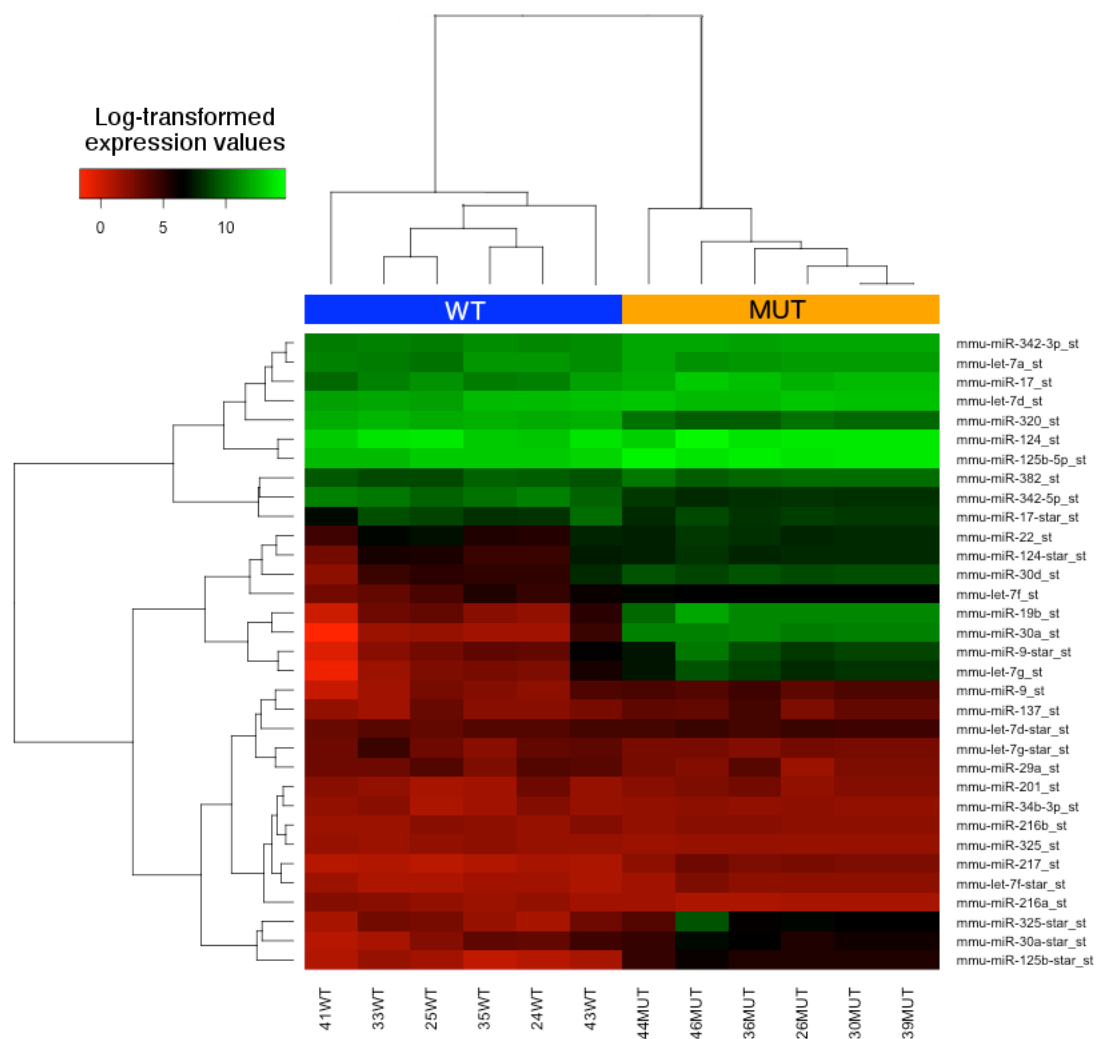


Figure 3.7. Expression and Clustering of MicroRNAs Previously Identified As Foxp2 Targets In Embryonic Mouse Brains.

Shown here is an expression heatmap and hierarchal clustering of 33 microRNAs in wildtype and mutant Foxp2.S321X mouse ganglionic eminences. These 33 microRNAs are produced from precursors that were previously found (Vernes *et al.* 2011) to be transcriptional targets of Foxp2 in the mouse brain.

Table 3.3. Expression and Clustering of MicroRNAs Previously Identified As Foxp2 Targets In Embryonic Mouse Brains.

miR	log ₂ Fold Change	Average Normalized Expression	Unadjusted p-value	Predicted or shown to target Foxp2?
mmu-miR-125b-2-3p	4.45	3.24	7.88E-03	
mmu-miR-320-3p	-2.17	11.05	8.53E-03	
mmu-miR-19b-3p	8.36	6.74	3.60E-02	✓
mmu-miR-342-3p	1.04	11.34	3.63E-02	
mmu-miR-30a-3p	9.19	6.11	3.86E-02	
mmu-miR-217-5p	1.71	1.70	4.00E-02	
mmu-miR-342-5p	-2.08	9.20	4.75E-02	
mmu-miR-125b-5p	1.13	13.50	1.19E-01	
mmu-let-7g-5p	5.95	5.30	1.73E-01	✓
mmu-miR-17-5p	1.69	11.69	1.86E-01	
mmu-miR-9-3p	5.84	5.91	2.02E-01	
mmu-miR-325-5p	4.45	4.32	2.21E-01	
mmu-miR-30d-5p	4.10	7.02	2.44E-01	
mmu-miR-30a-5p	3.31	4.24	2.70E-01	
mmu-let-7f-1-3p	0.74	1.59	2.91E-01	
mmu-miR-124-5p	2.74	6.55	3.29E-01	
mmu-let-7d-3p	0.78	4.08	3.32E-01	
mmu-miR-216a-5p	-0.70	1.52	3.40E-01	
mmu-let-7f-5p	2.13	5.52	3.42E-01	✓
mmu-miR-382-5p	0.70	9.67	3.76E-01	
mmu-miR-22-3p	1.80	7.05	4.30E-01	
mmu-miR-9-5p	1.85	3.05	4.31E-01	✓
mmu-miR-137-3p	1.14	2.81	4.36E-01	
mmu-let-7a-5p	0.64	11.21	4.81E-01	✓
mmu-let-7d-5p	0.50	12.43	5.58E-01	✓
mmu-miR-29a-3p	-0.77	2.94	6.14E-01	
mmu-let-7g-3p	-0.66	2.97	6.58E-01	
mmu-miR-124-3p	0.53	13.74	6.61E-01	
mmu-miR-201-5p	0.46	2.12	7.06E-01	
mmu-miR-325-3p	-0.10	1.71	7.57E-01	
mmu-miR-216b-5p	0.16	1.95	8.02E-01	
mmu-miR-34b-3p	0.19	1.83	8.14E-01	
mmu-miR-17-3p	-0.12	8.51	9.50E-01	

This table shows the log₂ fold changes of the same 33 microRNAs in mutant Foxp2.S321X mouse GEs, relative to wildtype GEs. Listed alongside these are the average log-transformed expression values of the microRNA in the samples, as well as the associated p-values. The table also indicates if the microRNA is predicted (by at least three algorithms), or shown experimentally, to target Foxp2.

3.2.3 Axon Guidance is A Central Theme of Predicted Targets of Top 8 MicroRNAs

To determine the genes targeted by 8 microRNAs significantly regulated by *Foxp2* (see Table 3.2), I queried four microRNA target prediction algorithms (microT-CSD, PicTar2, TargetScan Mouse, and miRanda). The different approaches used by each algorithms result in different sets of predicted targets for a given microRNA. A pooling approach, where only targets predicted by multiple algorithms are considered, increases the sensitivity (correctly predicted true targets) of the predictions (Alexiou *et al.* 2009, Witkos *et al.* 2011). Thus, I only considered target predictions that were shared by at least three algorithms. A total of 2913 microRNA-mRNA interactions (involving 2326 genes) were predicted as targets of the top 8 microRNAs by at least 3 prediction algorithms (Figure 3.8).

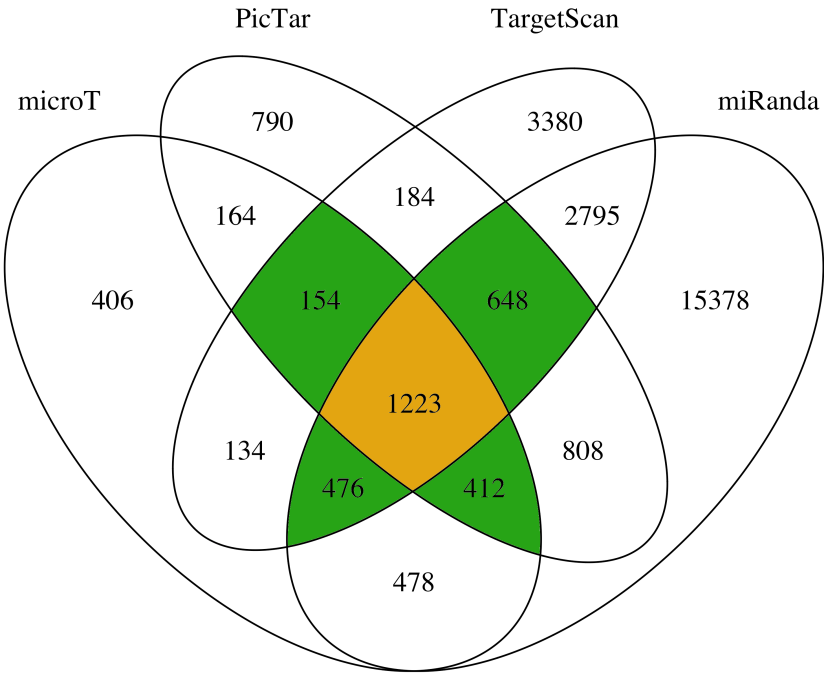


Figure 3.8. Overlap of Predicted Gene Targets Across Four Prediction Algorithms.

Four commonly used microRNA target prediction algorithms were used to obtain genes targeted by the top 8 miRs that were dysregulated by in mutant S321X.*Foxp2* mutant mouse brains. This Venn diagram depicts the extent of overlap across the target predictions. 2326 gene targets (involved in 2913 predicted interactions with the 8 microRNAs) were predicted by at least three algorithms (indicated by the green and tan regions), and were employed in subsequent gene ontology and pathway enrichment analyses.

One of the top molecular pathways over-represented in the targeted genes (Figure 3.9A) was axon guidance (adjusted p-value = 4.83×10^{-30}). Related pathways, such as neurotrophin signalling (adjusted p-value = 4.51×10^{-24}) and regulation of actin cytoskeleton (adjusted p-value = 3.20×10^{-17}) were also in the top 10 enriched pathways. Gene ontology (GO) enrichment analysis (Figure 3.9B) found that these target genes were significantly localised to the synapse (adjusted p-value = 7.21×10^{-6}) and the axon (adjusted p-value = 9.00×10^{-4}). Taken together, this suggests that the formation of neuronal networks in the mouse brain is a significant biological theme of the targets of the top 8 microRNAs.

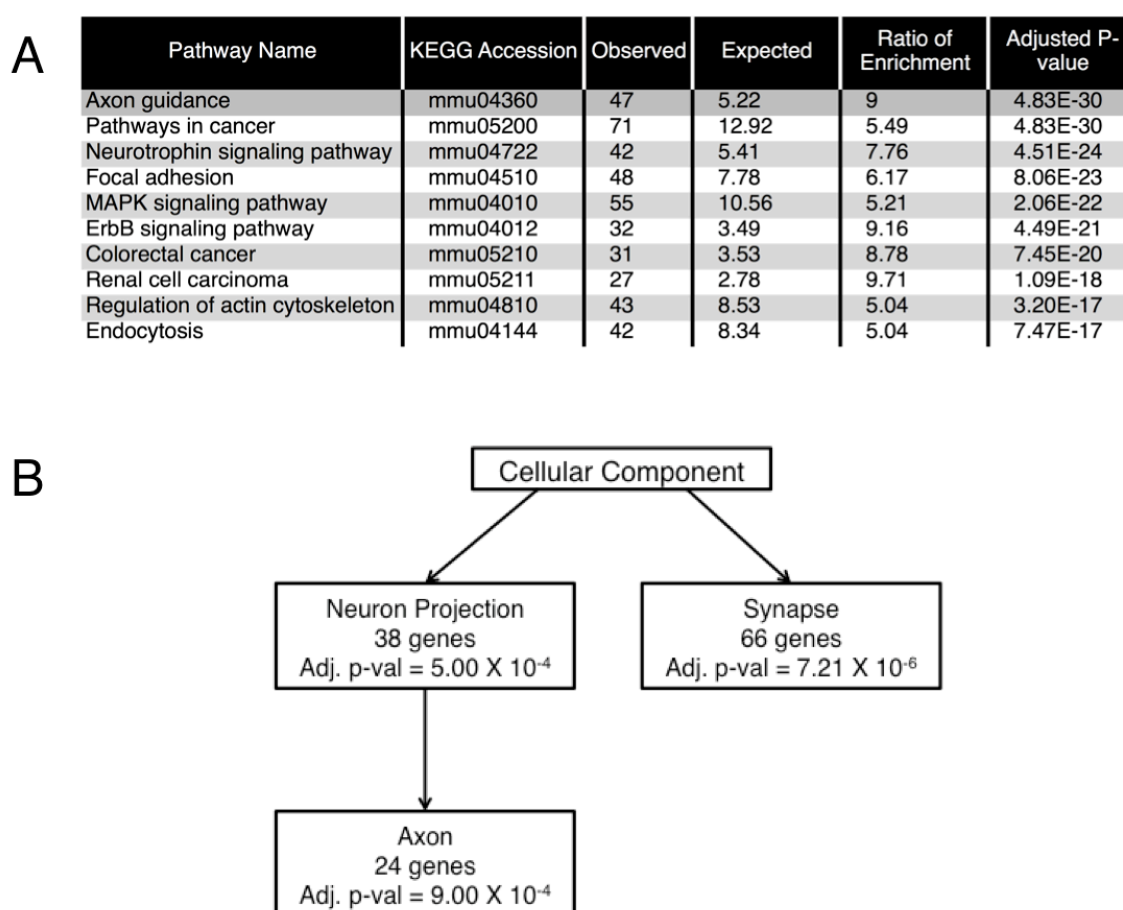


Figure 3.9. Predicted MicroRNA Targets of Foxp2 Are Involved In Axon Guidance, And Are Localised To The Synapse.

The predicted gene targets of the top 8 microRNAs that were differentially expressed in mutant Foxp2.S321X mouse brains were studied in relation to over-representation in **(A)** biological pathways and their **(B)** gene ontology.

GO analysis also revealed a significant enrichment of genes in regulatory processes, such as those involved the regulation of metabolism, as well as gene expression (Table 3.4). Over-enriched molecular functions included protein binding (adjusted p-value = 9.42×10^{-42}) and nucleic-acid binding (adjusted p-value = 1.35×10^{-20}). These results suggest that Foxp2 alters levels of microRNAs that target regulatory proteins, and which may have broad cell-wide effects on gene transcription.

Table 3.4. Gene Ontology (GO) Analysis of Predicted MicroRNA Targets Suggest Foxp2-regulated microRNAs Fine-tune Gene Regulation.

A	Biological Process	GO Accession	Observed	Expected	Enrichment Ratio	Adjusted p-value
	regulation of metabolic process	GO:0019222	545	304.57	1.79	1.72E-48
	regulation of primary metabolic process	GO:0080090	507	276.12	1.84	4.05E-48
	regulation of cellular metabolic process	GO:0031323	520	287.1	1.81	7.36E-48
	regulation of macromolecule metabolic process	GO:0060255	505	278.24	1.81	1.55E-46
	regulation of gene expression	GO:0010468	458	251.16	1.82	6.77E-42

B	Molecular Function	GO Accession	Observed	Expected	Enrichment Ratio	Adjusted p-value
	binding	GO:0005488	1524	1255.46	1.21	2.64E-48
	protein binding	GO:0005515	906	634.38	1.43	9.42E-42
	transcription regulator activity	GO:0030528	263	129.24	2.03	6.68E-30
	transcription factor activity	GO:0003700	175	82.67	2.12	3.02E-21
	DNA binding	GO:0003677	335	203.78	1.64	1.35E-20

Shown here are the top 5 over-represented gene ontologies in **(A)** biological processes and **(B)** molecular functions of the predicted targets of the top 8 differentially regulated microRNAs in mutant Foxp2.S321X mouse brains.

Because the embryonic mouse brain samples were previously assayed for differences in mRNA levels (Vernes *et al.* 2011), I was able to identify possible feedforward loops formed by Foxp2 and the microRNAs it regulates by overlapping the list of direct Foxp2 targets with the list of 2326 predicted microRNA targets. I observed significant overlap (41 genes, hypergeometric p-value = 0.003) between the list of genes with Foxp2-dependent expression

changes in mouse GEs, and the predicted targets of the top 8 microRNAs (Figure 3.10, Supplementary Table 3.1). Four of these 41 genes were also previously identified as direct targets of Foxp2 (Vernes *et al.* 2011): *Ebf2*, *Mapk8ip3*, *Negr1*, and *Tgfbi*.

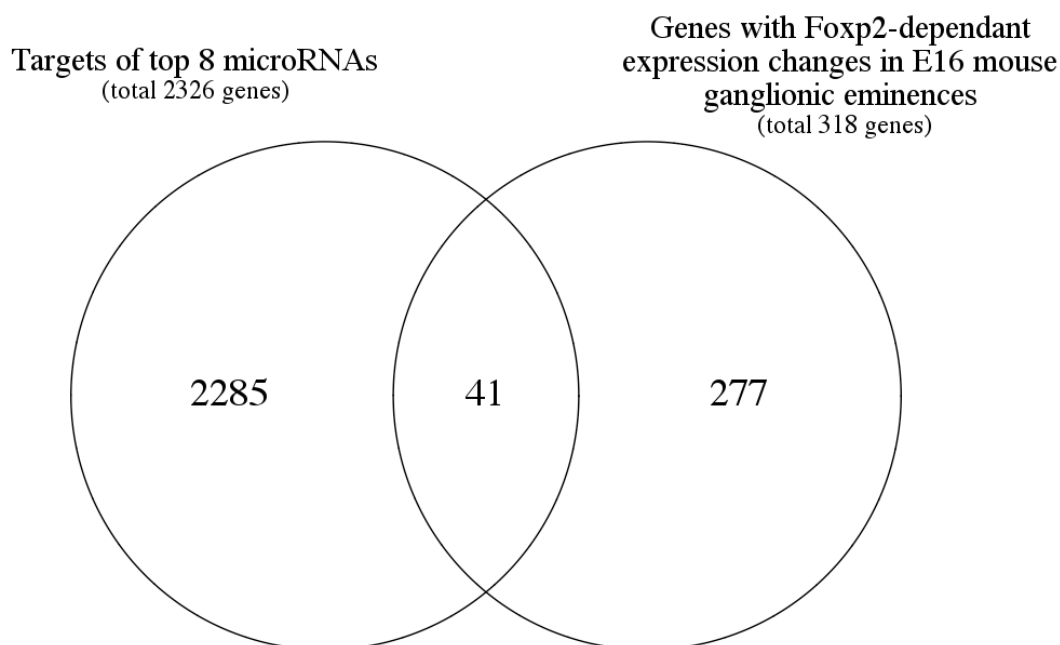


Figure 3.10. Overlap of Predicted Targets of Foxp2-regulated microRNAs with Genes Showing Foxp2-dependent Expression in Mouse Ganglionic Eminences.

This Venn diagram illustrates the extent of overlap between the targets of the top 8 microRNAs showing significant expression differences in Foxp2.S321X mutant mouse brains, and with genes showing Foxp2-dependent expression in E16 mouse ganglionic eminences (Vernes *et al.* 2011). Genes were deemed as targets of Foxp2-regulated microRNAs only if it was predicted by at least three of four target prediction algorithms.

3.2.4 Validation of Microarray Results *in vitro*

To validate *in vitro* the microRNA expression differences observed between the wildtype and Foxp2.S321X mutant mouse brains, I used Neuro2A murine cell lines stably-transfected with a pcDNA3.1 vector carrying either Foxp2 or mutant R552H.Foxp2. I also used SH-SY5Y human cell lines stably-transfected with with a pcDNA3.1 vector carrying either FOXP2 or mutant R553H.FOXP2. The transcript and protein levels of these lines have been assessed (P. Devanna, unpublished data), and FOXP2 levels are comparable with that of R553H.FOXP2 levels in these cells, allowing the direct effects of wildtype FOXP2 to be compared to that of R553H.FOXP2. Comparison of expression level differences

observed *in vivo* with these *in vitro* models will enable prioritization of microRNAs that should undergo further functional investigation. I will also be able to compare the *in vitro* effects of murine Foxp2 on microRNA levels with that of the human version – this will reveal if Foxp2 regulation of microRNAs is conserved in humans.

These neuronal cell lines have been used extensively to study the molecular functions of Foxp2, as well as the mutant KE form (Vernes *et al.* 2007, 2008, and 2011). As noted in the Introduction (Chapter 1), the KE mutation disrupts the DNA-binding properties of FOXP2 (Vernes *et al.* 2007, Nelson *et al.* 2013) and is likely to cause abolish regulation of Foxp2 transcriptional targets. Mice carrying the R552H mutation exhibit abnormal striatal activity and plasticity during motor skill learning (French *et al.* 2012). Thus, if microRNAs that show Foxp2-dependent expression changes in Foxp2.S321X mice brains are also dysregulated by the KE version in a mouse cell-line, they are likely to be strong candidates for microRNAs involved in neural functions associated with Foxp2, such as motor learning.

In addition to the top 8 microRNAs (unadjusted p-value < 0.01) identified here, I also studied *in vitro* expression levels of 4 ChIP-chip targets of Foxp2 (Vernes *et al.* 2011): mmu-let-7d-5p, mmu-miR-9-5p, mmu-miR-137-3p, and mmu-miR-124-3p. These 4 microRNAs were selected because they play important roles in brain function and development. Both miR-9-5p and miR-124-3p are brain-specific (Miska *et al.* 2003, Sempere *et al.* 2004.), while miR-137-3p and let-7d-5p are brain-enriched (Smrt *et al.* 2010, Zhao *et al.* 2013). Let-7d-5p, miR-9-5p, and miR-137-3p promote neuronal differentiation and migration (Zhao *et al.* 2009, Sun *et al.* 2010, Zhao *et al.* 2013). miR-137-3p also reduces spine density and dendrite growth (Smrt *et al.* 2010). MiR-124-3p plays a role in long-term synaptic plasticity (Rajasethupathy *et al.* 2009), and is down regulated in the zebra finch auditory forebrain during song exposure (Gunaratne *et al.* 2010).

Table 3.5 summarises the list of 12 microRNAs I selected for validation in murine and human neuronal cell-lines. Ultimately, *in vitro* expression data was obtained

for 10 of these candidates. Not all these microRNAs were successfully investigated, due to either very low expression (threshold cycle values very near that of negative controls) or unspecific melting curve peaks.

Table 3.5. Overview of qPCR Validation Candidates.

microRNA	Differentially expressed in S321X mutant brains ($p < 0.01$)	ChIP-chip target (Vernes <i>et al.</i> 2011)	qPCR validation in Neuro2A cells	qPCR validation in SH-SY5Y cells
mmu-miR-20a-5p	√		√	√
mmu-miR-30c-5p	√		√	√
mmu-miR-25-3p	√		√	√
mmu-miR-125b-2-3p	√	√	Not Detected	Not Detected
mmu-miR-181b-5p	√		√	Not Detected
mmu-miR-615-5p	√		√	√
mmu-miR-320-3p	√	√	√	Not Detected
mmu-miR-1195	√		Not Detected	Mouse-specific miR; not present in human genome
mmu-let-7d-5p		√	√	√
mmu-miR-9-5p		√	√	√
mmu-miR-137-3p		√	√	√
mmu-miR-124-3p		√	√	Not Detected

Shown here are the top 8 microRNAs differentially expressed in Foxp2.S321X mutant mouse brains, and 4 microRNAs of interest that were previously identified as Foxp2 targets. A tick (√) in the third and fourth columns indicates that the primer used for that microRNA was specific and able to detect expression.

To enable comparison of these *in vitro* effects with the *in vivo* microarray data, where Foxp2 protein expression was abolished in mutant mice, the log₂ fold changes of the 10 candidate microRNAs in wildtype E16 ganglionic eminences were computed relative to Foxp2.S321X (Figure 3.11). The microRNA levels in Neuro2A cells were normalised to mmu-miR-378-3p, as GENorm analysis (Vandesompele *et al.* 2002) showed it exhibited greater expression stability across in Neuro2A cell-lines than the commonly-used normaliser U6 (P. Rodenas-Cuadrado, unpublished data). Log-transformed fold expression changes for N2A.FOX cells (which stably express wildtype Foxp2) and N2A.KE cells (which stable express mutant R552H version with the homologous KE mutation) were computed relative to that of N2A.EMPTY cells, which do not express Foxp2, and are stably transfected with an empty vector.

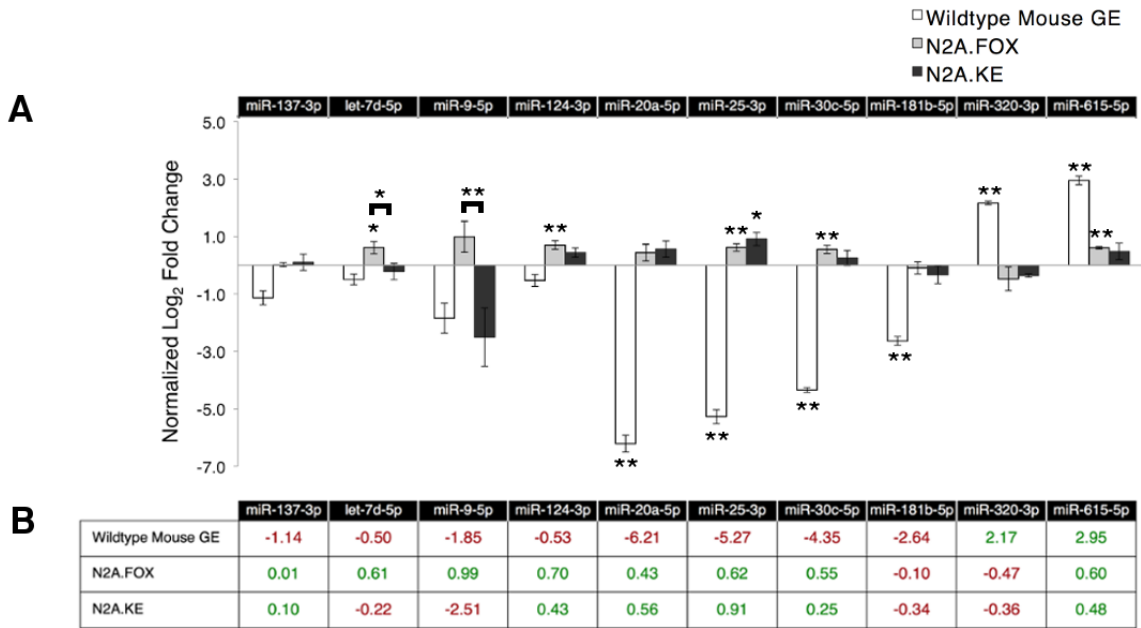


Figure 3.11. qPCR Validation of MicroRNA Targets In Murine Neuro2A Cells.

Shown here are normalised log₂ fold changes of 4 ChIP-chip targets, and 6 of the top 8 microRNAs from the expression array study, in stably transfected Neuro2A cells. **(A)** A histogram of log₂ fold changes of microRNA levels in N2A.FOX and N2A.KE, relative to that in N2A.EMPTY cells. The microarray log₂ fold changes of these microRNAs in wildtype mouse brains (relative to Foxp2.S321X ones) are shown as well. Significance of Tukey's post-hoc test is denoted as follows: P < 0.05 – *, P < 0.01 – **. Error bars denote standard error of the mean (SEM). **(B)** Shown in this table are the log-fold changes. Negative fold changes are in red, while positive ones are in green.

Five microRNAs showed a significant expression difference (Tukey's post-hoc p-value < 0.01) between N2A.FOX and N2A.EMPTY cell-lines. These are mmu-let-7d-5p, mmu-miR-124-3p, mmu-miR-25-3p, mmu-miR-30c-5p, and mmu-miR-615-5p, and all were up-regulated in N2A.FOX cells (Figure 3.10). Only the fold change of mmu-miR-615-5p *in vitro* is concordant with that observed *in vivo*; the other four do not show the same directionality.

The expression levels for mmu-let-7d-5p and mmu-miR-9-5p were significantly different (p-value < 0.01) between N2A.FOX and N2A.KE cell-lines. Both these microRNAs had higher expression in N2A.FOX cells compared to N2A.KE cells.

The importance of FOXP2 to human brain development and function has been demonstrated in several studies. Individuals carrying a non-functional copy of FOXP2 have grey matter abnormalities (Vargha-Khadem *et al.* 1998, Watkins *et al.* 2002b, Belton *et al.* 2003.), as well as aberrant brain activation patterns during language tasks (Vargha-Khadem *et al.* 1998, Liégeois *et al.* 2003, Liégeois *et al.* 2011). It is therefore of great interest to examine if the microRNAs regulated by Foxp2 in the developing murine brain are also regulated by human FOXP2. The expression of these microRNAs was investigated in SH-SY5Y human neuronal cell-lines that expressed wildtype FOXP2, the KE mutant form, or an empty vector (Vernes *et al.* 2007, Vernes *et al.* 2008).

A well-studied feature of SH-SY5Y cells is their ability to shift from a native proliferative state towards a more differentiated state, via exposure to differentiation compounds and/or growth factors (Agholme *et al.* 2010). Differentiated SH-SY5Y cells more closely mirror the morphological and biochemical features of mature neurons (Agholme *et al.* 2010, Lopes *et al.* 2010). Therefore, I employed a differentiation protocol (Encinas *et al.* 2000) where SH-SY5Y cells are treated with retinoic acid (RA) in low serum (2% foetal calf serum) media for 5 days, followed by a 2 successive treatments to brain-derived neurotrophic factor (BDNF) in the absence of RA or serum for 5 days each (Figure 3.12). RNA was harvested from undifferentiated cells, as well as at the

end of each of these treatments, giving four time-points of measurement: Undifferentiated, Day 5, Day 10, and Day 15.

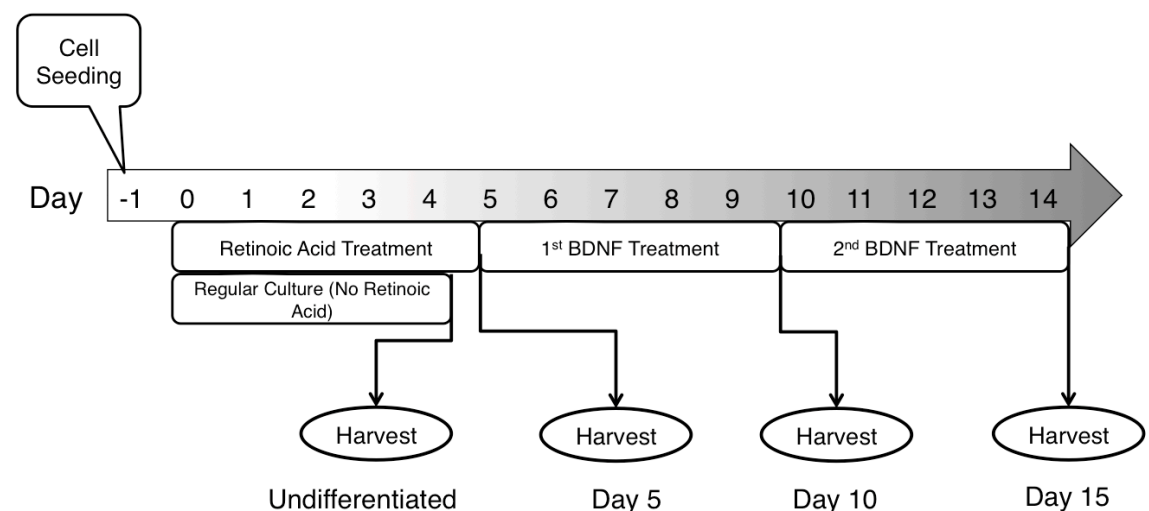


Figure 3.12. Differentiation Protocol Applied to Stably Transfected SH-SY5Y Cells.

This diagram depicts the protocol used to differentiate SH-SY5Y cells. Cells were treated for 5 days with retinoic acid, followed by two successive treatments with BDNF, each lasting 5 days as well. Cells were harvested at the end of each stage of treatment as indicated above, and were used for RNA extraction and subsequent qPCR analysis.

Given the unsuitability of U6 as a normaliser in N2A cells, as well as evidence suggesting that U6 is not an ideal normaliser because of its variability across biological replicates, and its different biogenesis pathway from that of microRNAs (Peltier and Latham 2008, Mestdagh *et al.* 2009, Meyer *et al.* 2010), I looked at the stability of U6 expression in SH-SY5Y cells alongside to a panel of 10 other candidate normalisers.

These 10 microRNAs (Figure 3.13) were selected with the GENorm algorithm (Vandesompele *et al.* 2002) from a separate microRNA expression study in SH-SY5Y cells (Ho and Rodenas-Cuadrado, unpublished) as having a relatively high expression, as well as being the most stable, across all three cell-lines (SH.EMPTY, SH.FOX, and SH.KE). Their expression stability was assessed alongside that of U6 at each of the four harvest timepoints by computing an M value with GENorm, which is the average pairwise variation of a candidate

normalizer gene with all other candidate normalizers. Therefore, the lower the M value of a candidate normalizer, the greater its expression stability across all samples. I found that, at all four timepoints, U6 was not the most stable small RNA in the panel. Therefore, the geometric mean of the two normalization candidates with the lowest GENorm M values was used as a normaliser.

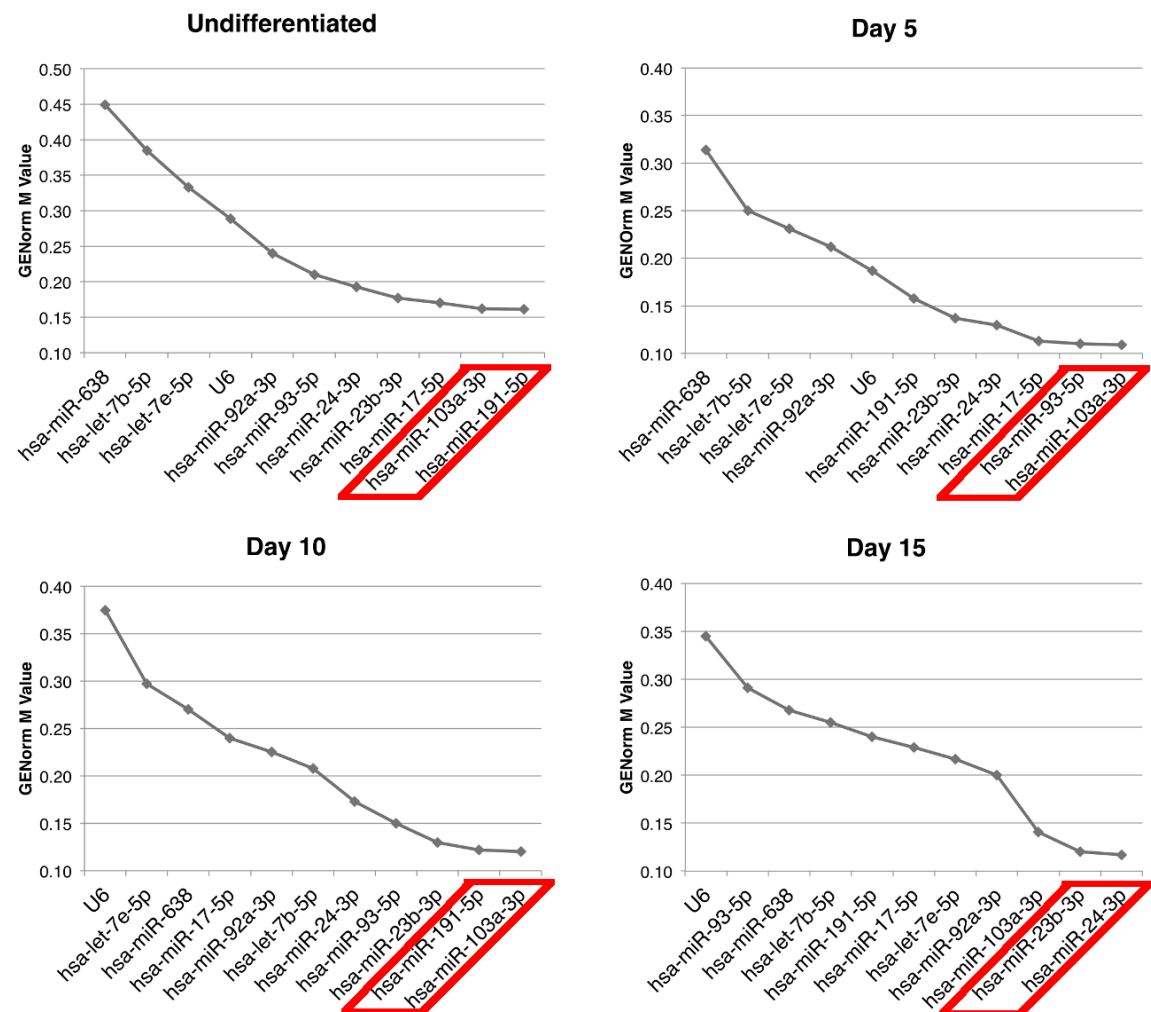


Figure 3.13. Selection of Stable MicroRNA Normalisers For qRT-PCR in SH-SY5Y cells.

Shown here are the M values (as determined using the GENorm algorithm) for a panel of 10 candidate normalisers, alongside U6. Four time points are depicted, one for every point of the differentiation protocol during which cells were harvested for RNA. The two microRNA normalization candidates with the lowest GENorm M values for each time point are boxed in red.

The log expression fold changes of 7 microRNAs of interest, relative to control cell lines (SH.EMPTY) are summarised in Figure 3.14 and in Table 3.6. None of the microRNAs were differentially expressed in undifferentiated SH.FOX cells. In undifferentiated SH.KE cells, three microRNAs – miR-20a-5p, miR-25-3p, and miR-615-5p – were significantly down-regulated relative to SH.EMPTY cells. MiR-25-3p and miR-615-5 were also significantly lower in SH.KE cells when compared to SH.FOX cells.

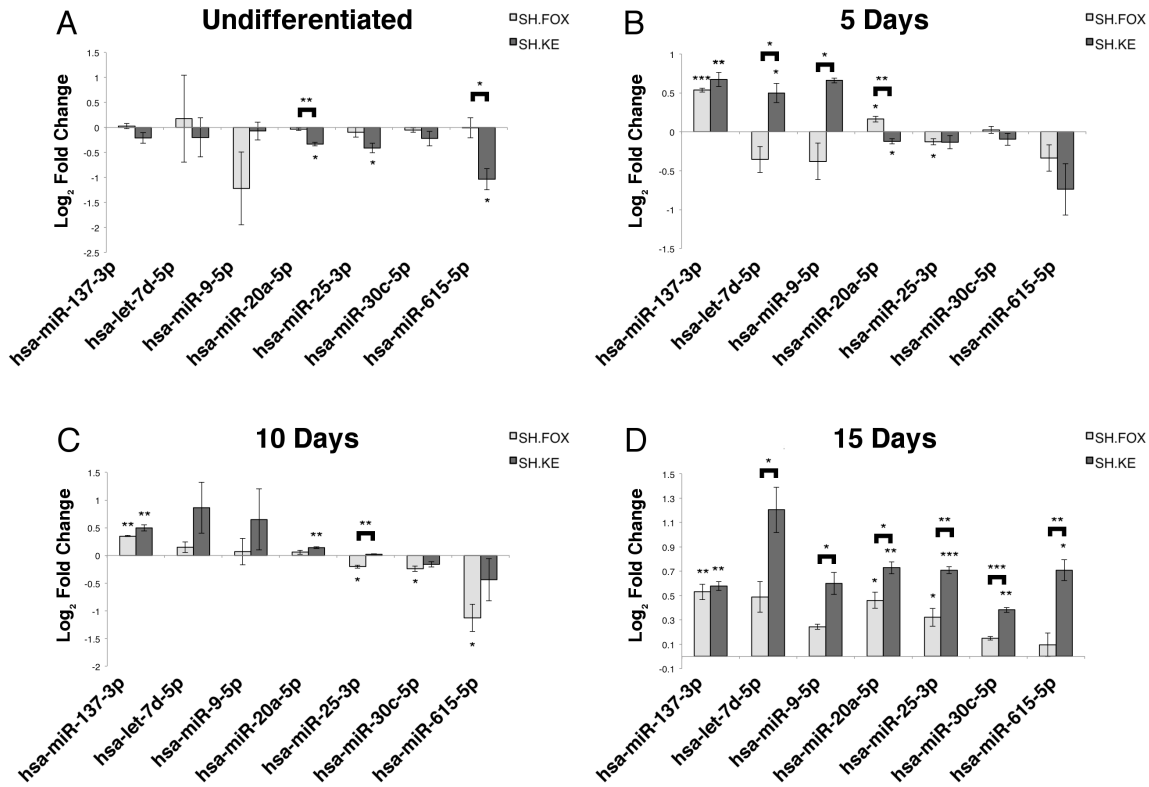


Figure 3.14. qPCR Validation of MicroRNA Targets In Human Neuron-like SH-SY5Y Cells.

The normalised log₂ fold changes of microRNAs of interest are shown here, in stably transfected SH-SY5Y cells. The levels of these 7 microRNAs were assessed for each period of a differentiation time course. The log₂ fold changes of microRNA levels in SH.FOX and SH.KE cells are shown relative to that in SH.EMPTY control cells. Significance of Tukey's post-hoc test is denoted as follows: P < 0.05 – *, P < 0.01 – **, P < 0.001 – ***. Error bars denote standard error of the mean.

MiR-137-3p was up-regulated 1.5-fold ($\log_2$ fold upregulation of ~ 0.5 -fold), in both SH.FOX and SH.KE cells, across all stages of differentiation. At Day 5, besides miR-137-3p, only miR-20a-5p and miR-25a-3p showed significant differential expression in SH.FOX cells relative to SH.EMPTY cells, with miR-20a-5p being up-regulated, while miR-25-3p was down-regulated. In SH.KE cells at Day 5, only miR-20a-5p had differential expression in SH.KE cells relative to SH.EMPTY ones, exhibiting a modest 1.1-fold decrease. Comparing microRNA levels of SH.KE cells relative to SH.FOX cells at Day 5 revealed that only two microRNAs – let-7d-5p and miR-9-5p – exhibited differential regulation, with both these microRNAs exhibiting a twofold up-regulation in SH.KE cells.

At Day 10, miR-25-3p, miR-30c-5p, and miR-615-5p were down-regulated in SH.FOX cells relative to SH.EMPTY cells. MiR-25-3p and miR-30c-5p are both 1.25-fold lower, while miR-615-5p is twofold lower. MiR-25-3p is also significantly lower in SH.FOX cells relative to SH.KE cells. In SH.KE cells, only miR-137-3p and miR-20a-5p are significantly different, relative to SH.EMPTY cells. Both these microRNAs are up-regulated by the KE mutant version of FOXP2.

At Day 15, three microRNAs had significant differential expression in SH.FOX cells relative to SH.EMPTY cells: miR-137-3p, miR-20a-5p, and miR-25-3p, with all three undergoing up-regulation by FOXP2. There was a striking trend of microRNA up-regulation in SH.KE cells (Figure 3.14D). All but 2 microRNAs (let-7d-5p and miR-9-5p) tested showed statistically significant up-regulation in SH.KE cells. The levels of all these microRNAs, except that of miR-137, were also significantly higher in SH.KE cells relative to SH.FOX cells.

To see if the regulation of these microRNAs by Foxp2/FOXP2 is conserved, I compared the qPCR results in SH-SH5Y cells with the microarray results from developing mouse striatum (Table 3.6). Only two microRNAs – miR-25-3p and miR-30c-5p – were regulated in the same direction in both developing mouse striatum and SH.FOX cells. MiR-25-3p was repressed in SH.FOX cells at Day 5 and Day 10 of the differentiation regime, while miR-30c-5p was repressed in

SH.FOX cells at only Day 10. This is concordant with the repressive effect of Foxp2 in the E16 murine ganglionic eminences.

Table 3.6. Summary of Significant Expression Fold Changes of MicroRNA Targets In Human Neuron-like SH-SY5Y Cells.

microRNA	Effect of Foxp2 in E16 mouse brain on Log ₂ Fold Expression	Log ₂ Fold Change (v.s. SH.EMPTY unless otherwise stated)	Undifferentiated	5 Days	10 Days	15 Days
hsa-miR-137-3p	-1.14	SH.FOX		0.54	0.35	0.53
		SH.KE		0.67	0.50	0.58
		SH.KE v.s. SH.FOX				
hsa-let-7d-5p	-0.50	SH.FOX				
		SH.KE		0.50		
		SH.KE v.s. SH.FOX		0.86		0.72
hsa-miR-9-5p	-1.85	SH.FOX				
		SH.KE				
		SH.KE v.s. SH.FOX		1.04		0.36
hsa-miR-20a-5p	-6.21	SH.FOX		0.16		0.46
		SH.KE	-0.33	-0.12	0.14	0.73
		SH.KE v.s. SH.FOX	-0.30	-0.29		0.27
hsa-miR-25-3p	-5.27	SH.FOX		-0.13	-0.20	0.32
		SH.KE	-0.41			0.71
		SH.KE v.s. SH.FOX			0.22	0.39
hsa-miR-30c-5p	-4.35	SH.FOX			-0.24	
		SH.KE				0.38
		SH.KE v.s. SH.FOX				0.23
hsa-miR-615-5p	2.95	SH.FOX			-1.13	
		SH.KE	-1.04			0.71
		SH.KE v.s. SH.FOX	-1.03			0.62

Shown here are significant (Tukey's post-hoc test p-value < 0.05) log₂ expression fold changes between the three stably-transfected SH-SY5Y cell-lines, alongside that observed in the microarray data from embryonic mouse brains. The directions of the significant fold changes are indicated above, for each timepoint of the differentiation protocol.

3.3. Discussion

FOXP2 is a transcription factor that is mutated in an inherited form of verbal dyspraxia (Lai *et al.* 2001). Mice with only one functional copy of *Foxp2* have abnormal cortico-striatal synaptic plasticity and impaired motor-skill learning (Groszer *et al.* 2008, French *et al.* 2012). *In vivo* work using ChIP-chip screens for *Foxp2*-binding in the embryonic mouse brain identified a set of microRNA targets (Vernes *et al.* 2011). In this chapter, I sought to expand upon this previous *in vivo* screen by determining the full repertoire of microRNAs regulated by *Foxp2*/FOXP2. I identified 8 microRNAs showing significant dysregulation in mutant *Foxp2*.S321X embryonic mouse brains. To identify biological themes associated with these 8 microRNAs, I obtained predicted targets of these microRNAs, and studied their gene ontologies and enrichment in biological pathways. Validation of the microarray findings was performed *in vitro* using well-studied cell-lines that overexpress FOXP2/*Foxp2* and the KE mutant form.

3.3.1. *Foxp2* regulation of microRNAs: Genetic Networks Underlying Neural Function and Development

This microarray study found that 8 microRNAs whose expression was significantly disrupted in embryonic mouse striatum lacking functional *Foxp2*. The number of significantly-regulated microRNAs seems small for a expression study, but it is worth noting that a recent study utilising next-generation sequencing to profile the effects of song exposure on miRNA levels in the zebra finch auditory cortex uncovered only five miRNAs that showed significant and consistent change across three biological replicates (Gunaratne *et al.* 2011). As none of the significantly-regulated microRNAs survived correction for multiple testing, a threshold was set with a nominal p-value of 0.01 to identify candidate microRNAs for *in vitro* validation and testing.

The promoters of two of these 8 microRNAs (mmu-miR-125b-2-3p and mmu-miR-320-3p) were previously identified as targets of *Foxp2* (Vernes *et al.* 2011). However, the same study did not find evidence of *Foxp2* occupancy at the promoters of four other microRNAs (miR-30c-5p, miR-25-3p, miR-615-5p, and

miR-181b-5p). It is unclear if Foxp2 occupies the promoters of miR-20a and miR-1195, as the ChIP-chip platform employed did not carry probes for these promoters. As with all ChIP-chip studies, the set of promoters being investigated was necessarily limited by the capacity and design of the arrays used. It is possible that Foxp2 does directly bind to all these microRNA loci, but the bound regions lie outside of classically defined promoters. A notable example is the well-validated Foxp2 target *Cntnap2*: Foxp2 does not bind to the *Cntnap2* promoter region near its transcription start site, but instead binds to an intronic region (Vernes *et al.* 2008, Vernes *et al.* 2011).

Six of the top eight microRNAs are involved in neural functions, summarised in Table 3.7. This suggests that microRNA regulation by Foxp2 contributes to brain function and development. Both miR-30c and miR-25 have been implicated in neurodegenerative diseases affecting motor control. MiR-30c is up-regulated in the striatum of individuals with Huntington's disease (Martí *et al.* 2010), while miR-25-3p is up-regulated in the striatum of mice with mutant *LRRK2*, a well-studied model of Parkinson's disease (Dorval *et al.* 2014). Interestingly, this microarray study found that miR-30c and miR-25 are up-regulated in the developing striatum of S321X.Foxp2 mutant mouse embryos. Mice with a KE mutant version of Foxp2 also exhibit aberrant striatal synaptic plasticity (Groszer *et al.* 2008) and aberrant striatal firing activity during motor learning in the striatum (French *et al.* 2012). Foxp2 might thus exert its effects on striatal-dependent motor learning through repression of miR-30c and miR-25.

Furthermore, the conserved analog of miR-25-3p was up-regulated in the auditory forebrains of both male and female zebra finches after song exposure (Gunaratne *et al.* 2011). Given the importance of FoxP2 in zebra finch vocal learning (Teramitsu and White 2006, Haesler *et al.* 2007), Foxp2-mediated regulation of miR-25-3p might be a functional link between motor and auditory aspects of vocal learning.

Table 3.7. Known Neuronal-related Functions of the Top 8 MicroRNAs.

microRNA	Function in the brain	References
mmu-miR-30c	Up-regulated in frontal cortices and striatum of patients with Huntington's.	Martí <i>et al.</i> 2010
mmu-miR-25	Song exposure causes up-regulation of miR-25-3p in both male and female zebra finches; also up-regulated in striatums of mouse models of Parkinson's disease with deficient LRRK2.	Gunaratne <i>et al.</i> 2011, Dorval <i>et al.</i> 2014.
mmu-miR-1195	Currently not known.	--
mmu-miR-20a	Along with miR-20b and miR-23a, miR-20a is involved in a feedback loop with cyclin-D1 in neural progenitor cells, playing an important role in cortical neurogenesis.	Ghosh <i>et al.</i> 2014
mmu-miR-615-5p	Currently not known.	--
mmu-miR-125b-2-3p	The related miR-125b-5p that is produced from the same precursor affects dendritic spine morphology of mouse neurons, and induces neuronal differentiation in SH-SY5Y cells.	Edbauer <i>et al.</i> 2010, Le <i>et al.</i> 2009
mmu-miR-181b	Up-regulated in temporal cortex of individuals with schizophrenia; and also up-regulated in cell-lines derived from autistic individuals relative to that unaffected siblings cell-lines.	Beveridge <i>et al.</i> 2008, Seno <i>et al.</i> 2010.
mmu-miR-320	Highly expressed in mouse and human cortex; induces neurite outgrowth in N2A cells.	Saba <i>et al.</i> 2008, Nelson <i>et al.</i> 2010, White and Giffard 2012.

Mmu-miR-20a reduces neural differentiation by increasing the number of neural progenitors in the ventricular and sub-ventricular zones, at the expense of neurons in the intermediate zone (Ghosh *et al.* 2014). The ganglionic eminences are part of the sub-ventricular zone (Brazel *et al.* 2003), which eventually gives rise to the adult striatum (Olsson *et al.* 1995, Marín *et al.* 2000). Foxp2 promotes the differentiation of neural progenitors in the ganglionic eminences into medium spiny neurons that comprise the mature striatum (Chiu *et al.* 2014). This microarray study showed that miR-20a is up-regulated in the ganglionic eminences of mice lacking functional Foxp2. Given the opposing effects of Foxp2 and miR-20a on neural precursor differentiation, Foxp2-mediated repression of

miR-20a expression might thus play an important role in the differentiation of sub-ventricular zone precursors that eventually develop into the mature striatum. Affected KE members exhibit structural abnormalities in the striatum: relative to healthy individuals, they have reduced grey matter in the caudate nucleus and increased grey matter in the putamen (Watkins *et al.* 2002b). It is possible that dysregulated levels of miR-20a might contribute to the aberrant grey matter density in affected KE members. Future research can investigate if knock-down of miR-20a in animal models can rescue FoxP2-related phenotypes, such as deficient motor-skill learning in Foxp2 mutant mice (Groszer *et al.* 2008, French *et al.* 2012.) and altered vocal learning and behaviours in songbirds (Haesler *et al.* 2007, Murugan *et al.* 2013.).

Mir-181b is enriched in the mammalian brain (Ryan *et al.* 2006), and is expressed in neurons (Kim *et al.* 2004). The human analog of mmu-miR-181b is up-regulated in the temporal cortices of individuals with schizophrenia (Beveridge *et al.* 2008), as well as in lymphoblastoid cell-lines derived from autistic individuals relative to that of cell-lines from unaffected siblings (Seno *et al.* 2010). Two validated targets of miR-181b are ionotropic glutamate receptor 2 (GRIA2) (Beveridge *et al.* 2008), and cylindromatosis tumor suppressor (CYLD), a deubiquitinase (Iliopoulos *et al.* 2010.). These two genes play important roles in glutamatergic signalling at the synapse (Passafaro *et al.* 2003, Dosemeci *et al.* 2013, Thein *et al.* 2014). Therefore, by regulating the neuronal levels of *GRIA2* and *CYLD*, miR-181b is likely to be important for synaptic function and neuronal circuit formation. In the microarray study presented here, miR-181b is up-regulated in the developing striatums of mutant mice lacking functional Foxp2. Foxp2 might regulate development and formation of striatal circuits by fine-tuning expression of relevant genetic networks via repression of miR-181b levels.

MiR-320-3p and miR-125b-2-3p are known to play a role in neurite outgrowth. Mmu-miR-320-3p is highly expressed in the mouse brain (Saba *et al.* 2008), and overexpression of pri-mir-320 induces neurite outgrowth in N2A cells (White and Giffard 2012). While very little is known about the function of miR-125b-2-3p, the

microRNA miR-125b-5p that is produced from the same precursor is involved in neuronal function. It alters spine morphology in the mouse hippocampus via regulation of the NMDA receptor subunit 2A (Edbaeur *et al.* 2010). In SH-SY5Y cells, miR-125b-5p induces neuronal differentiation and neurite outgrowth (Le *et al.* 2009). The effects of Foxp2/FOXP2 on neurite outgrowth *in vitro* and *in vivo* are well studied. Overexpression of Foxp2 in Neuro2A cells (Vernes *et al.* 2011) and in SH-SY5Y cells (Devanna *et al.* 2014) increases neurite outgrowth. Primary neurons from the developing striatum of E16 homozygous Foxp2.R552H mutant mice exhibit decreased branching and shorter neurites (Vernes *et al.* 2011). Given that miR-320-3p and miR-125b-2-3p are both ChIP-chip targets of Foxp2 (Vernes *et al.* 2011), the microarray study here supports their status as direct transcriptional targets of Foxp2. It is possible that Foxp2/FOXP2 regulates neurite outgrowth by controlling levels of miR-320-3p and miR-125b-2-3p. Future experiments can manipulate levels of these two microRNAs to see if they can rescue the neurite-outgrowth deficits observed in Foxp2-deficient *ex vivo* neurons (Vernes *et al.* 2011).

A limitation of this study is that the set of microRNAs being profiled, while extensive, remains limited by the capacity of the microarray platform used. Next-generation sequencing techniques (e.g. total RNA-seq and small RNA-seq) can obtain expression profiles of small non-coding RNAs with true genome-wide coverage. In particular, total RNA-seq can profile mRNA and microRNA expression simultaneously, permitting the correlation of microRNA levels with putative target mRNA levels. The future application of total RNA-seq to Foxp2 mice brains lacking functional Foxp2 will reveal the full range of microRNAs regulating, as well as uncover the exact molecular networks (co-)regulated by these microRNAs and Foxp2.

3.3.2. Foxp2 regulation of microRNAs: Wiring and Building Neuronal Networks

Axon guidance pathway genes are over-represented in the predicted targets of the top 8 microRNAs (Figure 3.9), an observation that is concordant with previous

observations that Foxp2 and FOXP2 ChIP targets, as well as genes showing Foxp2-dependent expression in embryonic mouse striatum, were enriched in genes involved in axon guidance (Spiterl *et al.* 2007, Vernes *et al.* 2007, Vernes *et al.* 2011). Axon guidance genes predicted to be targets of the top 8 microRNAs include semaphorins, ephrins, and Robo genes, and are summarised in Table 3.8. Some of these axon guidance genes were also Foxp2 ChIP targets in the embryonic mouse brain (Vernes *et al.* 2011), such as Ephrin-B2 and Semaphorin-6D. This is further evidence for previously established functional links between Foxp2 and neuronal connectivity (Enard *et al.* 2009, Vernes *et al.* 2011, Reimers-Kipping *et al.* 2011).

Table 3.8. Families of Axon Guidance Ligands and Receptors are Targeted by Foxp2-regulated MicroRNAs.

Gene Family	Gene	Foxp2 ChIP target?	Differentially Expressed in Foxp2 mutant brains?	Targeted by
Sempahorin	Sema3A	√		mmu-miR-30c-5p, mmu-miR-320-3p
	Sema4B			mmu-miR-20a-5p
	Sema4C			mmu-miR-615-5p
	Sema4G		√	mmu-miR-181b-5p
	Sema5A			mmu-miR-20a-5p
	Sema6B			mmu-miR-30c-5p
	Sema6D	√		mmu-miR-30c-5p, mmu-miR-320-3p
	Sema7A			mmu-miR-20a-5p
Ephrin	Efna3			mmu-miR-30c-5p
	Epha4			mmu-miR-20a-5p, mmu-miR-181b-5p
	Epha5			mmu-miR-181b-5p, mmu-miR-20a-5p
	Epha7			mmu-miR-20a-5p
	Ephb2	√		mmu-miR-30c-5p
	Ephb4			mmu-miR-20a-5p
Robo	Robo1			mmu-miR-25-3p
	Robo2	√		mmu-miR-25-3p, mmu-miR-181b-5p

MicroRNAs are ideal for rapid control of local protein synthesis at the synapse (McNeill and Van Vactor 2012), especially in the context of axon guidance. The nascent synapse grows progressively distant from the cell body, yet requires gene expression to remain responsive to external cues (such as neurotransmitters or neurotrophic gradients). Thus, the synapse has a pool of mRNA transcripts, allowing regulation of protein synthesis at the level of individual synapse, in response to synaptic activity and environmental growth cues. MicroRNAs are localised to the axonal and synaptic regions as well (Lugli

et al. 2008), and are able to fine-tune the localised expression of this synaptic mRNA pool. Indeed, microRNAs have been shown to affect synaptic development (Schratt *et al.* 2006, Edbauer *et al.* 2010) and axonal guidance (Baudet *et al.* 2011, Shibata *et al.* 2011).

Motor-skill learning, which involves close monitoring of desired motor output versus actual outcome of movement, is a well-studied aspect of Foxp2 function. Mice lacking functional Foxp2 have deficits in motor-skill learning (Groszer *et al.* 2008, French *et al.* 2012). At the molecular level, this involves widespread changes in gene expression in the striatum and the motor cortex (D'Amours *et al.* 2010, Cheung *et al.* 2013). At the neuronal level, motor-skill learning requires precise connections to develop between relevant neuronal populations. Mice with aberrant axonal pathfinding in midbrain dopaminergic neurons possess delays in motor-skill maturation (Shelag *et al.* 2013). Axonal guidance is thus a critical molecular pathway underlying motor-skill learning. The work presented here suggests that Foxp2 regulates genetic pathways underlying axonal guidance, and fine-tunes their expression through the action of microRNAs. It will be of great interest to see if manipulating levels of the microRNAs identified here can rescue or recapitulate Foxp2-related axon outgrowth phenotypes in appropriate animal models of Foxp2 function, such as mutant mice (Groszer *et al.* 2007, French *et al.* 2012) or zebra finches with FoxP2 knocked down (Haesler *et al.* 2007, Schulz *et al.* 2010).

The top 8 microRNAs were predicted across multiple algorithms to target 14.8% of genes (41 out of 318) that exhibit Foxp2-dependent expression in the developing mouse striatum (Vernes *et al.* 2011). This overlap was significantly higher than that expected by chance (hyper-geometric p-value < 0.005), and suggests that Foxp2 regulates some of its targets through the action of microRNAs. Four of these 41 co-targeted genes were also direct ChIP-chip targets of Foxp2 (Supp Table 3.1), representing the first evidence that Foxp2 participates in feedforward loops with microRNAs (Figure 3.2).

Only one of the top 8 microRNAs, miR-25-3p, is predicted to target Foxp2. As noted previously, miR-25-3p was up-regulated in the zebra finch auditory forebrain during song exposure (Gunaratne *et al.* 2011). Given that FoxP2 is not expressed in the auditory forebrain in adult male zebra finches (Teramitsu *et al.* 2004), it is possible that miR-25 represses FoxP2 expression in the auditory forebrain. A recent report showed that miR-9 and miR-140, which are up-regulated in Area X during undirected singing, repress zebra finch FoxP2 (Shi *et al.* 2013). The feedback loops formed by FoxP2 and these microRNAs thus represent functional links between song-related pathways in the zebra finch brain. Future work focusing on FoxP2 regulation of microRNAs in the songbird brain will shed light on these pathways.

3.3.3. A Complex Picture of MicroRNA Regulation by Foxp2: Comparing In Vivo and In Vitro Effects

Of the 10 microRNAs tested in N2A cells, 5 showed significant differences between N2A.FOX and N2A.EMPTY cell-lines, but only mmu-miR-615-5p showed a significant regulation by Foxp2 in the same direction as in embryonic mouse brains (Figure 3.11). The inability to observe the microRNA differences seen in E16 mouse brains in murine cell-lines could be attributed to the between *in vitro* and *in vivo* contexts. I chose to validate the *in vivo* microarray results with qPCR in well-studied stable Neuro2A cell-lines as a springboard for future experimental work in these cell-lines, but this was not satisfactorily achieved. This strongly suggests that validation of the microarray results should be performed in an independent batch of litter-matched mouse brains, so as to better identify which microRNAs had consistent and robust changes.

It is possible that the repertoire of binding partners and cofactors available for interaction with Foxp2 differs between N2A cells and the developing mouse striatum, leading to differences in microRNA regulation by Foxp2. There might be repressive factors that bind to the microRNA promoter regions *in vitro*, but not *in vivo*. Also, transcription factors can exhibit differential binding in the genome in different cell-types, depending on the presence of cell-line-specific co-factors

(Mullen *et al.* 2011). In the mouse brain, Foxp1 and Foxp2 are co-expressed in the developing ganglionic eminences, as well as in the adult striatum (Ferland *et al.* 2003). Foxp1 and Foxp2 form heterodimers with each other (Li *et al.* 2004, Deriziotis *et al.* 2014a, Sin *et al.* 2014), and the *in vitro* effects of FOXP2 on some of its transcriptional targets (such as *NEUROD2*) can be reversed or potentiated by over-expressing FOXP1 (Sin *et al.* 2014). This suggests that the transcriptional action of FOXP2 is sensitive to the presence (and amount) of co-factors and binding partners.

Secondly, both murine and human *in vitro* cell-lines used here do not endogenously express Foxp2/FOXP2. The induced stable expression of Foxp2/FOXP2 might produce relative levels that do not reflect *in vivo* conditions. For example, the ratio of endogenous Foxp1 to non-endogenous Foxp2 in N2A.FOX cells might be different from that of the Foxp1-Foxp2 ratio in embryonic mouse striatum. As noted above, a change in the levels of Foxp2 relative to its binding partners or co-factors can disturb its effects on transcriptional targets. This might also explain the differences in the transcriptional effects of Foxp2 on microRNA levels.

The microRNAs miR-137-3p, miR-20a-5p, and miR-25-3p were up-regulated in SH.FOX cells relative to SH.EMPTY cells, after being subject to the full differentiation regime. MiR-137 showed consistent up-regulation by FOXP2 in all cells subjected to the differentiation treatment. Given the high expression of miR-137 in the brain and its well-documented roles in neuronal maturation and dendritic development (Smrt *et al.* 2010), this finding suggests a possible mechanism by which FOXP2 shapes neuronal networks underlying human speech and language. Both Foxp2 and miR-25-5p have been implicated in vocal learning and response to song exposure respectively in the zebra finch brain (Haesler *et al.* 2007, Gunaratne *et al.* 2010, Murugan *et al.* 2013). In addition, recent reports have independently implicated Foxp2 and miR-20a in cortical neurogenesis (Tsui *et al.* 2013, Ghosh *et al.* 2014). The results presented here thus provide the first functional links between Foxp2 and the microRNAs miR-25-

5p and miR-20a-5p, and also suggests neuronal processes and molecular pathways that might be disrupted in human speech and language deficits.

3.3.4. The R553H KE Mutation: Dominant Negative Effects on MicroRNA Regulation?

The heterozygous R553H mutation in FOXP2 was first found in the three-generation family known as KE, in all members who suffered from a severe speech and language deficit (Lai *et al.* 2001). Individuals with this mutation have abnormal grey matter thickness in several brain regions such as the striatum and cerebellum (Watkins *et al.* 2002b), and exhibit atypical patterns of neural activation during language processing tasks (Liégeois *et al.* 2003). Mice carrying the same heterozygous mutation have impairments in motor learning tasks and auditory-motor learning tasks (Groszer *et al.* 2008, French *et al.* 2011, Kurt *et al.* 2012).

The studies cited above indicate the widespread effects produced by this heterozygous point mutation. At a molecular level, the mutation abolishes the ability of FOXP2 to bind DNA in general (Nelson *et al.* 2013), rather than a specific loss of its ability to recognise a consensus FOXP motif (Vernes *et al.* 2006). The mutation also affects the nuclear localisation of the protein (Vernes *et al.* 2006, Mizutani *et al.* 2006) although the exact molecular mechanisms behind this remain unclear. Importantly, the KE mutant can still dimerise with wildtype FOXP2 and FOXP1, and can also interact with FOXP2-binding partners such as CTBP1 (Deriziotis *et al.* 2014a). Thus, it is thought that the KE mutant exerts a dominant negative effect through its dimerisation and complex formation with endogenous wildtype proteins. While the repertoire of genes regulated by FOXP2/Foxp2 has been well studied (Vernes *et al.* 2007, Spiteri *et al.* 2007, Vernes *et al.* 2011), little is known about the genes dysregulated by the KE mutation.

The qPCR results presented here suggest that the KE mutation has a complex effect on FOXP2/Foxp2 function. Firstly, the KE mutation can produce loss-of-

function null effects on microRNA regulation. This is seen, for instance, in the levels of let-7d across the N2A cell lines: the up-regulation of let-7d observed in N2A.FOX cells (relative to N2A.EMPTY cells) is abolished in N2A.KE cell-lines (Figure 3.11). Given that Foxp2 was previously identified to directly bind to the let-7d promoter in the embryonic mouse brain (Vernes *et al.* 2011), this strongly suggests that Foxp2 regulates let-7d by directly binding to a promoter locus, which the KE mutant is unable to do, causing a loss of Foxp2-mediated up-regulation.

Secondly, the KE mutant was also observed to have gain-of-function effects. This is exemplified in the levels of miR-20a-5p, miR-25-3p, and miR-615-5p in undifferentiated SH-SY5Y cells (Figure 3.14A). These microRNAs were found to be significantly lower in SH.KE cells, relative to and SH.FOX cells. A gain-of-function effect is also strikingly seen in the SH-SY5Y cells that underwent the full 15 days of differentiation: 6 out of 7 microRNAs were higher in SH.KE cell-lines relative to SH.FOX cell-lines (Figure 3.14D, Table 3.6). The KE mutant protein is unable to bind to DNA, regardless of the sequence (Nelson *et al.* 2013). Thus, it is likely that FOXP2 occupancy at these microRNA loci is abolished. As a result, other factors that are unable to displace the wildtype FOXP2 might thus be able to bind, resulting in the observed gain-of-function transcriptional effects.

Lastly, the KE mutant has the same regulatory effects as the wildtype protein on some microRNAs, i.e. there was no dysregulation of the microRNA. This was seen in the case of miR-137 in differentiated SH-SY5Y cells (Figure 3.14B, C and D), where both the wildtype and KE mutant versions of FOXP2 increased the levels of miR-137. This suggests that FOXP2 regulation of this microRNA might be indirect and independent of the DNA-binding forkhead-box domain, such as through dimerization with another FOXP member or interaction with another co-factor. The KE mutant does not disrupt homo- and hetero- dimerization of FOXP2 (Deriziotis *et al.* 2014a, Deriziotis *et al.*, in prep). Therefore, it is possible that both wildtype and KE mutant FOXP2 regulate miR-137 in SH-SY5Y cells by forming a

protein complex between its binding partners that results in transcriptional regulation.

Taken together, these results provide evidence for a dominant negative effect of the KE mutation, although it remains difficult to classify it as gain-of-function: the exact behaviour of the KE mutant protein seems dependent on the microRNA locus, and possibly the specific co-factors interacting at the locus. It is important to remember that affected KE members have one functional version of FOXP2 alongside the KE mutant version. To fully understand how dysregulation of microRNAs contributes to the verbal dyspraxia in affected KE members, the results here need to be further evaluated in a heterozygous situation. Nonetheless, these results suggest that the phenotype associated with the KE mutation is a combination of loss of DNA binding (the KE mutant does not bind and regulate the target microRNA at all) and dominant-negative effects (the KE mutant has a gain-of-function which disrupts the normal function of the wildtype). These effects on microRNA levels are likely to cause widespread downstream gene dysregulation, and might be responsible for the impaired behaviour in mutant mouse models, such as diminished motor learning abilities.

3.3.5 Conclusion

The microarray data presented here has uncovered 8 microRNAs potentially regulated by *Foxp2* during brain development. Several of these microRNAs play important roles in the brain, such as neural differentiation. These microRNAs are also predicted to regulate genetic pathways involved in axonal guidance. This suggests that *Foxp2* fine-tunes the expression of axonal guidance pathways via microRNAs. However, divergence between the *in vivo* and *in vitro* effects of *Foxp2*/*FOXP2* on microRNA regulation was observed, strongly indicating a need for further validation in independent *in vivo* samples. The consequences of the KE mutation on the *in vitro* expression of these microRNAs suggest a dominant negative effect on transcriptional regulation. Future work studying the *in vivo* effects of these microRNAs on *Foxp2*-related phenotypes, such as neurite

outgrowth and motor-skill learning, will highlight the specific molecular pathways that mediate the phenotypic effects of *Foxp2*.

Chapter 4: Investigation of A Putative Genetic Network – Beyond FOXP2 and Verbal Dyspraxia?

4.1 Introduction

4.1.1. FOXP2 Target Genes and Other Disorders: Glimpses of a Common Underlying Genetic Network

FOXP2 is the first gene to be implicated in an inherited form of developmental verbal dyspraxia (DVD). As a transcription factor, the downstream target genes that it regulates are also likely to be risk factors, and to contribute to the aetiology of DVD. Interestingly, a number of validated *FOXP2* targets have been implicated in a range of other neurodevelopmental disorders, such as schizophrenia and autism, and are discussed in detail below. These functional genetic links point to a common genetic network underlying distinct neurological conditions.

4.1.2. *CNTNAP2* and Neurodevelopmental Disorders: Autism, Intellectual Disability, Epilepsy, and Language Impairment

The contactin-associated protein-like 2 gene (*CNTNAP2*) is one of the largest genes in the human genome, spanning 2.3Mb on chromosome 7q35. It codes for CASPR2, a transmembrane protein that is a member of the neurexin superfamily (Nakabayashi and Scherer 2001). Neurexin proteins mediate extracellular interactions between cells in the nervous system (Dean and Dresbach 2006). *CNTNAP2* has a restricted expression pattern in the cortex, amygdala, dorsal thalamus, and striatum of the foetal human brain (Abrahams *et al.* 2007, Alarcón *et al.* 2007), overlapping with the cortico-striatal-thalamic pathway important for human speech and language. Within the cortex, it is expressed in layers II through to V (Bakkaloglu *et al.* 2008). Early studies demonstrated that *CNTNAP2* is involved in the localization of potassium channels in juxtaparanodes of myelinated axons (Poliak *et al.* 2003, Horresh *et al.* 2008). Subsequent work

implicated *CNTNAP2* in neuronal migration (Strauss *et al.* 2006, Peñagarikano *et al.* 2011) and synaptic activity (Peñagarikano *et al.* 2011, Anderson *et al.* 2012).

Individuals with mutations in *CNTNAP2* have a complex phenotype with four common features: intellectual disability (ID), autistic behaviours, epileptic seizures, and language problems (Strauss *et al.* 2006, Friedman *et al.* 2007, Bakkaloglu *et al.* 2008, Gregor *et al.* 2011). Intellectual disability (also known as mental retardation) is a developmental disorder with impairments of general cognitive abilities and deficits in adaptive behaviours, such as social interactions, language ability, and learning ability. Autism spectrum disorders (ASD) refer to a range of developmental conditions characterised by impaired social behaviour, reduced verbal communication, stereotyped and repetitive behavior, and a narrow set of interests.

The presentation of both ID and autism in individuals with *CNTNAP2* mutations can range from severe to mild. The language impairments in affected individuals also have variable manifestations, such as delayed speech onset (O’Roak *et al.* 2011), or total absence of speech (Gregor *et al.* 2011). Epileptic seizures may or may not manifest alongside ID and speech delay (Gregor *et al.* 2011). The phenotypic heterogeneity of *CNTNAP2* mutations might be due to different effects on protein function: it is unclear if pathogenic mutations are null, or deleterious with dominant negative effects (Falivelli *et al.* 2012). Taken together, these results suggest that the genetic networks that employ *CNTNAP2* do not simply subserve human speech and language, and are likely to play roles in cognitive function as well.

This notion has been further strengthened by several association studies and genetic studies. Disruptions and mutations of *CNTNAP2* have been associated with autism (Alarcón *et al.* 2007, Arking *et al.* 2008, Bakkaloglu *et al.* 2008, Rossi *et al.* 2008, Jackman *et al.* 2009, Anney *et al.* 2012), dyslexia (Peter *et al.* 2010), schizophrenia (Ji *et al.* 2013), attention-deficit hyperactivity disorder (Bisgaard *et al.* 2006, Elia *et al.* 2010), bipolar disorder (Ji *et al.* 2013), mental retardation (Zweier *et al.* 2009), and SLI (Vernes *et al.* 2008, Newbury *et al.* 2011). A short

tandem repeat variant in the 5' promoter region of *CNTNAP2* was associated with an older age of first spoken word in individuals with ASD (Chiocchetti *et al.* 2014). The same study also identified rare mutations in the *CNTNAP2* promoter region, and demonstrated their effects on transcriptional regulation. This implies that altered regulation of *CNTNAP2* may confer ASD risk as well. A set of SNPs in the exon 13-15 region of *CNTNAP2* was also associated with early language acquisition in normal individuals (Whitehouse *et al.* 2011), suggesting that *CNTNAP2* variants might account for differences in language development in the general population.

Mice without two functional copies of *Cntnap2* (Peñagarikano *et al.* 2011) have neurological and behavioural deficits that mirror the epilepsy and autistic spectrum behaviors seen in humans carrying *CNTNAP2* mutations. Six-month-old mutant pups have spontaneous seizures, a phenotype shared with children bearing homozygous *CNTNAP2* mutations (Strauss *et al.* 2006, Jackman *et al.* 2009). *Cntnap2* knockout mice also displayed reduced social behavior, increased stereotypy, and perseveration, all of which are clinical features of ASD. Interestingly, *Cntnap2* mutant mice also emitted less ultrasonic vocalizations (Peñagarikano *et al.* 2011), but caution must be taken when comparing this phenotype with the language impairments observed in individuals carrying *CNTNAP2* mutations.

CNTNAP2 is one of the best-characterised direct downstream targets of FOXP2. Both the human and murine versions of *Cntnap2* have a site in intron 1 that is occupied by FOXP2/Foxp2 (Vernes *et al.* 2008, Vernes *et al.* 2011). Overexpression of FOXP2 in human neuron-like cell-lines produces a robust decrease in *CNTNAP2* mRNA levels (Vernes *et al.* 2008). Therefore, the results of patient studies, association studies, and mouse models implicate FOXP2 and *CNTNAP2* in genetic networks that underlie several different neurodevelopmental disorders.

4.1.3. *MIR137* and Schizophrenia

MiR-137 is a brain-specific microRNA (Smrt *et al.* 2010) that plays an important role in neuronal development, promoting neuronal differentiation while reducing neuronal proliferation (Sun *et al.* 2011, Jiang *et al.* 2013). MiR-137 is also involved in the formation of neuronal networks. It is enriched at the synapse (Siegel *et al.* 2009) and it has been found to decrease dendrite length (Smrt *et al.* 2010). Several gene transcripts are differentially expressed following changes in miR-137 levels, and are enriched in pathways for neurogenesis, neuronal differentiation, and synaptic transmission (Collins *et al.* 2014, Hill *et al.* 2014, Strazisar *et al.* 2014, Vallés *et al.* 2014), suggesting that miR-137 exerts its effects on neurodevelopment by regulating genes involved in these pathways.

Schizophrenia is a mental disorder with experiences of altered reality (such as delusions and hallucinations), deficits in reasoning and cognition, and impaired social skills. Genomewide studies have associated 2 intronic SNPs in the *MIR137* transcript (rs1625579 and rs1198588) with schizophrenia (Ripke *et al.* 2011, Ripke *et al.* 2013). These studies also found that genes with miR-137 target sites were enriched for lower genome-wide association p-values, strongly suggesting that pathways regulated by miR-137 are disrupted in schizophrenia. The risk allele of the SNP rs1625579 is also associated with endophenotypes of schizophrenia, such as more severe cognitive and psychotic disturbances, earlier age-of-onset, larger ventricles, smaller hippocampal volumes, and reduced white matter tracts (Cummings *et al.* 2013, Green *et al.* 2013, Lett *et al.* 2013). This same SNP was also found to have pleiotropic effects on ASD as well (Cross-Disorder Group of the Psychiatric Genomics Consortium 2013). Deletions spanning the *MIR137* locus have also been implicated in intellectual disability (Willemsen *et al.* 2012). Taken together with the transcriptional effects of miR-137 (Collins *et al.* 2014, Hill *et al.* 2014), this suggests that miR-137 regulates genetic networks underlying higher cognitive function.

In Chapter 3 of this thesis, I showed that *MIR137* was transcriptionally regulated *in vitro* by FOXP2. This validates the putative interaction suggested by ChIP-chip

data in embryonic mouse brains (Vernes *et al.* 2011) as well as ChIP-seq data in human neuron-like cells (ENCODE Project Consortium, 2011). Given that schizophrenic patients often exhibit impaired language ability (Bleich-Cohen *et al.* 2012), a FOXP2-MIR137 interaction might underlie language dysfunction in different neurodevelopmental disorders.

4.1.4. *TCF4* and Pitt-Hopkins Syndrome

Transcription factor 4 (*TCF4*) is a member of the basic helix-loop-helix (bHLH) superfamily of transcription factors. *TCF4* binds to Ephrussi box (E-Box) elements (5'-CANNTG-3'), leading to transcriptional activation (Pscherer *et al.* 1996). *TCF4* is also known as *E2-2* (in relation to its E-box binding ability), immunoglobulin transcription factor 2 (*ITF2*), and SL3-3 enhancer factor 2 (*SEF2*) in the literature. Importantly, *TCF4* should not be confused with the gene transcription factor 7-like 2 (*TCF7L2*) that is also known as T cell factor 4. *TCF7L2* is thus sometimes referred to as *TCF4* in the literature, but it is a totally different gene product.

Homo- and heterodimerisation are crucial to the function of bHLH proteins (Murre *et al.* 1989). The regulatory effects of *TCF4* are dependent on co-factors (Flora *et al.* 2007, Forrest *et al.* 2012), and this combinatorial action gives *TCF4* lineage- and context-dependent roles. *TCF4* is expressed across several different adult tissues, including the brain, spinal cord, heart, lung, and kidney (Fu *et al.* 2009, Navarette *et al.* 2013). In non-neural tissues, *TCF4* plays an important role in several processes, such as lymphocyte development (Bergqvist *et al.* 2000), and the transition of epithelial cells to mesenchymal identities (Sobrado *et al.* 2009).

Within the brain, *TCF4* is widely expressed, but shows enrichment in the cortex, striatum, thalamus, hippocampus and cerebellum (Fu *et al.* 2009). *TCF4* plays several roles during neural development, and is implicated in neural differentiation (Schmidt-Edelkraut *et al.* 2014.), pontine development (Flora *et al.* 2007) and oligodendrocyte maturation (Fu *et al.* 2009). Knockdown of *TCF4* in vitro resulted in altered expression of genes involved in neurogenesis, and in

genes mutated in mental retardation (Forrest *et al.* 2013). This suggests that *TCF4* exerts its effects on brain development and function by controlling the levels of genes involved in neurodevelopmental processes.

Mutations in *TCF4* cause Pitt-Hopkins Syndrome (PTHS) (Brockschmidt *et al.* 2007, Zweier *et al.* 2007, Amiel *et al.* 2007). PTHS is a rare neurological condition (Peippo and Ignatius 2012) and has a complex presentation (Pitt and Hopkins 1978, Brockschmidt *et al.* 2007). Affected individuals exhibit intellectual disability, epilepsy, autistic behaviours, delays in language development, and deficits in motor learning. Several of these features are also seen in patients with *CNTNAP2* mutations (Zweier *et al.* 2009). Indeed, *TCF4* binds to the promoter region of *CNTNAP2* in a luciferase assay, activating the expression of the reporter gene (Forrest *et al.* 2012). Taken together, these findings suggest that perturbation of the *TCF4-CNTNAP2* interaction is likely to underlie neurological disorders observed in PTHS.

Genomewide studies have found that common variants in *TCF4* confer susceptibility to schizophrenia (Ripke *et al.* 2011, Steinberg *et al.* 2011, Ripke *et al.* 2013). Mutant mice with postnatal overexpression of *Tcf4* in the brain displayed deficits in associative fear conditioning and sensorimotor gating, mirroring behavioral deficits frequently observed in psychiatric disorders (Brzózka *et al.* 2010). This strongly implicates *TCF4* dysfunction in schizophrenia. *TCF4* is also a target of the schizophrenia risk gene miR-137 (Kwon *et al.* 2011), suggesting that a MIR137-*TCF4* pathway is disrupted in schizophrenia.

In the embryonic mouse brain, *Foxp2* binds to an intronic region of *Tcf4*, suggesting that *Tcf4* is a direct transcriptional target of *Foxp2* (Vernes *et al.* 2011). *TCF4* variants have been associated with verbal fluency (Wirgenes *et al.* 2012) and verbal memory (Lennertz *et al.* 2011) in schizophrenic patients. This supports a hypothesis where a FOXP2-TCF4 regulatory interaction might underlie language impairment across distinct neurodevelopmental disorders.

4.1.5. A Mini-network Comprising Four Genes Involved Across A Range of Neurological Disorders

Besides being bound by *Foxp2* in embryonic mouse brain (Vernes *et al.* 2011), *Cntnap2*, *Tcf4*, and *Mir137* interact with each other as well. Reporter gene assays suggest that TCF4 can bind to the *CNTNAP2* promoter and activate its transcription (Forrest *et al.* 2012). Mature miR-137 is able to down-regulate a reporter gene transcript carrying either one of the 2 miR-137 binding sites found in the *TCF4* 3' UTR (Kwon *et al.* 2011). Thus, one is able to construct a genetic mini-network based on these lines of experimental evidence, functionally linking these four genes together (Figure 4.1).

Given that *TCF4*, *CNTNAP2*, *FOXP2*, and *MIR137* are mechanistically linked, and that *TCF4* and *CNTNAP2* mutations produce highly congruous cognitive phenotypes, it is possible that these four genes are involved in a common genetic network that is disrupted in disorders such as schizophrenia and language impairment. Furthermore, *FOXP2* has been shown to inhibit *DISC1* expression (Walker *et al.* 2012), the latter being a robustly characterised schizophrenia susceptibility gene. This is additional evidence for a shared underlying genetic pathway in these disparate neurological disorders.

FOXP2, *TCF4*, and hsa-miR-137 do not just co-regulate each other: as transcriptional regulators, each one regulates several hundred different transcripts (Vernes *et al.* 2007, Spiteri *et al.* 2007, Forrest *et al.* 2013, Collins *et al.* 2014, Hill *et al.* 2014). Thus, this mini-network proposed here might be a “hub of hubs”, centrally regulating a myriad of molecular pathways that underlie distinct neurodevelopmental disorders.

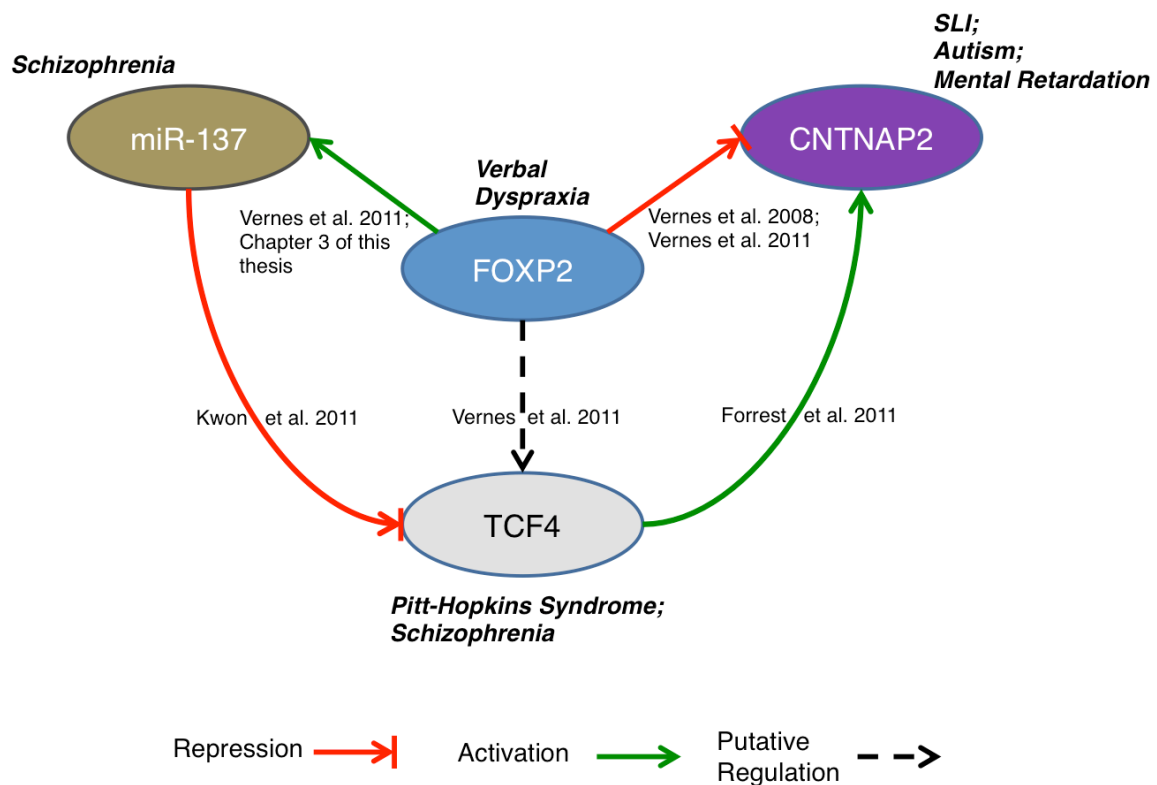


Figure 4.1. A Putative Mini-network of Genes Implicated In A Range of Neurodevelopmental Disorders.

The genes *FOXP2*, *TCF4*, *CNTNAP2*, and *MIR137* are each associated with different neurological disorders. This figure shows the direction of functional interactions (repression or activation), as well as the studies that demonstrated these relationships.

4.1.6. Aims of This Chapter

In this chapter, I will investigate the molecular mechanisms underlying this putative genetic network. Using stably-transfected SH-SY5Y cell-lines, I will test if *TCF4* is regulated by *FOXP2*. I will use published ChIP-seq data to identify regions in the loci of *TCF4*, *CNTNAP2*, and *MIR137* that are occupied by *FOXP2*, and I will employ a luciferase reporter assay to test if these regions are sufficient for *FOXP2*-mediated regulation.

4.2 Results

4.2.1. Regulation of Direct Foxp2 Murine Targets in Human Neuron-like Cell-lines

Previous ChIP studies (Vernes *et al.* 2011) found that murine Foxp2 occupied the genetic loci of *Cntnap2*, *Tcf4*, and *mmu-mir-137* in embryonic mouse brain (Figure 4.2A – C). For *Cntnap2*, ChIP-PCR revealed that a region in the first intron was enriched in wildtype embryonic mouse brains relative to Foxp2.S321X mutant brains that do not express any Foxp2 (Figure 4.2A). This region is orthologous to a FOXP2-bound region in the intron of human *CNTNAP2* (Vernes *et al.* 2008, Vernes *et al.* 2011). For *Tcf4* (Figure 4.2B) and *mmu-mir-137* (Figure 4.2C), ChIP-chip experiments found that the promoter regions for these genes were enriched in wildtype brains relative to Foxp2.S321X mutant brains.

To test if these direct Foxp2 targets were regulated by human FOXP2, transcript levels for each gene were measured in SH-SY5Y cells stably-transfected with a pcDNA3.1 vector carrying the wildtype FOXP2 coding sequence (See Section 2.1). *CNTNAP2* was strongly down-regulated by FOXP2 (Figure 4.2D), with a fivefold repression in undifferentiated cells (p-value < 0.005), and 2.5-fold repression in cells differentiated via treatment with retinoic acid (p-value < 0.05). This confirms previous observations of repression by FOXP2 in human neuron-like cell-lines, as well as in human fetal cortex (Vernes *et al.* 2008). *TCF4* showed a significant 1.6-fold up-regulation by FOXP2 in undifferentiated SH-SY5Y cells (p-value < 0.005) but not in retinoic acid-treated cells (Figure 4.2E). This suggests that *TCF4* is a direct target of FOXP2 in the human brain.

Lastly, the primary transcript of hsa-mir-137 (pri-mir-137), was up-regulated almost 4-fold by FOXP2 (p-value < 0.01) in undifferentiated SH-SY5Y cells, but showed no significant difference in differentiated cells (Figure 4.2F). Interestingly, the opposite pattern of regulation was seen for the mature microRNA (Figure 4.2G): mature hsa-miR-137 is up-regulated 1.5-fold (p-value < 0.005) in differentiated SH-SY5Y cells, but not in undifferentiated cells.

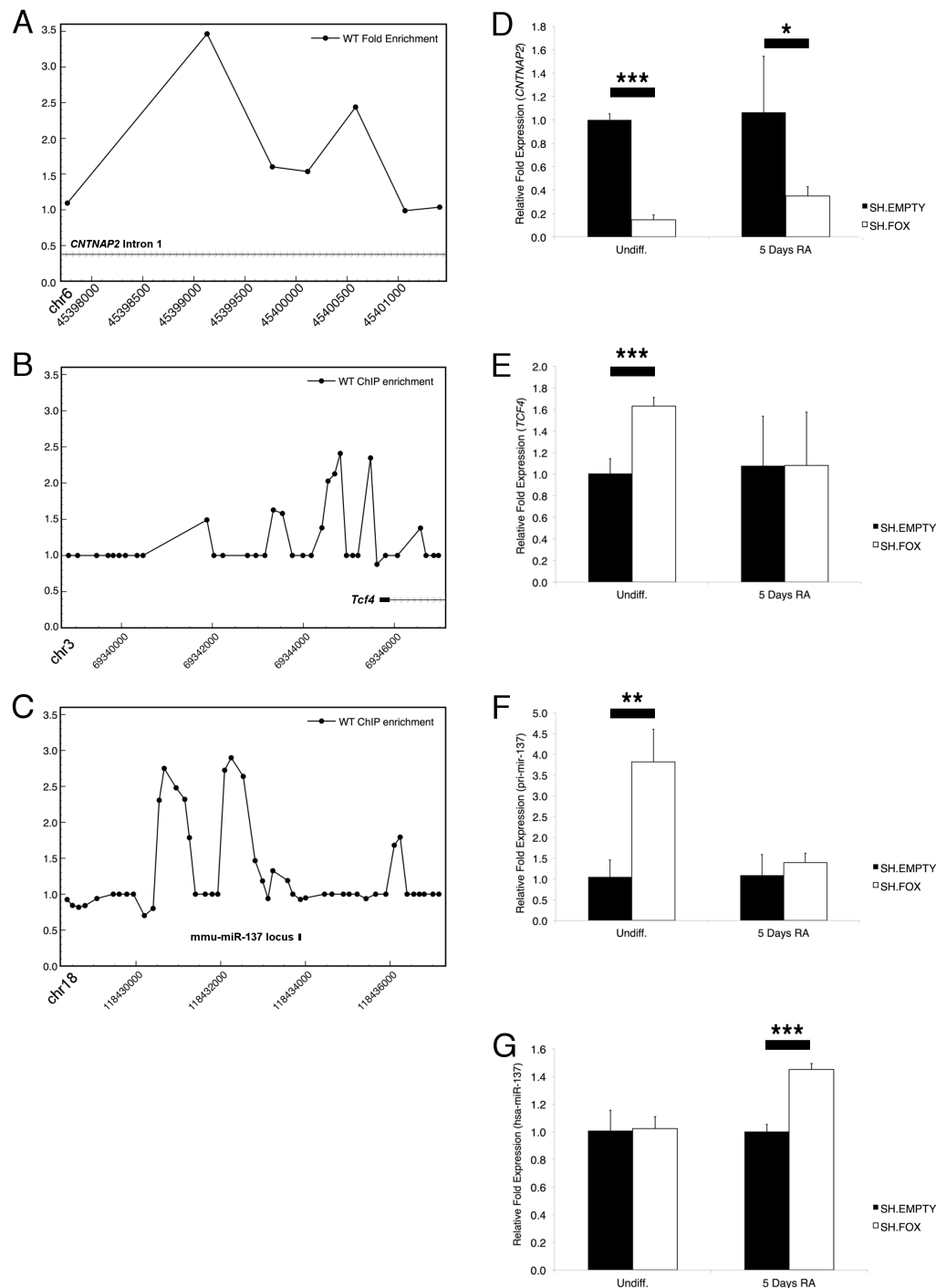


Figure 4.2. FOXP2 Directly Regulates *CNTNAP2*, *TCF4*, and miR-137.

Direct binding of Foxp2 in embryonic mouse brain was observed through previous ChIP studies (Vernes *et al.* 2011) at various loci, including **(A)** *Cntnap2*, **(B)** *Tcf4*, and **(C)** *mmu-miR-137*. The binding data shown here was adapted from the data first presented in Vernes *et al.* 2011. The mRNA levels of these genes in SH-SY5Y cells revealed significant regulation by FOXP2. **(D)** *CNTNAP2* showed a robust down-regulation in undifferentiated and retinoic acid-treated cells, while **(E)** *TCF4* was up-regulated only in undifferentiated cells. **(F)** The primary miR-137 transcript was up-regulated only in undifferentiated cells, but **(G)** mature miR-137 showed an opposite pattern. Significance

(as assessed by the unpaired two-tailed student's t-test) is indicated as follows: $P < 0.05$ – * ; $P < 0.01$ – ** ; $P < 0.005$ – *** .

4.2.2. Expression Patterns Suggest Genetic Interactions of *MIR137*, *CNTNAP2*, *FOXP2*, and *TCF4*

Having confirmed *CNTNAP2*, *TCF4*, and *hsa-miR-137* as direct transcriptional targets of *FOXP2*, I went on to investigate *in vivo* expression patterns of these four genes. Immunofluorescence of primary neurons from E16 mouse cortex and striatum showed *Tcf4* expression in *Map2*-positive neurons, but not in glial cells that do not express *Map2* (Figure 4.3). This is concordant with previous observations that did not observe *Tcf4* staining even after overexpression in rodent glial cells (Sepp *et al.* 2011). Given that *Foxp2* is expressed in both cortex and striatum as well during this developmental time-point (Shu *et al.* 2001, Vernes *et al.* 2011), this provides additional support for a functional genetic interaction between *Foxp2* and *Tcf4* in the brain.

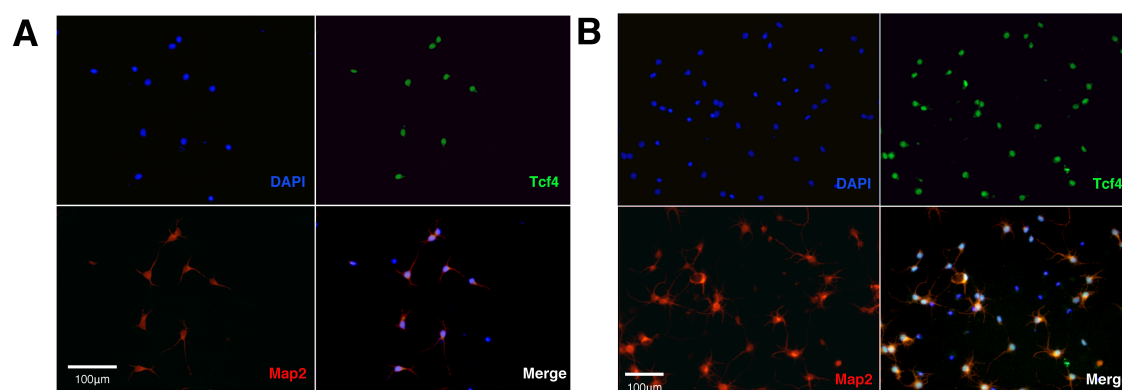


Figure 4.3. *Tcf4* is Expressed in Embryonic Mouse Brain.

Immunofluorescence was performed for *Tcf4* in embryonic mouse cortex (**A**) and striatum (**B**). *Tcf4* is expressed in the nuclei of neurons (which show staining with a *Map2* antibody), but not in that of glial or astrocytic cells (which do not express *Map2*).

By identifying any correlated gene expression patterns in the regions of the human brain with high *FOXP2* expression, we can infer the existence of genetic networks involving the four genes of interest here. To investigate this, I obtained microarray expression data from the Allen Brain Atlas (Hawrylycz *et al.* 2012) for these 4 genes in the striatum, across 6 adult human brain samples. It must be

noted that intra-regional differences in expression will not be discernible in the Allen Brain Atlas data. Thus, if the genetic interactions suggested in Figure 4.1 exist in specific neuronal subtypes within a brain region, the Brain Atlas data might not reflect it.

The Allen Brain Atlas microarray data showed that *FOXP2* was highly expressed in the striatum, which consists of the caudate nucleus and the putamen, sites of structural and functional abnormalities in affected KE members (Watkins *et al.* 2002b, Liégeois *et al.* 2003). I calculated the Pearson pairwise correlations of the normalized log-transformed expression levels of the 4 genes of interest, so as to ascertain the level of correlation between these genes in the striatum (Figure 4.4).

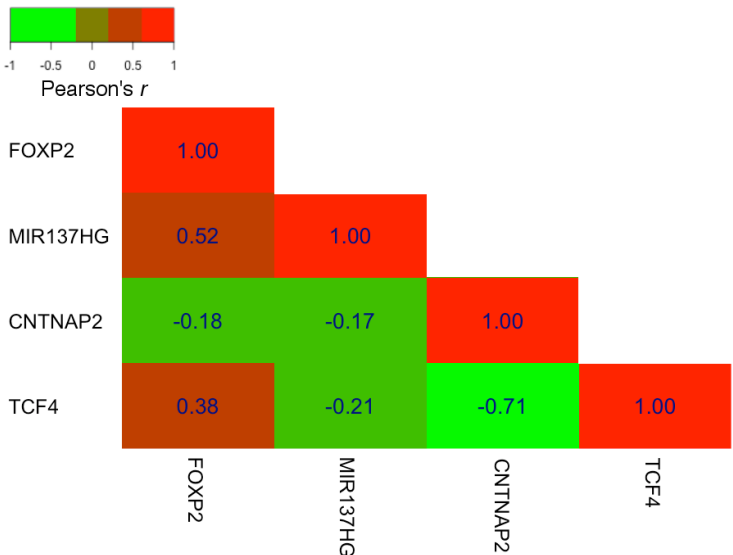


Figure 4.4. Correlation Heatmap of Expression Patterns of *MIR137*, *CNTNAP2*, *FOXP2*, and *TCF4* in Adult Human Striatum.

This heatmap depicts the correlation of log-transformed expression intensities of *MIR137*, *CNTNAP2*, *FOXP2*, and *TCF4*, in the striatums of 6 human adult brains. Each pairwise Pearson's *r* correlation is listed in the relevant cell. Expression data obtained from the Allen Brain Atlas (<http://human.brain-map.org/>, Hawrylycz *et al.* 2012).

Levels of *CNTNAP2* are relatively lower in the striatum and exhibit a weak negative correlation with *FOXP2* expression (Pearson's *r* = -0.18), while *MIR137* and *TCF4* levels both have moderate positive correlations with *FOXP2* expression (Pearson's *r* = 0.52 and 0.38 respectively). The pattern of pairwise correlations in the striatum, a region of high *FOXP2* expression, is suggestive of

possible genetic interactions as hypothesised in the proposed genetic network in Figure 4.1.

TCF4 is ubiquitously expressed across the adult brain, (Sepp *et al.* 2011), and has been shown to activate the *CNTNAP2* promoter *in vitro* (Forrest *et al.* 2012). Forrest *et al.* noted that overexpression of *TCF4* increased reporter gene expression at the *CNTNAP2* promoter up to a certain point, beyond which expression began falling (Forrest *et al.* 2012). This indicates a saturation effect at work, and might explain the relatively high negative correlation of *TCF4* and *CNTNAP2* expression (Pearson's $r = -0.71$) in the adult striatum.

4.2.3. The FOXP2-bound Intronic Region of *CNTNAP2* Is Insufficient To Produce Transcript Down-regulation

To test if direct binding of FOXP2 at the first intron of *CNTNAP2* is sufficient to produce repression, part of this intronic region was cloned upstream of a luciferase construct (Figure 4.5A–B). This region corresponds to that first identified through shotgun sequencing of DNA isolated from FOXP2-ChIP (Vernes *et al.* 2008). Overexpression of FOXP2 was previously shown to significantly repress the SV40 promoter, via a FOXP consensus binding motif (Vernes *et al.* 2006). Thus, I employed the native luciferase reporter driven by the SV40 promoter as a positive control for this assay.

In HEK cells, both the native luciferase reporter construct (p-value < 0.001), and the construct with the intronic FOXP2 binding site cloned upstream (p-value < 0.001) are repressed twofold when FOXP2 is overexpressed (Figure 4.5C). Thus, the intronic site, despite being bound by FOXP2, is not sufficient to produce additional repression of the reporter transcript. Interestingly, the binding site itself elevates the basal transcription of the construct by 1.5-fold relative to the empty construct, regardless of whether FOXP2 is being overexpressed (p-value < 0.01).

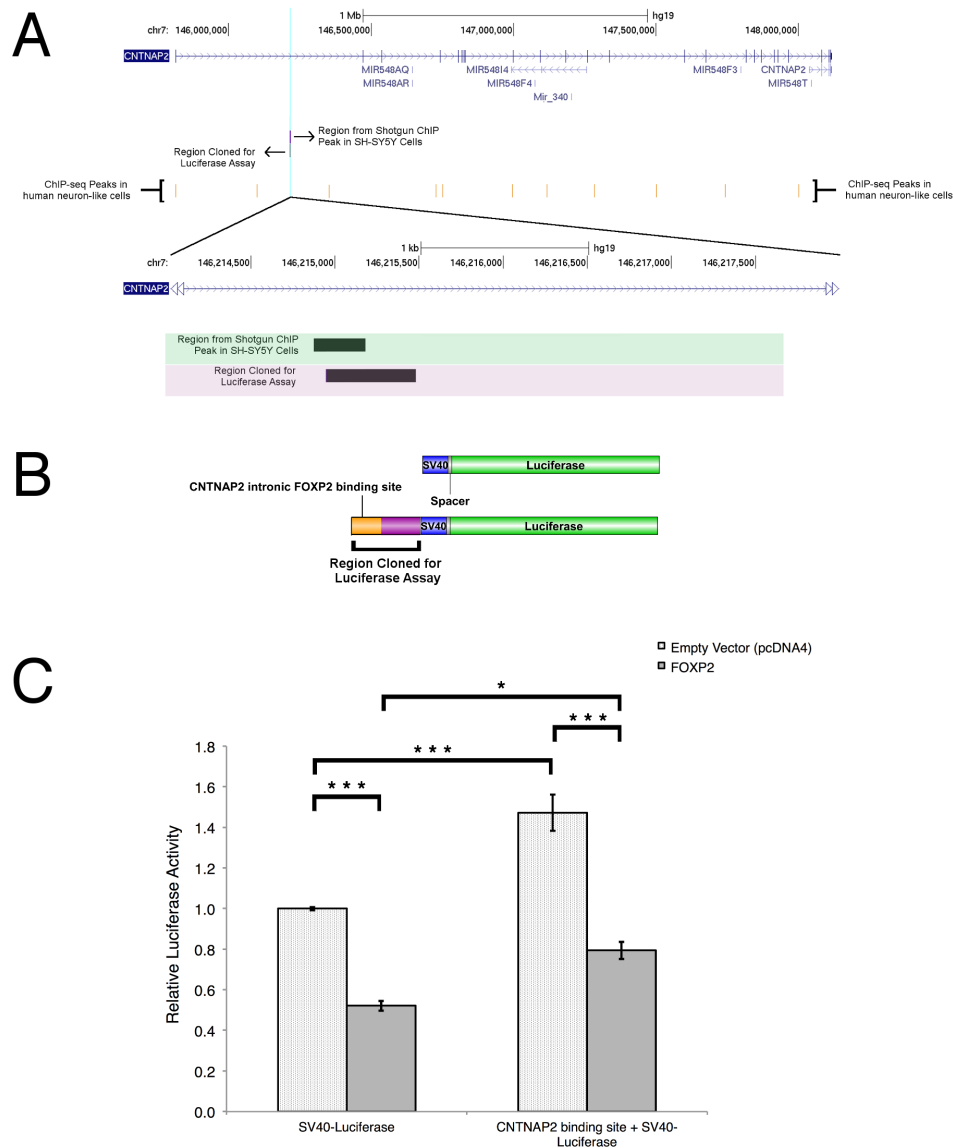


Figure 4.5. FOXP2 Regulation of CNTNAP2 Binding Site in HEK Cells.

(A) Genomic location of ChIP peak, and region cloned for luciferase assay. **(B)** Luciferase constructs were created, with the FOXP2 binding site cloned upstream of the firefly luciferase gene, but under the control of the SV40 promoter. **(C)** Luciferase results indicate over-expression of FOXP2 reduces reporter gene levels twofold, in both the native construct and the construct with a CNTNAP2 binding site cloned upstream. Thus, the binding site did not produce additional repression of the reporter gene. Significance (as determined by Tukey's post-hoc ANOVA) is indicated as follows: $P < 0.01$ – * ; $P < 0.005$ – ** ; $P < 0.001$ – ***

Table 4.1. Putative Transcription Factor Binding in the *CNTNAP2* Binding Site.

FactorName	Sequence	Position	Strand	Core Score	Matrix Score	Matrix
ARID3A	TTAATt	283	-	1	1	V\$DRI1_01
ARID3A	TTAATt	483	-	1	1	V\$DRI1_01
ARID3A	aATTAA	492	+	1	1	V\$DRI1_01
HDX	tttaagcAATCAagtgt	98	+	1	0.919	V\$HDX_01
HDX	ctgtaggAATCAgatga	438	+	1	0.913	V\$HDX_01
HMX1	ttaagcAATCAagtgt	99	+	0.822	0.864	V\$HMX1_02
HMX1	taaaccAATAAagtttg	524	+	0.814	0.814	V\$HMX1_02
HOMEZ	ttttaAACTAtgaccc	252	-	0.783	0.789	V\$HOMEZ_01
HOXB13	aaaccaATAAAgtttg	525	+	1	0.917	V\$HOXB13_01
IRX2	agaatACCTGtaggaat	431	+	0.803	0.851	V\$IRX2_01
IRX2	aagaataCCTGTaggaa	430	-	0.803	0.846	V\$IRX2_01
IRX2	tcatttCATTTaccca	158	-	0.788	0.837	V\$IRX2_01
Myogenin	CAGCTg	29	+	1	1	V\$MYOGENIN_Q6_01
Myogenin	cAGCTG	29	-	1	1	V\$MYOGENIN_Q6_01
POU2F1	tcattTACATt	158	+	0.96	0.972	V\$POU2F1_Q6
RUSH-1alpha	gcATAAGgtg	194	-	1	0.999	V\$RUSH1A_02
SIX1	tgaacgtTATCAtttca	365	+	1	0.849	V\$SIX1_01
SRY	TTGTTt	390	+	1	1	V\$SRY_Q6
ZNF333	ATAAT	36	+	1	1	V\$ZNF333_01
ZNF333	ATTAT	46	-	1	1	V\$ZNF333_01
ZNF333	ATAAT	225	+	1	1	V\$ZNF333_01
ZNF333	ATAAT	306	+	1	1	V\$ZNF333_01
ZNF333	ATAAT	461	+	1	1	V\$ZNF333_01

Shown here is a table of transcription factors (listed in the column ‘FactorName’) that are predicted to bind to the FOXP2-occupied region in *CNTNAP2*. The motif sequence predicted to be bound by the factor in the region of interest is listed in “Sequence”, with capital letters indicating the positions that match the core sequence of the matrix. The position of the match within the cloned region of interest, and the strand on which it is located, are listed in “Position” and “Strand” respectively. “Core Score” lists the similarity of the sequence to the core motif. “Matrix Score” indicates the similarity of the sequence to the TRANSFAC matrix, which is listed in “Matrix”.

The luciferase results suggest that this region acts as an enhancer, and the elevated basal transcription is likely due to other factors binding to this region. Thus, transcription factor binding sites were identified using the MATCH algorithm that employs binding site matrices annotated in the TRANSFAC database (Kel *et al.* 2003). Analysis of transcription factor binding sites in the *CNTNAP2* region of interest reveals that there are three sites for ARID3A binding. ARID3A (AT-rich interactive domain-containing protein 3A) can act as a transcriptional activator (Lin *et al.* 2007, Lestari *et al.* 2012), and thus might account for the elevated basal transcription of the reporter gene from the *CNTNAP2* binding site construct.

4.2.4. The FOXP2-bound Promoter/Intronic Region of *TCF4* Is Insufficient To Produce Transcript Up-regulation

Having confirmed the putative regulation of *TCF4* by FOXP2 *in vitro*, I went on to investigate if direct binding by FOXP2 is sufficient for up-regulation of the transcript (Figure 4.6A). I identified a FOXP2 binding peak in FOXP2 ChIP-seq data from the ENCODE project (ENCODE Project Consortium, 2011).

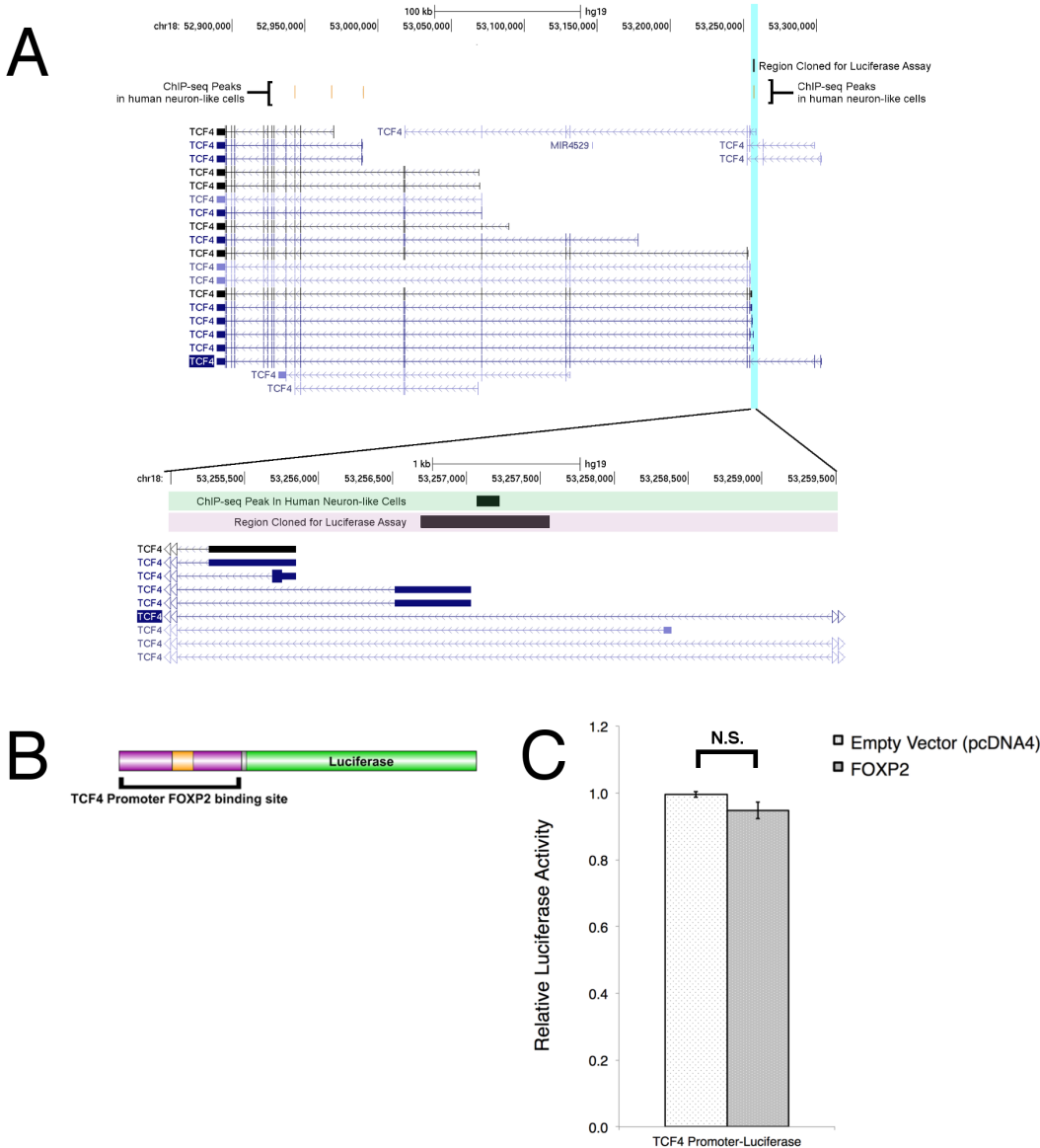


Figure 4.6. FOXP2 Regulation of TCF4 Binding Site in HEK Cells.

(A) Genomic location of ChIP peak, and region cloned for luciferase assay. **(B)** A luciferase construct was created, with the FOXP2 binding site / promoter region (indicated in gold) cloned upstream of the firefly luciferase gene. **(C)** Luciferase reporter assays indicate that FOXP2 overexpression does not modulate the transcriptional activity of this TCF4 promoter region.

Because this binding site overlaps with the TSS of a *TCF4* splice variant, it is likely to be a bona fide promoter, and was cloned in the place of the SV40 promoter used previously to drive luciferase transcription (Figure 4.6B). FOXP2 overexpression, however, did not produce a change in luciferase levels (Figure 4.6C). This is incongruous with the up-regulation of the transcript earlier observed in SH-SY5Y cell-lines (Figure 4.2E). Thus, this binding site is insufficient for FOXP2 to up-regulate the *TCF4* transcript.

4.2.5. The FOXP2-bound Regions in the *MIR137* Locus Are Insufficient To Produce Transcript Up-regulation

To test if direct binding by FOXP2 at the *MIR137* locus is sufficient for the observed transcriptional regulation, I again identified FOXP2 ChIP-seq binding peaks in the *MIR137* locus from the ENCODE project (Figure 4.7A). The first binding peak (Peak A) lies within the first intron of *MIR137*, and is 2.1 kb upstream of the mature miR-137, while the second one (Peak B) lies 4.4 kb upstream of the TSS. Both these peaks were each cloned upstream of a luciferase reporter gene driven by the SV40 promoter (Figure 4.7B).

Instead of up-regulating the luciferase reporter, as was expected after observing up-regulation of the miR137 primary transcript in SH.FOX cell-lines (Fig. 4.2), FOXP2 overexpression led to repression of both the Peak A ($P < 0.01$) and Peak B ($P < 0.001$) constructs (Fig. 4.7C). This suggests that both peaks are insufficient to produce up-regulation of *MIR137*, and are unable to overcome the repressive effect of FOXP2 on the native SV40 promoter.

Peak A has a strong repressive effect on basal transcription levels, resulting in a four- to five-fold reduction in luciferase reporter levels, with ($P < 0.01$) or without FOXP2 overexpression ($P < 0.001$). Peak B has a mild repressive effect on basal transcription that is not significantly different from that of the native reporter construct. This suggests that Peak A acts as a silencer element, which raises the possibility of repressive factors binding to Peak A and reducing the transcription of the luciferase reporter.

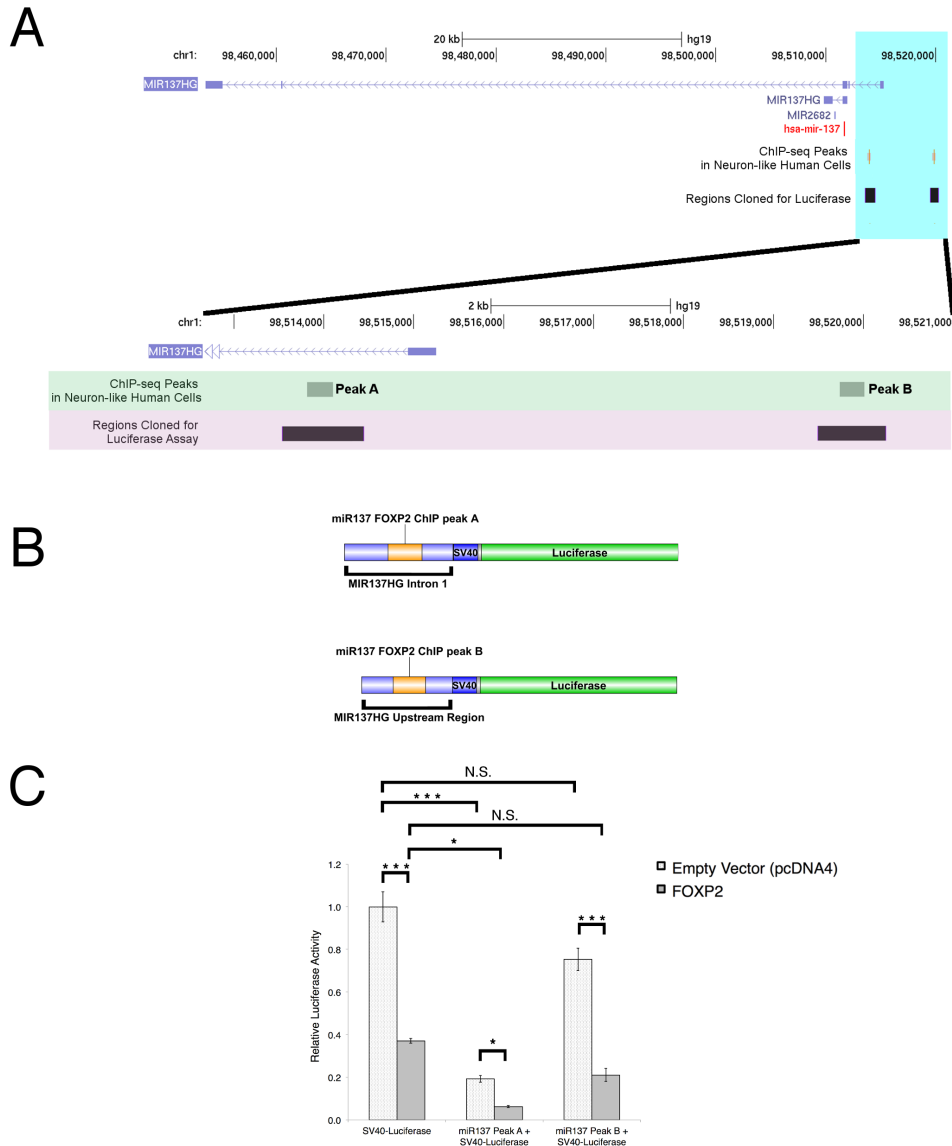


Figure 4.7. FOXF2 Regulation of MIR137 Binding Site in HEK Cells.

(A) Two ChIP-seq peaks – one in the intronic region of the *MIR137* gene, the other ~3.5kb upstream – are cloned for luciferase. **(B)** Luciferase constructs were created, with the FOXF2 binding sites (indicated in gold) cloned upstream of the firefly luciferase gene, but under the control of the SV40 promoter. **(C)** Luciferase results indicate that FOXF2 represses each of the reporter constructs by 2.5-fold. Significance (as determined by Tukey's post-hoc ANOVA) is indicated as follows: $P < 0.01$ – * ; $P < 0.005$ – * * ; $P < 0.001$ – * * * .

Again, TRANSFAC was used to uncover putative binding sites for transcription factors. A motif for the transcriptional repressor REST (Repressor Element-1 Silencing Transcription factor) was found in Peak A (Table 4.2). REST ChIP-seq data from ENCODE (Fig. 4.8) confirmed that this motif was bound by REST. This

suggests that REST might bind to Peak A with a higher affinity than FOXP2, exerting a silencing effect that counteracts FOXP2-mediated up-regulation.

FactorName	Sequence	Position	Strand	CoreScore	MatrixScore	Matrix
ARID5A	ccttttaATATtag	859	-	1	0.985	V\$ARID5A_Q3
BEN	ctcCGCTG	169	-	1	0.983	V\$BEN_Q1
BEN	CAGCGgca	357	+	1	0.986	V\$BEN_Q1
BPTF	ccgcAAAAACacacc	65	+	0.978	0.945	V\$FAC1_Q1
CDX2	TTTATa	802	+	1	1	V\$CDX2_Q5_Q2
CPBP	CCCCCtg	57	+	1	1	V\$CPBP_Q6
CPBP	aaGGGGGG	290	-	1	1	V\$CPBP_Q6
CPBP	gtGGGGGG	310	-	1	1	V\$CPBP_Q6
CPBP	GCCCCGc	403	+	1	1	V\$CPBP_Q6
DRI1	TTAAtI	645	-	1	1	V\$DRI1_Q1
E2F	tgccCGCAAaac	62	+	0.949	0.927	V\$E2F_Q6_Q1
GR	aggaaatggtGTTCTc	375	+	1	0.85	V\$GRE_C
HMX1	ttatttTAATTgtgagg	640	-	1	0.906	V\$HMX1_Q2
HOMEX	tgaccaTCCTTtgagt	786	+	0.76	0.794	V\$HOMEX_Q1
HOXB13	attctaATAAAcatgg	109	+	1	0.876	V\$HOXB13_Q1
HOXB13	tatatTTTATtttaat	634	-	1	0.875	V\$HOXB13_Q1
HOXC13	gattctaATAAAcatg	108	+	0.973	0.843	V\$HOXC13_Q1
HOXD12	tagtttcTTAAAGcgtg	593	+	0.848	0.804	V\$HOXD12_Q1
IKZF2	gtiTTCCcttt	854	-	1	0.996	V\$HELIOSA_Q2
ING4	TGGTGg	515	-	1	1	V\$ING4_Q1
KAISO	cIttGCAGGgag	125	-	1	0.991	V\$KAISO_Q1
KLF4	aaGGAGG	331	-	1	1	V\$GKLF_Q4
KLF4	CCTCCct	779	+	1	1	V\$GKLF_Q4
MAZ	gaaGGGGGGggagggg	289	+	0.981	0.956	V\$MAZ_Q6_Q1
MRF2	tgtGTATTgggggtg	731	-	1	0.943	V\$MRF2_Q1
MYB	ctgggaaCCGTCaatca	3	+	0.916	0.932	V\$MYB_Q5
NFATC2	tTTTCC	256	-	1	1	V\$NFAT1_Q4
NFATC2	GGAAAA	537	+	1	1	V\$NFAT1_Q4
NFATC2	tTTTCC	855	-	1	1	V\$NFAT1_Q4
NR5A1	gcggtggCCTTGag	23	-	1	0.993	V\$SF1_Q5
NR5A1	gaaagcaTATCAactgc	569	+	1	0.826	V\$SIX1_Q1
NR5A2	tggCCTTGaga	27	-	1	0.995	V\$LRH1_Q5_Q1
PBX1	gaaccgtCAATC	7	-	1	0.943	V\$PBX_Q3
REST	tccaggTGCTGaa	500	+	1	0.941	V\$REST_Q5
SIX1	gattcTGATGccctgtg	718	-	0.808	0.828	V\$SIX1_Q1
SOX10	aACAAAg	141	+	1	1	V\$SOX10_Q6_Q1
SP100	ggAGTCGatgaacaa	131	+	0.927	0.902	V\$SP100_Q4
SRY	TTGTTt	267	+	1	1	V\$SRY_Q6
SRY	TTGTTt	852	+	1	1	V\$SRY_Q6

Figure 4.8. The FOXP2 ChIP-seq Peak Overlaps with a REST Binding Peak.

4.3 Discussion

In this chapter, I validated *TCF4* and *MIR137* as targets of FOXP2, and provided evidence for a possible genetic mini-network between *FOXP2*, *CNTNAP2*, *TCF4*, and *MIR137* with co-expression data. Using luciferase reporter assays, I found that FOXP2 binding sites are, by themselves, insufficient to produce the observed transcriptional regulation of *CNTNAP2*, *TCF4*, and *MIR137* by FOXP2. Despite previous *in vivo* (Vernes *et al.* 2011) and *in vitro* (Vernes *et al.* 2008, (ENCODE Project Consortium 2012) evidence for FOXP2 binding at the loci of *CNTNAP2*, *TCF4*, and *MIR137*, FOXP2 overexpression did not produce the expected transcriptional regulation when these binding sites were cloned upstream of a reporter gene. The results of this chapter are summarised in Table 4.3.

Table 4.3. Summary of Luciferase Results.

Gene	Regulation by FOXP2	Effect of Binding Site on Transcription	<i>Effect of FOXP2 binding</i>
<i>CNTNAP2</i>	Down	Enhancer	Insufficient
<i>TCF4</i>	Up	Promoter/TSS	None
<i>MIR137</i>	Up	Peak A: Strong Silencer/repressor	Insufficient
		Peak B: Weak Silencer/repressor	Insufficient

4.3.1. Network Motifs in Proposed Mini-network

The network motif formed by the interactions of FOXP2, *TCF4*, and *CNTNAP2* is an incoherent feedforward one, where the direct and indirect (via *TCF4*) effects of a FOXP2 on *CNTNAP2* are antagonistic (Mangan and Alon 2003, see Figure 3.2). Such an interaction might act to preserve low basal levels of *CNTNAP2* in regions of high FOXP2 expression. In Area X of zebra finch, *FoxP2* transcript levels are higher (Teramitsu and White 2004) while *Cntnap2* transcript levels are

lower (Panitof *et al.* 2010) relative to the surrounding striatal tissue, suggesting a conserved repressive effect of FoxP2. However, Condro and White (2013) found that Cntnap2 protein levels were enriched in Area X relative to surrounding tissue. A feedforward network motif involving Tcf4 represents an attractive hypothesis to account for the low mRNA and high protein levels of Cntnap2 in Area X. Future work on the expression and function of this mini-network in the song circuit of the zebra finch will prove to be informative regarding the genetic networks underlying learned vocalizations.

4.3.2. Insufficiency of Binding Sites to Regulate Reporter Expression Suggests a Role of Chromatin Loops in *FOXP2* Function

A possible explanation for the discrepancy between FOXP2 occupancy and transcriptional regulation of the genes, and the insufficiency of the binding sites to produce reporter gene regulation, might lie in chromatin loops. Such loops arise from physical interactions between distal enhancer/repressor elements, promoter elements, DNA-binding factors, and cellular transcriptional machinery (Tan-Wong *et al.* 2008, Moabbi *et al.* 2012). Changes in chromatin loop conformation can be induced by transcription factor binding, and can increase or decrease gene expression (Jing *et al.* 2008, Nolis *et al.* 2009). It is possible that the FOXP2 binding sites studied here produce transcriptional regulation in a similar fashion, by altering looped chromatin conformation in response to FOXP2 binding, and therefore require a certain genomic context (such as distance from promoter or enhancer sites) to function.

Future work can identify the role of chromatin loops in FOXP2-mediated regulation by employing chromosome conformation capture (abbreviated as 3C, and first described in Dekker *et al.* 2002). 3C with a FOXP2-specific antibody will identify distal regions involved in long-range loop-forming interactions with FOXP2-bound regions. Proteins that bind to these distal fragments might assist in producing chromatin loops by interacting with FOXP2, and thus represent

putative binding partners. These proteins can be identified with bioinformatic prediction, and confirmed with DNA-binding assays.

To fully study how *Foxp2* interacts with the three-dimensional structure of these target genes, transgenic mouse lines, where the *Foxp2* binding site of interest is deleted or mutated, can be created using a bacterial artificial chromosome (BAC) (Heintz 2001). This will provide unambiguous information about the *in vivo* function of the binding sites. Given the substantial costs of generating such mice, however, an alternative is to test position effects of binding sites by replicating or imitating their context in a reporter construct. Rather than conventional cloning of these sites upstream of a promoter (as was performed here), future work could introduce the *FOXP2* binding site as an intron, or distal to the reporter gene, to allow for loop formation. Future experiments along these proposed lines will reveal more about the regulation of *CNTNAP2*, *TCF4*, and *MIR137*, and will shed light on the molecular mechanisms underlying the neurological disorders associated with these genes.

4.3.3. Conclusion

The results presented here point to a complexity of *FOXP2* regulation beyond conventional models where *FOXP2* directly binds to a promoter region and effects transcriptional changes. As regulatory relationships continue to be uncovered, the challenge will be functionally integrating individual interactions into a coherent network. It will be interesting to determine what functions the regulatory mini-network presented here have in the neural pathways important for vocal learning. Furthermore, the effects of pathogenic mutations in *FOXP2*, *TCF4*, and *MIR137* on the mini-network can be studied in cell-based models, and will provide insight into how disease-associated mutations disrupt genetic pathways underlying a range of neurodevelopmental disorders.

Chapter 5: Are MicroRNAs Risk Factors For Specific Language Impairment?

5.1 Introduction

5.1.1 Defining Specific Language Impairment

Specific Language Impairment (SLI) is a disorder in children who have delayed language acquisition despite the absence of possible causes, such as deafness, insufficient educational opportunity, and abnormalities in the physical apparatus for speech. Its presentation usually includes a limited vocabulary relative to peers, and difficulty in understanding complex speech (Bishop 2006). Diagnosis of SLI is often differential, and seeks to exclude conditions that might explain the presence of language impairment, such as low non-verbal intelligence, epileptic aphasia, selective mutism, and autism (Bishop and Norbury 2008). The development of quantitative assessment tests that score a child on various domains of receptive and expressive language abilities have enabled standardised diagnosis (Tomblin *et al.* 1996). Generally, children with scores 1 – 1.5 standard deviations (SD) below the age-appropriate mean for either expressive or receptive language abilities, despite having normal nonverbal IQ scores, are given a diagnosis of SLI. It must be noted, however, that despite these quantitative means of diagnosis, there is no standard set of criteria that definitively warrant a diagnosis of SLI.

Up to 7% of English-speaking pre-primary schoolchildren in the UK are diagnosed with SLI (Law *et al.* 1998). Another study of kindergarten children in the United States estimated a prevalence rate of 7.4% (Tomblin *et al.* 1997). As such, SLI is one of the most common developmental disorders in children (Bishop 2010). Children with SLI are also at higher risk of exhibiting behavioural disorders, such as hyperactivity, social anxiety, and aggressive social interactions (Tomblin *et al.* 2000). They also have persistent language problems during adolescence and adulthood (Beitchman *et al.* 1994), poorer post-secondary

educational outcomes (Snowling *et al.* 2001), and suffer disproportionately higher rates of psychiatric and developmental disorders (Cantwell and Baker 1987).

Endophenotypes of SLI have been shown to enhance diagnostic power and contribute to theoretical understanding (Gathercole and Baddeley 1990, Bishop 2006). An endophenotype of SLI that has been closely studied is nonword repetition (NWR). It is a measure of phonological short-term memory, which is a type of working memory used in processing novel speech sounds during the articulatory loop (Gathercole and Baddeley 1990). Tests of NWR entail a series of nonsense words, such as “blonterstaping” or “woogalimic”, being spoken to the subject, who then has to repeat the words. In normal school children, performance on the test was strongly correlated to language ability (Gathercole *et al.* 1994). Several studies have found that children with SLI have poorer NWR performance compared to normal language controls. (Weismer *et al.* 2000, Ellis Weismer *et al.* 2000, Conti-Ramsden *et al.* 2003). Individuals whose overt language difficulties resolved subsequently continue to perform poorly at NWR (Bishop *et al.* 1996, Stothard *et al.* 1998). Taken together, this is strong evidence that poor NWR is a core deficit of SLI, and is a robust endophenotype that persists even when learned compensatory strategies produce near-normal language performance.

5.1.2 The Search for Genetic Factors Underlying SLI

SLI is a thought to be a complex disorder with heterogeneous factors contributing to it, including genetic, epigenetic, and environmental ones (Bishop *et al.* 1995, Stromwold 2006). Nevertheless, several lines of evidence have established a clear genetic component underlying SLI. It shows strong familial aggregation: affected individuals are more likely to have first-degree relatives who also are affected (Neils and Aram 1986, Tallal *et al.* 1989, Choudhury and Benasich 2003, Conti-Ramsden *et al.* 2007). Several twin studies (Lewis and Thompson 1992, Bishop *et al.* 1995, Tomblin and Buckwater 1998) have shown that monozygotic

twins tend to have a higher concordance rate (70% – 96%) compared to that of dizygotic twins (46% – 69%).

In light of these findings suggesting a genetic basis for SLI, several groups have employed whole-genome linkage methods, and targeted association studies, to identify genetic regions that confer susceptibility for SLI. Their findings are summarised in Table 5.1, and are discussed in detail below.

Table 5.1. Genome-wide Studies Have Uncovered 5 Loci and Several Candidate Genes for SLI.

Study	Cohort	Study Design	Findings
SLI Consortium 2002	473 individuals with SLI (including a total of 219 sib pairs) from 98 families were recruited through two centres in the UK. Each proband was assessed to have language skills ≥ 1.5 SD below the mean scores for their age group.	Genome-wide family-based univariate QTL linkage	Two regions (chr16q and chr19q, SLI1 and SLI2 respectively) were identified. The 16q locus showed significant linkage to nonword repetition, while the 19q locus had significant linkage with expressive language scores.
Bartlett et al. 2002	86 individuals collected from several branches of five Canadian families (two nuclear, and three extended) with Celtic ancestry.	Genome-wide categorical linkage	Linkage of reading impairment to the chr13q21 locus, known as SLI3 , was observed under a recessive model.
SLI Consortium 2004	840 subjects (393 sib pairs) were recruited from 184 families by four separate centres. This is an expansion of the cohort in their 2002 paper.	Replication family-based univariate QTL linkage	An amalgamated cohort (consisting of original sibpairs with newer ones) replicated the linkage of the SLI1 locus to nonword repetition (and found additional linkage to reading-related measures), while the SLI2 locus was linked now to both nonword repetition and expressive language scores.
Bartlett et al. 2004	The cohort from Bartlett et al. 2002 was used, but included 22 nuclear and extended families from the US consisting of 19 Caucasian and 3 Hispanic families.	Replication family-based QTL linkage	The newer US cohort replicated the linkage of SLI3 to reading impairment.
Monaco and the SLI Consortium 2007	Same as pooled cohort in SLI Consortium 2004.	Genome-wide family-based multivariate QTL linkage	Multivariate analysis supported previously observed linkage of NWR to SLI1, as well as linkage of both expressive and receptive language traits to SLI2. A putative novel QTL was also identified at chromosome 10q26.
Vernes et al. 2008	184 families from the cohort previously reported in SLI Consortium 2004 were used.	Family-based candidate gene association and <i>in vitro</i> characterization	9 intronic SNPs in <i>CNTNAP2</i> were found to be significantly associated with nonword repetition. The <i>CNTNAP2</i> locus is also referred to as SLI4 .
Newbury et al. 2009	806 individuals from 211 families recruited by the SLI Consortium (SLIC), followed with a replication sample of 490 cases from the ALSPAC cohort.	Family-based QTL association screen of SLI1 locus, followed by replication in a language impaired cohort selected from general population.	Association to 2 genes in the SLI1 locus was identified - CMIP and ATP2C2.
Wiszniewski et al. 2013	15 unrelated families of predominantly Southeast Asian descent.	Array CGH, followed by family-based segregation and linkage analysis	A complex 4kb deletion in the <i>TM4SF20</i> gene (SLI5) was identified that was fully penetrant for early childhood language delay in all the families studied. This region was linked to language delay in 6 pedigrees with a LOD score of 3.1.
Nudel et al. 2014	Same cohort as used in Newbury et al. 2009, but with an additional 49 families first reported in this study.	Targeted quantitative association analysis and case-control association analysis of SNPs and imputed haplotypes of the HLA locus on chromosome 6p	Quantitative association analyses identified the HLA-A A1 allele as associated with NWR, and the HLA-A A3 allele as associated with ELS. Case-control analysis identified two protective alleles and one risk allele in the HLA-DRB1 gene.

A study by the SLI Consortium recruited 473 probands (including 219 sib pairs) from 98 families (SLI Consortium 2002). These families were identified either through clinical referral or through a long-term epidemiological study of speech and language difficulties in a community-based setting. Three quantitative

measure of language ability were employed in this study – an expressive language score (ELS), a receptive language score (RLS), and a NWR score. The first two were obtained with the Clinical Evaluation of Language Fundamentals–Revised (CELF-R, Semel *et al.* 1992), while the nonword repetition test first published by Gathercole *et al.* (1994) was used to obtain NWR scores. Probands were selected on the basis of their ELS, RLS, and nonverbal IQ scores: each had language scores 1.5 SD below the norm for their age group, and nonverbal IQ scores within the normal population range (> 80). Probands were also screened to exclude cases with low IQ, deafness, and autism. All three measures were obtained for all siblings, both affected and unaffected.

Two quantitative-trait loci, or QTL, were identified on chromosome 16q (SLI1, MIM 606711) and chromosome 19q (SLI2, MIM 606712). Significant linkage to NWR score was identified at the 16q locus, while the ELS was significantly linked to the 19q locus. The linkage of SLI1 to NWR was replicated in a second cohort of 86 families, who were recruited with the same criteria and phenotyped under the same test battery (SLI Consortium, 2004). The SLI2 locus was significantly linked with NWR, but not ELS, in this second cohort. When the newer cohort was combined with the previous cohort (resulting in a set with 840 subjects across 393 sib pairs in 184 families), a more significant linkage to NWR at SLI1 was identified, but no significant linkage of NWR or ELS to SLI2 was seen. The authors suggested that the discrepancy of results at the SLI2 locus might be due to the diversity of the cohort, which consisted of probands across 4 different clinical centers.

The SLI Consortium followed up these 2 studies with a genome-wide multivariate linkage analysis on the pooled cohort (Monaco and the SLI Consortium, 2007). A total of 13 traits were investigated, including expressive and receptive language ability, reading ability, and non-verbal IQ. This study provided additional support for the previous linkage of chromosomes 16q (SLI1) and 19q (SLI2) to language impairment. This multivariate study also identified a putative QTL for SLI at chromosome 10q26.

Bartlett and colleagues looked at a cohort of 84 probands that were collected across five extended Canadian pedigrees (Bartlett *et al.* 2002). These families were first identified through a linkage study on schizophrenia, and were observed to have a history of reading and language difficulties. Therefore, individuals with schizophrenia or schizophrenia spectrum disorders were excluded. The remaining subjects were classified into three non-exclusive qualitative categories: language impaired, reading impaired, and clinically impaired. A subject was classified as language impaired if their score on the Spoken Language subtest of the Test of Language Development (abbreviated TOLD, Hammill and Newcomer 1988) more than one SD below the mean. A subject was classified as reading impaired if their nonword reading score on the Woodcock Reading Mastery Test (Woodcock 1987) was one SD below their performance IQ. Lastly, a subject was classified as clinically impaired if they were language impaired or reading impaired as above, or if the subject had a history of language difficulty for at least two years during childhood. Individuals who scored one SD below the mean of three individual subtests of the TOLD, or scored ≤ 85 on a receptive language test, were classified as clinically impaired as well. Linkage to 13q21 was found for reading impairment under a recessive model of inheritance (SLI3, MIM607134). They were able to replicate this linkage in an expanded sample that included 22 clinically-referred American families diagnosed with the same criteria as before (Bartlett *et al.* 2004).

Several follow-up studies also employed the family-based cohort from the SLI Consortium in candidate association studies. Newbury *et al.* (2009) performed a higher-resolution screen of the SLI1 locus and its association with nonword repetition. An expanded cohort was studied that included cohorts in the 2002 and 2004 SLI Consortium studies, as well as additional nuclear families selected with the same criteria as before. This brought the final cohort to a total of 211 families. They identified significant association of NWR score to variants lying in the genes *CMIP* and *ATP2C2*. This finding was replicated in a general-population cohort of 490 cases with language scores significantly lower than the mean.

A second study by Vernes and colleagues (Vernes *et al.* 2008) performed a candidate gene study to assess if *CNTNAP2* was associated with speech and language deficits. Vernes and colleagues first showed that it was a direct downstream target of FOXP2. *CNTNAP2* is expressed in the human brain (Abrahams *et al.* 2007, Alarcón *et al.* 2007) and was previously shown to be involved in a range of neurodevelopmental disorders affecting language ability, such as autism and intellectual disability (Strauss *et al.* 2006, Bakkaloglu *et al.* 2008). Thus, they selected it as a candidate for association to language ability in the SLIC cohort. Across 39 SNPs in *CNTNAP2*, 9 intronic SNPs (lying between exons 13 and 15) were significantly associated with lower nonword repetition (SLI4, MIM 612514). Taken together, this provided a strong case that *FOXP2* and its downstream targets are involved in genetic networks underlying human speech and language.

More recently, Wiszniewski *et al.* (2013) studied 15 unrelated families (a total of 52 individuals) of Southeast Asian descent. In all these families, a heterozygous complex 4kb deletion at chr 2q36.3 was found that disrupted the *TM4SF20* gene (SLI5, MIM 615404), causing the removal of the penultimate 3rd exon as well as introducing a premature stop codon in the terminal 4th exon. The deletion was fully penetrant for early childhood language delay, and was incompletely penetrant for white matter abnormalities. Linkage analysis was performed on 6 of these pedigrees, yielding a LOD score of 3.1 for the language delay phenotype. Interestingly, this deletion allele had a frequency of 1% in the Vietnamese population, and was shown to be an ancestral haplotype in Southeast Asian populations. Wiszniewski *et al.* also performed *in vitro* studies, and found that the mutant *TM4SF20* transcript was stable and did not undergo nonsense-mediated decay. The mutant gene product, however, was consistently mislocalised to the cytoplasm, in contrast to wildtype localization to the cell membrane.

A third study (Nudel *et al.* 2014) and performed targeted association of the human leukocyte antigen (HLA) locus with SLI, and included 49 additional British families in addition to the cohort studied by Newbury *et al.* 2009. Both quantitative

association analysis and case-control analysis were employed. The quantitative association analyses identified the *HLA-A* A1 allele as associated with NWR, and the *HLA-A* A3 allele as associated with ELS. The case-control analysis identified two protective alleles and one risk allele in the *HLA-DRB1* gene. Therefore, the study suggested a possible contribution of the HLA locus to SLI risk, but its findings await replication in independent cohorts.

5.1.3 Exome Sequencing and Neurodevelopmental Disorders

Recent advances in DNA sequencing technologies have greatly transformed the search for genetic factors underlying neurodevelopmental disorders. Previously, researchers used sparsely-spaced polymorphic markers across the genome to identify loci associated with the disorders. Today, however, they are able to interrogate the vast majority of coding sequences in the human genome for mutations in affected individuals. This is known as exome sequencing, and uses tagged oligonucleotide probes complementary to exomic sequences, which are hybridised against sheared genomic DNA. The captured genomic fragments are then sequenced using high-throughput platforms, followed by mapping to reference genomes, culminating in the identification of possible causal variants (Ng *et al.* 2009).

Geneticists have successfully used exome sequencing to identify rare Mendelian mutations highly penetrant for familial forms of neurological disorders. Recent examples include the identification of recessive variants in *WDR62* in an inherited form of severe brain malformation (Bilgüvar *et al.* 2010), and two different heterozygous mutations in *TGM6* in two pedigrees with familial spinocerebellar ataxia (Wang *et al.* 2010). These approaches are complementary to classical linkage methods, and are often employed to confirm the findings of exome approaches.

Exome sequencing has also been used to sequence families with sporadic cases (affected individuals who have unaffected parents) of complex neurological disorders, such as autism (Neale *et al.* 2012, O'Roak *et al.* 2011, O'Roak *et al.*

2012, Sanders *et al.* 2012, lossifov *et al.* 2012, lossifov *et al.* 2014), schizophrenia (Girard *et al.* 2011, Xu *et al.* 2012, Gulsuner *et al.* 2013, Fromer *et al.* 2014, McCarthy *et al.* 2014), and intellectual disability (de Ligt *et al.* 2012, Gilissen *et al.* 2014). This approach has proven useful in detecting *de novo* protein-coding variants. *De novo* variants are mutations possessed by the proband, but not by the unaffected parents, and are thus likely to confer substantial genetic risk to these sporadic cases. Such a family-based approach has revealed evidence for a “multiple-hit” model of sporadic autism, where a *de novo* variant in a susceptibility gene occurs together with an inherited variant in a different risk gene (O’Roak *et al.* 2011).

Exome sequencing has also been applied on cohorts of affected unrelated individuals in neurological disorders, such as autism (Lim *et al.* 2013), childhood apraxia of speech (CAS) (Worthey *et al.* 2013), Schinzel-Giedion syndrome (Hoischen *et al.* 2010), schizophrenia (Purcell *et al.* 2014), and SLI (Chen *et al.*, in prep.). This approach has identified novel variants in the same genes shared by the study cohort. Importantly, it is able to identify novel risk factors that escaped identification by previous association efforts (such as *SETBP1* in Schinzel-Giedion syndrome). Exome sequencing has also uncovered variants in genes that are risk factors in related phenotypes, providing evidence for shared genetic networks underlying different neurological disorders: This is exemplified in the results of Worthey *et al.* 2013, which found, among others, an individual with CAS carrying a nonsynonymous variant in *KIAA0319*, a gene previously associated with dyslexia (Paracchini *et al.*, 2006). However, this finding does not necessarily imply that *KIAA0319* is a risk factor for CAS as well as dyslexia, and highlights the need to perform functional testing of exome-derived mutations.

5.1.4 From Variants To Function: Validating Exome-derived Coding Mutations

While progress has been made in uncovering rare risk variants for neurological disorders, it remains difficult to ascertain the extent these variants have on protein structure and function. This further compounds the difficulty in attributing

causative risk to these variants. Computational predictions of coding changes – such as SiFT (Kumar *et al.* 2008) and PolyPhen (Adzhubei *et al.* 2010) – do not easily extrapolate any resultant deleterious effects on biochemical properties. Flanagan and colleagues have shown that both SiFT and PolyPhen have high sensitivity but low specificity in predicting pathogenicity (Flanagan *et al.* 2010). Both tools were also better at predicting loss-of-function than gain-of-function mutations.

In lieu of traditional molecular studies, analysis of co-expression and protein-protein interaction networks has been used (O’Roak *et al.*, 2012). In the case of O’Roak and colleagues, this approach was appropriate as they sought to understand 126 *de novo* mutations across 209 nuclear families with cases of sporadic autism. While this network approach has revealed genetic pathways that might underlie complex neurological disorders, classical functional studies have yet to be performed.

Some successful validation efforts take advantage of the fact that novel variants lie in genes previously associated to disease, and use well-studied experimental models to demonstrate pathogenicity (O’Roak *et al.* 2011). Other studies are able to show loss-of-function in novel disease genes first uncovered from exome data (Lee *et al.* 2011, Veeramah *et al.* 2013). However, there may not be a disease-relevant property of the protein that can be easily measured in the lab, even if the mutation produces an obvious drastic effect, such as a premature stop codon or a frameshift (Deriziotis and Fisher 2013).

5.1.5 From Variants To Function: Validating Noncoding Mutations

In terms of time and cost, it can be argued that noncoding changes predicted to affect microRNA binding are easier to investigate, compared to coding mutations whose function on the protein might not be immediately apparent. Functional investigation of such noncoding changes involves a straightforward reporter gene assay, where a 3’ UTR of interest is cloned downstream of a reporter gene (typically the firefly luciferase). This construct is then transfected into an *in vitro*

model of choice, along with the mature form of the microRNA of interest, or another construct that over-expresses it. This is a rapid and direct test of the effect of a variant (which could lie in a microRNA seed region itself, or in a predicted binding site in the 3' UTR of a gene) on microRNA binding. It is still necessary to perform functional assays – even if the point mutation lies in a predicted seed region, and even when computational simulation of RNA hybridization shows a divergence in the binding abilities of the wildtype and variant binding sites – for the same reasons that limit the usefulness of predicted effects of protein-coding point mutations.

Despite being expressed as part of gene transcripts, several intronic miRNAs, as well as both 5' and 3' UTRs, were not captured by initial versions of exome-sequencing platforms. More recent iterations, however, have the capacity to target these noncoding regions. Given the ease of functional validation, I am interested in identifying noncoding variants in the 3' UTR that lie in predicted microRNA binding sites, and then rapidly testing them *in vitro* for functional relevance.

5.1.6 Genetic Mutations in 3' UTR That Affect MicroRNA Binding Can Cause Human Disease

There is a growing number of direct links between microRNA target site polymorphisms in 3'UTRs of genes, resultant altered expression of that gene, and the occurrence of human diseases. Two patients with Tourette's syndrome were found to possess a rare variant in the 3'-UTR of the *SLITRK1* gene (Abelson *et al.* 2005). This variant created a binding site of greater affinity for hsa-miR-189 relative to the wild-type allele, leading to enhanced repression of *SLITRK1* expression. A genetic association study linked a variant of the hsa-miR-433 target site in the *FGF20* gene to Parkinson's disease (Wang *et al.* 2008). More recently, a mutation in the 3'-UTR of the *HDAC6* gene was found to cause a

dominant X-linked form of chondrodysplasia, by abolishing the ability of hsa-miR-433 to regulate the *HDAC6* transcript (Simon *et al.* 2010).

Common variation in microRNA binding sites can also affect target gene levels and have a phenotypic effect. A recent study by Li *et al.* (2012) showed that SNPs in the 3' UTRs of *SMNDC1*, *SLC25A19*, and *TYRP1* affected regulation of the gene transcript by mature miR-329, miR-122, and miR-155 respectively. This raises the possibility of screening the 3' UTRs of known genetic risk factors for variants affecting microRNA binding and function, and also testing such variation as candidate risk factors.

5.1.7 Genetic Mutations in MicroRNAs Can Affect Cellular Function, And Cause Human Disease

Classical linkage methods have uncovered point mutations in the seed region of microRNAs. At the time of writing, only three published studies have found linkage to a point mutation in the seed region of a microRNA. Men  cia and colleagues published the first report of 2 Spanish pedigrees each with a novel mutation in the seed region of hsa-miR-96 that segregated with autosomal dominant inherited progressive hearing loss (Men  cia *et al.* 2009). They also showed that the mutant pre-mir-96 affects biogenesis of the mature form, and thus affects downstream down-regulation of miR-96 target genes. A second study by Hughes and colleagues followed up a previous linkage of familial severe keratoconus (in a Northern Irish pedigree) to chr15 q22-q25 by performing targeted resequencing of all genes in that region (Hughes *et al.* 2011). They found a mutation in the seed region of hsa-miR-184 in affected family members that prevented mature miR-184 from competing with miR-205 for target sites on the 3' UTR of two target genes, resulting in hyper-repression of the target genes. Another group of researchers found the same mutation in another pedigree that segregated with a familial occurrence of a set of ophthalmological syndromes known as EDICT (Iliff *et al.* 2012).

Variation outside of the mature seed region can also have biological effects. Such variants – which might lie in the hairpin of the precursor, or outside of it but within the primary transcript – are hypothesised to affect secondary structure, thermodynamic stability, and processing by microRNA biogenesis enzymes such as Drosha and Dicer (Li and Zhang 2013). It was recently shown that SNPs lying in the hairpin regions of pre-miR-612 and pre-mir-1206 had an effect on the biogenesis and resultant levels of mature miR-612 and mature miR-1206, respectively (Kim *et al.*, 2012). The exact molecular mechanisms by which such variation might affect microRNA biogenesis and maturation, however, remain poorly understood. The downstream effects of such variants on genes targeted by microRNAs are also not well documented.

5.1.8 Aims of This Chapter

Given that FOXP2 regulates a set of microRNAs, I hypothesised that microRNAs might be genetic risk factors for related speech and language impairments, in the same fashion that a FOXP2 target, *CNTNAP2*, is associated with nonword repetition in an SLI cohort (Vernes *et al.* 2008).

To test this hypothesis, the cohort collected by the SLI Consortium was used. In addition to previous families studied in the past (SLI Consortium 2002, SLI Consortium 2004, Vernes *et al.* 2008, Newbury *et al.* 2009), this cohort consists of 278 nuclear families with at least one SLI proband. A related study selected 48 probands from this cohort for whole-exome capture with the Agilent SureSelect v2 platform, followed by high-throughput sequencing with the SOLiD 5500xl System (Chen *et al.*, in prep.). Both the family-based genome-wide association data and the exome data from SLI probands were analysed in this chapter.

I first performed candidate SNP association analysis on the family-based cohort using a list of SNPs lying in FOXP2-regulated primary microRNA transcripts. As variation in target sites in the 3' UTRs can also affect microRNA-mediated gene regulation, I also identified variants lying in these regions from the exome sequences of 48 SLI probands. The variants I uncovered were subsequently

used as candidate SNPs to test for QTL association in the larger SLI cohort. Lastly, I subjected these identified variants to a reporter gene assay to test if they truly do exert an effect on microRNA regulation of the gene transcript.

5.2 Results

5.2.1 Candidate Association of SNPs lying in 13 FOXP2-regulated MicroRNAs to SLI

The SLIC cohort used here has been previously described by Nudel *et al.* 2014a and 2014b, and the relevant statistics of the score distributions for the children in the SLIC cohort studied here are listed below Figure 5.1, as reported by Nudel *et al.* 2014b. These scores are standardized to a general-population calibration sample with mean 100 and SD 15 (SLI Consortium 2002).

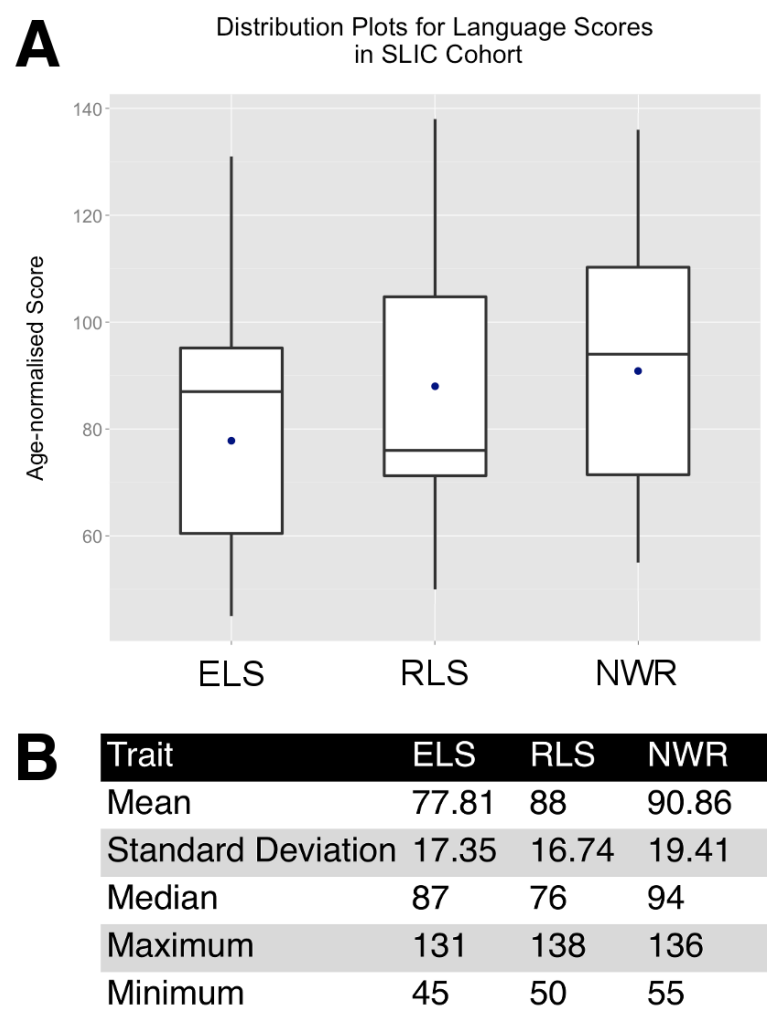


Figure 5.1. Distribution of Language Scores for Children.

(A) Boxplots for the distribution of the three language traits in the children from the SLIC cohort used here, as reported by Nudel *et al.* 2014b. The line inside each boxplot indicates the median score, while the blue dot indicates the mean score. The white box indicates ± 1 standard deviation of scores around the mean. **(B)** The score distributions

are shown here in tabular format. ELS – Expressive Language Score; RLS – Receptive Language Score; NWR – Nonword Repetition Score.

Recent work has shown that SNPs in miRNA primary transcripts (Ripke *et al.* 2012) and in miRNA precursor regions (Kim *et al.* 2012) are associated with human disorders and have biological effects. Thus, the 13 primary transcripts, which give rise to the 12 microRNAs previously identified as FOXP2 targets in Chapter 3, were used as candidate regions for association in the SLI cohort (hsa-miR-181b is produced from 2 precursors in the human genome, known as mir-181b-1 and mir-181b-2.).

The coordinates of these primary transcripts (Table 5.2) were obtained from literature (Saini *et al.* 2008, Marson *et al.* 2006). If the microRNA was located in a protein-coding host gene, the entire genomic region of the gene was taken as the primary transcript for that microRNA. The transcript for hsa-mir-124-3 was not covered in the cohort, resulting in the final list of transcripts spanning a total size of 365.9kb. These transcripts ranged from 2kb to 105kb, with an average length of 30.5kb. Within these 13 loci, 726 SNPs were either genotyped or imputed in the SLIC SLI cohort. The average SNP density across these 13 loci was 1.9 SNPs/kb. These SNPs were employed as candidates for association for language ability measured in three domains: expressive language, receptive language, and nonword repetition.

Table 5.2. Genomic Coordinates of 13 Primary MicroRNA Transcripts Selected As Candidates for Association with SLI.

miRNA	Chr	Strand	Transcript Start (bp)	Transcript End (bp)	Size (kB)	No of SNPs	Density (SNP/kB)	Average SNP Spacing (bp)
hsa-mir-137	1	-	98,453,556	98,515,249	61.7	147	2.4	407
hsa-mir-181b-1	1	-	198,801,622	198,906,558	104.9	181	1.7	575
hsa-mir-30c-1	1	+	41,204,508	41,237,276	32.8	73	2.2	456
hsa-mir-9-2	5	-	87,960,264	87,969,157	8.9	14	1.6	511
hsa-mir-30c-2	6	-	72,054,323	72,126,178	71.9	194	2.7	374
hsa-mir-25	7	-	99,690,351	99,699,563	9.2	27	2.9	365
hsa-mir-320a	8	-	22,090,478	22,102,672	12.2	31	2.5	414
hsa-let-7d	9	+	96,928,697	96,941,816	13.1	30	2.3	435
hsa-mir-181b-2	9	+	127,420,715	127,460,907	40.2	11	0.3	3843
hsa-mir-615	12	+	54,426,832	54,429,144	2.3	6	2.6	572
hsa-mir-20a	13	+	92,000,074	92,006,829	6.8	11	1.6	690
hsa-mir-124-3	20	+	61,806,258	61,814,938	8.7	0	0.0	NA
hsa-mir-125b-2	21	+	17,960,907	17,962,926	2.0	1	0.5	NA

The genome coordinates correspond to the Genome Reference Consortium Human Build 37 (GRCh37/hg19) assembly.

5.2.2 Association of SNPs in FOXP2-regulated MicroRNAs to Measures of Language Ability

A total of 4 SNPs were significantly associated with ELS (unadjusted p-value < 0.01, Table 5.3). The top SNP is rs78807223 (unadjusted p-value = 3.02×10^{-4}), which lies in the hsa-mir-137 primary transcript (Figure 5.2A).

Table 5.3. Association Results of SNPs in Primary Transcripts of Top 12 MicroRNAs for SLI.

MicroRNA Transcript	Chr	SNP	Allele 1	Allele 2	MAF in SLIC cohort	MAF in GBR subpopulation	Hardy-Weinberg p-value	Trait	Effect Size	Standard Error	H ²	LOD	p-value
hsa-mir-137	1	rs78807223	T	C	0.021	0.22	1	ELS	-12.899	3.57	2.88	2.83	0.0003029
hsa-mir-25	7	rs2293481	C	T	0.38	0.39	0.075	ELS	-3.92	1.311	2.62	1.94	0.002784
hsa-mir-25	7	rs1527423	A	G	0.41	0.44	0.303	ELS	-3.782	1.326	2.50	1.77	0.004331
hsa-mir-25	7	rs1534310	A	T	0.35	0.32	0.366	ELS	3.684	1.421	2.28	1.46	0.009554
hsa-mir-30c-2	6	rs9342831	A	C	0.033	0.03	0.028	NWR	-10.96	3.726	2.22	1.88	0.003267
hsa-mir-181b-1	1	rs12140000	C	G	0.15	0.17	0.31	NWR	-5.246	1.805	1.98	1.83	0.003662
hsa-mir-181b-1	1	rs11807765	C	G	0.15	0.17	0.3124	NWR	-5.208	1.812	1.95	1.79	0.00405
hsa-mir-20a	13	rs4284505	G	A	0.34	0.43	0.0129	NWR	-4.248	1.497	2.20	1.75	0.004532
hsa-mir-20a	13	rs1428	A	C	0.30	0.37	0.5401	NWR	-4.244	1.495	2.09	1.75	0.004533
hsa-mir-181b-1	1	rs12132527	T	G	0.15	0.17	0.3204	NWR	-5.137	1.816	1.90	1.74	0.00468
hsa-mir-181b-1	1	rs77091809	G	A	0.030	0.07	1	NWR	8.815	3.174	1.46	1.68	0.005474

Shown here are candidate SNPs with a p-value < 0.01 – and in ascending order of p-value – for association to expressive language scores (ELS) and nonword repetition scores (NWR) in the family-based SLI cohort (SLI Consortium 2002, SLI Consortium 2004). This table lists the alleles for each SNP (Allele 1 and Allele 2, with the latter being the minor allele), the minor allele frequency in the SLIC cohort and in a Great British population sample, the p-value of the test for Hardy-Weinberg equilibrium in the SLIC cohort, the trait and the effect size of Allele 1 on trait scores (in normalized score units),

standard error of the effect size estimate, H^2 (proportion of total variance accounted for by the SNP), the LOD score of association, and the association p-value.

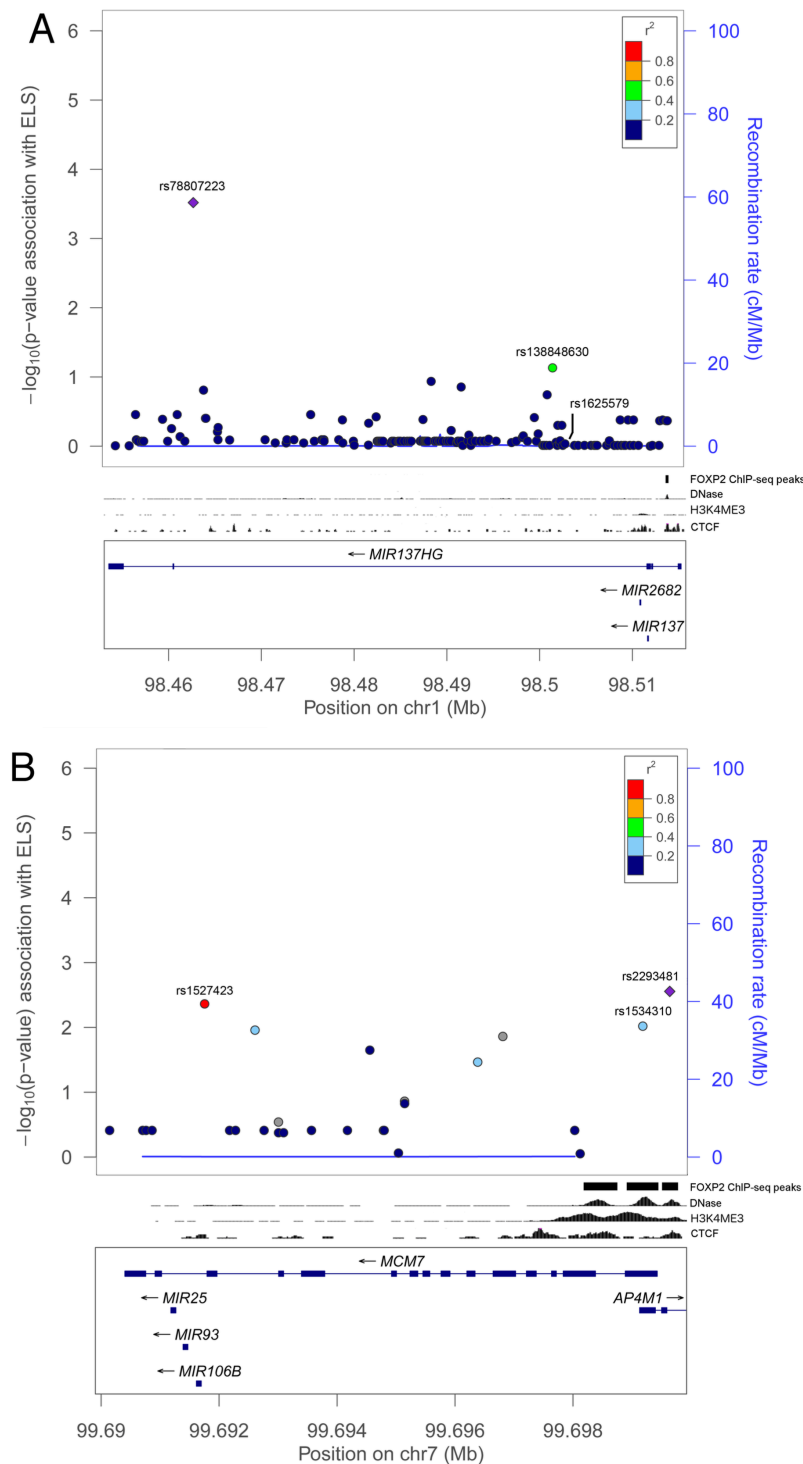


Figure 5.2. Regional Association Plots for MicroRNA Transcripts Associated With Expressive Language Scores.

Shown here are regional association plots for the primary transcripts of **(A)** hsa-mir-137 and **(B)** hsa-mir-25. The SNP with the highest $-\log_{10}$ p-value is plotted as a diamond, while the rest of the SNPs are plotted as circles, and their colour (as indicated by the r^2

legend) depends on their individual pairwise LD with it. that lie inside these transcripts are plotted with their corresponding $-\log_{10}$ p-values as a function of chromosomal position. A gene diagram plot (lower panel) shows the exon-intron structure and orientation of the genes in this region. Signal peaks are shown for DNase hypersensitivity (which indicates open chromatin regions) alongside occupancy peaks for H3K4ME3 and CTCF (which denote promoter regions) in neuroblastoma cell-lines.

The other three significant SNPs (rs2293481, rs1527423, and rs1534310) lie in or near the *MCM7* locus, which harbours hsa-mir-25 in its penultimate intron. The strongest association signal in the locus comes from rs2293481 (p-value = 0.00278), which is ~200bp upstream of the *MCM7* transcription start site, and 8.6kb from the mir-25 locus (Figure 5.2B). It is in strong LD with the rs1527423 (pairwise $r^2 = 0.841$), and in weak LD with rs1534310 (pairwise $r^2 = 0.382$).

Rs2293481 and rs1534310 both lie inside 2 regions bound by *FOXP2* (Fig 5.2B), as detected by an ENCODE ChIP-seq study (ENCODE Project Consortium, 2011). The top SNP rs2293481 also lies within a DNaseI hypersensitivity peak (a mark of transcriptionally-active chromatin), a H3K4ME3 ChIP-seq peak (used previously to indicate the TSS of microRNA transcripts by Marson *et al.* 2008), and a CTCF promoter occupancy peak (an indicator of promoter regions). These three pieces of evidence indicate that the region is likely to function as a promoter for this locus.

A total of 7 SNPs in 3 microRNA primary transcripts were associated with NWR (nominal p-value < 0.01, Table 5.3). The top SNP is rs9342831 (nominal p-value = 0.00327) and it lies in the 71.9kb predicted primary transcript of hsa-mir-30c-2 (Figure 5.3A). It is 32.7kb downstream from the mir-30c-2 precursor and is not in LD with SNPs in its vicinity. The SNPs closest to the mir-30c-2 precursor are neither significantly associated to NWR, nor do they lie in putative *FOXP2* binding sites.

Four SNPs in the primary transcript of hsa-mir-181b-1 (rs12140000, rs11807765, rs12132527, and rs77091809) are associated with NWR (unadjusted p-value < 0.01). The first 3 SNPs lie almost 22kb downstream from the hsa-mir-181b-1 precursor, and are in strong LD with each other (Figure 5.3B). The fourth SNP (rs77091809) is not in LD with these three SNPs, but is as strongly associated

with NWR. The SNPs lying closest to hsa-mir-181b-1 are not significantly associated with NWR in the SLI cohort, and do not exhibit LD with the 4 SNPs that are. None of these four SNPs lie in known FOXP2 binding sites.

Lastly, 2 SNPs (rs4284505 and rs1428) in the primary transcript of hsa-mir-20a (known as *MIR17HG*, which also contains the precursors of hsa-mir-17, hsa-mir-18a, and hsa-mir-19a) are associated with NWR with p-values 0.00453 and 0.00452 (Figure 5.3C). Linkage disequilibrium analysis reveals that these two SNPs are in moderate LD with each other (pairwise $r^2 = 0.56$). There are two FOXP2 ChIP-seq peaks in this transcript: the first at the 5' end overlapping the first exon, and the second one lying in the first intron, which harbours the most highly-associated SNP rs4284505. ChIP-seq data (ENCODE Project Consortium, 2011) also reveals that rs4284505 lies in a promoter region that is occupied by methylation factors and CTCF.

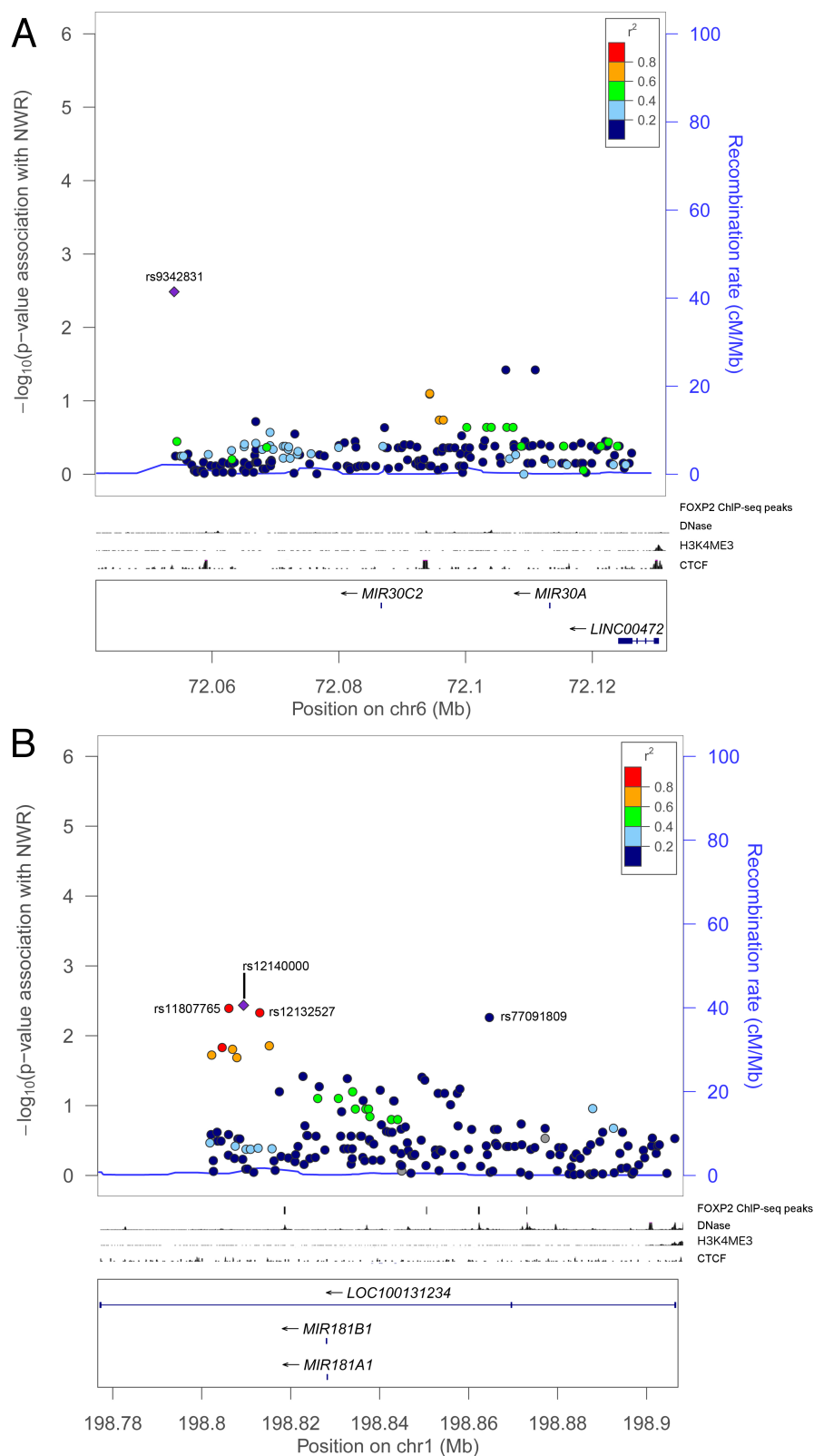


Figure 5.3. Regional Association Plots for MicroRNA Transcripts Associated With Nonword Repetition Scores.

Shown here are regional association plots for the primary transcripts of **(A)** hsa-mir-30c-2, **(B)** hsa-mir-181b-1, and **(C)** hsa-mir-20a. (Con't on next page.)

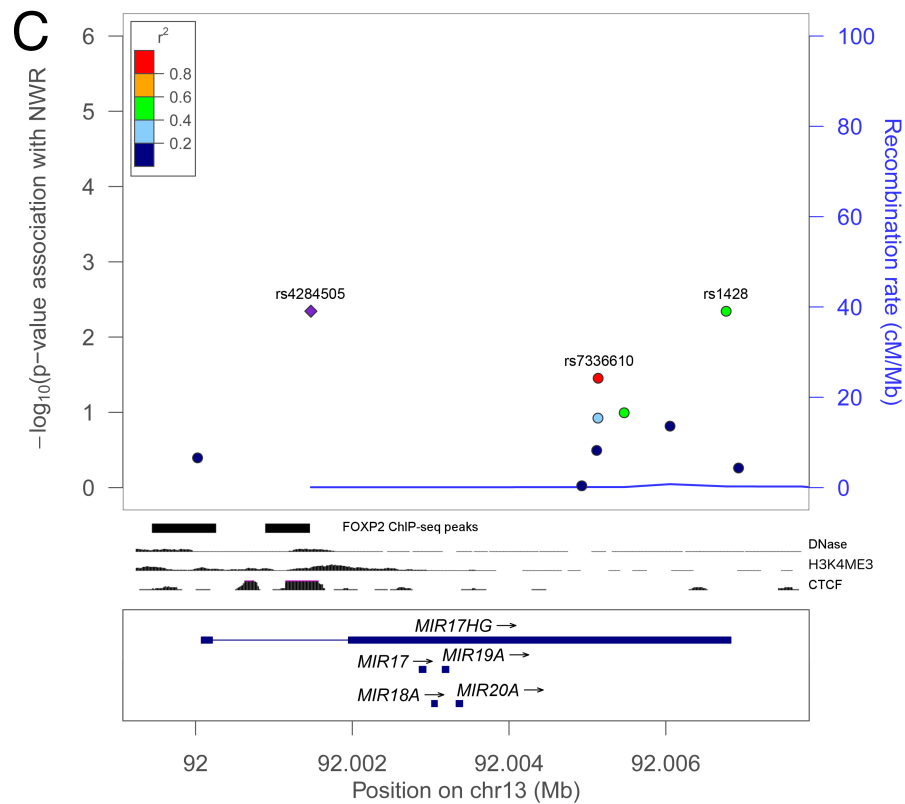


Figure 5.3. Regional Association Plots for MicroRNA Transcripts Associated With Nonword Repetition Scores (Con't).

The top SNP (rs4284505) in the MIR17HG host transcript (**C**) lies in a region bearing the hallmarks of transcriptional activity, and is a ChIP-seq target of FOXP2.

5.2.3 Identifying Variants in MicroRNA Binding Sites From Exome Data

Thus far, I have shown that SNPs found in primary microRNA transcripts regulated by FOXP2 might contribute towards risk for SLI. It is important to note that variants in microRNA loci are not the only types of variants that can affect microRNA-mediated gene regulation. Variants in 3' UTR might themselves be risk factors, as they might disrupt regulation of the gene transcript by microRNAs (Saunders *et al.* 2007, Li *et al.* 2012, Bulik-Sullivan *et al.* 2013). To identify such variants that might be functionally relevant in the context of SLI, a systematic filter was performed to pinpoint 3' UTR variants in microRNA binding sites in the exome data. These variants were then further investigated for association to quantitative language-related traits in families affected by SLI. To study the functional testing of the effects of these variants, a reporter gene assay was performed. The workflow that was used is summarised in Figure 5.4.

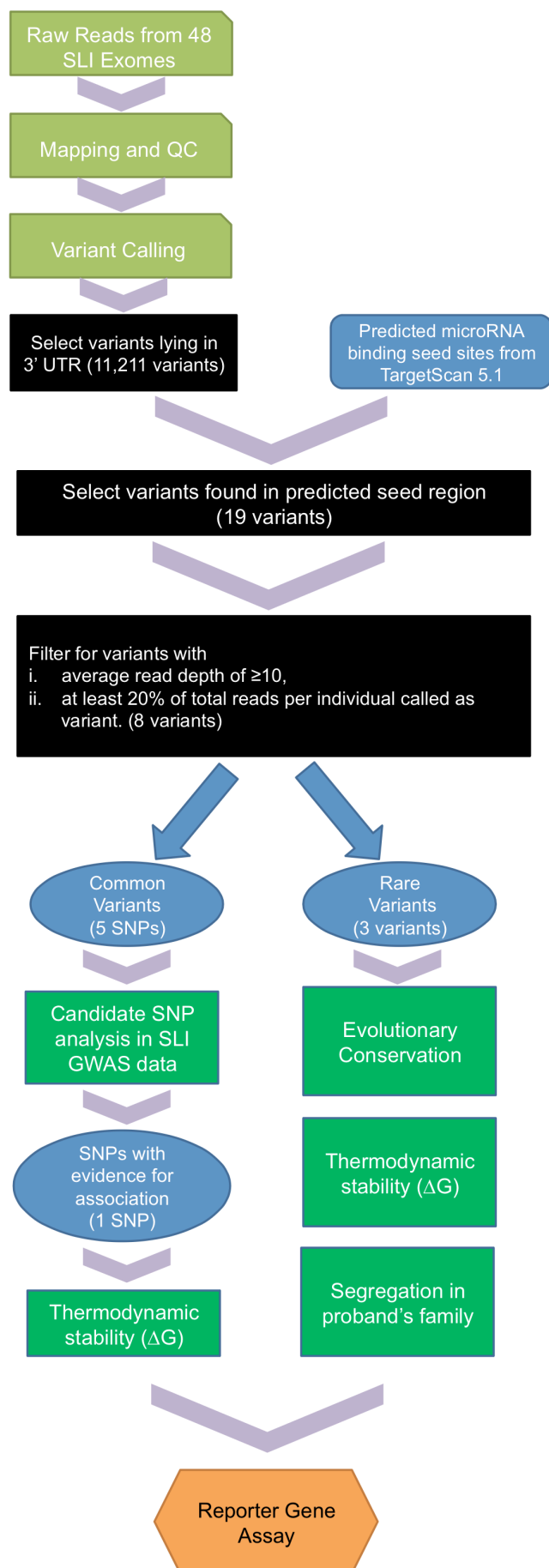


Figure 5.4. Variant Identification in 3' UTRs from Exome Data of 48 SLI Probands: A Summary Workflow.

By applying the filters as indicated in the black rectangles, a list of 3' UTR seed-region variants were obtained from the exome data of 48 SLI probands. The variants left after each filter was applied is noted in parenthesis. The common variants were used as candidate SNPs for association in a larger SLI family-based cohort, while the rare variants were assessed for evolutionary conservation, thermodynamic stability, and segregation in the proband's pedigree. The four seed-region variants in the final shortlist were subjected to a reporter gene assay to directly determine their functional effects on microRNA targeting and repression.

The exome capture platform used (Agilent SureSelect v2) was designed to capture 6.7% of total 3' UTR length. In the exome data studied here (Chen *et al.*, in prep.), 10% of total 3' UTR length was covered at a read depth ≥ 10 . Read depth can be expected to fall the further a basepair is from the start of the 3' UTR, as the input DNA is sheared to ~ 300 bp before undergoing exome capture. In this exome data, the average read depth fell below 10 after ~ 200 bp from the start of the 3' UTR (Figure 5.5A). This high-coverage region in the 3' UTR corresponded to a peak in the average density of microRNA binding sites (predicted by the TargetScan 5.1 algorithm) in the set of captured 3' UTRs (Figure 5.5B). Overall, 62.7% of RefSeq-annotated 3' UTRs were covered in the exome data at a read depth ≥ 10 for the first 200 bases.

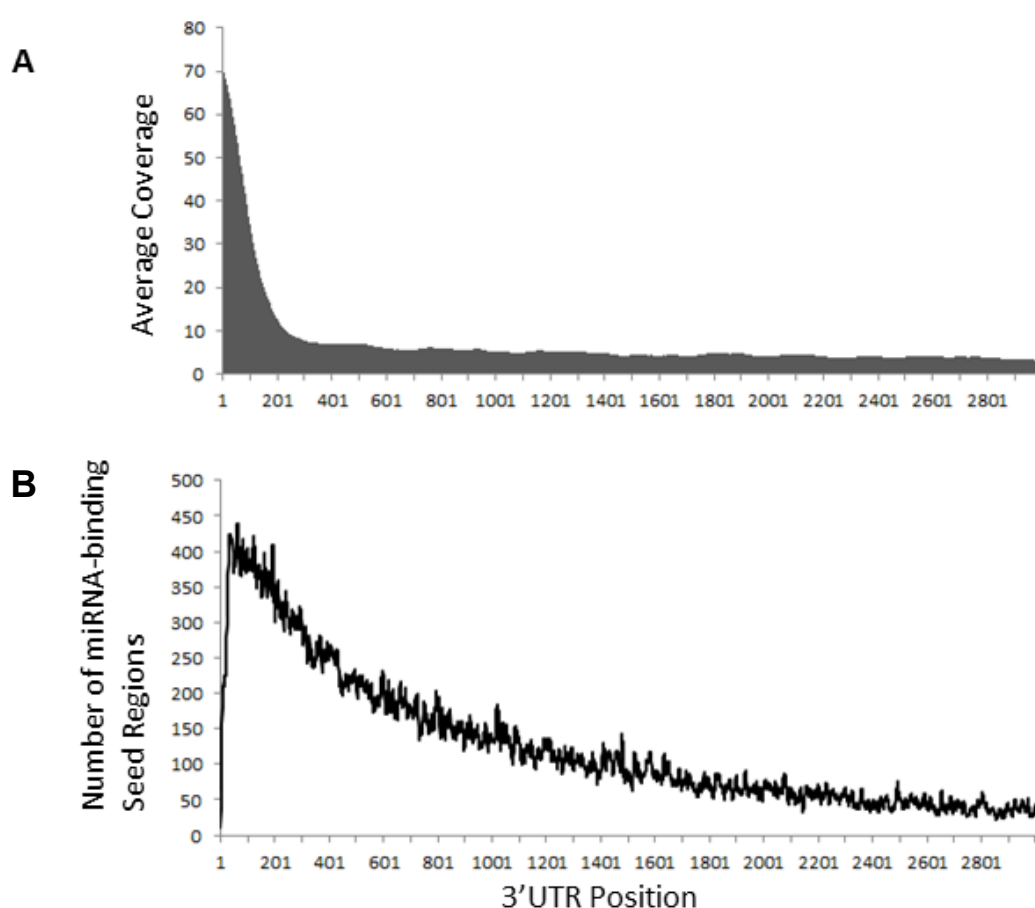


Figure 5.5. Coverage and miRNA Binding Site Density in 3' UTR Regions of 48 SLI Exomes.

(A) Average Read Depth Profile Along The 3' UTR In SLI Exome Data **(B)** Distribution of predicted miRNA binding sites along the 3' UTR. Data from Chen *et al.*, in prep.

A total of 11,211 single-nucleotide variants (SNVs) were found in 3' UTR regions in the exome data. This list was further refined by selecting for those variants that lie in microRNA binding site seed regions (as predicted by the TargetScan 5.1 algorithm), had an average read depth of ≥ 10 , and with at least 20% of total reads per individual called as variant. Five known SNPs (listed in Table 5.4; defined as a variant documented in dbSNP with a unique reference SNP, or "rs" number, and with a minor allele frequency $> 1\%$) and three rare variants (listed in Table 5.5, defined as variants not found in dbSNP) were identified.

Table 5.4. Five Common Variants Lying in Predicted 3' UTR MicroRNA Binding Sites are Present in the Exomes of 48 SLI Probands.

Chr	Position	dbSNP ID	Ref	Alt	MAF in SLI cohort	MAF in GBR subpopulation	Gene	Predicted microRNA binding partner	Average Read Depth	Genotype		
										Ref/Ref	Ref/Alt	Alt/Alt
chr9	35661943	rs72727021	A	C	0.099	0.121	C9orf100	miR-192/215	27.27	36	12	0
chr16	79245820	rs383362	G	T	0.37	0.429	WWOX	miR-134, miR-758	14.85	9	25	14
chr16	87436764	rs1054528	A	G	N.A.	0.083	MAP1LC3B	miR-204/211	12.98	1	26	21
chr19	6751293	rs1049232	T	G	0.194	0.187	TRIP10	miR-17-5p	18.94	32	15	1
chr19	53086946	rs190191374	G	T	N.A.	0	ZNF701	miR-199/199-5p	20.46	47	1	0

Shown here are the genomic coordinates (hg19) and the dbSNP IDs of the five common variants – along with the reference ("REF") and alternate ("ALT") alleles – found in predicted 3' UTR binding sites in the exomes of 48 SLI probands), the minor allele frequency (MAF) in the SLIC cohort, as well as that in a Great British population sample, along with the corresponding gene, and the microRNA binding to a seven-nucleotide seed region harbouring the SNP, as predicted by the TargetScan 5.1 algorithm. Each common variant had an average read depth of ≥ 10 across all 48 exomes. The genome coordinates correspond to the Genome Reference Consortium Human Build 37 (GRCh37/hg19) assembly. If the SNP was neither genotyped or imputed in the SLI cohort, the MAF could not be reported, and is indicated with an "N.A.".

Table 5.5. Three Rare Variants Lying in Predicted 3' UTR MicroRNA Binding Sites are Found in the Exomes of 48 SLI Probands.

Chr	Position	Ref	Alt	Gene	Predicted microRNA binding partner	Average Read Depth	Probands With Ref. Allele	Probands With Variant
chr13	25457299	G	C	BTN2A1	miR-342	16.21	47	1
chr22	30421860	G	T	CENPJ	miR-433	33.73	47	1
chr6	26469054	C	T	MTMR3	miR-346	13.29	47	1

Shown here are the genomic coordinates (hg19) of the three rare variants found in predicted 3' UTR binding sites in the exomes of 48 SLI probands, along with the corresponding gene, and the microRNA binding to a seven-nucleotide seed region

harbouring the SNP, as predicted by the TargetScan 5.1 algorithm. Each of the rare variants had an average read depth of ≥ 10 across all exomes, indicating good coverage of these basepairs in the dataset, even though these 3' UTRs are not directly captured by the exome platform used. The genome coordinates correspond to the Genome Reference Consortium Human Build 37 (GRCh37/hg19) assembly.

5.2.4 Association of Exome-Derived Variants in MicroRNA Binding Sites With SLI

To ascertain the relevance of the five common SNPs identified from the exome data to SLI, they were employed as candidate SNPs for association to three quantitative SLI traits in the SLI cohort. Three of these SNPs (rs72727021, rs383362, and rs1049232) were genotyped and/or imputed in the SLIC dataset, and were directly tested for association with language ability (Table 5.6). The SNP rs72727021 was found to be significantly associated with nonword repetition ability in the cohort ($p\text{-value} = 2.75 \times 10^{-4}$). It lies in a predicted binding site for miR-192 and miR-215 in the 3' UTR of the gene *C9orf100* (also known as *ARHGEF39*). This association remains significant even after employing a conservative Bonferroni correction for multiple testing (which assumes that each trait-wise test for each SNP is independent and corrects based on the total number of tests performed, adjusted $p\text{-value} = 2.5 \times 10^{-3}$). Out of 5000 gene-dropping simulations performed in MERLIN (Abecasis *et al.* 2001), an association $p\text{-value}$ lower than 2.75×10^{-4} at this SNP was observed only 5 times, giving a false discovery rate of 0.1%.

Table 5.6. The 3' UTR SNP rs72727021 Is Significantly Associated With Nonword Repetition.

Chr	Pos	dbSNP ID	Ref	Alt	Gene	Predicted microRNA binding partner	Candidate SNP Association Test in SLI Cohort			
							Trait	Effect	H ²	p-value
9	35661943	rs72727021	A	C	C9orf100	miR-192, miR-215	NWR	7.919	2.986	0.0002752
							ELS	3.536	0.801	0.06773
							RLS	4.853	1.369	0.01755
16	79245820	rs383362	G	T	WWOX	miR-134, miR-758	NWR	1.477	0.273	0.321
							ELS	0.702	0.083	0.5881
							RLS	1.191	0.216	0.3858
19	6751293	rs1049232	T	G	TRIP10	miR-17-5p	NWR	0.145	0.002	0.9352
							ELS	-0.29	0.009	0.8554
							RLS	0.093	0.001	0.9559

Three SNPs (rs72727021, rs383362, and rs1049232) were genotyped or imputed in the SLIC dataset, and were tested directly for association for three language-measures: nonword repetition (NWR), expressive language (ELS) and receptive language (RLS). For every SNP, the following measures were identified for each of the three traits: the effect of the SNP on the quantitative trait normalized score, heritability of the trait (a measure of the percentage of the variation in the trait accounted for by the SNP, denoted here at H²), and corresponding p-value of the association. The SNP rs72727021, lying in a predicted binding site for miR-192 and miR-215 in the 3' UTR of the gene C9orf100, was found to be significantly associated with nonword repetition ability in the cohort (p-value = 2.75×10^{-4}). The genome coordinates correspond to the Genome Reference Consortium Human Build 37 (GRCh37/hg19) assembly.

The gene product of *C9orf100* is localised to the cell membrane, and its overexpression promotes cellular proliferation (Wang *et al.* 2012). A search of the Allen Brain Atlas reveals strong expression in adult human cerebellum, while expression of miR-215 and miR-192 has been observed in the adult brain (Kaya *et al.* 2010). Co-expression in the brain is good evidence that the microRNAs might interact with *C9orf100*.

The other two common SNPs from the exome data (rs1054528 in the 3' UTR of *MAP1LC3B*, and rs190191374 in the 3' UTR of *ZNF701*) were neither genotyped nor imputed in the SLIC dataset. To evaluate if they might be associated with language ability in the SLIC cohort, SNPs that were in LD with rs1054528 and rs190191374 in British population samples from the 1000 Genomes Project (The 1000 Genomes Consortium 2012) were obtained. These SNPs served as a proxy for the exome-derived microRNA binding site SNPs, and enabled me to test indirectly if rs1054528 and rs190191374 were associated with language ability in the SLIC dataset.

The SNP rs1054528 in 3' UTR of *MAP1LC3B* was found to be in LD with two SNPs: rs8051218 (pairwise $r^2 = 0.904$) and rs1054521 (pairwise $r^2 = 0.697$). No significant association was observed between these 2 SNPs and three measures of language ability in the SLIC dataset (Table 5.7). Pairwise LD data was not present in the 1000 Genomes data for SNP rs190191374, but inspection of a 2kb window around the SNP revealed that it lay in an LD block defined by 4 SNPs: rs62117248, rs62117249, rs369009, and rs11274688 (Figure 5.6A). These 4 SNPs were imputed in the SLIC dataset, and were tested for association with SLI (Figure 5.6B), and no significant association was observed.

Table 5.7. SNPs in LD with the 3' UTR Exome Variant rs1054528 Are Not Significantly Associated with Language Ability.

Chr	Pos	SNP	Pairwise r^2 with rs1054528	SLI Association (p-value)		
				ELS	RLS	NWR
16	87436463	rs8051218	0.904	0.798	0.462	0.8816
16	87436720	rs1054521	0.697	0.923	0.5262	0.7819

The exome-derived variant rs1054528 in *MAP1LC3B* was not captured in the SLIC dataset, so 2 SNPs in LD with it (in British population samples from 1000 Genomes haplotype data) were analysed for association. No significant association to ELS, RLS, or NWR was observed. The genome coordinates correspond to the Genome Reference Consortium Human Build 37 (GRCh37/hg19) assembly.

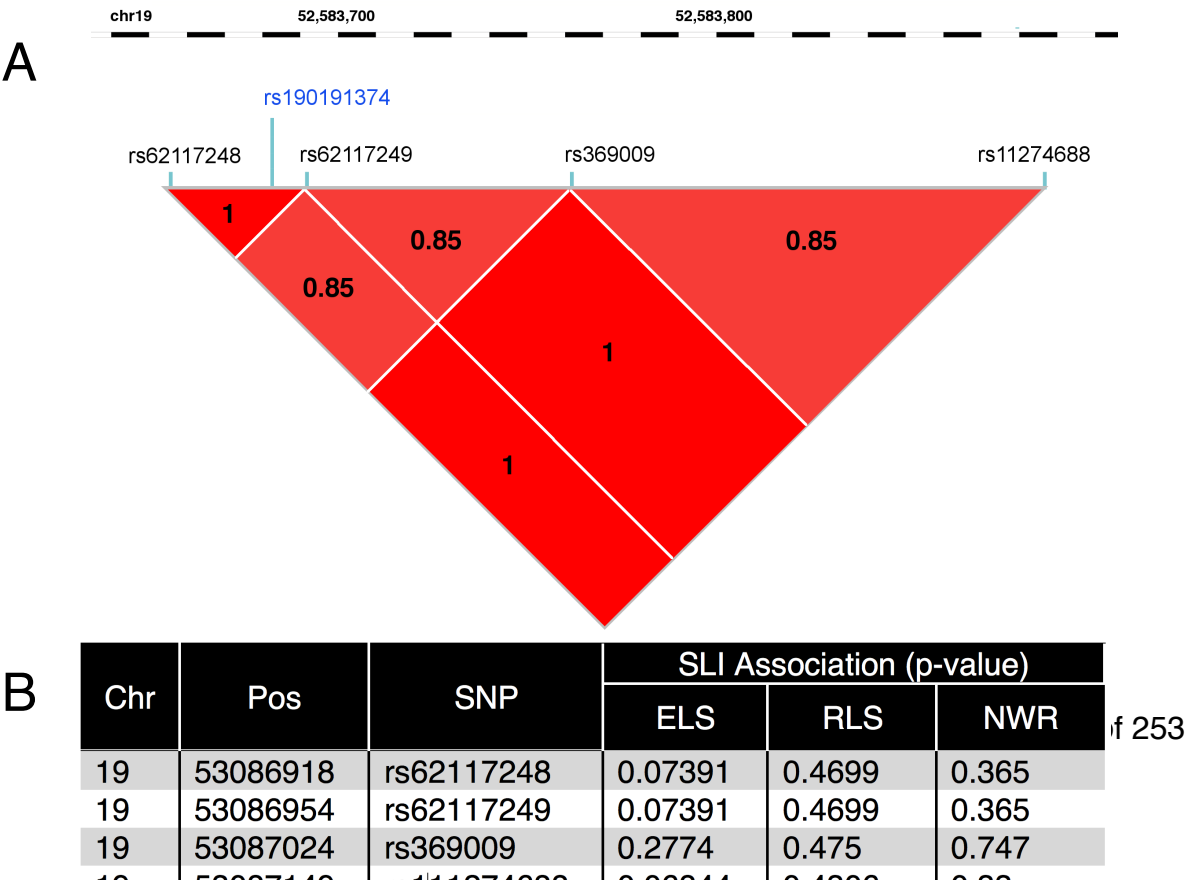


Figure 5.6. SNPs in LD with the 3' UTR Exome Variant 1 Are Not Significantly Associated with Language Ability.

(A) Shown here is the linkage disequilibrium structure for at chr19: 53,086,918-53,087,149. The exome-derived variant rs190191374 (indicated by blue text) is neither genotyped nor imputed in the SLI dataset, but lies in an LD block (in British population samples from 1000 Genomes haplotype data) with four other SNPs that are imputed in the dataset. The pairwise r^2 between these 4 SNPs are indicated inside the LD block. (B). These 4 SNPs were tested analysed for association. No significant association to ELS, RLS, or NWR was observed. The genome coordinates correspond to the Genome Reference Consortium Human Build 37 (GRCh37/hg19) assembly.

5.2.5 Assessing The Conservation and Known Function of Exome Variants in MicroRNA Binding Sites

All the rare SNVs were called as heterozygous and were found in only one individual among the 48, and were not reported in publicly-available control exomes. The presence of these variants in the relevant probands was validated with bi-directional Sanger sequencing (Figure 5.7). Each rare variant was assessed for their evolutionary conservation, as target sites that are highly conserved across genomes are thought to be under selection and thus likely to be functional (Lewis *et al.* 2003). This was done using scores generated by the Genomic Evolutionary Rate Profiling (GERP) algorithm (Davydov *et al.* 2010, Goode *et al.* 2010).

As briefly mentioned in the Methods (see Section 2.12), the GERP score computes rejected substitution at a region of interest as the difference between the expected number of substitutions for at a nucleotide of interest (under an assumption of neutral evolution), and the actual observed number of substitutions at that basepair in 36 mammalian genomes. A higher positive GERP score indicates that the nucleotide has less observed substitutions than expected, and is thus conserved to a greater extent. Across the genome, GERP scores range from -15 to 7, with a GERP score of > 2 highly indicative of true constraint (Goode *et al.* 2010). GERP has been used in previous whole-exome studies (such as O'Roak *et al.* 2011) to prioritise conserved rare variants for further study.

The variant in the miR-342 target site on the Butyrophilin subfamily 2, member A1 (*BTN2A1*) 3' UTR has a GERP score of 1.60 (Figure 5.7A), indicative of moderate constraint at this site. This basepair, as well the target site, is highly conserved in mammals, as well as in the mouse, horse, and dog genomes. This is evidence of moderate constraint at this target site, and supports the hypothesis that the rare exome variant is likely to have a functional effect. *BTN2A1* is a cell surface glycoprotein that is ubiquitously expressed in the body, with strong expression observed in the brain as well as the spinal cord (Malcherek *et al.* 2007), and is thought to be involved in lipid metabolism (Taz-Ahnini *et al.* 1997). Mature miR-342 expression has been observed in the mouse brain, and its dysregulation has been implicated in mouse models of Alzheimer's disease (Sun *et al.* 2014). The high expression of both *BTN2A1* and miR-342 in the brain suggests that they might indeed interact.

The variant lying in the miR-433 target site on the Centromere protein J (*CENPJ*) 3' UTR, is highly conserved with a GERP score of 4.83 (Figure 5.7B), and thus is likely to have a functional effect. The gene product is involved in centromere division, and is ubiquitously expressed in the developing embryo but displays strong neural and spinal expression (Bond *et al.* 2005). Mutations in *CENPJ* cause a form of autosomal recessive primary microcephaly (Bond *et al.* 2005, Gul *et al.* 2006, Darvish *et al.* 2010) as well as Seckel syndrome (Al-Dosari *et al.* 2010). Mature miR-433 is expressed highly in the brain and muscle (Davis *et al.* 2005) and has been implicated in dominant X-linked chondrodysplasia (Simon *et al.* 2010) and risk for Parkinson's disease (Wang *et al.* 2008). Considered together, the predicted targeting of the *CENPJ* by miR-433 is an attractive candidate for functional confirmation.

The variant in the miR-346 binding site of Myotubularin-related protein 3 (*MTMR3*) 3' UTR, is highly conserved as well, with a GERP score of 4.98 (Figure 5.7C). The gene product is a phosphatase that acts upon specific phosphatidylinositol species (Walker *et al.* 2001). It is widely expressed across different tissues and brain regions, including the pre-frontal cortex and

hypothalamus (Su *et al.* 2004). Mature miR-346 is also expressed in the pre-frontal cortex (Zhu *et al.* 2009), and is thought to have an expression pattern similar to its host gene, glutamate receptor δ 1 (GRID1) which is an important in post-synaptic signaling. The overlapping domains of expression of *MTMR3* and miR-346 indicate that they might interact.

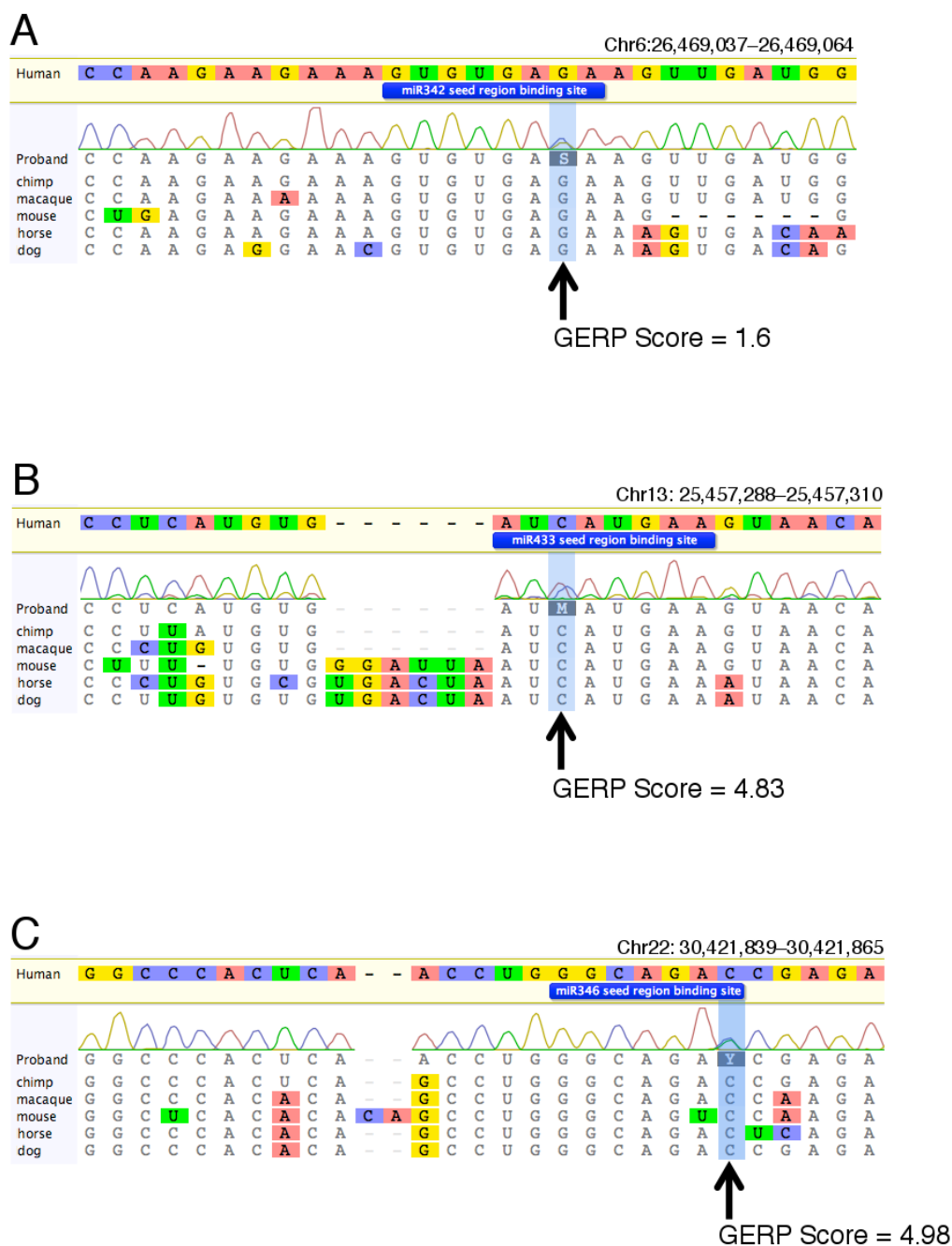


Figure 5.7. The 3 Rare Variants are Highly Conserved Nucleotides.

All the three rare variants in 3' UTRs of **(A)** *BTN2A1*, **(B)** *CENPJ*, and **(C)** *MTMR3* were confirmed as heterozygous. For each 3' UTR, the probands' genomic sequence is shown aligned to the reference human genome, alongside the relevant genomic regions from chimp, macaque, mouse, horse, and dog genomes. The hg19 coordinates of the aligned human genomic segment are found on the upper-right corner of each alignment. Nucleotides that are not conserved in the human genome are black against a coloured background. Dashes indicate the presence of gaps in the alignment. The location of the rare variant is highlighted by the light blue rectangle, and the GERP score of the nucleotide is indicated beneath it. The genome coordinates correspond to the Genome Reference Consortium Human Build 37 (GRCh37/hg19) assembly.

5.2.6 Assessing The Thermodynamic Properties of Exome Variants in MicroRNA Binding Sites

To obtain more evidence that these variants have a functional effect, I assessed the thermodynamic stability of the relevant microRNA-mRNA duplexes. This is done by calculating the free energy change (ΔG) when a microRNA hybridises to the 3' UTR. The more negative the ΔG of the duplex, the more thermodynamically stable the microRNA-mRNA duplex is. Therefore, the ΔG values for microRNA-mRNA duplex formation were obtained for both the wildtype and the variant 3' UTRs. This was performed for all the rare variants as well as the common variant rs72727021, which was significantly associated with NWR in the SLIC cohort (Table 5.4). The STarMiR algorithm (Liu *et al.* 2013, available online at <http://sfold.wadsworth.org>) utilises a statistical sampling method to obtain the most stable Boltzmann equilibrium structure for a microRNA-3'UTR hybrid of interest, and is able to furnish visualization of the most stable secondary structure and predicted ΔG values.

The three rare variants are predicted to produce a less-stable microRNA-3' UTR duplex, relative to the wildtype basepair (Figure 5.8A – C). Thus, all three are predicted to reduce microRNA binding to the respective 3' UTRs, suggesting these variants might disrupt on microRNA targeting, and thus disturb microRNA-mediated translational repression of these transcripts.

The major 'A' allele ('T' in the *C9ORF100* mRNA transcript) of the common SNP rs72727021, which is significantly associated with higher nonword repetition scores (Table 5.6), produces a more stable duplex compared to the less prevalent C allele ('G' in the *C9ORF100* mRNA transcript) ($\Delta G_{A \text{ allele}} = -18.6$ kcal/mol, $\Delta G_{C \text{ allele}} = -16.7$ kcal/mol, Figure 5.8D). It is possible that the 'A' allele results in a 3' UTR with a greater affinity for miR-192 and miR-215, resulting in lower translation of the *C9ORF100* gene transcript.

C9orf100, however, the more widespread T allele has a more stable configuration with miR-215 compared to the less common G allele.

5.2.7 Segregation of Rare 3' UTR Variants in MicroRNA Binding Sites

To determine if the three rare variants identified here are *de novo* mutations, I sequenced the relevant variants in the DNA of the probands' siblings and parents. None of the variants were *de novo*; all of them are inherited from one parent (Figure 5.9).

The mir-342 binding site variant in the 3' UTR of *BTN2A1* is passed from the mother to two affected daughters, but is not possessed by her third unaffected daughter (Figure 5.9A). The mir-433 binding site variant in the *CENPJ* 3' UTR is also inherited from the mother, and her 2 offspring – both of whom are affected individuals – also possess the variant. (Figure 5.9B). The language ability scores and SLI diagnosis results for the parents were unavailable for analysis. As such, it is possible that the parents transmitting the rare variant do have language impairment. If so, the variants in the 3' UTRs of *BTN2A1* and *CENPJ* segregate with SLI in these two pedigrees, and represent potential risk factors.

The segregation of the 2 variants in these 2 pedigrees with SLI status is a piece of evidence suggesting that these rare variants might be aetiological. However, segregation might have arisen through other means: the variant might be in LD with a real causal mutation, or the variant might segregate in these small pedigrees simply through chance.

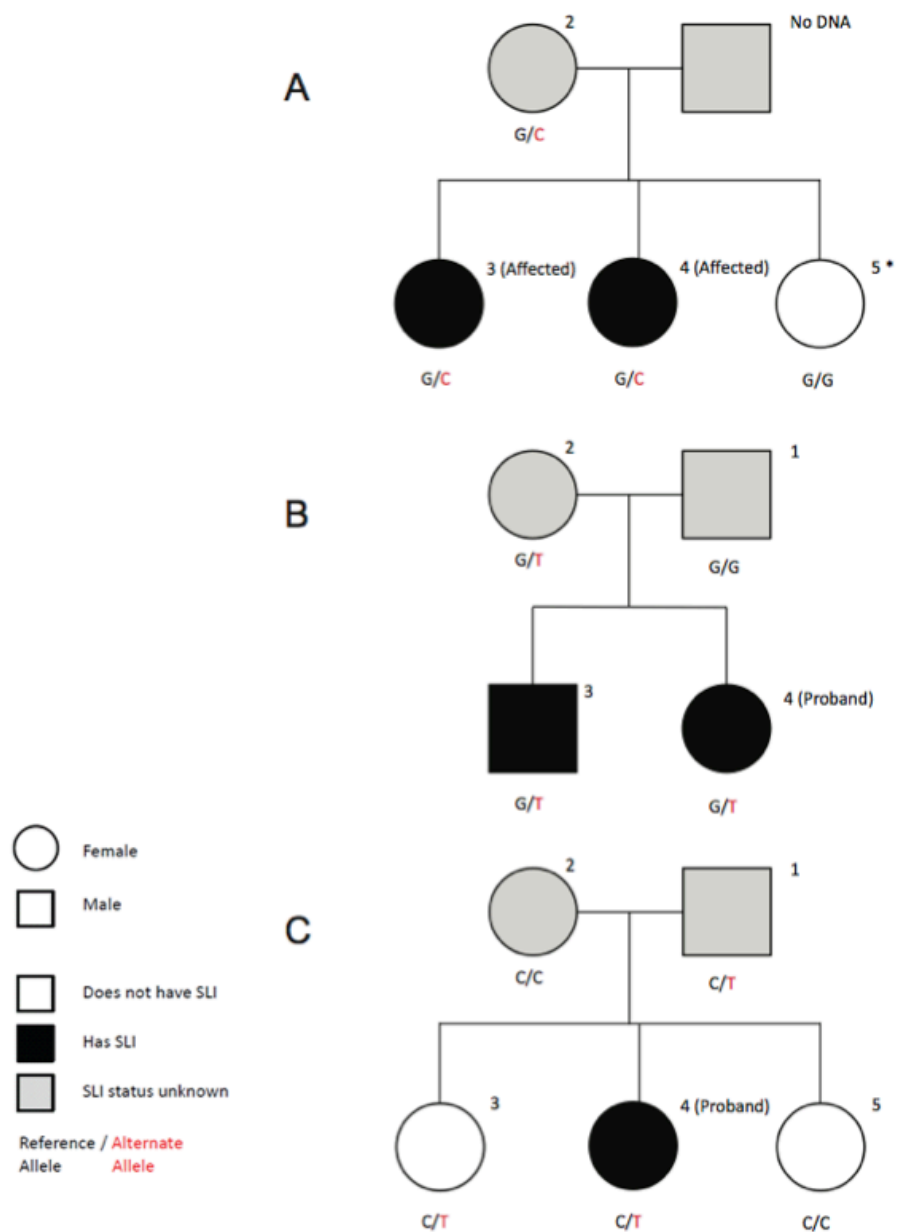


Figure 5.9. The 3 Rare Variants are Inherited Mutations.

(A) The variant in the 3' UTR of *BTN2A1* is inherited from the mother, and is shared by two affected offspring, but is not present in a third unaffected daughter. **(B)** The variant in the 3' UTR of *CENPJ* is inherited from the mother as well, and is found in each of her two offspring, both of whom have SLI. **(C)** The variant in the 3' UTR of *MTMR3* is passed from the father to the proband, as well as to an unaffected daughter.

The last rare variant in the *MTMR3* 3' UTR in the mir-346 target site, however, is passed from father to two of his daughters. Only one of these daughters is an affected individual while the other is not (Figure 5.9C). The language scores (obtained as part of the SLIC study) for the proband and her sister were analysed. The proband has normal scores for receptive and expressive language (93 and 91 respectively), but her nonword repetition score is 70, more than 1 SD below from mean of 100. The unaffected sister who is heterozygous for the variant has normal language scores for all three measures (101, 95 and 94 for receptive language, expressive language, and nonword repetition respectively). The presence of this rare variant in an unaffected sibling can be explained by reduced penetrance. It can also suggest, however, that the variant might not a real risk factor. Functional testing is thus required to clarify its aetiological significance.

5.2.8 Functional Analysis of Exome Variants in MicroRNA Binding Sites With Reporter Gene Assays

To functionally assess if the three rare microRNA binding site variants, and the SNP rs72727021, do disrupt microRNA binding to the 3' UTR, a luciferase reporter assay was employed (Figure 5.9). The aim here is twofold: to confirm the 3' UTR of interest is a true target of the microRNA, and to test if the variant has an effect on gene repression.

A fragment of the 3' UTR of interest that harbours the relevant microRNA binding site was cloned between a luciferase gene and a polyA termination region (Figure 5.9A). This produces a transcript containing the luciferase gene with a chimeric 3' UTR; the expression of this luciferase gene transcript should thus be repressed by microRNAs that target the chimeric 3' UTR. In addition, the microRNA of interest was overexpressed, by cloning the microRNA stem-loop precursor of interest (with ~100bp genomic flanks) immediately downstream of the human U6 promoter of the pLKO.1 vector (Figure 5.9B).

I generated luciferase reporter constructs with fragments of the 3' UTRs of *BTN2A1*, *CENPJ*, *MTMR3*, and *C9orf100* (Figure 5.9C), as well as overexpression constructs for miR-342, miR-433, miR-346, and miR-215 (Figure 5.9D). Co-transfection of empty pLKO.1 and luciferase vectors served as the baseline condition, and a luciferase vector with a relevant microRNA sponge cloned into the 3' UTR region (Devanna and Vernes 2014, modified from Kluiver *et al.* 2012) was used as a control to demonstrate expression of the microRNA of interest, and also to indicate that the assay is performing as expected.

Figure 5.10. Schematic of Reporter Constructs and Over-expression Constructs.

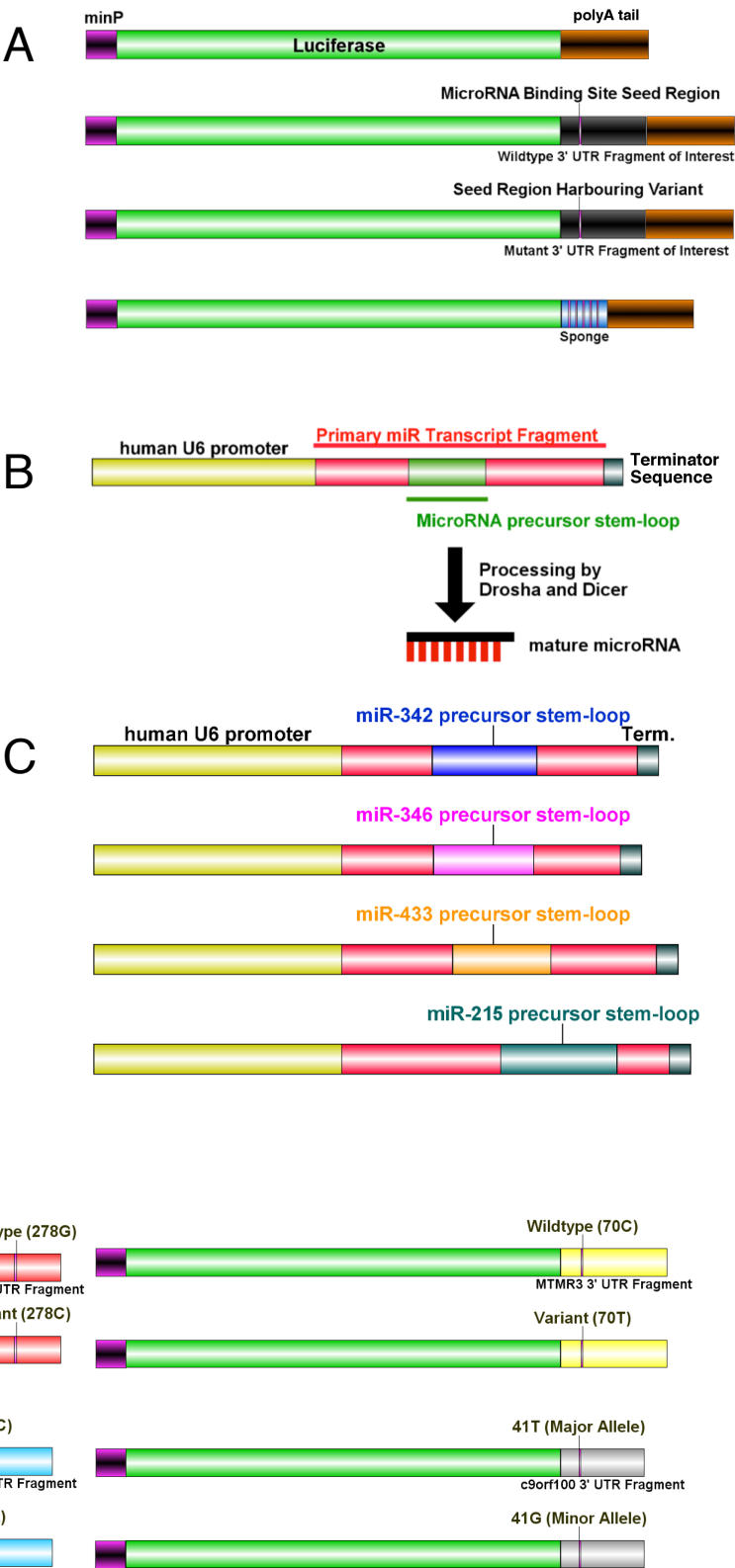
A luciferase reporter assay was employed in conjunction with a microRNA over-expression system to functionally investigate the effect of the identified variants on microRNA-mediated repression.

(A) The pGL4.23 luciferase reporter vector was employed to generate reporter constructs. A fragment of the 3' UTR of interest (containing the relevant microRNA binding site) is cloned between the luciferase gene sequence and the polyA termination sequence.

(B) The pLKO.1 vector was used to overexpress the microRNAs of interest. Primary miR transcripts were cloned downstream of the human U6 promoter.

(C) Four different pLKO.1 vectors were generated to over-express the following microRNAs: miR-342, miR-346, miR-433, and miR-215.

(D) Four pairs of pGL4.23 luciferase constructs (a total of 8) were generated, for each of the 3' UTRs of interest - BTN2A1, CENPJ, MTMR3, and C9orf100. Each pair consisted of a fragment of the wild-type 3' UTR cloned into the luciferase 3' UTR, as well as the variant-harboursing fragment.



The mutant *BTN2A1* 3' UTR fragment does not produce a significant down-regulation of the luciferase reporter relative to the wild-type one, and both the wildtype and mutant constructs are not responsive to overexpression of miR-342 (Figure 5.11A).

The mutant *CENPJ* 3' UTR fragment produces significantly lower levels of the luciferase reporter compared to that with the wildtype fragment (Tukey's post-hoc p-value < 0.01, Figure 5.11B). The difference, however, between the reporter levels of wild-type and variant constructs is modest (normalised luminescence ratios of 1.07 ± 0.08 for the wildtype construct compared to 0.87 ± 0.13 for the mutant construct). Moreover, neither the wildtype 3' UTR nor the mutant 3' UTR show any significance response to overexpression of mature miR-433.

The mutant *MTMR3* 3' UTR fragment produces almost twice the amount of the reporter gene (Tukey's post-hoc p-value < 0.001, Figure 5.11C) compared to that of the wildtype 3' UTR. Nonetheless, the mature miR-346 has no effect on either the wildtype or the mutant fragment.

The common variant rs72727021 in the 3' UTR of C9orf100 consists of the major 'T' allele, and a minor 'G' allele. The presence of either allele in the C9orf100 3' UTR of the reporter transcript did not make a significant difference to reporter gene levels in response to overexpression of miR-215, and there are no significant differences as well between a reporter transcript with the 'T' allele and another with the 'G' allele (Figure 5.11D).

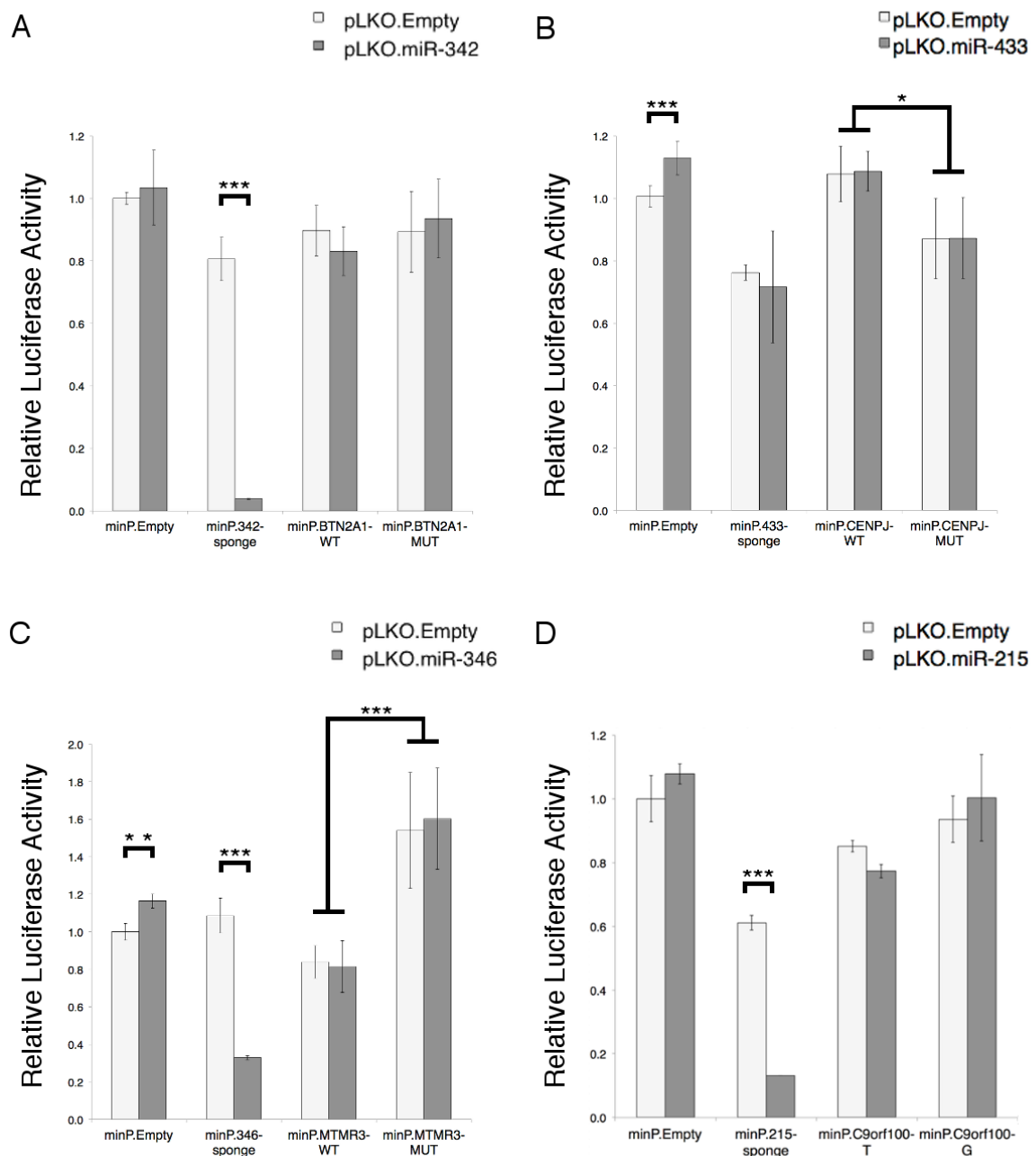


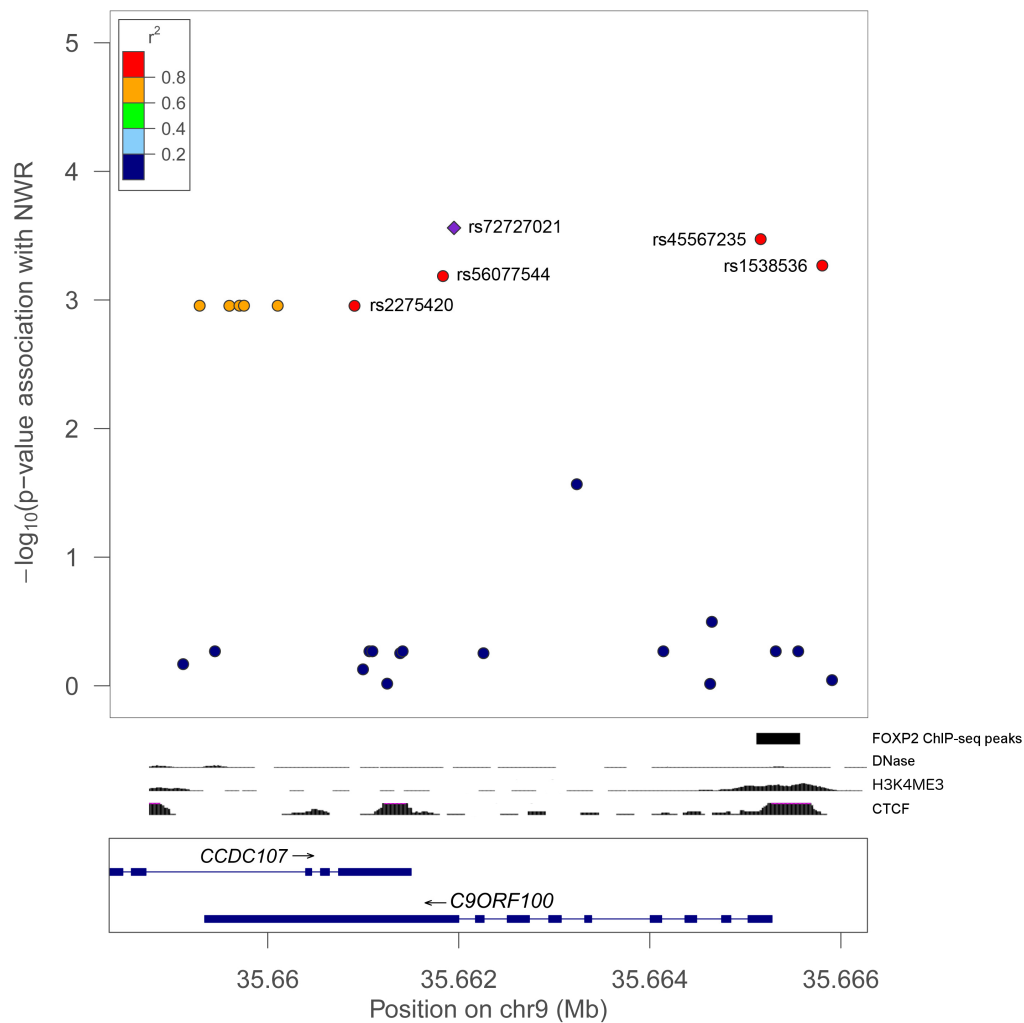
Figure 5.11. Exome-derived 3' UTR Seed-region Variants Are Unlikely To Be Functional.

Normalised luciferase luminescence values were obtained for reporter constructs with fragments of the 3' UTRs of **(A)** *BTN2A1*, **(B)** *CENPJ*, **(C)** *MTMR3*, and **(D)** *C9orf100* carrying the binding site of interest. With the exception of *C9orf100*, the histograms summarise at least two independent experiments. The y-axis for all four histograms charts depict normalised luminescence ratios. To determine statistical significance between different means, a one-way ANOVA was performed, followed by a post-hoc Tukey's test. Significance is indicated as follows: $P < 0.001$ – ***, $P < 0.005$ – **, and $P < 0.01$ – *.

In light of the fact that rs72727021 does not affect the regulation of C9orf100 by miR-215, the regional LD of the C9orf100 locus was investigated in the SLI cohort to account for the association of this 3' UTR miR target site SNP to NWR (Figure 5.12A). This regional association plot reveals that rs72727021 is the most strongly associated with NWR, and is in strong LD with several SNPs spanning the C9orf100 locus. These SNPs (such as rs45567235 and rs1538536) are themselves associated with NWR to the same extent as rs72727021 (Figure 5.12B). Notably, rs45567235 lies in a region shown by *in vitro* ChIP-seq studies to be bound by *FOXP2*. This binding site is also occupied by methylation factors as well as CTCF, pointing to a possible promoter function.

Variants in the 3' UTR have also been shown to disrupt secondary structure (and thus localization and translation) of the mRNA transcript (Chen *et al.* 2006). These mechanisms, rather than microRNA-mediated regulation, might explain the observed difference in reporter gene levels between the wildtype and mutant 3' UTRs for *CENPJ* and *MTMR3*. To test this, the impact of the variants of interest on RNA secondary structure was evaluated using the Vienna RNAfold program (Hofacker and Stadler 2006). The rare variants in *CENPJ* and *MTMR3*, as well as rs72727021 in the c9orf100, but not the rare variant in *BTN2A1*, caused altered folding of the 3' UTR (Figure 5.13).

A



B

SNP	Pairwise r^2 with rs72727021	P-value (Assoc. To NWR)
rs45567235	0.966	0.000336
rs1538536	0.822	0.000539
rs56077544	0.966	0.000650
rs10441726	0.800	0.001105
rs12350788	0.800	0.001105
rs1890593	0.800	0.001105
rs7029668	0.800	0.001105
rs10441725	0.794	0.001107
rs2275420	0.801	0.001107

Figure 5.12. The SNP rs72727021 is in Linkage Disequilibrium With Other Highly Associated SNPs in the C9orf100 Locus.

(A) This figure displays a regional association plot of chr9: 35658341-35666278, which spans the C9orf100 locus with a 1kb flank on either side. The top SNP rs72727021 is in LD with another significantly associated SNP, rs45567235, which lies in a promoter region bound by FOXP2. (B). This table lists the pairwise r^2 and association p-values to NWR for the 4 SNPs in strongest LD with rs72727021.

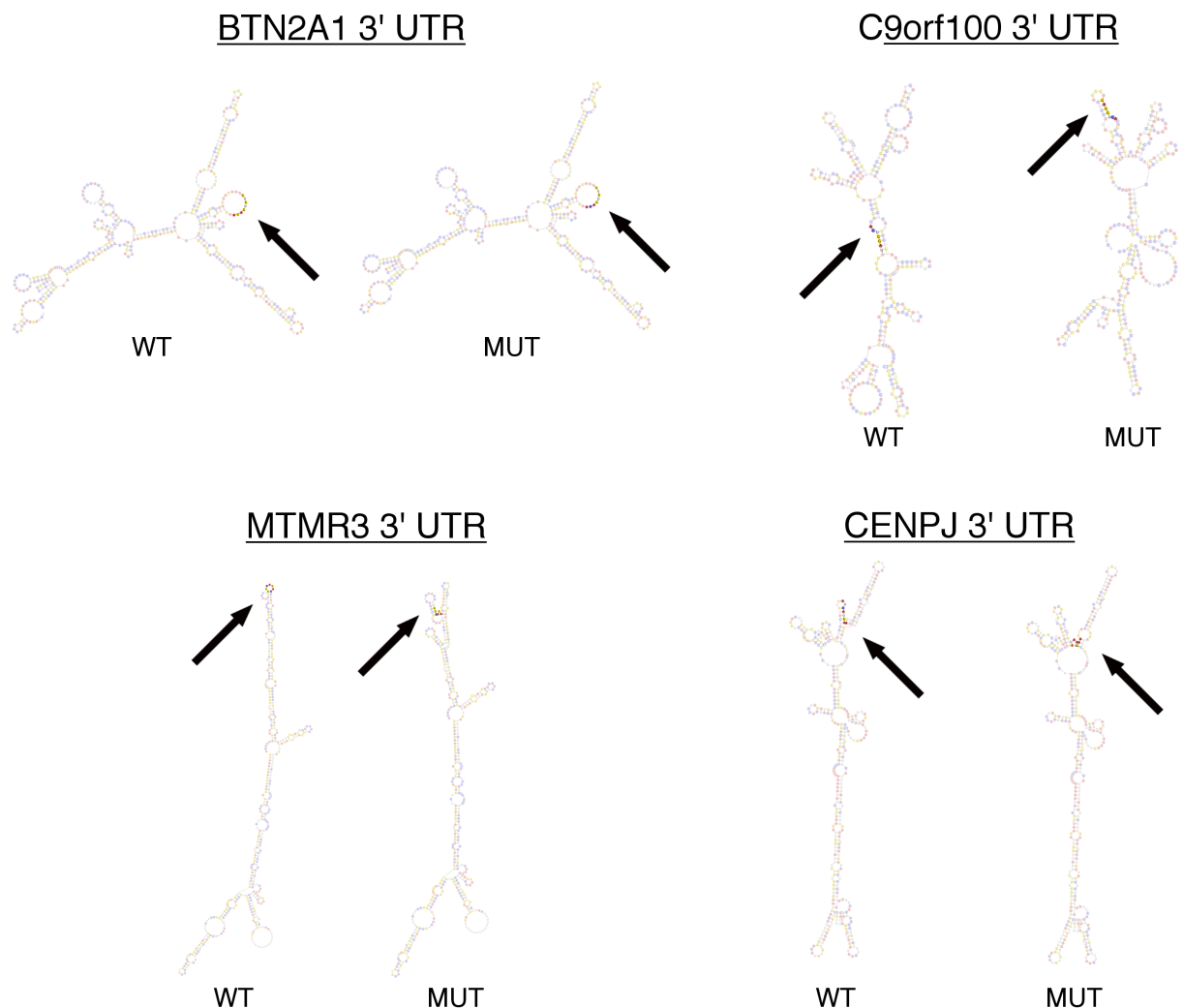


Figure 5.13. Effect of 3' UTR Variants on RNA Folding.

Shown here are the secondary structures of the 3' UTRs cloned for investigation with a reporter gene assay. Highlighted basepairs, as well as black arrows, indicate the seed region of interest. The variants identified in *CENPJ* and *MTMR3*, as well as rs72727021 in *c9orf100*, alter the secondary structure of the 3' UTR, but the variant in *BTN2A1* does not.

All sponge constructs were strongly and reliably down-regulated when the relevant microRNA was overexpressed, with the exception of the miR-433 sponge (Figure 5.11B). This is indicative of either a failure of the construct to overexpress mature miR-433, or a failure of the sponge construct to properly bind miR-433 and induce repression of the construct. To test if the microRNAs were being expressed as desired, HEK cells were transfected with constructs overexpressing hsa-mir-346 and hsa-mir-433. Quantitative PCR was then

performed on cDNA obtained from these cells to detect the presence of the over-expressed mature microRNA.

The miR-346 amplification curves of cDNA from cells transfected with pLKO.miR-346 exhibit an expected leftward shift relative to untransfected cells, and cells transfected with the empty vector, and melting peak analysis indicates that the product being amplified is specific (Figure 5.14A). While the miR-433 amplification curves of cells transfected with pLKO.miR-433 exhibit the same leftward shift in amplification curves, melting peak analysis reveals a profile with two melting peaks (Figure 5.14B), with a major peak at 84°C, and a smaller melting peak at around 75°C (typical for a polyA-tailed mature microRNA amplification product, as is seen in the pLKO.miR-346 melt curves). This points to amplification of a larger polyA-tailed product, likely the transfected plasmid itself, rather than the polyA-tailed mature miR-433. The latter is unlikely to be expressed at detectable levels.

Previous work (Song and Wang 2009) has identified a 1.8kb transcript from which miR-433 is produced. This is substantially larger than the 282bp fragment (containing the hsa-mir-433 stem-loop precursor) that was cloned here to overexpress mature miR-433 (Figure 5.14C). Production of mature miR-433 might require the presence of recognition sites and/or binding elements lying inside in the full-length transcript, which were absent in the shorter fragment used here. Over-expression of mature miR-433 might thus necessitate cloning of the full-length fragment.

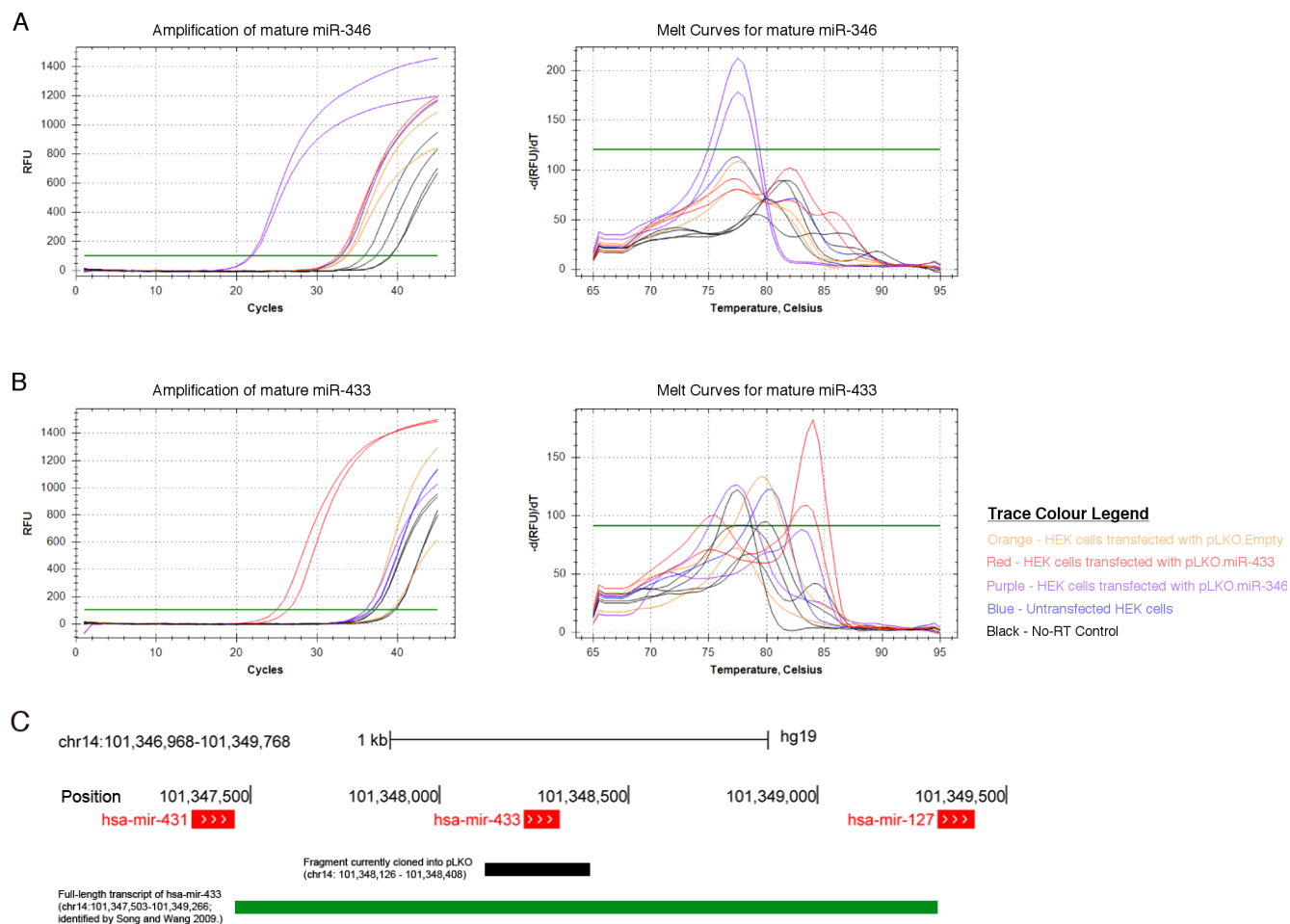


Figure 5.14 The pLKO.mir-433 Plasmid Used is Insufficient to Drive Artificial Overexpression of Mature hsa-miR-433.

Shown here are amplification curves and melt curves of qPCRs performed on cDNA obtained from HEK cells that were transfected with **(A)** pLKO.miR-346 and **(B)** pLKO.miR-433 alongside cells transfected with the empty vector, untransfected cells, and a control without reverse transcriptase added to the RNA sample. The left panels show amplification curves as measured by relative fluorescence units (RFUs) generated by the SYBR Green dye, while the right panels chart the melting peaks of the amplified products, as measured by $-\frac{d(RFU)}{dT}$. The qPCR amplification curves of cDNA from cells transfected with pLKO.miR-346 produce an expected leftward shift, and melting peak analysis shows amplification of a specific product. While cells transfected with pLKO.miR-433 exhibit the same leftward shift, the presence of multiple melting peaks – with the major peak at 84°C rather than a characteristic melting peak at ~75°C for a polyA-tailed mature microRNA amplification product, as is seen in the pLKO.miR-346 melt curves – points to the presence of unspecific product. **(C)** It might be necessary to clone the full-length transcript (as identified by Song and Wang 2009) as it contains promoter elements and other possible motifs necessary for proper processing and production of the mature microRNA.

5.3 Discussion

Figure 5.15 summarises the approaches used to study SNPs that might affect FOXP2-related microRNA biogenesis and function in two datasets: a genome-wide dataset from a family-based SLI cohort and a set of 48 exomes from probands in that cohort. I employed FOXP2-regulated primary microRNA transcripts as candidates in genome-wide association. I also identified SNPs in microRNA precursors, as well as rare and common variants in microRNA binding sites, in exome data, and I employed these as candidates for association in the larger cohort. I showed here that variants gathered from exome data could be integrated with candidate association approaches, as well as reporter gene assays, to determine functional relevance of variants in primary microRNA transcripts, and in microRNA target sites.

5.3.1 Variation in MicroRNA Primary Transcripts and Association to SLI

Previous evidence (Vernes *et al.* 2011) and work in this thesis (Chapter 3) show that FOXP2 regulates microRNA levels, and that FOXP2 target genes are risk factors for SLI (Vernes *et al.* 2008). Recent work relied on data-mining approaches to suggest a possible link between SLI and microRNAs (Rudov *et al.* 2013), but failed to perform formal genetic tests to validate their hypothesis. The work performed here is the first to test if microRNAs that are regulated by FOXP2 are associated with language impairment.

The promoter regions and primary transcripts of microRNAs remain poorly annotated, despite previous efforts (Saini *et al.* 2008, Marson *et al.* 2008, Barski *et al.* 2009, Yang *et al.* 2013). These studies focus on a specific cell-type, or on a handful of well-documented transcription factor binding sites as proxies for promoter regions and transcription start sites. Curated sequence data for primary microRNA transcripts does not currently exist as well.

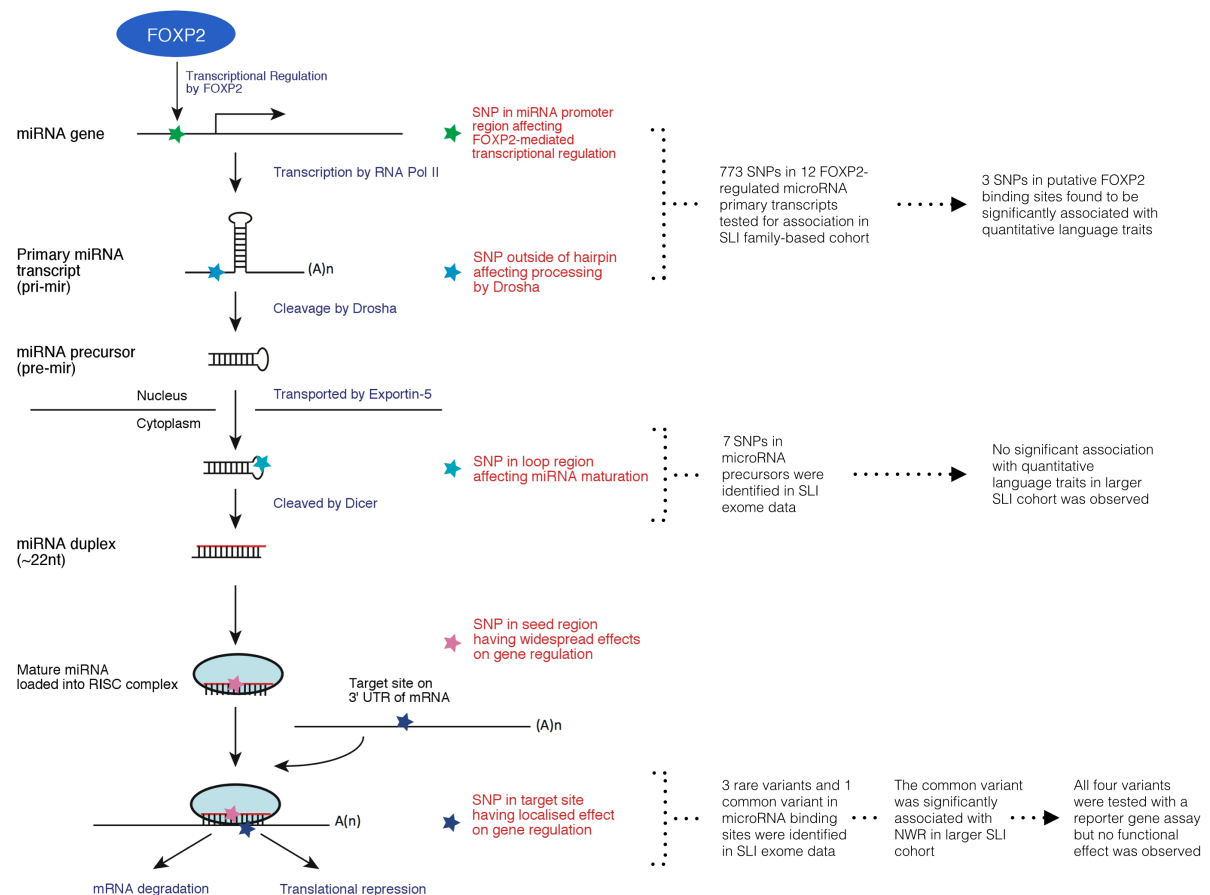


Figure 5.15. Are MicroRNAs Risk Factors For SLI? Integrating Exome Data With Association Analysis.

This figure summarises the approaches employed to study exome data from 48 SLI probands in conjunction with a larger family-based genome-wide SLI dataset, followed by their respective results. Modified after Li and Zhang 2013.

Thus the delineation of primary microRNA transcript coordinates was performed through a combination of literature review and identification of host-gene transcripts, and produced a sizable combined genomic locus for candidate association (773 SNPs across 376.4kb of the genome). While nominal association was observed for ELS and NWR (p -value < 0.01), after correction for multiple testing, no associations were found to be significant. One possible reason for this is the fact that single nucleotide variation in microRNA primary transcripts are likely to have widespread, though modest, individual effects on several targets (Selbach *et al.* 2008). Any genetic risk conferred by these microRNA transcript SNPs is likely to be modulated by genetic heterogeneity, and have reduced penetrance.

The strongest association to ELS was seen at rs78807223, in pri-mir-137. While this SNP does not lie in any known functional region, it has a moderate effect on language ability (the major 'T' allele is predicted to cause a decrease of 12.9 points, or a drop of 0.86 standard deviations, in expressive language scores), and accounts 2.8% of the total variance in the trait. Another SNP in pri-mir-137, rs1625579, was significantly associated with schizophrenia in a genome-wide study (Ripke *et al.* 2011), but it is not in LD with rs78807223, and is not associated with language ability in the SLIC cohort. This suggests that microRNA-137 might be a genetic factor in both schizophrenia and SLI, but in different molecular contexts.

Two other SNPs that showed association with ELS are rs2293481 and rs1534310, and lie in the hsa-mir-25 transcript. Both these SNPs are located in transcriptionally active regions bound by FOXP2, as identified through *in vitro* ChIP-seq (ENCODE Project Consortium, 2011). Both these SNPs are almost 8.5kb upstream from the hsa-mir-25 precursor, and are not in LD with it. Thus, it is variation at the FOXP2 binding site, and not the precursor itself, that is associated with ELS. It cannot be discounted here, however, that the true transcriptional effect of rs2293481 is not on MCM7/hsa-mir25, but on the gene *AP4M1*, given that the promoter of interest overlaps with the 5' UTR of *AP4M1*.

Another SNP, rs4284505, that was associated to NWR, also lies in a region putatively bound by FOXP2. The SNP is 1.89kb upstream from the hsa-mir-20a precursor, and is not in LD with the precursor. Again, association to NWR is driven by variation at the FOXP2 binding site, and not at the precursor region. These results lend support to a model whereby polymorphism in the regulatory region of a microRNA transcript affects FOXP2-mediated expression, contributing risk to SLI (See Figure 5.15, Li and Zhang 2013).

It is important to note here that effect sizes estimated from significant association in a GWAS are likely to be inflated due to ascertainment bias (also known as the "winner's curse") (Göring *et al.* 2001, Zöllner and Pritchard 2007). This arises from the fact both statistical significance (as measured by p-values) and the

estimated effect sizes are determined from the same samples. As such, these two related, but separate measures are correlated, and are not independent. Thus, given a genotype that is significantly associated with a phenotype, the effect size of the genotype on the phenotype will tend to be biased upwards (Zöllner and Pritchard 2007). To practically and reasonably reduce inflation of effect size due to ascertainment bias, the effect sizes of risk alleles identified here should be re-evaluated in an independent cohort.

It is also worth noting that a recent study of variation in microRNA precursor regions (Quach *et al.* 2012) only examined a 1.2kb region flanking the precursors of 117 miRNAs of interest; such a cutoff would have not yielded the FOXP2-binding site polymorphisms that are associated with SLI in this study. This suggests that considering longer-range cis-regulatory regions (that might affect primary transcript levels) could enrich current views of how microRNA precursor variation affects human traits and disease.

The polymorphism rates (Saunders *et al.* 2006) and mutation rates (Quach *et al.* 2009) of SNPs in microRNA stem-loop precursors, and their contribution to human disease (Bulik-Sullivan *et al.* 2013), have been studied at a population level, but a similar analysis for primary transcripts of microRNAs (and their putative regulatory regions) has not been performed. This will require better annotation of primary microRNA transcripts, and higher-resolution capture of the genome, and identification of regulatory motifs and transcription factor binding sites in these transcripts. SNPs in primary microRNA transcripts can then be studied at a population level, and meta-analysed in the context of previous genome-wide disease association studies (Bulik-Sullivan *et al.* 2013). Such work will be invaluable in understanding how variation might affect microRNA transcription and processing, and thus contribute to complex human diseases such as SLI.

5.3.2 Functional Investigation of Exome-Derived 3' UTR Binding Site Variants

In addition to variants in microRNA precursor regions, I also identified both 3' UTR variants from the exomes of 48 SLI probands, followed by candidate association analysis of these variants. Out of 5 SNPs present in 48 exomes of SLI probands, only rs72727021, lying in the miR-215/192 binding site of the *C9orf100* 3' UTR, was significantly associated with nonword repetition in the larger SLIC cohort. Three rare microRNA binding site variants were also identified. These were all heterozygous mutations occurring in one exome each, and lay in the genes *BTN2A1*, *CENPJ*, and *MTMR3*.

None of the four 3' UTRs studied here were responsive to over-expression of the cognate microRNAs, and thus none of the predicted binding sites are likely to be real. This suggests that binding site conservation and thermodynamic properties are insufficient to assess if the binding sites are truly functional, or if the identified variants might disrupt microRNA targeting. It is important to note that, currently, there does not exist an empirical threshold for conservation or thermodynamic stability to distinguish true binding sites from falsely predicted ones. The rare variants occurred at conserved basepairs (as measured by GERP scores), and the target sites identified were conserved in vertebrates; considered in isolation, this provided good evidence of functionality. Also, for four variants, there was a difference in the stability of microRNA-mRNA duplexes (as measured by the free energy ΔG) of the wildtype and mutant duplexes. Despite this, none of the 3' UTRs could be demonstrated to be down-regulated by the predicted microRNAs.

Why did rs72727021 in the 3' UTR of *C9orf100* not show a significant effect on microRNA binding, despite association to nonword repetition? Firstly, the effect size of the risk allele of this SNP might be inflated due to ascertainment bias, as mentioned above in Section 5.3.1. A second explanation is that it might be difficult to observe robust biological effects caused by common variation in microRNA binding sites at a single-gene level. Li *et al.* (2012), employing the

same target database (TargetScan) used here, demonstrated allele-dependent microRNA regulation in only 3 out of 7 candidate common variants in predicted microRNA binding sites. Lin and colleagues (2012) found that a common SNP in a microRNA binding site produced modest allelic difference in reporter gene assays. A large-scale study that measured protein levels with mass spectrometry in response to microRNA overexpression observed widespread, but relatively small, repressive effects (Selbach *et al.* 2008). A third reason is that rs72727021 is in LD with an upstream FOXP2 binding site that the true locus of association. *C9ORF100* remains an attractive candidate that affects nonword repetition ability in this SLI cohort, but the work here does not shed any light on a possible molecular mechanism.

The inability to validate four putative microRNA binding sites here suggests that the microRNA prediction algorithm employed here lacks specificity. Work done by Selbach and colleagues found that, across 9 microRNA target prediction algorithms, 40% to 61% of predicted targets showed protein repression (Selbach *et al.* 2008). A more recent investigation of target prediction algorithms (using reporter gene assays), however, found poorer performance across the board: the highest accuracy rate observed was 27%, with a 6% false negative rate (Zhou *et al.* 2013). It has been previously suggested that integrating expression levels of transcripts might help identify bona-fide microRNA targets (Saunders *et al.* 2006). A few algorithms have attempted to take such information into account (such as Stempor *et al.* 2012 and Gennarino *et al.* 2009) but formal assessment of their performance is yet to be done. While evidence exists that the four microRNA-genes in this chapter are expressed in the brain, the exact spatial-temporal expression overlap of cognate microRNA-mRNA pairs remains unclear, and might have pointed towards a lack of functional interaction. In addition, I cannot rule out here that the failure to show an interaction between miR-433 and *CENPJ* was due to an inability to successfully over-express miR-433.

The field of microRNA biology is heavily reliant on *in vitro* reporter gene assays to validate computational predictions. Given the growing collection of validated microRNA-mRNA interactions, such as Tarbase (Vergoulis *et al.* 2011), target

predictions could be refined by integrating empirical data. Recent work using *in vitro* methods to cross-link microRNA target transcripts to Argonaut proteins, followed by immunoprecipitation and high-throughput sequencing (abbreviated as CLIP-seq) have identified RISC-bound microRNA target sites in an unbiased manner (Hafner *et al.* 2010), and attempts have been made to integrate these datasets with computational target predictions (Liu *et al.* 2013). Such improvements to microRNA predictions will enable better discrimination of true functional 3' UTR variants.

Of the three rare 3' UTR variants tested here, only *BTN2A1* did not exhibit any difference in reporter gene levels between the wildtype and mutant 3' UTRs. The other two 3' UTRs, however, had a microRNA-independent difference in reporter gene levels between the wildtype and mutant. The mutant 3' UTR of *CENPJ* produces a significant, but modest, decrease in reporter gene levels relative to the wildtype 3' UTR. The mutant 3' UTR of *MTMR3* produced almost twofold reporter gene levels relative to the wildtype 3' UTR. This pattern was also observed in the predicted effect of the variant on RNA secondary structure: the secondary structures of the 3' UTRs of *MTMR3* and *CENPJ*, but not *BTN2A1*, were altered by the presence of the rare variant. These differences in reporter gene levels might thus be caused by differences in RNA secondary structure. Previous work has shown that disease-associated 3' UTR variants can cause changes to RNA folding (Reamon-Buettner *et al.* 2007). This suggests that 3' UTR variants should be not be evaluated solely in terms of microRNA targeting, but also on impact on RNA folding.

The results here reinforce the basic requirement to study segregation of putative risk variants in pedigrees of interest. The rare variants in the 3' UTRs of *BTN2A1* and *CENPJ* segregated perfectly with SLI affection status within the offspring in the pedigrees, providing good support for the variant having a functional effect. The rare variant in the 3' UTR of *MTMR3*, however, was present in the proband as well as an unaffected sibling. This variant might thus have reduced

penetrance, or a “multi-hit” genetic model might underlie this particular form of SLI, where the inherited variant in *MTMR3* might interact with an unidentified *de novo* variant in this proband to produce the observed language impairment (as in O’Roak *et al.* 2011).

This work represents initial efforts at identifying the contribution of noncoding regions to human speech and language disorders. Whole-genome sequencing has been applied to identify causative variants in complex brain disorders (Lupski *et al.* 2010), and will overcome the limited ability of whole-exome approaches to identify noncoding variation. Coupling whole-genome capture with better annotation of primary microRNA transcripts, followed with functional testing, will present a better picture of the genetic networks underlying human speech and language, and other complex human traits.

Chapter 6: Conclusion

6.1 Summary of Findings of This Thesis

FOXP2 is the first gene identified as a monogenic cause of developmental verbal dyspraxia in humans (Lai *et al.* 2001, MacDermot *et al.* 2005). It encodes a transcription factor that is highly expressed in the central nervous system during development, as well as throughout adulthood. In zebra finches, it is actively regulated during singing behaviour (Teramitsu and White, 2006, Miller *et al.* 2008, Teramitsu *et al.* 2010), and knockdown of FoxP2 in brain regions required for song learning disrupts song imitation and song variability (Haesler *et al.* 2007, Murugan *et al.* 2013). Mice heterozygous for mutant R552H.Foxp2 (homologous to the mutation first seen in humans) have impaired motor-skill learning (Groszer *et al.* 2008, French *et al.* 2012).

MicroRNAs are a growing family of short noncoding genes that act to regulate gene transcription by base-pairing with the 3' UTR of gene transcripts. Given this mechanism of gene regulation, a single microRNAs is able to potentially regulate the expression of several thousand gene transcripts. As such, microRNAs play an important role in the fine-tuning of gene networks, and have been shown to play crucial roles in neuronal processes such as neurogenesis and differentiation (Conaco *et al.* 2006, Visvanathan *et al.* 2007). In this thesis, I investigated the regulation of microRNA networks by Foxp2/FOXP2, and assessed the risk contributed by FOXP2-regulated microRNAs to specific language impairment (SLI).

I identified microRNAs regulated by Foxp2 in the embryonic mouse brain, and then validated a number of these microRNAs in relevant neuronal cell-lines. This revealed that *MIR137* was a conserved target of Foxp2/FOXP2. In addition, I put forward the hypothesis that FOXP2 and *MIR137* formed a genetic mini-network with 2 other FOXP2 targets, *TCF4* and *CNTNAP2*. I evaluated this hypothesis *in vitro*, validating *TCF4* as a transcriptional target of FOXP2. I also studied the molecular mechanisms of FOXP2 regulation of this mini-network. Because

FOXP2 targets (namely *CNTNAP2*) have been associated with SLI, I tested a set of FOXP2-regulated microRNAs for SLI risk in a family-based cohort, and found that they did not represent significant genetic risk for SLI. I also demonstrated a novel use of exome data to identify and study microRNA binding sites mutations in SLI probands, and I performed functional investigation of their transcriptional effects.

6.2 An Expanded Reach of FOXP2 via MicroRNAs

In this thesis, microarray analysis revealed that *Foxp2* significantly regulated 8 microRNAs in the embryonic mouse brain. This suggests that, via the action of microRNAs, *Foxp2* indirectly regulates a secondary set of genes, in addition to its direct targets. Axon guidance was a key biological theme that was over-represented in the predicted gene targets of the top 8 microRNAs. This corroborates previously established links between *Foxp2* and neural connectivity (Enard *et al.* 2009, Vernes *et al.* 2011, Reimers-Kipping *et al.* 2011). Previous work showed that direct targets of *Foxp2*, as well as genes showing *Foxp2*-dependent expression, are enriched for genes involved in axon guidance (Vernes *et al.* 2011). Therefore, it is likely that the microRNAs targeted by *Foxp2*/FOXP2 act to fine-tune the expression of these genes. Future work will characterise these microRNA-mediated indirect effects of *Foxp2*, by investigating if key microRNAs and their gene targets can rescue reduced neurite outgrowth following loss of *Foxp2* function (Vernes *et al.* 2011).

It is important to acknowledge two limitations of the microRNA expression profiling performed here. Firstly, while the microarray platform used captures several thousand microRNAs, it is by no means the definitive set of known microRNAs. Given that the list of miRNAs in mammalian genomes continues to grow, the work presented here is leading work that was based upon developing technologies. Secondly, microRNA expression in other brain regions that express *Foxp2* during development, such as the Purkinje cells in the cerebellum, and the cortical plate (Lai *et al.* 2003), were not studied here. Future work can remedy these two areas, by obtaining the transcriptomes of specific *Foxp2*-expressing

cellular populations, using RNA sequencing (RNA-seq) in both mutant and wildtype mice. This will allow an unbiased interrogation of microRNA regulation by Foxp2, and also shed light on the different functions of Foxp2 in different brain regions.

6.3 Foxp2/FOXP2 Function, Genetic Background And Cell-type Differences

The differences between *in vitro* and *in vivo* regulation of microRNAs observed in Chapter 3 might be modulated by genetic background. Independent *in vivo* studies have yielded conflicting observations regarding the phenotypic effects of mutant Foxp2, suggesting that genetic background plays a part in determining phenotypic outcomes of Foxp2 mutations (Fisher and Scharff 2009, Scharff and Petri 2011). Shu *et al.* (2005) assessed null heterozygote mutants derived on a mixed background (without backcrossing), and observed modest developmental delays and abnormal gross cerebellar morphology. Groszer *et al.* (2008) and French *et al.* (2007 and 2012), however, studied null mutants backcrossed onto founder backgrounds, and found instead that heterozygotes had normal motor development and normal gross brain morphology. To understand how genetic backgrounds modulate the effects of Foxp2, these mutant mice from different genetic backgrounds can be subject in parallel to molecular studies of Foxp2, electrophysiological experiments, and behavioural characterization. Such an effort would require significant collaborative efforts, but would provide a valuable opportunity to examine Foxp2 function.

The specific molecular effects of Foxp2 in the different neuronal subtypes that express it (such as cortical neurons versus striatal neurons, Reimers-Kipping *et al.* 2011) might also be explained by different molecular profiles possessed by each neuronal subtype. These effects of cell-type might also account for the discordance between FOXP2 binding (as observed with chromatin immunoprecipitation experiments in human neuron-like cell-lines) and reporter gene regulation by FOXP2 (as observed through luciferase assays in HEK293 cell-lines) across different cell-lines. FOXP2 binding at regulatory regions, as well as

the resultant downstream transcriptional effects, might be modulated not only by genomic context (such as the presence of chromatin loops), but also by cell-type specific differences. Such differences include variable composition of the bound regulatory complexes between cell-types (Mullen *et al.* 2011, Arvey *et al.* 2012), and chromatin accessibility (Håndstad *et al.* 2012). These can result in differential binding of transcription factors at the same locus across different cell-lines.

One way to identify the effects of cell-type on FOXP2 is to synthesise the ChIP data already available across different neuronal cell-lines and brain regions. Vernes *et al.* (2007) performed ChIP-chip on SH-SY5Y cells that stably expressed FOXP2, while Spiteri *et al.* (2007) performed ChIP-chip on foetal human brain (namely basal ganglia and inferior frontal cortex). The ENCODE project (ENCODE Project Consortium, 2012) contains ChIP-seq data from 2 neuronal cell-lines (SK-N-MC and PFSK-1 cells lines) that natively express FOXP2. Indeed, Nelson and colleagues (2013) analyzed the ENCODE FOXP2 ChIP-seq data and identified 71 high-confidence peaks that occurred in all replicates of both cell-lines. Such an approach might be extended to other FOXP2/Foxp2 occupancy data, with close attention paid to the experimental methodologies and cellular models used in each study. Another future experimental approach is to generate neuronal cells from induced pluripotent stem cells (iPSCs) derived from KE family members, as well as from individuals carrying other identified mutations, such as the R328X nonsense mutation (MacDermot *et al.* 2005).

6.4 FOXP2 and Genetic Networks Underlying Different Neurodevelopmental Disorders

In the literature, the biological notion of how genes are linked into a network has three facets. Firstly, genes can be linked via sequential function (such as an enzyme-catalysed metabolic pathway, or a signal cascade involving protein phosphorylation). Secondly, a set of genes that interact, either at a genetic level or a physical level, is also termed a genetic network. Thirdly, a genetic network can be defined as genes that are co-expressed in a certain tissue, or after a

certain experimental treatment (Oldham *et al.* 2006). All these concepts describing how genes function together are central to the analysis of genetic networks. The outcome of such analysis is twofold: the distillation of large gene lists into a few relevant biological pathways, and the identification of key biological themes that can generate hypotheses regarding the fundamental biology behind the condition or disorder of interest (Khatri *et al.* 2012).

These concepts of genetic pathways are relevant to studies of transcriptional regulators, such as FOXP2, whose main function is to control the levels of several different genes. FOXP2 is often described as a window into speech and language impairments (Fisher 2007, Fisher and Scharff 2009), as its target genes (such as *CNTNAP2*, *DISC1*, and *SPRX2*) as well as its dimerization partners (such as *CTBP1*, *FOXP1*, and *TBR1*) are implicated across a wide range of neurodevelopmental disorders involving language problems, such as intellectual disability and autism (Li *et al.* 2004, Vernes *et al.* 2008, Hamdan *et al.* 2010, Roll *et al.* 2010, O’Roak *et al.* 2011, Walker *et al.* 2012, Deriziotis *et al.* 2014b). The recurrence of several genes that are functionally related to FOXP2 in genetic studies of different neurodevelopmental disorders implies that FOXP2-related pathways might underlie the language deficits seen in these neurodevelopmental disorders.

In this thesis, I hypothesized the existence of a genetic network (defined as a set of genetic and physical interactions) involving FOXP2, *TCF4*, *CNTNAP2*, and *MIR137*. I showed that *TCF4* is up-regulated *in vitro* by FOXP2. This represents a novel genetic connection between distinct neurodevelopmental disorders that affect speech and language. Mutations in *TCF4* result in Pitt–Hopkins Syndrome (PTHS), an autism-spectrum disorder with features of developmental delay and intellectual disability (Sweatt 2013), and common variants in *TCF4* are associated with schizophrenia (Ripke *et al.* 2011, Ripke *et al.* 2013). Thus, the FOXP2-*TCF4* regulatory relationship is additional evidence for the role of FOXP2 and its targets in genetic networks that play roles across neurodevelopmental disorders where language is disrupted.

MicroRNAs have also been implicated in genetic networks underlying complex neurodevelopmental disorders. MiR-137 has emerged as a risk factor across multiple genome-wide scans for schizophrenia risk loci (Ripke *et al.* 2011, Ripke *et al.* 2013, Strazisar *et al.* 2014). It has also been implicated in the molecular networks associated with autism, as it regulates the autism candidate gene ROR-alpha by binding to multiple sites in the ROR-alpha 3' UTR (Devanna and Vernes 2014). In this thesis I showed that microRNAs are involved in the molecular networks underlying human speech and language. Firstly, I showed that a number of microRNAs are differentially regulated in developing mice brains lacking functional Foxp2. Secondly, I showed that miR-137 is up-regulated by FOXP2 *in vitro*. Thirdly, I screened a set of 13 FOXP2-regulated microRNAs for risk in a large and well-studied family-based SLI cohort, but did not observe significant risk for those loci tested. Even if they do not contribute significant risk for SLI, in light of their functional relationship with FOXP2, these microRNAs might be risk factors for other neurodevelopmental disorders, and should not be discounted as candidates for future investigation.

6.5 The Functional Consequences of MicroRNA Variation: Moving Beyond Seed Region and Binding Site Variants

Exome sequencing has proven valuable in identifying protein-coding mutations underlying for neurodevelopmental disorders that are polygenic or genetically heterogenous in nature (such as autism). However, noncoding regions, such as microRNA transcripts, and microRNA binding sites, remain under-studied as putative risk factors. Emerging work has begun to investigate common variation in microRNA transcripts, outside of the seed region, and their effects on microRNA levels (Strasizar *et al.* 2014, Warburton *et al.* 2014). This suggests that noncoding variation might account for some of the genetic heterogeneity in various neurodevelopmental disorders.

I show here that it is possible to identify noncoding mutations in microRNA binding sites from exome data. This can serve to motivate re-analysis of current exome data sets to identify noncoding variation, which might have been filtered

out during initial analysis. Future exome sequencing projects should also refrain from excluding microRNA variants and 3' UTR binding site variants. Any variants lying in microRNA binding sites can then be subject to reporter gene assays, as were performed in Chapter 5.

As technological advancements continue, whole-genome sequencing of probands will become routine in both clinical and research settings. This will uncover a larger number of disease-associated mutations that lie in noncoding regions. In the context of primary microRNA (pri-mir) transcripts, more variants lying outside of the seed region and the precursor region will be revealed. Thus, future work will move beyond functional testing of seed region variants and 3' UTR variants in microRNA binding sites, for which the molecular biology is relatively well understood. Transcription factor occupancy data and RNA folding predictions will allow putative function to be assigned to non-seed region variants in pri-mir transcripts, and will point towards possible diagnostic candidates. Future experimental directions on functional variation in pri-mir transcripts are exemplified in very recent work from Strasizar *et al.* (2014) and Warburton *et al.* (2014). Both these groups functionally investigated the effects of common variants in the pri-mir-137 transcript lying outside the seed region on promoter activity, and on subsequent mature miR-137 expression.

6.6 Genetic Networks, and Human Speech and Language: The Ways Forward

Human speech and language requires the functional integration of neural pathways involved in cognitive and motor processes. To build these neural pathways, several different genes must interact in appropriate cellular contexts and developmental time-points. The transcription factor *FOXP2*, which causes a severe deficit in speech and language when disrupted, is a unique and effective entry-point into studying these genetic networks. MicroRNAs have gained widespread attention for their post-transcriptional effects on gene expression, but the factors governing their upstream regulation remain under-studied.

In this thesis, I identified several microRNAs under the regulation of FOXP2/Foxp2, and I tested their risk contribution to SLI in a family-based cohort. While much of the field has focussed on the function of wildtype FOXP2 / Foxp2, here I study the R533H mutation on FOXP2 regulation of microRNAs, revealing its complex effects. The work described here is the first, to my knowledge, to identify 3' UTR variants from exome data, and to subject them to *in vitro* functional testing. This suggests that the growing repository of exome data (across studies of various human phenotypes) harbours noncoding mutations that might provide functional insight. In addition, I functionally investigated a mini-network involving FOXP2 target genes *MIR137*, *TCF4*, and *CNTNAP2*, each of which are implicated in a range of neurodevelopmental disorders. MicroRNAs such as miR-9, miR-34a, and miR-140 have recently been shown to regulate FOXP2/Foxp2 expression (Clovis *et al.* 2012, Shi *et al.* 2013, Cuiffo *et al.* 2014). Future studies will continue functional validation and manipulation of FOXP2/Foxp2 microRNA targets (and putative feedback loops) identified here in appropriate cell-lines and animal models. Functional validation can proceed in parallel with candidate association studies and variant identification in relevant cohorts, as demonstrated in this thesis. The integration of these different analyses will shed more light on the genetic networks and neural pathways underlying human speech and language.

References

- Abecasis, G.R., Cherny, S.S., Cookson, W.O., and Cardon, L.R. (2001). Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 30, 97–101.
- Abelson, J.F., Kwan, K.Y., O’Roak, B.J., Baek, D.Y., Stillman, A.A., Morgan, T.M., Mathews, C.A., Pauls, D.L., Rasin, M.-R., Gunel, M., et al. (2005). Sequence variants in *SLITRK1* are associated with Tourette's syndrome. *Science* 310, 317–320.
- Abrahams, B.S., Tentler, D., Perederiy, J.V., Oldham, M.C., Coppola, G., and Geschwind, D.H. (2007). Genome-wide analyses of human perisylvian cerebral cortical patterning. *Proc Natl Acad Sci USA* 104, 17849–17854.
- Abu-Elneel, K., Liu, T., Gazzaniga, F.S., Nishimura, Y., Wall, D.P., Geschwind, D.H., Lao, K., and Kosik, K.S. (2008). Heterogeneous dysregulation of microRNAs across the autism spectrum. *Neurogenetics* 9, 153–161.
- Ackermann, H. (2007). Cerebellar contributions to speech production and speech perception: psycholinguistic and neurobiological perspectives. *Trends Neurosci* 31, 265–272.
- Addis, L., Friederici, A.D., Kotz, S.A., Sabisch, B., Barry, J., Richter, N., Ludwig, A.A., Rübsamen, R., Albert, F.W., Pääbo, S., et al. (2010). A locus for an auditory processing deficit and language impairment in an extended pedigree maps to 12p13.31-q14.3. *Genes, Brain and Behavior* 9, 545–561.
- Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. *Nat Meth* 7, 248–249.
- Agholme, L., Lindström, T., Kågedal, K., Marcusson, J., and Hallbeck, M. (2010). An in vitro model for neuroscience: differentiation of SH-SY5Y cells into cells with morphological and biochemical characteristics of mature neurons. *J. Alzheimers Dis.* 20, 1069–1082.
- Al-Dosari, M.S., Shaheen, R., Colak, D., and Alkuraya, F.S. (2010). Novel CENPJ mutation causes Seckel syndrome. *Journal of Medical Genetics* 47, 411–414.
- Alarcón, M., Abrahams, B.S., Stone, J.L., Duvall, J.A., Perederiy, J.V., Bomar, J.M., Sebat, J., Wigler, M., Martin, C.L., Ledbetter, D.H., et al. (2007). Linkage, association, and gene-expression analyses identify *CNTNAP2* as an autism-susceptibility gene. *The American Journal of Human Genetics* 82, 150–159.
- Alario, F.X., Chainay, H., Lehericy, S., and Cohen, L. (2006). The role of the supplementary motor area (SMA) in word production. *Brain Research Reviews* 1076, 129–143.

- Alcock, K.J., Passingham, R.E., Watkins, K.E., and Vargha-Khadem, F. (2000). Oral dyspraxia in inherited speech and language impairment and acquired dysphasia. *Brain Lang* 75, 17–33.
- Alexander, M.P., Naeser, M.A., and Palumbo, C. (1990). Broca's area aphasias: aphasia after lesions including the frontal operculum. *Neurology* 40, 353–362.
- Alon, U. (2007). Network motifs: theory and experimental approaches. *Nat Rev Genet* 8, 450–461.
- Amunts, K., and Zilles, K. (2012). Architecture and organizational principles of Broca's region. *Trends Cogn Sci (Regul Ed)* 16, 418–426.
- Anders, G., Mackowiak, S.D., Jens, M., Maaskola, J., Kuntzagk, A., Rajewsky, N., Landthaler, M., and Dieterich, C. (2011). doRiNA: a database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res* 40, D180–D186.
- Anderson, G.R., Galfin, T., Xu, W., Aoto, J., Malenka, R.C., and Sudhof, T.C. (2012). Candidate autism gene screen identifies critical role for cell-adhesion molecule CASPR2 in dendritic arborization and spine development. *Proceedings of the National Academy of Sciences* 109, 18120–18125.
- Anney, R., Klei, L., Pinto, D., Almeida, J., Bacchelli, E., Baird, G., Bolshakova, N., Bolte, S., Bolton, P.F., Bourgeron, T., et al. (2012). Individual common variants exert weak effects on the risk for autism spectrum disorders. *Hum Mol Genet* 21, 4781–4792.
- Arcila, M.L., Betizeau, M., Cambronne, X.A., Guzman, E., Doerflinger, N., Bouhallier, F., Zhou, H., Wu, B., Rani, N., Bassett, D.S., et al. (2014). Novel Primate miRNAs Coevolved with Ancient Target Genes in Germinal Zone-Specific Expression Patterns. *Neuron*.
- Arking, D.E., Cutler, D.J., Brune, C.W., Teslovich, T.M., West, K., Ikeda, M., Rea, A., Guy, M., Lin, S., Cook, E.H., Jr., et al. (2008). A Common Genetic Variant in the Neurexin Superfamily Member CNTNAP2 Increases Familial Risk of Autism. *The American Journal of Human Genetics* 82, 160–164.
- Arora, S., Rana, R., Chhabra, A., Jaiswal, A., and Rani, V. (2013). miRNA-transcription factor interactions: a combinatorial regulation of gene expression. *Mol Genet Genomics* 288, 77–87.
- Arriaga, G., and Jarvis, E.D. (2012). Mouse vocal communication system: are ultrasounds learned or innate? *Brain Lang* 124, 96–116.
- Arvey, A., Agius, P., Noble, W.S., and Leslie, C. (2012). Sequence and chromatin determinants of cell-type-specific transcription factor binding. *Genome Research* 22, 1723–1734.
- Aschrafi, A., Schwechter, A.D., Mameza, M.G., Natera-Naranjo, O., Gioio, A.E.,

and Kaplan, B.B. (2008). MicroRNA-338 Regulates Local Cytochrome c Oxidase IV mRNA Levels and Oxidative Phosphorylation in the Axons of Sympathetic Neurons. *J Neurosci* 28, 12581–12590.

Ayub, Q., Yngvadottir, B., Chen, Y., Xue, Y., Hu, M., Vernes, S.C., Fisher, S.E., and Tyler-Smith, C. (2013). FOXP2 targets show evidence of positive selection in European populations. *Am J Hum Genet* 92, 696–706.

Badcock, N.A., Bishop, D.V.M., Hardiman, M.J., Barry, J.G., and Watkins, K.E. (2012). Brain & Language. *Brain Lang* 120, 310–320.

Bagga, S., Bracht, J., Hunter, S., Massirer, K., Holtz, J., Eachus, R., and Pasquinelli, A.E. (2005). Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell* 122, 553–563.

Bakkaloglu, B., O'Roak, B.J., Louvi, A., Gupta, A.R., Abelson, J.F., Morgan, T.M., Chawarska, K., Klin, A., Ercan-Sencicek, A.G., Stillman, A.A., et al. (2008). Molecular Cytogenetic Analysis and Resequencing of Contactin Associated Protein-Like 2 in Autism Spectrum Disorders. *The American Journal of Human Genetics* 82, 165–173.

Ballantyne, A.O., Spilkin, A.M., and Trauner, D.A. (2007). The Revision Decision: Is Change Always Good? A Comparison of CELF-R and CELF-3 Test Scores in Children With Language Impairment, Focal Brain Damage, and Typical Development. *Language, Speech, and Hearing Services in Schools* 38, 182–189.

Ballarino, M., Pagano, F., Girardi, E., Morlando, M., Cacchiarelli, D., Marchioni, M., Proudfoot, N.J., and Bozzoni, I. (2009). Coupled RNA Processing and Transcription of Intergenic Primary MicroRNAs. *Mol Cell Biol* 29, 5632–5638.

Bamshad, M.J., Ng, S.B., Bigham, A.W., Tabor, H.K., Emond, M.J., Nickerson, D.A., and Shendure, J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 12, 745–755.

Barbas, H., García-Cabezas, M.Á., and Zikopoulos, B. (2013). Frontal-thalamic circuits associated with language. *Brain Lang* 126, 49–61.

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

Barski, A., Jothi, R., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., and Zhao, K. (2009). Chromatin poises miRNA- and protein-coding genes for expression. *Genome Research* 19, 1742–1751.

Bartel, D.P. (2004). MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297.

Bartel, D.P., and Chen, C.-Z. (2004). Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nat Rev Genet* 5, 396–

Bartlett, C.W., Flax, J.F., Logue, M.W., Smith, B.J., Vieland, V.J., Tallal, P., and Brzustowicz, L.M. (2004). Examination of potential overlap in autism and language loci on chromosomes 2, 7, and 13 in two independent samples ascertained for specific language impairment. *Urol Int* 57, 10–20.

Bartlett, C.W., Flax, J.F., Logue, M.W., Vieland, V.J., Bassett, A.S., Tallal, P., and Brzustowicz, L.M. (2002). A major susceptibility locus for specific language impairment is located on 13q21. *Am J Hum Genet* 71, 45–55.

Baskerville, S., and Bartel, D.P. (2005). Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *Rna* 11, 241–247.

Baudet, M.-L., Zivraj, K.H., Abreu-Goodger, C., Muldal, A., Armisen, J., Blenkiron, C., Goldstein, L.D., Miska, E.A., and Holt, C.E. (2011). miR-124 acts through CoREST to control onset of Sema3A sensitivity in navigating retinal growth cones. *Nat Neurosci* 15, 29–38.

Beitchman, J.H., Brownlie, E.B., Inglis, A., Wild, J., Mathews, R., Schachter, D., Kroll, R., Martin, S., Ferguson, B., and Lancee, W. (1994). Seven-year follow-up of speech/language-impaired and control children: speech/language stability and outcome. *J Am Acad Child Adolesc Psychiatry* 33, 1322–1330.

Bell, M.L., Buvoli, M., and Leinwand, L.A. (2010). Uncoupling of Expression of an Intronic MicroRNA and Its Myosin Host Gene by Exon Skipping. *Mol Cell Biol* 30, 1937–1945.

Belton, E., Salmond, C.H., Watkins, K.E., Vargha-Khadem, F., and Gadian, D.G. (2003). Bilateral brain abnormalities associated with dominantly inherited verbal and orofacial dyspraxia. *Hum. Brain Mapp.* 18, 194–200.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B* 57, 289–300.

Berdnik, D., Fan, A.P., Potter, C.J., and Luo, L. (2008). MicroRNA Processing Pathway Regulates Olfactory Neuron Morphogenesis. *Curr Biol* 18, 1754–1759.

Berezikov, E., Chung, W.-J., Willis, J., Cuppen, E., and Lai, E.C. (2006). Mammalian Mirtron Genes. *Mol Cell* 28, 328–336.

Bergqvist, I., Eriksson, M., Saarikettu, J., Eriksson, B., Corneliussen, B., Grundström, T., and Holmberg, D. (2000). The basic helix-loop-helix transcription factor E2-2 is involved in T lymphocyte development. *Eur. J. Immunol.* 30, 2857–2863.

Bernal, B., and Altman, N. (2010). The connectivity of the superior longitudinal

fasciculus: a tractography DTI study. *Magnetic Resonance Imaging* 28, 217–225.

Betel, D., Wilson, M., Gabow, A., Marks, D.S., and Sander, C. (2008). The microRNA.org resource: targets and expression. *Nucleic Acids Res* 36, D149–D153.

Betel, D., Koppal, A., Agius, P., Sander, C., and Leslie, C. (2010). Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol* 11, R90.

Beveridge, N.J., and Cairns, M.J. (2012). MicroRNA dysregulation in schizophrenia. *Neurobiology of Disease* 46, 263–271.

Beveridge, N.J., Tooney, P.A., Carroll, A.P., Gardiner, E., Bowden, N., Scott, R.J., Tran, N., Dedova, I., and Cairns, M.J. (2008). Dysregulation of miRNA 181b in the temporal cortex in schizophrenia. *Hum Mol Genet* 17, 1156–1168.

Bilgüvar, K., Öztürk, A.K., Louvi, A., Kwan, K.Y., Choi, M., Tatlı, B., Yalnızoğlu, D., Tüysüz, B., Çağlayan, A.O., Gökben, S., et al. (2010). Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature* 467, 207–210.

Bin Xu, Roos, J.L., Dexheimer, P., Boone, B., Plummer, B., Levy, S., Gogos, J.A., and Karayiorgou, M. (2011). Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nature Publishing Group* 43, 864–868.

Bisgaard, A.-M., Rackauskaite, G., Thelle, T., Kirchhoff, M., and Bryndorf, T. (2006). Twins with mental retardation and an interstitial deletion 7q34q36.2 leading to the diagnosis of long QT syndrome. *Am. J. Med. Genet.* 140A, 644–648.

Bishop, D.V.M. (2002). The role of genes in the etiology of specific language impairment. *J Commun Disord* 35, 311–328.

Bishop, D.V.M. (2013). Problems with tense marking in children with specific language impairment: not how but when. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369, 20120401–20120401.

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

Bishop, D.V., North, T., and Donlan, C. (1995). Genetic basis of specific language impairment: evidence from a twin study. *Dev Med Child Neurol* 37, 56–71.

Bishop, D.V., North, T., and Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: evidence from a twin study. *J Child Psychol Psychiatry* 37, 391–403.

- Bishop, D.V.M. (2006a). What Causes Specific Language Impairment in Children? *Curr Dir Psychol Sci* 15, 217–221.
- Bishop, D.V.M. (2006b). Developmental cognitive genetics: how psychology can inform genetics and vice versa. *Q J Exp Psychol (Hove)* 59, 1153–1168.
- Bishop, D.V.M. (2010). Which Neurodevelopmental Disorders Get Researched and Why? *PLoS ONE* 5, e15112.
- Bishop, D.V. (1997). *Uncommon Understanding* (Psychology Press).
- Bishop, D.V., and Norbury, C.F. (2008). Speech and language disorders. *Rutter's Child and Adolescent Psychiatry, Fifth Edition* 782–801.
- Bleich-Cohen, M., Kupchik, M., Gruberger, M., Kotler, M., and Hendler, T. (2012). Never resting region--mPFC in schizophrenia. *Schizophrenia Research* 140, 155–158.
- Bohnsack, M.T., Czaplinski, K., and Gorlich, D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *Rna* 10, 185–191.
- Bond, J., Roberts, E., Springell, K., Lizarraga, S.B., Lizarraga, S., Scott, S., Higgins, J., Hampshire, D.J., Morrison, E.E., Leal, G.F., et al. (2005). A centrosomal mechanism involving CDK5RAP2 and CENPJ controls brain size. *Nat Genet* 37, 353–355.
- Bonkowski, J.L., and Chien, C.-B. (2005). Molecular cloning and developmental expression of foxP2 in zebrafish. *Dev Dyn* 234, 740–746.
- Borchert, G.M., Lanier, W., and Davidson, B.L. (2006). RNA polymerase III transcribes human microRNAs. *Nat Struct Mol Biol* 13, 1097–1101.
- Bostan, A.C., Dum, R.P., and Strick, P.L. (2013). Cerebellar networks with the cerebral cortex and basal ganglia. *Trends Cogn Sci (Regul Ed)* 17, 241–254.
- Botting, N. (2005). Non-verbal cognitive development and language impairment. *Journal of Child Psychology and Psychiatry* 46, 317–326.
- Bowers, J.M., Perez-Pouchoulen, M., Edwards, N.S., and McCarthy, M.M. (2013). Foxp2 Mediates Sex Differences in Ultrasonic Vocalization by Rat Pups and Directs Order of Maternal Retrieval. *J Neurosci* 33, 3276–3283.
- Bralten, L.B.C., Gravendeel, A.M., Kloosterhof, N.K., Sacchetti, A., Vrijenhoek, T., Veltman, J.A., van den Bent, M.J., Kros, J.M., Hoogenraad, C.C., Sillevius Smitt, P.A.E., et al. (2010). The CASPR2 cell adhesion molecule functions as a tumor suppressor gene in glioma. *Oncogene* 29, 6138–6148.
- Braun, C.J., Zhang, X., Savelyeva, I., Wolff, S., Moll, U.M., Schepeler, T., Orntoft,

T.F., Andersen, C.L., and Dobbelstein, M. (2008). p53-Responsive MicroRNAs 192 and 215 Are Capable of Inducing Cell Cycle Arrest. *Cancer Research* 68, 10094–10104.

Brazel, C.Y., Romanko, M.J., Rothstein, R.P., and Levison, S.W. (2003). Roles of the mammalian subventricular zone in brain development. *Progress in Neurobiology* 69, 49–69.

Brevin, K., and Esquela-Kerscher, A. (2010). The complexities of microRNA regulation: mirandering around the rules. *Int J Biochem Cell Biol* 42, 1316–1329.

Broca, P.P. (1861). Remarques sur le siège de la faculté du langage articulé, suivies d’une observation d’aphémie (perte de la parole). *Bulletin De La Société Anthropologique* 330–357.

Brockschmidt, A., Todt, U., Ryu, S., Hoischen, A., Landwehr, C., Birnbaum, S., Frenck, W., Radlwimmer, B., Lichter, P., Engels, H., et al. (2007). Severe mental retardation with breathing abnormalities (Pitt-Hopkins syndrome) is caused by haploinsufficiency of the neuronal bHLH transcription factor TCF4. *Hum Mol Genet* 16, 1488–1494.

Brown, K.M., Chu, C.-Y., and Rana, T.M. (2005). Target accessibility dictates the potency of human RISC. *Nat Struct Mol Biol* 12, 469–470.

Brzózka, M.M., Radyushkin, K., Wichert, S.P., Ehrenreich, H., and Rossner, M.J. (2010). Cognitive and sensorimotor gating impairments in transgenic mice overexpressing the schizophrenia susceptibility gene Tcf4 in the brain. *Biol Psychiatry* 68, 33–40.

Buchner, D.A., Geisinger, J.M., Glazebrook, P.A., Morgan, M.G., Spiezio, S.H., Kaiyala, K.J., Schwartz, M.W., Sakurai, T., Furley, A.J., Kunze, D.L., et al. (2012). The juxtaparanodal proteins CNTNAP2 and TAG1 regulate diet-induced obesity. *Mamm Genome* 23, 431–442.

Buckner, R.L. (2013). Perspective. *Neuron* 80, 807–815.

Bulik-Sullivan, B., Selitsky, S., and Sethupathy, P. (2013). Prioritization of genetic variants in the microRNA regulome as functional candidates in genome-wide association studies. *Hum. Mutat.* 34, 1049–1056.

Burk, U., Schubert, J., Wellner, U., Schmalhofer, O., Vincan, E., Spaderna, S., and Brabletz, T. (2008). A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. *EMBO Reports* 9, 582–589.

Cai, X., Hagedorn, C.H., and Cullen, B.R. (2004). Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *Rna* 10, 1957–1966.

- Campbell, P., Reep, R.L., Stoll, M.L., Ophir, A.G., and Phelps, S.M. (2009). Conservation and diversity of Foxp2 expression in muroid rodents: Functional implications. *J. Comp. Neurol.* *512*, 84–100.
- Cantwell, D.P., and Baker, L. (1987). Prevalence and type of psychiatric disorder and developmental disorders in three speech and language groups. *J Commun Disord* *20*, 151–160.
- Chandra, S., Gallardo, G., Fernández-Chacón, R., Schlüter, O.M., and Südhof, T.C. (2005). α -Synuclein Cooperates with CSP α in Preventing Neurodegeneration. *Cell* *123*, 383–396.
- Chang, T.-C., Yu, D., Lee, Y.-S., Wentzel, E.A., Arking, D.E., West, K.M., Dang, C.V., Thomas-Tikhonenko, A., and Mendell, J.T. (2008). Widespread microRNA repression by Myc contributes to tumorigenesis. *Nat Genet* *40*, 43–50.
- Chen, D., Fu, L.Y., Zhang, Z., Li, G., Zhang, H., Jiang, L., Harrison, A.P., Shanahan, H.P., Klukas, C., Zhang, H.Y., et al. (2014). Dissecting the chromatin interactome of microRNA genes. *Nucleic Acids Res* *42*, 3028–3043.
- Chen, J.-M., Férec, C., and Cooper, D.N. (2006). A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes II: the importance of mRNA secondary structure in assessing the functionality of 3' UTR variants. *Hum Genet* *120*, 301–333.
- Chen, K., and Rajewsky, N. (2006). Natural selection on human microRNA binding sites inferred from SNP data. *Nat Genet* *38*, 1452–1456.
- Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., and Shiekhattar, R. (2005). TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* *436*, 740–744.
- Cheng, H.-Y., and Clayton, D.F. (2004). Activation and habituation of extracellular signal-regulated kinase phosphorylation in zebra finch auditory forebrain during song presentation. *J Neurosci* *24*, 7503–7513.
- Cheung, V.C.K., DeBoer, C., Hanson, E., Tunesi, M., D'Onofrio, M., Arisi, I., Brandi, R., Cattaneo, A., and Goosens, K.A. (2013). Gene Expression Changes in the Motor Cortex Mediating Motor Skill Learning. *PLoS ONE* *8*, e61496.
- Chiang, H.R., Schoenfeld, L.W., Ruby, J.G., Auyeung, V.C., Spies, N., Baek, D., Johnston, W.K., Russ, C., Luo, S., Babiarz, J.E., et al. (2010). Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes & Development* *24*, 992–1009.
- Chiocchetti, A.G., Kopp, M., Waltes, R., Haslinger, D., Duketis, E., Jarczok, T.A., Poustka, F., Voran, A., Graab, U., Meyer, J., et al. (2014). Variants of the CNTNAP2 5' promoter as risk factors for autism spectrum disorders: a genetic

and functional approach. *Mol Psychiatry*.

Chiu, Y.-C., Li, M.-Y., Liu, Y.-H., Ding, J.-Y., Yu, J.-Y., and Wang, T.-W. (2014). *Foxp2* regulates neuronal differentiation and neuronal subtype specification. *Dev Neurobiol* *74*, 723–738.

Choudhury, N., and Benasich, A.A. (2003). A Family Aggregation Study: The Influence of Family History and Other Risk Factors on Language Development. *J Speech Lang Hear Res* *46*, 261–272.

Clark, K.L., Halay, E.D., LAI, E.S., and BURLEY, S.K. (1993). Co-Crystal Structure of the Hnf-3/Fork Head Dna-Recognition Motif Resembles Histone-H5. *Nature* *364*, 412–420.

Clovis, Y.M., Enard, W., Marinaro, F., Huttner, W.B., and 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.

Collins, A.L., Kim, Y., Bloom, R.J., Kelada, S.N., Sethupathy, P., and Sullivan, P.F. (2014). Transcriptional targets of the schizophrenia risk gene. *Translational Psychiatry* *4*, e404–e408.

Conaco, C., Otto, S., Han, J.-J., and Mandel, G. (2006). Reciprocal actions of REST and a microRNA promote neuronal identity. *Proc Natl Acad Sci USA* *103*, 2422–2427.

Condro, M.C., and White, S.A. (2013). Distribution of language-related *Cntnap2* protein in neural circuits critical for vocal learning. *J. Comp. Neurol.* *522*, 169–185.

Consortium, 1.G., Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* *491*, 56–65.

Consortium, C.-D.G.O.T.P.G. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *The Lancet* *381*, 1371–1379.

Conti-Ramsden, G., Botting, N., and Faragher, B. (2001a). Psycholinguistic markers for specific language impairment (SLI). *J Child Psychol Psychiatry* *42*, 741–748.

Conti-Ramsden, G., Falcaro, M., Simkin, Z., and Pickles, A. (2007). Familial loading in specific language impairment: patterns of differences across proband characteristics, gender and relative type. *Genes, Brain and Behavior* *6*, 216–228.

Conti-Ramsden, G., Botting, N., and Knox, E. (2001b). Follow-up of children

attending infant language units: outcomes at 11 years of age. *J Commun Disord* 36, 207–219.

Conti-Ramsden, G., St Clair, M.C., Pickles, A., and Durkin, K. (2012). Developmental Trajectories of Verbal and Nonverbal Skills in Individuals With a History of Specific Language Impairment: From Childhood to Adolescence. *J Speech Lang Hear Res* 55, 1716.

Coop, G., Bullaughey, K., Luca, F., and Przeworski, M. (2008). The Timing of Selection at the Human FOXP2 Gene. *Molecular Biology and Evolution* 25, 1257–1259.

Corcoran, D.L., Pandit, K.V., Gordon, B., Bhattacharjee, A., Kaminski, N., and Benos, P.V. (2009). Features of mammalian microRNA promoters emerge from polymerase II chromatin immunoprecipitation data. *PLoS ONE* 4, e5279.

Crosson, B. (2013). Thalamic mechanisms in language: A reconsideration based on recent findings and concepts. *Brain Lang* 126, 73–88.

Cuiffo, B.G., Campagne, A., Bell, G.W., Lembo, A., Orso, F., Lien, E.C., Bhasin, M.K., Raimo, M., Hanson, S.E., Marusyk, A., et al. (2014). MSC-Regulated MicroRNAs Converge on the Transcription Factor FOXP2 and Promote Breast Cancer Metastasis. *Cell Stem Cell* 15, 762–774.

Cummings, E., Donohoe, G., Hargreaves, A., Moore, S., Fahey, C., Dinan, T.G., McDonald, C., O'Callaghan, E., O'Neill, F.A., Waddington, J.L., et al. (2013). Mood congruent psychotic symptoms and specific cognitive deficits in carriers of the novel schizophrenia risk variant at MIR-137. *Neurosci Lett* 532, 33–38.

Darvish, H., Esmaeeli-Nieh, S., Monajemi, G.B., Mohseni, M., Ghasemi-Firouzabadi, S., Abedini, S.S., Bahman, I., Jamali, P., Azimi, S., Mojahedi, F., et al. (2010). A clinical and molecular genetic study of 112 Iranian families with primary microcephaly. *Journal of Medical Genetics* 47, 823–828.

Davis, E., Caiment, F., Tordoir, X., Cavaillé, J., Ferguson-Smith, A., Cockett, N., Georges, M., and Charlier, C. (2005). RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus. *Curr Biol* 15, 743–749.

Davis, T.H., Cuellar, T.L., Koch, S.M., Barker, A.J., Harfe, B.D., McManus, M.T., and Ullian, E.M. (2008). Conditional loss of Dicer disrupts cellular and tissue morphogenesis in the cortex and hippocampus. *J Neurosci* 28, 4322–4330.

Davydov, E.V., Goode, D.L., Sirota, M., Cooper, G.M., Sidow, A., and Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP. *PLoS Comput Biol* 6, e1001025.

de Ligt, J., Willemsen, M.H., van Bon, B.W.M., Kleefstra, T., Yntema, H.G., Kroes, T., Silfhout, A.T.V.-V., Koolen, D.A., de Vries, P., Gilissen, C., et al.

(2012). Diagnostic exome sequencing in persons with severe intellectual disability. *New England Journal of Medicine* 367, 1921–1929.

de Pontual, L., Mathieu, Y., Golzio, C., Rio, M., Malan, V., Boddaert, N., Soufflet, C., Picard, C., Durandy, A., Dobbie, A., et al. (2009). Mutational, functional, and expression studies of the TCF4 gene in Pitt-Hopkins syndrome. *Hum. Mutat.* 30, 669–676.

Dean, C., and Dresbach, T. (2006). Neuroligins and neuroligins: linking cell adhesion, synapse formation and cognitive function. *Trends Neurosci* 29, 21–29.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science* 295, 1306–1311.

Denisenko-Nehrbass, N., Oguievetskaia, K., Goutebroze, L., Galvez, T., Yamakawa, H., Ohara, O., Carnaud, M., and Girault, J.-A. (2003). Protein 4.1B associates with both Caspr/paranodin and Caspr2 at paranodes and juxtaparanodes of myelinated fibres. *European Journal of Neuroscience* 17, 411–416.

Denli, A.M., Tops, B.B.J., Plasterk, R.H.A., Ketting, R.F., and Hannon, G.J. (2004). Processing of primary microRNAs by the Microprocessor complex. *Nature* 432, 231–235.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Publishing Group* 43, 491–498.

Deriziotis, P., and Fisher, S.E. (2013). Neurogenomics of speech and language disorders: the road ahead. *Genome Biol* 14, 204.

Deriziotis, P., Graham, S.A., Estruch, S.B., and Fisher, S.E. (2014a). Investigating Protein-protein Interactions in Live Cells Using Bioluminescence Resonance Energy Transfer. *JoVE*.

Deriziotis, P., Roak, B.J.O.R., Graham, S.A., Estruch, S.B., Dimitropoulou, D., Bernier, R.A., Gerds, J., Shendure, J., Eichler, E.E., and Fisher, S.E. (2014b). De novo TBR1 mutations in sporadic autism disrupt protein functions. *5*, 1–12.

Devanna, P., and Vernes, S.C. (2014). A direct molecular link between the autism candidate gene RORa and the schizophrenia candidate MIR137. *Sci. Rep.* 4.

Devanna, P., Middelbeek, J., and Vernes, S.C. (2014). FOXP2 drives neuronal differentiation by interacting with retinoic acid signaling pathways. *Front. Cell. Neurosci.* 8, 305–305.

DeWitt, I., and Rauschecker, J.P. (2013). Wernicke's area revisited: Parallel streams and word processing. *Brain Lang* 127, 181–191.

Démonet, J.-F., Thierry, G., and Cardebat, D. (2004). Renewal of the neurophysiology of language: functional neuroimaging. *Physiol Rev* 85, 49–95.

Ding, X., Park, S.I., McCauley, L.K., and Wang, C.-Y. (2013). Signaling between transforming growth factor β (TGF- β) and transcription factor SNAI2 represses expression of microRNA miR-203 to promote epithelial-mesenchymal transition and tumor metastasis. *J Biol Chem* 288, 10241–10253.

Djuranovic, S., Nahvi, A., and Green, R. (2012). miRNA-Mediated Gene Silencing by Translational Repression Followed by mRNA Deadenylation and Decay. *Science* 336, 237–240.

Dorval, V., Mandemakers, W., Jolivet, F., Coudert, L., Mazroui, R., De Strooper, B., and Hébert, S.S. (2014). Gene and MicroRNA Transcriptome Analysis of Parkinson's Related LRRK2 Mouse Models. *PLoS ONE* 9, e85510.

Dosemeci, A., Thein, S., Yang, Y., Reese, T.S., and Tao-Cheng, J.-H. (2013). Biochemical and Biophysical Research Communications. *Biochem Biophys Res Commun* 430, 245–249.

Doupe, A.J., and Kuhl, P.K. (1999). Birdsong and human speech: common themes and mechanisms. *Annu Rev Neurosci* 22, 567–631.

D'Amours, G., Bureau, G., Boily, M.-J., and Cyr, M. (2010). Differential gene expression profiling in the mouse brain during motor skill learning: Focus on the striatum structure. *Behav Brain Res* 221, 108–117.

Ebert, M.S., and Sharp, P.A. (2012). Roles for microRNAs in conferring robustness to biological processes. *Cell* 149, 515–524.

Edbauer, D., Neilson, J.R., Foster, K.A., Wang, C.-F., Seeburg, D.P., Batterton, M.N., Tada, T., Dolan, B.M., Sharp, P.A., and Sheng, M. (2010). Regulation of Synaptic Structure and Function by FMRP-Associated MicroRNAs miR-125b and miR-132. *Neuron* 65, 373–384.

Elia, J., Gai, X., Xie, H.M., Perin, J.C., Geiger, E., Glessner, J.T., arcy, M.D.A., deBerardinis, R., Frackelton, E., Kim, C., et al. (2009). Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. *Mol Psychiatry* 15, 637–646.

Ellis Weismer, S., Tomblin, J.B., Zhang, X., Buckwalter, P., Chynoweth, J.G., and Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *J Speech Lang Hear Res* 43, 865–878.

Enard, W. (2011). FOXP2 and the role of cortico-basal ganglia circuits in speech and language evolution. *Curr Opin Neurobiol* 21, 415–424.

Enard, W., Gehre, S., Hammerschmidt, K., Hölter, S.M., Blass, T., Somel, M., Brückner, M.K., Schreiweis, C., Winter, C., Sohr, R., et al. (2009). A humanized

version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell* 137, 961–971.

Enard, W., Przeworski, M., Fisher, S.E., Lai, C.S.L., Wiebe, V., Kitano, T., Monaco, A.P., and Pääbo, S. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* 418, 869–872.

Encinas, M., Iglesias, M., Liu, Y., Wang, H., Muhaisen, A., Ceña, V., Gallego, C., and Comella, J.X. (2000). Sequential treatment of SH-SY5Y cells with retinoic acid and brain-derived neurotrophic factor gives rise to fully differentiated, neurotrophic factor-dependent, human neuron-like cells. *J Neurochem* 75, 991–1003.

ENCODE Project Consortium (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* 9, e1001046–e1001046.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74.

Falivelli, G., De Jaco, A., Favaloro, F.L., Kim, H., Wilson, J., Dubi, N., Ellisman, M.H., Abrahams, B.S., Taylor, P., and Comoletti, D. (2012). Inherited genetic variants in autism-related CNTNAP2 show perturbed trafficking and ATF6 activation. *Hum Mol Genet* 21, 4761–4773.

Felsenfeld, S., and Plomin, R. (1997). Epidemiological and Offspring Analyses of Developmental Speech Disorders Using Data From the Colorado Adoption Project. *J Speech Lang Hear Res* 40, 778–791.

Feng, Z., Zhang, C., Wu, R., and Hu, W. (2011). Tumor suppressor p53 meets microRNAs. *Journal of Molecular Cell Biology* 3, 44–50.

Ferland, R.J., Cherry, T.J., Preware, P.O., Morrissey, E.E., and Walsh, C.A. (2003). Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain. *J. Comp. Neurol.* 460, 266–279.

Feuk, L., Kalervo, A., Lipsanen-Nyman, M., Skaug, J., Nakabayashi, K., Finucane, B., Hartung, D., Innes, M., Kerem, B., Nowaczyk, M.J., et al. (2006). Absence of a paternally inherited FOXP2 gene in developmental verbal dyspraxia. *The American Journal of Human Genetics* 79, 965–972.

Fischer, J., and Hammerschmidt, K. (2010). Ultrasonic vocalizations in mouse models for speech and socio-cognitive disorders: insights into the evolution of vocal communication. *Genes, Brain and Behavior* 10, 17–27.

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

Fisher, S.E. (2007). Molecular windows into speech and language disorders. *Folia Phoniatica Et Logopaedica* : Official Organ of the International Association

of Logopedics and Phoniatrics (IALP) *59*, 130–140.

Fisher, S.E., and Marcus, G.F. (2006). The eloquent ape: genes, brains and the evolution of language. *Nat Rev Genet* *7*, 9–20.

Fisher, S.E., and Scharff, C. (2009). FOXP2 as a molecular window into speech and language. *Trends Genet* *25*, 166–177.

Fisher, S.E., Lai, C.S.L., and Monaco, A.P. (2003). Deciphering the genetic basis of speech and language disorders. *Annu Rev Neurosci* *26*, 57–80.

Flanagan, S.E.S., Patch, A.-M.A., and Ellard, S.S. (2010). Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genet Test Mol Biomarkers* *14*, 533–537.

Flora, A., Garcia, J.J., Thaller, C., and Zoghbi, H.Y. (2007). The E-protein Tcf4 interacts with Math1 to regulate differentiation of a specific subset of neuronal progenitors. *Proc Natl Acad Sci USA* *104*, 15382–15387.

Forrest, M.P., Hill, M.J., Quantock, A.J., Martin-Rendon, E., and Blake, D.J. (2014). The emerging roles of TCF4 in disease and development. *Trends in Molecular Medicine* *20*, 322–331.

Forrest, M.P., Waite, A.J., Martin-Rendon, E., and Blake, D.J. (2013). Knockdown of Human TCF4 Affects Multiple Signaling Pathways Involved in Cell Survival, Epithelial to Mesenchymal Transition and Neuronal Differentiation. *PLoS ONE* *8*, e73169.

Forrest, M., Chapman, R.M., Doyle, A.M., Tinsley, C.L., Waite, A., and Blake, D.J. (2012). Functional analysis of TCF4 missense mutations that cause Pitt-Hopkins syndrome. *Hum. Mutat.* *33*, 1676–1686.

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

French, C.A., Groszer, M., Preece, C., Coupe, A.-M., Rajewsky, K., and Fisher, S.E. (2007). Generation of mice with a conditional Foxp2 null allele. *Genesis* *45*, 440–446.

Fromer, M., Pocklington, A.J., Kavanagh, D.H., Williams, H.J., Dwyer, S., Gormley, P., Georgieva, L., Rees, E., Palta, P., Ruderfer, D.M., et al. (2014). De novo mutations in schizophrenia implicate synaptic networks. *Nature* *506*, 179–184.

Fujita, E., Tanabe, Y., Shiota, A., Ueda, M., Suwa, K., Momoi, M.Y., and 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 USA* *105*, 3117–3122.

- Gathercole, S.E., Willis, C.S., Baddeley, A.D., and Emslie, H. (1994). The Children's Test of Nonword Repetition: a test of phonological working memory. *Memory* 2, 103–127.
- Gathercole, S.E., and Baddeley, A.D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *J Mem Lang* 29, 336–360.
- Gaub, S., Groszer, M., Fisher, S.E., and Ehret, G. (2010). The structure of innate vocalizations in Foxp2-deficient mouse pups. *Genes, Brain and Behavior* 9, 390–401.
- Gennarino, V.A., Sardiello, M., Avellino, R., Meola, N., Maselli, V., Anand, S., Cutillo, L., Ballabio, A., and Banfi, S. (2009). MicroRNA target prediction by expression analysis of host genes. *Genome Research* 19, 481–490.
- Georges, S.A., Biery, M.C., Kim, S.Y., Schelter, J.M., Guo, J., Chang, A.N., Jackson, A.L., Carleton, M.O., Linsley, P.S., Cleary, M.A., et al. (2008). Coordinated Regulation of Cell Cycle Transcripts by p53-Inducible microRNAs, miR-192 and miR-215. *Cancer Research* 68, 10105–10112.
- Geschwind, N. (1970). The organization of language and the brain. *Science* 170, 940–944.
- Ghosh, T., Aprea, J., Nardelli, J., Engel, H., Selinger, C., Mombereau, C., Lemonnier, T., Moutkine, I., Schwendimann, L., Dori, M., et al. (2014). MicroRNAs Establish Robustness and Adaptability of a Critical Gene Network to Regulate Progenitor Fate Decisions during Cortical Neurogenesis. *CellReports* 7, 1779–1788.
- Girard, S.L., Gauthier, J., Noreau, A., Xiong, L., Zhou, S., Jouan, L., Dionne-Laporte, A., Spiegelman, D., Henrion, E., Diallo, O., et al. (2011). Increased exonic de novo mutation rate in individuals with schizophrenia. *Nature Publishing Group* 43, 860–863.
- Glasser, M.F., and Rilling, J.K. (2008). DTI Tractography of the Human Brain's Language Pathways. *Cerebral Cortex* 18, 2471–2482.
- Gong, J., Tong, Y., Zhang, H.-M., Wang, K., Hu, T., Shan, G., Sun, J., and Guo, A.-Y. (2011). Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. *Hum. Mutat.* 33, 254–263.
- Goode, D.L., Cooper, G.M., Schmutz, J., Dickson, M., Gonzales, E., Tsai, M., Karra, K., Davydov, E., Batzoglou, S., Myers, R.M., et al. (2010). Evolutionary constraint facilitates interpretation of genetic variation in resequenced human genomes. *Genome Research* 20, 301–310.

Gopnik, M. (1990). FEATURE-BLIND GRAMMAR AND DYSPHASIA. *Nature* 344, 715–715.

Gopnik, M. (1997). Language deficits and genetic factors. *Trends Cogn Sci (Regul Ed)* 1, 5–9.

Gopnik, M., and Crago, M.B. (1991). Familial aggregation of a developmental language disorder. *Cognition* 39, 1–50.

Gopnik, M., and Goad, H. (1997). What underlies inflectional error patterns in genetic dysphasia? *J Neurolinguistics* 10, 109–137.

Göring, H.H., Terwilliger, J.D., and Blangero, J. (2001). Large upward bias in estimation of locus-specific effects from genomewide scans. *The American Journal of Human Genetics* 69, 1357–1369.

Graunke, D.M., Fornace, A.J., and Pieper, R.O. (1999). Presetting of chromatin structure and transcription factor binding poise the human GADD45 gene for rapid transcriptional up-regulation. *Nucleic Acids Res* 27, 3881–3890.

Green, M.J., Cairns, M.J., Wu, J., Dragovic, M., Jablensky, A., Tooney, P.A., Scott, R.J., Carr, V.J., and Carr, V.J. (2013). Genome-wide supported variant MIR137 and severe negative symptoms predict membership of an impaired cognitive subtype of schizophrenia. *Mol Psychiatry* 18, 774–780.

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.Y., et al. (2010). A Draft Sequence of the Neandertal Genome. *Science* 328, 710–722.

Gregor, A., Albrecht, B., Bader, I., Bijlsma, E.K., Ekici, A.B., Engels, H., Hackmann, K., Horn, D., Hoyer, J., Klapceki, J., et al. (2011). Expanding the clinical spectrum associated with defects in CNTNAP2 and NRXN1. *BMC Med Genet* 12, 106.

Gregory, R.I., Chendrimada, T.P., Cooch, N., and Shiekhattar, R. (2005). Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell* 123, 631–640.

Gregory, R.I., Yan, K.-P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. *Nature* 432, 235–240.

Griffiths-Jones, S., Saini, H.K., van Dongen, S., and Enright, A.J. (2007). miRBase: tools for microRNA genomics. *Nucleic Acids Res* 36, D154–D158.

Grigorenko, E.L. (2009). Speaking genes or genes for speaking? Deciphering the genetics of speech and language. *Journal of Child Psychology and Psychiatry* 50, 116–125.

Grimson, A., Farh, K.K.-H., Johnston, W.K., Garrett-Engele, P., Lim, L.P., and Bartel, D.P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* 27, 91–105.

Groszer, M., Keays, D.A., Deacon, R.M.J., de Bono, J.P., Prasad-Mulcare, S., Gaub, S., Baum, M.G., French, C.A., Nicod, J., Coventry, J.A., et al. (2008). Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits. *Curr Biol* 18, 354–362.

Grüter, C., and Farina, W.M. (2009). The honeybee waggle dance: can we follow the steps? *Trends in Ecology & Evolution* 24, 242–247.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells. *Cell* 130, 77–88.

Gul, A., Hassan, M.J., Hussain, S., Raza, S.I., Chishti, M.S., and Ahmad, W. (2006). A novel deletion mutation in CENPJ gene in a Pakistani family with autosomal recessive primary microcephaly. *J Hum Genet* 51, 760–764.

Gulsuner, S., Walsh, T., Watts, A.C., Lee, M.K., Thornton, A.M., Casadei, S., Rippey, C., Shahin, H., Shahin, H., Shahin, H., et al. (2013). Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* 154, 518–529.

Gunaratne, P.H., Lin, Y.-C., Benham, A.L., Drnevich, J., Coarfa, C., Tennakoon, J.B., Creighton, C.J., Kim, J.H., Milosavljevic, A., Watson, M., et al. (2010). Song exposure regulates known and novel microRNAs in the zebra finch auditory forebrain. *BMC Genomics* 12, 277–277.

Guo, L., and Lu, Z. (2010). The fate of miRNA* strand through evolutionary analysis: implication for degradation as merely carrier strand or potential regulatory molecule? *PLoS ONE* 5, e11387.

Gurtan, A.M., and Sharp, P.A. (2013). The role of miRNAs in regulating gene expression networks. *Journal of Molecular Biology* 425, 3582–3600.

Haase, A.D., Jaskiewicz, L., Zhang, H., Lainé, S., Sack, R., Gatignol, A., and Filipowicz, W. (2005). TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Reports* 6, 961–967.

Haesler, S., Rochefort, C., Georgi, B., Licznarski, P., Osten, P., and Scharff, C. (2007). Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. *PLoS Biol* 5, e321.

Haesler, S., Wada, K., Nshdejan, A., Morrissey, E.E., Lints, T., Jarvis, E.D., and Scharff, C. (2004). FoxP2 expression in avian vocal learners and non-learners. *J*

Neurosci 24, 3164–3175.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.-C., Munschauer, M., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141, 129–141.

Hagoort, P. (2014). ScienceDirectNodes and networks in the neural architecture for language: Broca's region and beyond. *Curr Opin Neurobiol* 28, 136–141.

Hamdan, F.F., Daoud, H., Rochefort, D., Piton, A., Gauthier, J., Langlois, M., Foomani, G., Dobrzeniecka, S., Krebs, M.-O., Joober, R., et al. (2010). REPORTDe Novo Mutations in FOXP1 in Cases with Intellectual Disability, Autism, and Language Impairment. *The American Journal of Human Genetics* 87, 671–678.

Hammill, D.D., and Newcomer, P.L. (1988). Test of Language Development 2 - Intermediate Version (Pro-Ed).

Hammond, S.M., Bernstein, E., Beach, D., and Hannon, G.J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* 404, 293–296.

Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R., and Hannon, G.J. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* 293, 1146–1150.

Han, J., Lee, Y., Yeom, K.-H., Nam, J.-W., Heo, I., Rhee, J.-K., Sohn, S.Y., Cho, Y., Zhang, B.-T., and Kim, V.N. (2006). Molecular Basis for the Recognition of Primary microRNAs by the Drosha-DGCR8 Complex. *Cell* 125, 887–901.

Han, T.-U., Park, J., Domingues, C.F., Moretti-Ferreira, D., Paris, E., Sainz, E., Gutierrez, J., and Drayna, D. (2014). A study of the role of the FOXP2 and CNTNAP2 genes in persistent developmental stuttering. *Neurobiology of Disease* 69, 23–31.

Hawrylycz, M.J., Lein, E.S., Guillozet-Bongaarts, A.L., Shen, E.H., Ng, L., Miller, J.A., Lagemaat, L.N.V. de, Smith, K.A., Ebbert, A., Riley, Z.L., et al. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 489, 391–399.

Hayashita, Y., Osada, H., Tatematsu, Y., Yamada, H., Yanagisawa, K., Tomida, S., Yatabe, Y., Kawahara, K., Sekido, Y., and Takahashi, T. (2005). A polycistronic microRNA cluster, miR-17-92, is overexpressed in human lung cancers and enhances cell proliferation. *Cancer Research* 65, 9628–9632.

Håndstad, T., Rye, M., Močnik, R., Drabløs, F., and Sætrom, P. (2012). Cell-type specificity of ChIP-predicted transcription factor binding sites. *BMC Genomics* 13, 372.

- Heintz, N. (2001). BAC to the future: the use of bac transgenic mice for neuroscience research. *Nat Rev Neurosci* 2, 861–870.
- Hengst, U., Cox, L.J., Macosko, E.Z., and Jaffrey, S.R. (2006). Functional and selective RNA interference in developing axons and growth cones. *J Neurosci* 26, 5727–5732.
- Herranz, H., and Cohen, S.M. (2010). MicroRNAs and gene regulatory networks: managing the impact of noise in biological systems. *Genes & Development* 24, 1339–1344.
- Hickok, G., Bellugi, U., and Klima, E.S. (1998). The neural organization of language: evidence from sign language aphasia. *Trends Cogn Sci (Regul Ed)* 2, 129–136.
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat Rev Neurosci* 8, 393–402.
- Hickok, G., Kritchevsky, M., Bellugi, U., and Klima, E.S. (1996). The role of the left frontal operculum in sign language aphasia. *Neurocase* 2, 373–380.
- Hill, M.J., Donocik, J.G., Nuamah, R.A., Mein, C.A., Sainz-Fuertes, R., and Bray, N.J. (2014). Schizophrenia Research. *Schizophrenia Research* 153, 225–230.
- Hino, K., Tsuchiya, K., Fukao, T., Kiga, K., Okamoto, R., Kanai, T., and Watanabe, M. (2008). Inducible expression of microRNA-194 is regulated by HNF-1 during intestinal epithelial cell differentiation. *Rna* 14, 1433–1442.
- Hoerdli, F.J., Toigo, M., Schild, A., Götz, J., and Day, P.J. (2004). Reference genes identified in SH-SY5Y cells using custom-made gene arrays with validation by quantitative polymerase chain reaction. *Anal Biochem* 335, 30–41.
- Hofacker, I.L., and Stadler, P.F. (2006). Memory efficient folding algorithms for circular RNA secondary structures. *Bioinformatics* 22, 1172–1176.
- Hoischen, A., van Bon, B.W.M., Gilissen, C., Arts, P., van Lier, B., Steehouwer, M., de Vries, P., de Reuver, R., Wieskamp, N., Mortier, G., et al. (2010). De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat Genet* 42, 483–485.
- Hon, L.S., and Zhang, Z. (2007). The roles of binding site arrangement and combinatorial targeting in microRNA repression of gene expression. *Genome Biol* 8, R166.
- Hornstein, E., and Shomron, N. (2006). Canalization of development by microRNAs. *Nat Genet* 38, S20–S24.
- Horresh, I., Bar, V., Kissil, J.L., and Peles, E. (2010). Organization of Myelinated Axons by Caspr and Caspr2 Requires the Cytoskeletal Adapter Protein 4.1B.

Journal of Neuroscience 30, 2480–2489.

Horresh, I., Poliak, S., Grant, S., Bredt, D., Rasband, M.N., and Peles, E. (2008). Multiple Molecular Interactions Determine the Clustering of Caspr2 and Kv1 Channels in Myelinated Axons. *Journal of Neuroscience* 28, 14213–14222.

Hu, H.Y., Guo, S., Xi, J., Yan, Z., Fu, N., Zhang, X., Menzel, C., Liang, H., Yang, H., Zhao, M., et al. (2011). MicroRNA expression and regulation in human, chimpanzee, and macaque brains. *PLoS Genet* 7, e1002327.

Hughes, A.E., Bradley, D.T., Campbell, M., Lechner, J., Dash, D.P., Simpson, D.A., and Willoughby, C.E. (2011). Mutation Altering the miR-184 Seed Region Causes Familial Keratoconus with Cataract. *The American Journal of Human Genetics* 89, 628–633.

Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Bálint, E., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* 293, 834–838.

Iliff, B.W., Riazuddin, S.A., and Gottsch, J.D. (2012). A single-base substitution in the seed region of miR-184 causes EDICT syndrome. *Investigative Ophthalmology & Visual Science* 53, 348–353.

Iliopoulos, D., Jaeger, S.A., Hirsch, H.A., Bulyk, M.L., and Struhl, K. (2010). STAT3 Activation of miR-21 and miR-181b-1 via PTEN and CYLD Are Part of the Epigenetic Switch Linking Inflammation to Cancer. *Mol Cell* 39, 493–506.

Inda, M.C., DeFelipe, J., and Muñoz, A. (2006). Voltage-gated ion channels in the axon initial segment of human cortical pyramidal cells and their relationship with chandelier cells. *Proc Natl Acad Sci USA* 103, 2920–2925.

Iossifov, I., O’Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E., et al. (2014). The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 515, 216–221.

Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y.-H., Narzisi, G., Leotta, A., et al. (2012). De Novo Gene Disruptions in Children on the Autistic Spectrum. *Neuron* 74, 285–299.

Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., and Speed, T.P. (2003). Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 31, e15.

Iwai, N., and Naraba, H. (2005). Polymorphisms in human pre-miRNAs. *Biochem Biophys Res Commun* 331, 1439–1444.

Iwama, H. (2013). Coordinated networks of microRNAs and transcription factors with evolutionary perspectives. *Adv. Exp. Med. Biol.* 774, 169–187.

- Jackman, C., Horn, N.D., Molleston, J.P., and Sokol, D.K. (2009). Gene Associated with Seizures, Autism, and Hepatomegaly in an Amish Girl. *Pediatric Neurology* 40, 310–313.
- Jackson, B.C., Carpenter, C., Nebert, D.W., and Vasiliou, V. (2010). Update of human and mouse forkhead box (FOX) gene families. *Hum Genomics* 4, 345–352.
- Jeon, S.H., Lee, K., Lee, K.S., Kunkeaw, N., Johnson, B.H., Holthauzen, L.M.F., Gong, B., Leelayuwat, C., and Lee, Y.S. (2012). Characterization of the direct physical interaction of nc886, a cellular non-coding RNA, and PKR. *FEBS Letters* 586, 3477–3484.
- Ji, W., Li, T., Pan, Y., Tao, H., Ju, K., Wen, Z., Fu, Y., An, Z., Zhao, Q., Wang, T., et al. (2013). Psychiatry Research. *Psychiatry Research* 207, 225–228.
- Jiang, K., Ren, C., and Nair, V.D. (2013). MicroRNA-137 represses Klf4 and Tbx3 during differentiation of mouse embryonic stem cells. *Stem Cell Research* 11, 1299–1313.
- Jin, P., Zarnescu, D.C., Ceman, S., Nakamoto, M., Mowrey, J., Jongens, T.A., Nelson, D.L., Moses, K., and Warren, S.T. (2004). Biochemical and genetic interaction between the fragile X mental retardation protein and the microRNA pathway. *Nature Publishing Group* 7, 113–117.
- Jing, H., Vakoc, C.R., Ying, L., Mandat, S., Wang, H., Zheng, X., and Blobel, G.A. (2008). Exchange of GATA Factors Mediates Transitions in Looped Chromatin Organization at a Developmentally Regulated Gene Locus. *Mol Cell* 29, 232–242.
- Kaddis, J.S., Wai, D.H., Bowers, J., Hartmann, N., Baeriswyl, L., Bajaj, S., Anderson, M.J., Getts, R.C., and Triche, T.J. (2011). Influence of RNA labeling on expression profiling of microRNAs. *J Mol Diagn* 14, 12–21.
- Kaestner, K.H., Knochel, W., and Martinez, D.E. (2000). Unified nomenclature for the winged helix/forkhead transcription factors. *Genes & Development* 14, 142–146.
- Karolchik, D., Barber, G.P., Casper, J., Clawson, H., Cline, M.S., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., et al. (2014). The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res* 42, D764–D770.
- Karres, J.S., Hilgers, V., Carrera, I., Treisman, J., and Cohen, S.M. (2007). The Conserved microRNA MiR-8 Tunes Atrophin Levels to Prevent Neurodegeneration in *Drosophila*. *Cell* 131, 136–145.
- Kaufmann, E., and Knöchel, W. (1996). Five years on the wings of fork head. *Mech Dev* 57, 3–20.

- Kaya, K.D., Karakülah, G., Yakicier, C.M., Acar, A.C., and Konu, O. (2010). mESAdb: microRNA expression and sequence analysis database. *Nucleic Acids Res* *39*, D170–D180.
- Kel, A.E. (2003). MATCHTM: a tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res* *31*, 3576–3579.
- Keren-Happuch, E., Chen, S.-H.A., Ho, M.-H.R., and Desmond, J.E. (2014). A meta-analysis of cerebellar contributions to higher cognition from PET and fMRI studies. *Hum. Brain Mapp.* *35*, 593–615.
- Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes & Development* *15*, 2654–2659.
- Khatri, P., Sirota, M., and Butte, A.J. (2012). Ten Years of Pathway Analysis: Current Approaches and Outstanding Challenges. *PLoS Comput Biol* *8*, e1002375.
- Khvorova, A., Reynolds, A., and Jayasena, S.D. (2002). Functional siRNAs and miRNAs Exhibit Strand Bias. *Cell* *115*, 209–216.
- Kim, H.K., Prokunina-Olsson, L., and Chanock, S.J. (2012). Common genetic variants in miR-1206 (8q24.2) and miR-612 (11q13.3) affect biogenesis of mature miRNA forms. *PLoS ONE* *7*, e47454.
- Kim, J., Inoue, K., Ishii, J., Vanti, W.B., Voronov, S.V., Murchison, E., Hannon, G., and Abeliovich, A. (2007). A MicroRNA Feedback Circuit in Midbrain Dopamine Neurons. *Science* *317*, 1220–1224.
- Kim, J., and Bartel, D.P. (2009). Allelic imbalance sequencing reveals that single-nucleotide polymorphisms frequently alter microRNA-directed repression. *Nat Biotechnol* *27*, 472–477.
- Kim, J., Krichevsky, A., Grad, Y., Hayes, G.D., Kosik, K.S., Church, G.M., and Ruvkun, G. (2004). Identification of many microRNAs that copurify with polyribosomes in mammalian neurons. *Proc Natl Acad Sci USA* *101*, 360–365.
- Kim, Y.-K., and Kim, V.N. (2007). Processing of intronic microRNAs. *Embo J* *26*, 775–783.
- Kiriakidou, M., Tan, G.S., Lamprinaki, S., De Planell-Saguer, M., Nelson, P.T., and Mourelatos, Z. (2006). An mRNA m7G Cap Binding-like Motif within Human Ago2 Represses Translation. *Cell* *129*, 1141–1151.
- Klein, M.E., Lioy, D.T., Ma, L., Impey, S., Mandel, G., and Goodman, R.H. (2007). Homeostatic regulation of MeCP2 expression by a CREB-induced microRNA. *Nat Neurosci* *10*, 1513–1514.

- Kluiver, J., Gibcus, J.H., Hettinga, C., Adema, A., Richter, M.K.S., Halsema, N., Slezak-Prochazka, I., Ding, Y., Kroesen, B.-J., and Van Den Berg, A. (2012a). Rapid generation of microRNA sponges for microRNA inhibition. *PLoS ONE* 7, e29275.
- Kluiver, J., Slezak-Prochazka, I., Smigielska-Czepiel, K., Halsema, N., Kroesen, B.-J., and Van Den Berg, A. (2012b). Generation of miRNA sponge constructs. *Methods* 58, 113–117.
- Knecht, S., Dräger, B., Deppe, M., Bobe, L., Lohmann, H., Flöel, A., Ringelstein, E.B., and Henningsen, H. (2000). Handedness and hemispheric language dominance in healthy humans. *Brain* 123 Pt 12, 2512–2518.
- Kolodkin, A.L., and Tessier-Lavigne, M. (2011). Mechanisms and Molecules of Neuronal Wiring: A Primer. *Cold Spring Harbor Perspectives in Biology* 3, a001727–a001727.
- Konopka, G., Bomar, J.M., Winden, K., Coppola, G., Jonsson, Z.O., Gao, F., Peng, S., Preuss, T.M., Wohlschlegel, J.A., and Geschwind, D.H. (2009). Human-specific transcriptional regulation of CNS development genes by FOXP2. *Nature* 462, 213–217.
- Kovas, Y., Hayiou-Thomas, M.E., Oliver, B., Dale, P.S., Bishop, D.V.M., and Plomin, R. (2005). Genetic influences in different aspects of language development: the etiology of language skills in 4.5-year-old twins. *Child Dev* 76, 632–651.
- Kozomara, A., and Griffiths-Jones, S. (2013). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* 42, D68–D73.
- Krause, J., Lalueza-Fox, C., Orlando, L., Enard, W., Green, R.E., Burbano, H.A., Hublin, J.-J., Hänni, C., Fortea, J., la Rasilla, de, M., et al. (2007). The derived FOXP2 variant of modern humans was shared with Neandertals. *Curr Biol* 17, 1908–1912.
- Krek, A., Grün, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M., et al. (2005). Combinatorial microRNA target predictions. *Nat Genet* 37, 495–500.
- Krol, J., Busskamp, V., Markiewicz, I., Stadler, M.B., Ribi, S., Richter, J., Duebel, J., Bicker, S., Fehling, H.J., Schubeler, D., et al. (2010). Characterizing Light-Regulated Retinal MicroRNAs Reveals Rapid Turnover as a Common Property of Neuronal MicroRNAs. *Cell* 141, 618–631.
- Kumar, P., Henikoff, S., and Ng, P.C. (2008). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Biotechnol* 4, 1073–1081.

Kurt, S., Fisher, S.E., and Ehret, G. (2012). Foxp2 Mutations Impair Auditory-Motor Association Learning. *PLoS ONE* 7, e33130.

Kurt, S., Groszer, M., Fisher, S.E., and Ehret, G. (2009). Modified sound-evoked brainstem potentials in Foxp2 mutant mice. *Brain Research Reviews* 1289, 30–36.

Kwon, E., Wang, W., and Tsai, L.-H. (2011). Validation of schizophrenia-associated genes CSMD1, C10orf26, CACNA1C and TCF4 as miR-137 targets. *Mol Psychiatry*.

La Spada, A.R., and Taylor, J.P. (2010). Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nat Rev Genet* 11, 247–258.

Laffin, J.J.S., Raca, G., Jackson, C.A., Strand, E.A., Jakielski, K.J., and Shriberg, L.D. (2012). Novel candidate genes and regions for childhood apraxia of speech identified by array comparative genomic hybridization. *Genet Med* 14, 928–936.

Lai, C.S., Fisher, S.E., Hurst, J.A., Levy, E.R., Hodgson, S., Fox, M., Jeremiah, S., Povey, S., Jamison, D.C., Green, E.D., et al. (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.

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

Lai, C.S.L., Gerrelli, D., Monaco, A.P., Fisher, S.E., and Copp, A.J. (2003). FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. *Brain* 126, 2455–2462.

Lai, E., Prezioso, V.R., Tao, W.F., Chen, W.S., and Darnell, J.E. (1991). Hepatocyte nuclear factor 3 alpha belongs to a gene family in mammals that is homologous to the Drosophila homeotic gene fork head. *Genes & Development* 5, 416–427.

Lai, E., Clark, K.L., Burley, S.K., and Darnell, J.E.J. (1993). Hepatocyte Nuclear Factor 3/Fork Head or “Winged Helix” Proteins: A Family of Transcription Factors of Diverse Biologic Function. pp. 10421–10423.

Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A.O., Landthaler, M., et al. (2007). A Mammalian microRNA Expression Atlas Based on Small RNA Library Sequencing. *Cell* 129, 1401–1414.

Law, J., Boyle, J., Harris, F., Harkness, A., and Nye, C. (1998). Screening for speech and language delay: a systematic review of the literature. *Health Technol Assess* 2, 1–184.

- Le, M.T.N., Xie, H., Zhou, B., Chia, P.H., Rizk, P., Um, M., Udolph, G., Yang, H., Lim, B., and Lodish, H.F. (2009). MicroRNA-125b Promotes Neuronal Differentiation in Human Cells by Repressing Multiple Targets. *Mol Cell Biol* 29, 5290–5305.
- Lee, B.-K., Bhinge, A.A., and Iyer, V.R. (2011a). Wide-ranging functions of E2F4 in transcriptional activation and repression revealed by genome-wide analysis. *Nucleic Acids Res* 39, 3558–3573.
- Lee, K., Kunkeaw, N., Jeon, S.H., Lee, I., Johnson, B.H., Kang, G.-Y., Bang, J.Y., Park, H.S., Leelayuwat, C., and Lee, Y.S. (2011b). Precursor miR-886, a novel noncoding RNA repressed in cancer, associates with PKR and modulates its activity. *Rna* 17, 1076–1089.
- Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75, 843–854.
- Lee, W.-S., Kang, C., Drayna, D., and Kornfeld, S. (2011c). Analysis of mannose 6-phosphate uncovering enzyme mutations associated with persistent stuttering. *J Biol Chem* 286, 39786–39793.
- Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S., et al. (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature* 425, 415–419.
- Lee, Y., Jeon, K., Lee, J.-T., Kim, S., and Kim, V.N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. *Embo J* 21, 4663–4670.
- Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S.H., and Kim, V.N. (2004). MicroRNA genes are transcribed by RNA polymerase II. *Embo J* 23, 4051–4060.
- Lemaitre, H., Mattay, V.S., Sambataro, F., Verchinski, B., Straub, R.E., Callicott, J.H., Kittappa, R., Hyde, T.M., Lipska, B.K., Kleinman, J.E., et al. (2010). Genetic variation in FGF20 modulates hippocampal biology. *Journal of Neuroscience* 30, 5992–5997.
- Lennertz, L., Rujescu, D., Wagner, M., Frommann, I., Schulze-Rauschenbach, S., Schuhmacher, A., Landsberg, M.W., Franke, P., Möller, H.-J., Wölwer, W., et al. (2011). Novel Schizophrenia Risk Gene TCF4 Influences Verbal Learning and Memory Functioning in Schizophrenia Patients. *Neuropsychobiology* 63, 131–136.
- Lennon, P.A., Cooper, M.L., Peiffer, D.A., Gunderson, K.L., Patel, A., Peters, S., Cheung, S.W., and Bacino, C.A. (2007). Deletion of 7q31.1 supports involvement of FOXP2 in language impairment: Clinical report and review. *Am. J. Med. Genet.* 143A, 791–798.

Lestari, W., Ichwan, S.J.A., Otsu, M., Yamada, S., Iseki, S., Shimizu, S., and Ikeda, M.-A. (2012). Biochemical and Biophysical Research Communications. *Biochem Biophys Res Commun* 417, 710–716.

Lett, T.A., Chakravarty, M.M., Chakavarty, M.M., Felsky, D., Brandl, E.J., Tiwari, A.K., Gonçalves, V.F., Rajji, T.K., Daskalakis, Z.J., Meltzer, H.Y., et al. (2013). The genome-wide supported microRNA-137 variant predicts phenotypic heterogeneity within schizophrenia. *Mol Psychiatry* 18, 443–450.

Lewis, B.A., and Thompson, L.A. (1992). A study of developmental speech and language disorders in twins. *J Speech Hear Res* 35, 1086–1094.

Lewis, B.A., Cox, N.J., and Byard, P.J. (1993). Segregation analysis of speech and language disorders. *Behav Genet* 23, 291–297.

Lewis, B.A., Freebairn, L.A., Hansen, A., Gerry Taylor, H., Iyengar, S., and Shriberg, L.D. (2004). Family pedigrees of children with suspected childhood apraxia of speech. *J Commun Disord* 37, 157–175.

Lewis, B.P., Shih, I.-H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. *Cell* 115, 787–798.

Li, G., Wang, J., Rossiter, S.J., Jones, G., and Zhang, S. (2007). Accelerated FoxP2 evolution in echolocating bats. *PLoS ONE* 2, e900.

Li, J., and Zhang, Z. (2013). miRNA regulatory variation in human evolution. *Trends Genet* 29, 116–124.

Li, J., Liu, Y., Xin, X., Kim, T.S., Cabeza, E.A., Ren, J., Nielsen, R., Wrana, J.L., and Zhang, Z. (2012). Evidence for positive selection on a number of MicroRNA regulatory interactions during recent human evolution. *PLoS Genet* 8, e1002578.

Li, S., Weidenfeld, J., and Morrissey, E.E. (2004). Transcriptional and DNA Binding Activity of the Foxp1/2/4 Family Is Modulated by Heterotypic and Homotypic Protein Interactions. *Mol Cell Biol* 24, 809–822.

Li, X., Cassidy, J.J., Reinke, C.A., Fischboeck, S., and Carthew, R.W. (2009). A MicroRNA Imparts Robustness against Environmental Fluctuation during Development. *Cell* 137, 273–282.

Liang, Y., Ridzon, D., Wong, L., and Chen, C. (2007). Characterization of microRNA expression profiles in normal human tissues. *BMC Genomics* 8, 166.

Lieberman, P. (2002). On the nature and evolution of the neural bases of human language. *Am. J. Phys. Anthropol.* 119, 36–62.

Lieberman, P. (2007). The Evolution of Human Speech: Its Anatomical and Neural Bases. *Curr Anthropol* 48, 39–66.

- Liégeois, F., Morgan, A.T., Connelly, A., and Vargha-Khadem, F. (2011). Endophenotypes of. *European Journal of Paediatric Neurology* 15, 283–288.
- Liégeois, F., Baldeweg, T., Connelly, A., Gadian, D.G., Mishkin, M., and Vargha-Khadem, F. (2003). Language fMRI abnormalities associated with FOXP2 gene mutation. *Nat Neurosci* 6, 1230–1237.
- Lim, E.T., Raychaudhuri, S., Sanders, S.J., Stevens, C., Sabo, A., MacArthur, D.G., Neale, B.M., Kirby, A., Ruderfer, D.M., Fromer, M., et al. (2013). Rare Complete Knockouts in Humans: Population Distribution and Significant Role in Autism Spectrum Disorders. *Neuron* 77, 235–242.
- Lin, D., Ippolito, G.C., Zong, R.-T., Bryant, J., Koslovsky, J., and Tucker, P. (2007). Bright/ARID3A contributes to chromatin accessibility of the immunoglobulin heavy chain enhancer. *Mol Cancer* 6, 23.
- Lin, Y., Nie, Y., Zhao, J., Chen, X., Ye, M., Li, Y., Du, Y., Cao, J., Shen, B., and Li, Y. (2012). Genetic polymorphism at miR-181a binding site contributes to gastric cancer susceptibility. *Carcinogenesis* 33, 2377–2383.
- Liu, C., Mallick, B., Long, D., Rennie, W.A., Wolenc, A., Carmack, C.S., and Ding, Y. (2013). CLIP-based prediction of mammalian microRNA binding sites. *Nucleic Acids Res* 41, e138–e138.
- Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.-J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. *Science* 305, 1437–1441.
- Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25, 402–408.
- Loohuis, N.F.M.O., Kos, A., Martens, G.J.M., Van Bokhoven, H., Kasri, N.N., and Aschrafi, A. (2011). MicroRNA networks direct neuronal development and plasticity. *Cell Mol Life Sci* 69, 89–102.
- Lopes, F.M., Schröder, R., da Frota Júnior, M.L.C., Zanotto-Filho, A., Müller, C.B., Pires, A.S., Meurer, R.T., Colpo, G.D., Gelain, D.P., Kapczinski, F., et al. (2010). Comparison between proliferative and neuron-like SH-SY5Y cells as an. *Brain Research Reviews* 1337, 85–94.
- Lu, J., and Clark, A.G. (2012). Impact of microRNA regulation on variation in human gene expression. *Genome Research* 22, 1243–1254.
- Lugli, G., Torvik, V.I., Larson, J., and Smalheiser, N.R. (2008). Expression of microRNAs and their precursors in synaptic fractions of adult mouse forebrain. *J Neurochem* 106, 650–661.
- Lund, E., and Dhalberg, J.E. (2005). Substrate selectivity of exportin 5 and Dicer

in the biogenesis of microRNAs. *Cold Spring Harb Symp Quant Biol* 71, 59–66.

Lund, E., Güttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. (2004). Nuclear export of microRNA precursors. *Science* 303, 95–98.

Lupski, J.R., Reid, J.G., Gonzaga-Jauregui, C., Rio Deiros, D., Chen, D.C.Y., Nazareth, L., Bainbridge, M., Dinh, H., Jing, C., Wheeler, D.A., et al. (2010). Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med* 362, 1181–1191.

Macdermot, K.D., Bonora, E., Sykes, N., Coupe, A.-M., Lai, C.S.L., Vernes, S.C., Vargha-Khadem, F., McKenzie, F., Smith, R.L., Monaco, A.P., et al. (2005). Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. *The American Journal of Human Genetics* 76, 1074–1080.

Malcherek, G., Mayr, L., Roda-Navarro, P., Rhodes, D., Miller, N., and Trowsdale, J. (2007). The B7 homolog butyrophilin BTN2A1 is a novel ligand for DC-SIGN. *J Immunol* 179, 3804–3811.

Mangan, S., and Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci USA* 100, 11980–11985.

Maniataki, E., and Mourelatos, Z. (2005). A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Genes & Development* 19, 2979–2990.

Maragkakis, M., Alexiou, P., Papadopoulos, G.L., Reczko, M., Dalamagas, T., Giannopoulos, G., Goumas, G., Koukis, E., Kourtis, K., Simossis, V.A., et al. (2009). Accurate microRNA target prediction correlates with protein repression levels. *BMC Bioinformatics* 10, 295.

Margolis, R.L., Abraham, M.R., Gatchell, S.B., Li, S.H., Kidwai, A.S., Breschel, T.S., Stine, O.C., Callahan, C., McInnis, M.G., and Ross, C.A. (1997). cDNAs with long CAG trinucleotide repeats from human brain. *Hum Genet* 100, 114–122.

Maricic, T., Gunther, V., Georgiev, O., Gehre, S., Curlin, M., Schreiweis, C., Naumann, R., Burbano, H.A., Meyer, M., Lalueza-Fox, C., et al. (2013). A Recent Evolutionary Change Affects a Regulatory Element in the Human FOXP2 Gene. *Molecular Biology and Evolution* 30, 844–852.

Marín, O., Anderson, S.A., and Rubenstein, J.L. (2000). Origin and molecular specification of striatal interneurons. *J Neurosci* 20, 6063–6076.

Marler, P. (1967). Animal Communication Signals. *Science* 157, 769–774.

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., et al. (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* 134, 521–533.

- Martello, G., Rosato, A., Ferrari, F., Manfrin, A., Cordenonsi, M., Dupont, S., Enzo, E., Guzzardo, V., Rondina, M., Spruce, T., et al. (2010). A MicroRNA targeting dicer for metastasis control. *Cell* *141*, 1195–1207.
- Martinez, J., Patkaniowska, A., Urlaub, H., Lührmann, R., and Tuschl, T. (2002). Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* *110*, 563–574.
- Martí, E., Pantano, L., Bañez-Coronel, M., Llorens, F., Miñones-Moyano, E., Porta, S., Sumoy, L., Ferrer, I., and Estivill, X. (2010). A myriad of miRNA variants in control and Huntington's disease brain regions detected by massively parallel sequencing. *Nucleic Acids Res* *38*, 7219–7235.
- Matranga, C., Tomari, Y., Shin, C., Bartel, D.P., and Zamore, P.D. (2004). Passenger-Strand Cleavage Facilitates Assembly of siRNA into Ago2-Containing RNAi Enzyme Complexes. *Cell* *123*, 607–620.
- Mawhood, L., Howlin, P., and Rutter, M. (2000). Autism and developmental receptive language disorder--a comparative follow-up in early adult life. I: Cognitive and language outcomes. *J Child Psychol Psychiatry* *41*, 547–559.
- McCarthy, S.E., Gillis, J., Kramer, M., Lihm, J., Yoon, S., Berstein, Y., Mistry, M., Pavlidis, P., Solomon, R., Ghiban, E., et al. (2014). mutations in schizophrenia implicate chromatinremodeling and support a genetic overlap with autism andintellectual disability. *Mol Psychiatry* *19*, 652–658.
- McNeill, E., and Van Vactor, D. (2012). MicroRNAs Shape the Neuronal Landscape. *Neuron* *75*, 363–379.
- Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell* *15*, 185–197.
- Mellios, N., Huang, H.S., Grigorenko, A., Rogaev, E., and Akbarian, S. (2008). A set of differentially expressed miRNAs, including miR-30a-5p, act as post-transcriptional inhibitors of BDNF in prefrontal cortex. *Hum Mol Genet* *17*, 3030–3042.
- Mencía, A., Modamio-Høybjør, S., Redshaw, N., Morín, M., Mayo-Merino, F., Olavarrieta, L., Aguirre, L.A., Del Castillo, I., Steel, K.P., Dalmay, T., et al. (2009). Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. *Nat Genet* *41*, 609–613.
- Menenti, L., Gierhan, S.M.E., Segaert, K., and Hagoort, P. (2011). Shared Language: Overlap and Segregation of the Neuronal Infrastructure for Speaking and Listening Revealed by Functional MRI. *Psychological Science : a Journal of the American Psychological Society / APS* *22*, 1173–1182.

- Mestdagh, P., Van Vlierberghe, P., De Weer, A., Muth, D., Westermann, F., Speleman, F., and Vandesompele, J. (2009). A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biol* 10, R64.
- Meyer, S.U., Pfaffl, M.W., and Ulbrich, S.E. (2010). Normalization strategies for microRNA profiling experiments: a “normal” way to a hidden layer of complexity? *Biotechnol Lett* 32, 1777–1788.
- Meynert, A.M., Ansari, M., FitzPatrick, D.R., and Taylor, M.S. (2014). Variant detection sensitivity and biases in whole genome and exome sequencing. *BMC Bioinformatics* 15, 247.
- Middleton, F.A., and Strick, P.L. (1994). Anatomical evidence for cerebellar and basal ganglia involvement in higher cognitive function. *Science* 266, 458–461.
- Miller, J.E., Spiteri, E., Condro, M.C., Dosumu-Johnson, R.T., Geschwind, D.H., and White, S.A. (2008). Birdsong decreases protein levels of FoxP2, a molecule required for human speech. *J. Neurophysiol.* 100, 2015–2025.
- Miska, E.A., Alvarez-Saavedra, E., Townsend, M., Yoshii, A., Sestan, N., Rakic, P., Constantine-Paton, M., and Horvitz, H.R. (2003). Microarray analysis of microRNA expression in the developing mammalian brain. *Genome Biol* 5, R68–R68.
- Mizutani, A., Matsuzaki, A., Momoi, M.Y., Fujita, E., Tanabe, Y., and Momoi, T. (2007). Intracellular distribution of a speech/language disorder associated FOXP2 mutant. *Biochem Biophys Res Commun* 353, 869–874.
- Moabbi, A.M., Agarwal, N., Kaderi, El, B., and Ansari, A. (2012). Role for gene looping in intron-mediated enhancement of transcription. *Proc Natl Acad Sci USA* 109, 8505–8510.
- Mohr, J.P., Pessin, M.S., Finkelstein, S., Funkenstein, H.H., Duncan, G.W., and Davis, K.R. (1978). Broca aphasia: Pathologic and clinical. *Neurology* 28, 311–311.
- Monaco, A.P., the SLI Consortium (2007). Multivariate linkage analysis of specific language impairment (SLI). *Ann Hum Genet* 71, 660–673.
- Monteys, A.M., Spengler, R.M., Wan, J., Tecedor, L., Lennox, K.A., Xing, Y., and Davidson, B.L. (2010). Structure and activity of putative intronic miRNA promoters. *Rna* 16, 495–505.
- Morlando, M., Ballarino, M., Gromak, N., Pagano, F., Bozzoni, I., and Proudfoot, N.J. (2008). Primary microRNA transcripts are processed co-transcriptionally. *Nat Struct Mol Biol* 15, 902–909.
- Mueser, K.T., and McGurk, S.R. (2004). Schizophrenia. *Lancet* 363, 2063–2072.

- Mukamel, Z., Konopka, G., Wexler, E., Osborn, G.E., Dong, H., Bergman, M.Y., Levitt, P., and Geschwind, D.H. (2011). Regulation of MET by FOXP2, Genes Implicated in Higher Cognitive Dysfunction and Autism Risk. *J Neurosci* *31*, 11437–11442.
- Mullen, A.C., Orlando, D.A., Newman, J.J., Lovén, J., Kumar, R.M., Bilodeau, S., Reddy, J., Guenther, M.G., DeKoter, R.P., and Young, R.A. (2011). Master Transcription Factors Determine Cell-Type-Specific Responses to TGF- β Signaling. *Cell* *147*, 565–576.
- Murchison, E.P., and Hannon, G.J. (2004). miRNAs on the move: miRNA biogenesis and the RNAi machinery. *Current Opinion in Cell Biology* *16*, 223–229.
- Murugan, M., Harward, S., Scharff, C., and Mooney, R. (2013). Diminished FoxP2 Levels Affect Dopaminergic Modulation of Corticostriatal Signaling Important to Song Variability. *Neuron* *80*, 1464–1476.
- Nagpal, K., Watanabe, K.S., Tsao, B.P., and Tsokos, G.C. (2014). Transcription factor Ikaros Represses Protein Phosphatase 2A (PP2A) Expression through an Intronic Binding Site. *J Biol Chem* *289*, 13751–13757.
- Nakabayashi, K., and Scherer, S.W. (2001). The Human Contactin-Associated Protein-like 2 Gene (CNTNAP2) Spans over 2 Mb of DNA at Chromosome 7q35. *Genomics* *73*, 108–112.
- Natera-Naranjo, O., Aschrafi, A., Gioio, A.E., and Kaplan, B.B. (2010). Identification and quantitative analyses of microRNAs located in the distal axons of sympathetic neurons. *Rna* 1–15.
- Navarrete, K., Pedroso, I., De Jong, S., Stefansson, H., Steinberg, S., Stefansson, K., Ophoff, R.A., Schalkwyk, L.C., and Collier, D.A. (2013). TCF4 (e2-2; ITF2): a schizophrenia-associated gene with pleiotropic effects on human disease. *Am J Med Genet B Neuropsychiatr Genet* *162B*, 1–16.
- Neale, B.M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K.E., Sabo, A., Lin, C.-F., Stevens, C., Wang, L.-S., Makarov, V., et al. (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*.
- Neils, J., and Aram, D.M. (1986). Family history of children with developmental language disorders. *Percept Mot Skills* *63*, 655–658.
- Nelson, C.S., Fuller, C.K., Fordyce, P.M., Greninger, A.L., Li, H., and DeRisi, J.L. (2013). Microfluidic affinity and ChIP-seq analyses converge on a conserved FOXP2-binding motif in chimp and human, which enables the detection of evolutionarily novel targets. *Nucleic Acids Res* *41*, 5991–6004.
- Nelson, P.T., Dimayuga, J., and Wilfred, B.R. (2010). MicroRNA in Situ

Hybridization in the Human Entorhinal and Transentorhinal Cortex. *Front Hum Neurosci* 4, 7–7.

Newbury, D.F., and Monaco, A.P. (2009). Genetic Advances in the Study of Speech and Language Disorders. *Neuron* 68, 309–320.

Newbury, D.F., Bonora, E., Lamb, J.A., Fisher, S.E., Lai, C.S.L., Baird, G., Jannoun, L., Slonims, V., Stott, C.M., Merricks, M.J., et al. (2002). FOXP2 Is Not a Major Susceptibility Gene for Autism or Specific Language Impairment. *The American Journal of Human Genetics* 70, 1318–1327.

Newbury, D.F., Paracchini, S., Scerri, T.S., Winchester, L., Addis, L., Richardson, A.J., Walter, J., Stein, J.F., Talcott, J.B., and Monaco, A.P. (2011). Investigation of Dyslexia and SLI Risk Variants in Reading- and Language-Impaired Subjects. *Behav Genet* 41, 90–104.

Newbury, D.F., Bishop, D.V.M., and Monaco, A.P. (2005). Genetic influences on language impairment and phonological short-term memory. *Trends Cogn Sci (Regul Ed)* 9, 528–534.

Newbury, D.F., Fisher, S.E., and Monaco, A.P. (2010). Recent advances in the genetics of language impairment. *Genome Medicine* 2, 6.

Newbury, D.F., Winchester, L., Addis, L., Paracchini, S., Buckingham, L.-L., Clark, A., Cohen, W., Cowie, H., Dworzynski, K., Everitt, A., et al. (2009). CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *The American Journal of Human Genetics* 85, 264–272.

Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E., et al. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461, 272–276.

Nolis, I.K., McKay, D.J., Mantouvalou, E., Lomvardas, S., Merika, M., and Thanos, D. (2009). Transcription factors mediate long-range enhancer-promoter interactions. *Proc Natl Acad Sci USA* 106, 20222–20227.

Nordeen, K.W., and Nordeen, E.J. (1991). Auditory feedback is necessary for the maintenance of stereotyped song in adult zebra finches. *Behav Neural Biol* 57, 58–66.

Nottebohm, F., and Liu, W.-C. (2010). Brain & Language. *Brain Lang* 115, 3–17.

Nowicki, S., and Searcy, W.A. (2004). Song Function and the Evolution of Female Preferences: Why Birds Sing, Why Brains Matter. *Ann N Y Acad Sci* 1016, 704–723.

Nudel, R., Simpson, N.H., Baird, G., O'hare, A., Conti-Ramsden, G., Bolton, P.F., Hennessy, E.R., Monaco, A.P., Knight, J.C., Winney, B., et al. (2014a).

Associations of HLA alleles with specific language impairment. *J Neurodevelopmental Disord* 6, –1.

Nudel, R., Simpson, N.H., Baird, G., O'hare, A., Conti-Ramsden, G., Bolton, P.F., Hennessy, E.R., the SLI Consortium, Ring, S.M., Smith, G.D., et al. (2014b). Genome-wide association analyses of child genotype effects and parent-of-origin effects in specific language impairment. *Genes, Brain and Behavior*.

O'Donnell, K.A., Wentzel, E.A., Zeller, K.I., Dang, C.V., and Mendell, J.T. (2005). c-Myc-regulated microRNAs modulate E2F1 expression. *Nature* 435, 839–843.

O'Roak, B.J., Deriziotis, P., Lee, C., Vives, L., Schwartz, J.J., Girirajan, S., Karakoc, E., Mackenzie, A.P., Ng, S.B., Baker, C., et al. (2011). Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet* 43, 585–589.

O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D., et al. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*.

Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., and Lai, E.C. (2006). The Mirtron Pathway Generates microRNA-Class Regulatory RNAs in *Drosophila*. *Cell* 130, 89–100.

Oldham, M.C., Horvath, S., and Geschwind, D.H. (2006). Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc Natl Acad Sci USA* 103, 17973–17978.

Olena, A.F., and Patton, J.G. (2009). Genomic organization of microRNAs. *J. Cell. Physiol.* n/a–n/a.

Olsen, P.H., and Ambros, V. (1999). The lin-4 regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev Biol* 216, 671–680.

Olsson, M., Campbell, K., Wictorin, K., and Björklund, A. (1995). Projection neurons in fetal striatal transplants are predominantly derived from the lateral ganglionic eminence. *Neuroscience* 69, 1169–1182.

Omura, N., Li, C.-P., Li, A., Hong, S.-M., Walter, K., Jimeno, A., Hidalgo, M., and Goggins, M. (2008). Genome-wide profiling of methylated promoters in pancreatic adenocarcinoma. *Cancer Biol Ther* 7, 1146–1156.

Ono, M., Yaguchi, H., Ohkura, N., Kitabayashi, I., Nagamura, Y., Nomura, T., Miyachi, Y., Tsukada, T., and Sakaguchi, S. (2007). Foxp3 controls regulatory T-cell function by interacting with AML1/Runx1. *Nature* 446, 685–689.

Overdier, D.G., Porcella, A., and Costa, R.H. (1994). The DNA-binding specificity of the hepatocyte nuclear factor 3/forkhead domain is influenced by amino-acid

residues adjacent to the recognition helix. *Mol Cell Biol* 14, 2755–2766.

Ozsolak, F., Poling, L.L., Wang, Z., Liu, H., Liu, X.S., Roeder, R.G., Zhang, X., Song, J.S., and Fisher, D.E. (2008). Chromatin structure analyses identify miRNA promoters. *Genes & Development* 22, 3172–3183.

Palka, C., Alfonsi, M., Mohn, A., Cerbo, R., Franchi, P.G., Fantasia, D., Morizio, E., Stuppia, L., Calabrese, G., Zori, R., et al. (2012). Mosaic 7q31 Deletion Involving FOXP2 Gene Associated With Language Impairment. *Pediatrics* 129, e183–e188.

Panaitof, S.C., Abrahams, B.S., Dong, H., Geschwind, D.H., and White, S.A. (2010). Language-related Cntnap2 gene is differentially expressed in sexually dimorphic song nuclei essential for vocal learning in songbirds. *J. Comp. Neurol.* 518, 1995–2018.

Paracchini, S., Thomas, A., Castro, S., Lai, C., Paramasivam, M., Wang, Y., Keating, B.J., Taylor, J.M., Hacking, D.F., Scerri, T., et al. (2006). The chromosome 6p22 haplotype associated with dyslexia reduces the expression of KIAA0319, a novel gene involved in neuronal migration. *Hum Mol Genet* 15, 1659–1666.

Passafaro, M., Nakagawa, T., Sala, C., and Sheng, M. (2003). Induction of dendritic spines by an extracellular domain of AMPA receptor subunit GluR2. *Nature* 424, 677–681.

Peippo, M., and Ignatius, J. (2012). Pitt-Hopkins Syndrome. *Mol Syndromol* 2, 171–180.

Peltier, H.J., and Latham, G.J. (2008). Normalization of microRNA expression levels in quantitative RT-PCR assays: Identification of suitable reference RNA targets in normal and cancerous human solid tissues. *Rna* 14, 844–852.

Peñagarikano, O., Abrahams, B.S., Herman, E.I., Winden, K.D., Gdalyahu, A., Dong, H., Sonnenblick, L.I., Gruver, R., Almajano, J., Bragin, A., et al. (2011). Absence of CNTNAP2 Leads to Epilepsy, Neuronal Migration Abnormalities, and Core Autism-Related Deficits. *Cell* 147, 235–246.

Peter, B., Raskind, W.H., Matsushita, M., Lisowski, M., Vu, T., Berninger, V.W., Wijsman, E.M., and Brkanac, Z. (2010). Replication of CNTNAP2 association with nonword repetition and support for FOXP2 association with timed reading and motor activities in a dyslexia family sample. *J Neurodevelop Disord* 3, 39–49.

Petersen, C.P., Bordeleau, M.-E., Pelletier, J., and Sharp, P.A. (2005). Short RNAs Repress Translation after Initiation in Mammalian Cells. *Mol Cell* 21, 533–542.

Pinter, R., and Hindges, R. (2010). Perturbations of microRNA function in mouse

dicer mutants produce retinal defects and lead to aberrant axon pathfinding at the optic chiasm. *PLoS ONE* 5, e10021.

Piriyapongsa, J., Jordan, I.K., Conley, A.B., Ronan, T., and Smalheiser, N.R. (2011). Transcription factor binding sites are highly enriched within microRNA precursor sequences. *Biology Direct* 6, 61.

Poeppel, D., Emmorey, K., Hickok, G., and Pyllkanen, L. (2012). Towards a New Neurobiology of Language. *J Neurosci* 32, 14125–14131.

Poeppel, D. (2011). Genetics and language: a neurobiological perspective on the missing link (-ing hypotheses). *J Neurodevel Disord* 3, 381–387.

Poeppel, D. (2014). The neuroanatomic and neurophysiological infrastructure for speech and language. *Curr Opin Neurobiol* 28, 142–149.

Poliak, S., Gollan, L., Martinez, R., Custer, A., Einheber, S., Salzer, J.L., Trimmer, J.S., Shrager, P., and Peles, E. (1999). Caspr2, a new member of the neurexin superfamily, is localized at the juxtaparanodes of myelinated axons and associates with K⁺ channels. *Neuron* 24, 1037–1047.

Poliak, S., Gollan, L., Salomon, D., Berglund, E.O., Ohara, R., Ranscht, B., and Peles, E. (2001). Localization of Caspr2 in myelinated nerves depends on axon-glia interactions and the generation of barriers along the axon. *Journal of Neuroscience* 21, 7568–7575.

Poliak, S., Salomon, D., Elhanany, H., Sabanay, H., Kiernan, B., Pevny, L., Stewart, C.L., Xu, X., Chiu, S.-Y., Shrager, P., et al. (2003). Juxtaparanodal clustering of Shaker-like K⁺ channels in myelinated axons depends on Caspr2 and TAG-1. *The Journal of Cell Biology* 162, 1149–1160.

Powell, L.M., and Jarman, A.P. (2007). Context dependence of proneural bHLH proteins. *Curr Opin Genet Dev* 18, 411–417.

Price, C.J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann N Y Acad Sci* 1191, 62–88.

Price, C.J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* 62, 816–847.

Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 26, 2336–2337.

Pscherer, A., Dörflinger, U., Kirfel, J., Gawlas, K., Rüschhoff, J., Buettner, R., and Schüle, R. (1996). The helix-loop-helix transcription factor SEF-2 regulates the activity of a novel initiator element in the promoter of the human somatostatin

receptor II gene. *Embo J* 15, 6680–6690.

Ptak, S.E., Enard, W., Wiebe, V., Hellmann, I., Krause, J., Lachmann, M., and Pääbo, S. (2009). Linkage disequilibrium extends across putative selected sites in FOXP2. *Molecular Biology and Evolution* 26, 2181–2184.

Purcell, S.M., Moran, J.L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., O'Dushlaine, C., Chambert, K., Bergen, S.E., Kähler, A., et al. (2014). A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* 506, 185–190.

Qian, J., Zhang, Z., Liang, J., Ge, Q., Duan, X., Ma, F., and Li, F. (2011). The full-length transcripts and promoter analysis of intergenic microRNAs in *Drosophila melanogaster*. *Genomics* 97, 294–303.

Quach, H., Barreiro, L.B., Laval, G., Zidane, N., Patin, E., Kidd, K.K., Kidd, J.R., Bouchier, C., Veuille, M., Antoniewski, C., et al. (2009). Signatures of purifying and local positive selection in human miRNAs. *Am J Hum Genet* 84, 316–327.

Rajasethupathy, P., Fiumara, F., Sheridan, R., Betel, D., Puthanveetil, S.V., Russo, J.J., Sander, C., Tuschl, T., and Kandel, E. (2009). Characterization of Small RNAs in *Aplysia* Reveals a Role for miR-124 in Constraining Synaptic Plasticity through CREB. *Neuron* 63, 803–817.

Rasband, M.N., Park, E.W., Zhen, D., Arbuckle, M.I., Poliak, S., Peles, E., Grant, S.G.N., and Trimmer, J.S. (2002). Clustering of neuronal potassium channels is independent of their interaction with PSD-95. *The Journal of Cell Biology* 159, 663–672.

Raymer, A.M., Moberg, P., Crosson, B., Nadeau, S., and Rothi, L. (1995). Lexical–semantic deficits in two patients with dominant thalamic infarction. *Neuropsychologia* 35, 211–219.

Reamon-Buettner, S., Cho, S.-H., and Borlak, J. (2007). Mutations in the 3'-untranslated region of GATA4 as molecular hotspots for congenital heart disease (CHD). *BMC Med Genet* 8, 38.

Reczko, M., Maragkakis, M., Alexiou, P., Grosse, I., and Hatzigeorgiou, A.G. (2012). Functional microRNA targets in protein coding sequences. *Bioinformatics* 28, 771–776.

Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola, B., Briggs, A.W., Stenzel, U., Johnson, P.L.F., et al. (2011). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468, 1053–1060.

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

- Rice, G.M., Raca, G., Jakielski, K.J., Laffin, J.J., Iyama-Kurtycz, C.M., Hartley, S.L., Sprague, R.E., Heintzelman, A.T., and Shriberg, L.D. (2011). Phenotype of FOXP2 haploinsufficiency in a mother and son. *Am. J. Med. Genet.* *158A*, 174–181.
- Riley, J.R., Greggers, U., Smith, A.D., Reynolds, D.R., and Menzel, R. (2005). The flight paths of honeybees recruited by the waggle dance. *Nature* *435*, 205–207.
- Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J.L., Kähler, A.K., Akterin, S., Bergen, S.E., Collins, A.L., Crowley, J.J., Fromer, M., et al. (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet* *45*, 1150–1159.
- Ripke, S., Sanders, A.R., Kendler, K.S., Levinson, D.F., Sklar, P., Holmans, P.A., Lin, D.-Y., Duan, J., Ophoff, R.A., Andreassen, O.A., et al. (2011). Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* *43*, 969–976.
- Rodenas-Cuadrado, P., Ho, J., and Vernes, S.C. (2013). Shining a light on CNTNAP2: complex functions to complex disorders. *Eur J Hum Genet* *22*, 171–178.
- Rodriguez, A., Griffiths-Jones, S., Ashurst, J.L., and Bradley, A. (2004). Identification of mammalian microRNA host genes and transcription units. *Genome Research* *14*, 1902–1910.
- Roll, P., Vernes, S.C., Bruneau, N., Cillario, J., Ponsole-Lenfant, M., Massacrier, A., Rudolf, G., Khalife, M., Hirsch, E., Fisher, S.E., et al. (2010). Molecular networks implicated in speech-related disorders: FOXP2 regulates the SRPX2/uPAR complex. *Hum Mol Genet* *19*, 4848–4860.
- Rossi, E., Verri, A.P., Patricelli, M.G., Destefani, V., Ricca, I., Vetro, A., Ciccone, R., Giorda, R., Toniolo, D., Maraschio, P., et al. (2008). A 12 Mb deletion at 7q33eq35 associated with autism spectrum disorders and primary amenorrhea. *European Journal of Medical Genetics* *51*, 631–638.
- Ruby, J.G., Jan, C.H., and Bartel, D.P. (2007). Intronic microRNA precursors that bypass Drosha processing. *Nature* *448*, 83–86.
- Rudov, A., Rocchi, M.B.L., Accorsi, A., Spada, G., Procopio, A.D., Olivieri, F., Rippo, M.R., and Albertini, M.C. (2013). Putative miRNAs for the diagnosis of dyslexia, dyspraxia, and specific language impairment. *Epigenetics* *8*, 1023–1029.
- Ryan, D.G., Oliveira-Fernandes, M., and Lavker, R.M. (2005). MicroRNAs of the mammalian eye display distinct and overlapping tissue specificity. *Mol Vis* *12*, 1175–1184.

Saba, R., Goodman, C.D., Huzarewich, R.L.C.H., Robertson, C., and Booth, S.A. (2008). A miRNA Signature of Prion Induced Neurodegeneration. *PLoS ONE* 3, e3652.

Sachdeva, M., Zhu, S., Wu, F., Wu, H., Walia, V., Kumar, S., Elble, R., Watabe, K., and Mo, Y.-Y. (2009). p53 represses c-Myc through induction of the tumor suppressor miR-145. *Proc Natl Acad Sci USA* 106, 3207–3212.

Saini, H.K., Enright, A.J., and Griffiths-Jones, S. (2008). Annotation of mammalian primary microRNAs. *BMC Genomics* 9, 564.

Saini, H.K., Griffiths-Jones, S., and Enright, A.J. (2007). Genomic analysis of human microRNA transcripts. *Proc Natl Acad Sci USA* 104, 17719–17724.

Sakai, Y., Shaw, C.A., Dawson, B.C., Dugas, D.V., Al-Mohtaseb, Z., Hill, D.E., and Zoghbi, H.Y. (2011). Protein interactome reveals converging molecular pathways among autism disorders. *Science Translational Medicine* 3, 86ra49–86ra49.

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P.P. (2011). A ceRNA Hypothesis: The Rosetta Stone of a Hidden RNA Language? *Cell* 146, 353–358.

Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., Dilullo, N.M., Parikshak, N.N., Stein, J.L., et al. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*.

Santarelli, D.M., Beveridge, N.J., Tooney, P.A., and Cairns, M.J. (2011). Upregulation of dicer and microRNA expression in the dorsolateral prefrontal cortex Brodmann area 46 in schizophrenia. *Biol Psychiatry* 69, 180–187.

Saunders, M.A., Liang, H., and Li, W.-H. (2007). Human polymorphism at microRNAs and microRNA target sites. *Proc Natl Acad Sci USA* 104, 3300–3305.

Schanen, B.C., and Li, X. (2011). Transcriptional regulation of mammalian miRNA genes. *Genomics* 97, 1–6.

Scharff, C., and Petri, J. (2011). Evo-devo, deep homology and FoxP2: implications for the evolution of speech and language. *Philosophical Transactions of the Royal Society B: Biological Sciences* 366, 2124–2140.

Schmahmann, J.D., and Sherman, J.C. (1998). The cerebellar cognitive affective syndrome. *Brain* 121 (Pt 4), 561–579.

Schmidt-Edelkraut, U., Daniel, G., Hoffmann, A., and Spengler, D. (2014). Zac1 regulates cell cycle arrest in neuronal progenitors via Tcf4. *Mol Cell Biol* 34, 1020–1030.

Schratt, G. (2009). microRNAs at the synapse. *Nat Rev Neurosci* 10, 842–849.

Schratt, G.M., Tuebing, F., Nigh, E.A., Kane, C.G., Sabatini, M.E., Kiebler, M., and Greenberg, M.E. (2006). A brain-specific microRNA regulates dendritic spine development. *Nature* 439, 283–289.

Schulz, S.B., Haesler, S., Scharff, C., and Rochefort, C. (2010). Knock-down of FoxP2 alters spine density in Area X of the Zebra Finch. *Genes, Brain and Behavior* no–no.

Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the Assembly of the RNAi Enzyme Complex. *Cell* 115, 199–208.

Scott-Van Zeeland, A.A., Abrahams, B.S., Alvarez-Retuerto, A.I., Sonnenblick, L.I., Rudie, J.D., Ghahremani, D., Mumford, J.A., Poldrack, R.A., Dapretto, M., Geschwind, D.H., et al. (2010). Altered Functional Connectivity in Frontal Lobe Circuits Is Associated with Variation in the Autism Risk Gene CNTNAP2. *Science Translational Medicine* 2, 56ra80–56ra80.

Segaert, K., Menenti, L., Weber, K., Petersson, K.M., and Hagoort, P. (2012). Shared Syntax in Language Production and Language Comprehension--An fMRI Study. *Cerebral Cortex* 22, 1662–1670.

Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. *Nature* 455, 58–63.

Semel, E.M., Wiig, E.H., and Secord, W. (1992). CELF-R: Clinical Evaluation of Language Fundamentals - Revised (The Psychological Corporation).

Sempere, L.F., Freemantle, S., Pitha-Rowe, I., Moss, E., Dmitrovsky, E., and Ambros, V. (2004). Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. *Genome Biol* 5, R13.

Seno, M.M.G., Hu, P., Gwadry, F.G., Pinto, D., Marshall, C.R., Casallo, G., and Scherer, S.W. (2010). Gene and miRNA expression profiles in autism spectrum disorders. *Brain Research Reviews* 1380, 85–97.

Sepp, M., Pruunsild, P., and Timmusk, T. (2012). Pitt-Hopkins syndrome-associated mutations in TCF4 lead to variable impairment of the transcription factor function ranging from hypomorphic to dominant-negative effects. *Hum Mol Genet* 21, 2873–2888.

Sepp, M., Kannike, K., Eesmaa, A., Urb, M., and Timmusk, T. (2011). Functional Diversity of Human Basic Helix-Loop-Helix Transcription Factor TCF4 Isoforms Generated by Alternative 5' Exon Usage and Splicing. *PLoS ONE* 6, e22138.

Sethupathy, P., and Collins, F.S. (2008). MicroRNA target site polymorphisms

and human disease. *Trends Genet* 24, 489–497.

Sevcik, R.A., and Savage-Rumbaugh, E.S. (1994). Language comprehension and use by great apes. *Language & Communication* 14, 37–58.

Seyfarth, R., and Cheney, D. (1990). The assessment by vervet monkeys of their own and another species' alarm calls. *Animal Behaviour* 40, 754–764.

Seyfarth, R.M., and Cheney, D.L. (2010). Brain & Language. *Brain Lang* 115, 92–100.

Shah, R., Medina-Martinez, O., Chu, L.-F., Samaco, R.C., and Jamrich, M. (2006). Expression of FoxP2 during zebrafish development and in the adult brain. *Int. J. Dev. Biol.* 50, 435–438.

Shalgi, R., Lieber, D., Oren, M., and Pilpel, Y. (2007). Global and Local Architecture of the Mammalian microRNA–Transcription Factor Regulatory Network. *PLoS Comput Biol* 3, e131.

Sharma, M., Burré, J., and Südhof, T.C. (2010). CSP α promotes SNARE-complex assembly by chaperoning SNAP-25 during synaptic activity. *Nat Cell Biol* 13, 30–39.

Sheleg, M., Yochum, C.L., Wagner, G.C., Zhou, R., and Richardson, J.R. (2013). Behavioural Brain Research. *Behav Brain Res* 236, 139–147.

Shi, Y., Sawada, J.-I., Sui, G., Affar, E.B., Whetstine, J.R., Lan, F., Ogawa, H., Luke, M.P.-S., Nakatani, Y., and Shi, Y. (2003). Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature* 422, 735–738.

Shi, Z., Luo, G., Fu, L., Fang, Z., Wang, X., and Li, X. (2013). miR-9 and miR-140-5p Target FoxP2 and Are Regulated as a Function of the Social Context of Singing Behavior in Zebra Finches. *J Neurosci* 33, 16510–16521.

Shibata, M., Nakao, H., Kiyonari, H., Abe, T., and Aizawa, S. (2011). MicroRNA-9 Regulates Neurogenesis in Mouse Telencephalon by Targeting Multiple Transcription Factors. *Journal of Neuroscience* 31, 3407–3422.

Shu, W., Lu, M.M., Zhang, Y., Tucker, P.W., Zhou, D., and Morrissey, E.E. (2007). Foxp2 and Foxp1 cooperatively regulate lung and esophagus development. *Development* 134, 1991–2000.

Shu, W., Yang, H., Zhang, L., Lu, M.M., and Morrissey, E.E. (2001). Characterization of a New Subfamily of Winged-helix/Forkhead (Fox) Genes That Are Expressed in the Lung and Act as Transcriptional Repressors. *Journal of Biological Chemistry* 276, 27488–27497.

Shu, W., Cho, J.Y., Jiang, Y., Zhang, M., Weisz, D., Elder, G.A., Schmeidler, J., De Gasperi, R., Sosa, M.A.G., Rabidou, D., et al. (2005). Altered ultrasonic

vocalization in mice with a disruption in the Foxp2 gene. *Proc Natl Acad Sci USA* 102, 9643–9648.

Sia, G.M., Clem, R.L., and Hagan, R.L. (2013). The Human Language-Associated Gene SRPX2 Regulates Synapse Formation and Vocalization in Mice. *Science* 342, 987–991.

Siegel, G., Obernosterer, G., Fiore, R., Oehmen, M., Bicker, S., Christensen, M., Khudayberdiev, S., Leuschner, P.F., Busch, C.J.L., Kane, C., et al. (2009). A functional screen implicates microRNA-138-dependent regulation of the depalmitoylation enzyme APT1 in dendritic spine morphogenesis. *Nat Cell Biol* 11, 705–716.

Simon, D., Laloo, B., Barillot, M., Barnette, T., Blanchard, C., Rooryck, C., Marche, M., Burgelin, I., Couprie, I., Chassaing, N., et al. (2010). A mutation in the 3'-UTR of the HDAC6 gene abolishing the post-transcriptional regulation mediated by hsa-miR-433 is linked to a new form of dominant X-linked chondrodysplasia. *Hum Mol Genet* 19, 2015–2027.

Simpson, N.H., Addis, L., Brandler, W.M., Slonims, V., Clark, A., Watson, J., Scerri, T.S., Hennessy, E.R., Bolton, P.F., Conti-Ramsden, G., et al. (2013). Increased prevalence of sex chromosome aneuploidies in specific language impairment and dyslexia. *Dev Med Child Neurol* n/a–n/a.

Sin, C., Li, H., and Crawford, D.A. (2014). Transcriptional Regulation by FOXP1, FOXP2, and FOXP4 Dimerization. *J Mol Neurosci*.

Slezak-Prochazka, I., Durmus, S., Kroesen, B.J., and Van Den Berg, A. (2010). MicroRNAs, macrocontrol: Regulation of miRNA processing. *Rna* 1–10.

SLI Consortium (2002). A genomewide scan identifies two novel loci involved in specific language impairment. *Am J Hum Genet* 70, 384–398.

SLI Consortium (2004). Highly significant linkage to the SLI1 locus in an expanded sample of individuals affected by specific language impairment. *Am J Hum Genet* 74, 1225–1238.

Smrt, R.D., Szulwach, K.E., Pfeiffer, R.L., Li, X., Guo, W., Pathania, M., Teng, Z.-Q., Luo, Y., Peng, J., Bordey, A., et al. (2010). MicroRNA miR-137 Regulates Neuronal Maturation by Targeting Ubiquitin Ligase Mind Bomb-1. *Stem Cells* 28, 1060–1070.

Snowling, M.J., Adams, J.W., Bishop, D.V., and Stothard, S.E. (2001). Educational attainments of school leavers with a preschool history of speech-language impairments. *Int J Lang Commun Disord* 36, 173–183.

Sobrado, V.R., Moreno-Bueno, G., Cubillo, E., Holt, L.J., Nieto, M.A., Portillo, F., and Cano, A. (2009). The class I bHLH factors E2-2A and E2-2B regulate EMT.

Journal of Cell Science 122, 1014–1024.

Sohrabji, F., Nordeen, E.J., and Nordeen, K.W. (1990). Selective impairment of song learning following lesions of a forebrain nucleus in the juvenile zebra finch. *Behav Neural Biol* 53, 51–63.

Soldati, C., Bithell, A., Johnston, C., Wong, K.-Y., Stanton, L.W., and Buckley, N.J. (2013). Dysregulation of REST-regulated coding and non-coding RNAs in a cellular model of Huntington's disease. *J Neurochem* 124, 418–430.

Somel, M., Liu, X., Tang, L., Yan, Z., Hu, H., Guo, S., Jiang, X., Zhang, X., Xu, G., Xie, G., et al. (2011). MicroRNA-driven developmental remodeling in the brain distinguishes humans from other primates. *PLoS Biol* 9, e1001214.

Song, G., and Wang, L. (2009). A Conserved Gene Structure and Expression Regulation of miR-433 and miR-127 in Mammals. *PLoS ONE* 4, e7829.

Spiteri, E., Konopka, G., Coppola, G., Bomar, J., Oldham, M., Ou, J., Vernes, S.C., Fisher, S.E., Ren, B., and Geschwind, D.H. (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.

Stadler, P.F., Chen, J.J.-L., Hackermüller, J., Hoffmann, S., Horn, F., Khaitovich, P., Kretzschmar, A.K., Mosig, A., Prohaska, S.J., Qi, X., et al. (2009). Evolution of vault RNAs. *Molecular Biology and Evolution* 26, 1975–1991.

Steinberg, S., De Jong, S., De Jong, S., Andreassen, O.A., Werge, T., Børghlum, A.D., Mors, O., Mortensen, P.B., Gustafsson, O., Costas, J., et al. (2011). Common variants at VRK2 and TCF4 conferring risk of schizophrenia. *Hum Mol Genet* 20, 4076–4081.

Stempor, P.A., Cauchi, M., and Wilson, P. (2011). MMpred: functional miRNA--mRNA interaction analyses by miRNA expression prediction. *BMC Genomics* 13, 620–620.

Stoodley, C.J., and Schmahmann, J.D. (2009). Brain & Language. *Brain Lang* 110, 149–153.

Stoodley, C.J., Valera, E.M., and Schmahmann, J.D. (2012). Functional topography of the cerebellum for motor and cognitive tasks: An fMRI study. *Neuroimage* 59, 1560–1570.

Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 100, 9440–9445.

Stothard, S.E., Snowling, M.J., Bishop, D.V., Chipchase, B.B., and Kaplan, C.A. (1998). Language-impaired preschoolers: a follow-up into adolescence. *J Speech Lang Hear Res* 41, 407–418.

Strauss, K.A., Puffenberger, E.G., Huentelman, M.J., Gottlieb, S., Dobrin, S.E., Parod, J.M., Stephan, D.A., and Morton, D.H. (2006). Recessive symptomatic focal epilepsy and mutant contactin-associated protein-like 2. *New England Journal of Medicine* 354, 1370–1377.

Strazisar, M., Cammaerts, S., van der Ven, K., Forero, D.A., Lenaerts, A.-S., Nordin, A., Almeida-Souza, L., Genovese, G., Timmerman, V., Liekens, A., et al. (2014). MIR137 variants identified in psychiatric patients affect synaptogenesis and neuronal transmission gene sets. *Mol Psychiatry*.

Stromswold, K. (1998). Genetics of spoken language disorders. *Hum. Biol.* 70, 297–324.

Stromswold, K. (2001). The heritability of language: A Review and Metaanalysis of Twin, Adoption, and Linkage Studies. *Language* 77, 647–723.

Stromswold, K. (2006). Why aren't identical twins linguistically identical? Genetic, prenatal and postnatal factors. *Cognition* 101, 333–384.

Stroud, J.C., Wu, Y., Bates, D.L., Han, A., Nowick, K., Paabo, S., Tong, H., and Chen, L. (2006). Structure of the forkhead domain of FOXP2 bound to DNA. *Structure* 14, 159–166.

Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., et al. (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA* 101, 6062–6067.

Sun, G., Ye, P., Murai, K., Lang, M.-F., Li, S., Zhang, H., Li, W., Fu, C., Yin, J., Wang, A., et al. (2010). miR-137 forms a regulatory loop with nuclear receptor TLX and LSD1 in neural stem cells. *Nat Commun* 2, 529–529.

Sun, X., Wu, Y., Gu, M., and Zhang, Y. (2014). miR-342-5p Decreases Ankyrin G Levels in Alzheimer's Disease Transgenic Mouse Models. *CellReports* 6, 264–270.

Sweatt, J.D. (2013). Pitt-Hopkins Syndrome: intellectual disability due to loss of TCF4-regulated gene transcription. *Exp Mol Med* 45, e21.

Takahashi, K., Liu, F.-C., Hirokawa, K., and Takahashi, H. (2003). Expression of Foxp2, a gene involved in speech and language, in the developing and adult striatum. *J Neurosci Res* 73, 61–72.

Takahashi, K., Liu, F.-C., Oishi, T., Mori, T., Higo, N., Hayashi, M., Hirokawa, K., and Takahashi, H. (2008). Expression of FOXP2 in the developing monkey forebrain: comparison with the expression of the genes FOXP1, PBX3, and MEIS2. *J. Comp. Neurol.* 509, 180–189.

Tallal, P., R, R., and S, C. (1989). Familial aggregation in specific language

impairment. *J Speech Hear Res* 52, 167–173.

Tallal, P., Townsend, J., S, C., and Wulfeck, B. (1991). Phenotypic profiles of language-impaired children based on genetic/family history. *Brain Lang* 41, 81–95.

Tan-Wong, S.M., French, J.D., Proudfoot, N.J., and Brown, M.A. (2008). Dynamic interactions between the promoter and terminator regions of the mammalian BRCA1 gene. *Pnas* 105, 5160–5165.

Tay, Y., Rinn, J., and Pandolfi, P.P. (2014). The multilayered complexity of ceRNA crosstalk and competition. *Nature* 505, 344–352.

Tazi-Ahnini, R., Henry, J., Offer, C., Bouissou-Bouchouata, C., Mather, I.H., and Pontarotti, P. (1997). Cloning, localization, and structure of new members of the butyrophilin gene family in the juxta-telomeric region of the major histocompatibility complex. *Immunogenetics* 47, 55–63.

Teramitsu, I., Kudo, L., London, S., Geschwind, D., and White, S. (2004). Parallel FoxP1 and FoxP2 expression in songbird and human brain predicts functional interaction. *Journal of Neuroscience* 24, 3152.

Teramitsu, I., and White, S.A. (2006). FoxP2 regulation during undirected singing in adult songbirds. *J Neurosci* 26, 7390–7394.

Teramitsu, I., Poopatanapong, A., Torrisi, S., and White, S.A. (2010). Striatal FoxP2 Is Actively Regulated during Songbird Sensorimotor Learning. *PLoS ONE* 5, e8548.

Thein, S., Tao-Cheng, J.-H., Li, Y., Bayer, K.U., Reese, T.S., and Dosemeci, A. (2014). CaMKII Mediates Recruitment and Activation of the Deubiquitinase CYLD at the Postsynaptic Density. *PLoS ONE* 9, e91312.

Tomblin, J.B., and Buckwalter, P.R. (1998). Heritability of poor language achievement among twins. *J Speech Lang Hear Res* 41, 188–199.

Tomblin, J.B., Zhang, X., Buckwalter, P., and Catts, H. (2000). The association of reading disability, behavioral disorders, and language impairment among second-grade children. *J Child Psychol Psychiatry* 41, 473–482.

Tomblin, J.B., O'Brien, M., Shriberg, L.D., Williams, C., Murray, J., Patil, S., Bjork, J., Anderson, S., and Ballard, K. (2009). Language Features in a Mother and Daughter of a Chromosome 7;13 Translocation Involving FOXP2. *J Speech Lang Hear Res* 52, 1157.

Tomblin, J.B., Records, N.L., and Zhang, X. (1996). A System for the Diagnosis of Specific Language Impairment in Kindergarten Children. *J Speech Lang Hear Res* 39, 1284–1294.

Tomblin, J.B., Records, N.L., Buckwalter, P., Zhang, X., Smith, E., and O'Brien, M. (1997). Prevalence of Specific Language Impairment in Kindergarten Children. *J Speech Lang Hear Res* 40, 1245–1260.

Traka, M., Goutebroze, L., Denisenko, N., Bessa, M., Nifli, A., Havaki, S., Iwakura, Y., Fukamauchi, F., Watanabe, K., Soliven, B., et al. (2003). Association of TAG-1 with Caspr2 is essential for the molecular organization of juxtaparanodal regions of myelinated fibers. *The Journal of Cell Biology* 162, 1161–1172.

Tsang, J., Zhu, J., and van Oudenaarden, A. (2007). MicroRNA-Mediated Feedback and Feedforward Loops Are Recurrent Network Motifs in Mammals. *Mol Cell* 26, 753–767.

Tsui, D., Vessey, J.P., Tomita, H., Kaplan, D.R., and Miller, F.D. (2013). FoxP2 Regulates Neurogenesis during Embryonic Cortical Development. *J Neurosci* 33, 244–258.

Turner, S.J., Hildebrand, M.S., Block, S., Damiano, J., Fahey, M., Reilly, S., Bahlo, M., Scheffer, I.E., and Morgan, A.T. (2013). Small intragenic deletion in FOXP2 associated with childhood apraxia of speech and dysarthria. *Am. J. Med. Genet.* 161, 2321–2326.

Urduingio, R.G., Fernandez, A.F., Lopez-Nieva, P., Rossi, S., Huertas, D., Kulis, M., Liu, C.-G., Croce, C.M., and Calin, G.A. (2010). Disrupted microRNA expression caused by Mecp2 loss in a mouse model of Rett syndrome. *Epigenetics* 5, 656–663.

Vallès, A., Martens, G., De Weerd, P., Poelmans, G., and Aschrafi, A. (2014). MicroRNA-137 regulates a glucocorticoid receptor–dependent signalling network: implications for the etiology of schizophrenia. *J Psychiatry Neurosci* 39, 312–320.

van BALKOM, I.D.C., VUIJK, P.J., FRANSSENS, M., HOEK, H.W., and HENNEKAM, R.C.M. (2012). Development, cognition, and behaviour in Pitt-Hopkins syndrome. *Dev Med Child Neurol* 54, 925–931.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 3, RESEARCH0034.

Vargha-Khadem, F., Watkins, K.E., Price, C.J., Ashburner, J., Alcock, K.J., Connelly, A., Frackowiak, R.S., Friston, K.J., Pembrey, M.E., Mishkin, M., et al. (1998). Neural basis of an inherited speech and language disorder. *Proc Natl Acad Sci USA* 95, 12695–12700.

Vargha-Khadem, F., Watkins, K., Alcock, K., Fletcher, P., and Passingham, R. (1995). Praxic and nonverbal cognitive deficits in a large family with a genetically

transmitted speech and language disorder. *Proc Natl Acad Sci USA* 92, 930–933.

Vargha-Khadem, F., and Passingham, R. (1990). Speech and language defects. *Nature* 346, 226.

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

Veeramah, K.R., Johnstone, L., Karafet, T.M., Wolf, D., Sprissler, R., Salogiannis, J., Barth Maron, A., Greenberg, M.E., Stuhlmann, T., Weinert, S., et al. (2013). Exome sequencing reveals new causal mutations in children with epileptic encephalopathies. *Epilepsia* 54, 1270–1281.

Vera, J., Lai, X., Schmitz, U., and Wolkenhauer, O. (2012). MicroRNA-Regulated Networks: The Perfect Storm for Classical Molecular Biology, the Ideal Scenario for Systems Biology. *Adv. Exp. Med. Biol.* 774, 55–76.

Vergoulis, T., Vlachos, I.S., Alexiou, P., Georgakilas, G., Maragkakis, M., Reczko, M., Gerangelos, S., Koziris, N., Dalamagas, T., and Hatzigeorgiou, A.G. (2011). TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. *Nucleic Acids Res* 40, D222–D229.

Vernes, S.C., Newbury, D.F., Abrahams, B.S., Winchester, L., Nicod, J., Groszer, M., Alarcón, M., Oliver, P.L., Davies, K.E., Geschwind, D.H., et al. (2008). A functional genetic link between distinct developmental language disorders. *N Engl J Med* 359, 2337–2345.

Vernes, S.C., Nicod, J., Elahi, F.M., Coventry, J.A., Kenny, N., Coupe, A.-M., Bird, L.E., Davies, K.E., and Fisher, S.E. (2006). Functional genetic analysis of mutations implicated in a human speech and language disorder. *Hum Mol Genet* 15, 3154–3167.

Vernes, S.C., Oliver, P.L., Spiteri, E., Lockstone, H.E., Puliyadi, R., Taylor, J.M., Ho, J., Mombereau, C., Brewer, A., Lowy, E., et al. (2011). Foxp2 Regulates Gene Networks Implicated in Neurite Outgrowth in the Developing Brain. *PLoS Genet* 7, e1002145.

Vernes, S.C., Spiteri, E., Nicod, J., Groszer, M., Taylor, J.M., Davies, K.E., Geschwind, D.H., and Fisher, S.E. (2007). High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders. *Am J Hum Genet* 81, 1232–1250.

Villanueva, P., Newbury, D.F., Jara, L., Barbieri, Z.D., Mirza, G., Palomino, H.A.N.M., Fern Aacute Ndez, M.I.A.A.E.L., Cazier, J.-B., Monaco, A.P., and Palomino, H.A.N. (2011). Genome-wide analysis of genetic susceptibility to language impairment in an isolated Chilean population. *European Journal of Human Genetics* 19, 687–695.

Visvanathan, J., Lee, S., Lee, B., Lee, J.W., and Lee, S.K. (2007). The microRNA miR-124 antagonizes the anti-neural REST/SCP1 pathway during embryonic CNS development. *Genes & Development* *21*, 744–749.

Vlachos, I.S., Kostoulas, N., Vergoulis, T., Georgakilas, G., Reczko, M., Maragkakis, M., Paraskevopoulou, M.D., Prionidis, K., Dalamagas, T., and Hatzigeorgiou, A.G. (2012). DIANA miRPath v.2.0: investigating the combinatorial effect of microRNAs in pathways. *Nucleic Acids Res* *40*, W498–W504.

Vo, N., Klein, M.E., Varlamova, O., Keller, D.M., Yamamoto, T., Goodman, R.H., and Impey, S. (2005). A cAMP-response element binding protein-induced microRNA regulates neuronal morphogenesis. *Proc Natl Acad Sci USA* *102*, 16426–16431.

Wahl, M., Marzinzik, F., Friederici, A.D., Hahne, A., Kupsch, A., Schneider, G.-H., Saddy, D., Curio, G., and Klostermann, F. (2008). The Human Thalamus Processes Syntactic and Semantic Language Violations. *Neuron* *59*, 695–707.

Walker, D.M., Urbé, S., Dove, S.K., Tenza, D., Raposo, G., and Clague, M.J. (2001). Characterization of MTMR3, an inositol lipid 3-phosphatase with novel substrate specificity. *Curr Biol* *11*, 1600–1605.

Walker, R.M., Hill, A.E., Newman, A.C., Hamilton, G., Torrance, H.S., Anderson, S.M., Ogawa, F., Derizioti, P., Nicod, J., Vernes, S.C., et al. (2012). The DISC1 promoter: characterization and regulation by FOXP2. *Hum Mol Genet* *21*, 2862–2872.

Wang, B., Lin, D., Li, C., and Tucker, P. (2003). Multiple Domains Define the Expression and Regulatory Properties of Foxp1 Forkhead Transcriptional Repressors. *Journal of Biological Chemistry* *278*, 24259–24268.

Wang, G., van der Walt, J.M., Mayhew, G., Li, Y.-J., Züchner, S., Scott, W.K., Martin, E.R., and Vance, J.M. (2008). Variation in the miRNA-433 binding site of FGF20 confers risk for Parkinson disease by overexpression of alpha-synuclein. *Am J Hum Genet* *82*, 283–289.

Wang, H., Li, Y., Wang, Y., Han, Z.-G., and Cai, B. (2012). C9orf100, a new member of the Dbl-family guanine nucleotide exchange factors, promotes cell proliferation and migration in hepatocellular carcinoma. *Mol Med Report* *5*, 1169–1174.

Wang, J., Duncan, D., Shi, Z., and Zhang, B. (2013). WEB-based GENE SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res* *41*, W77–W83.

Wang, J.L.J., Yang, X.X., Xia, K.K., Hu, Z.M.Z., Weng, L.L., Jin, X.X., Jiang, H.H., Zhang, P.P., Shen, L.L., Guo, J.F.J., et al. (2010). TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. *Brain* *133*, 3510–3518.

- Warburton, A., Breen, G., Rujescu, D., Bubbs, V.J., and Quinn, J.P. (2014). Characterization of a REST-Regulated Internal Promoter in the Schizophrenia Genome-Wide Associated Gene MIR137. *Schizophrenia Bulletin*.
- Warren, W.C., Clayton, D.F., Ellegren, H., Arnold, A.P., Hillier, L.W., Künstner, A., Searle, S., White, S., Vilella, A.J., Fairley, S., et al. (2010). The genome of a songbird. *Nature* *464*, 757–762.
- Watkins, K.E., Dronkers, N.F., and Vargha-Khadem, F. (2002a). Behavioural analysis of an inherited speech and language disorder: comparison with acquired aphasia. *Brain* *125*, 452–464.
- Watkins, K.E., Vargha-Khadem, F., Ashburner, J., Passingham, R.E., Connelly, A., Friston, K.J., Frackowiak, R.S.J., Mishkin, M., and Gadian, D.G. (2002b). MRI analysis of an inherited speech and language disorder: structural brain abnormalities. *Brain* *125*, 465–478.
- Wayman, G.A., Davare, M., Ando, H., Fortin, D., Varlamova, O., Cheng, H.-Y.M., Marks, D., Obrietan, K., Soderling, T.R., Goodman, R.H., et al. (2008). An activity-regulated microRNA controls dendritic plasticity by down-regulating p250GAP. *Proc Natl Acad Sci USA* *105*, 9093–9098.
- Weigel, D., JURGENS, G., KUTTNER, F., SEIFERT, E., and JACKLE, H. (1989). The Homeotic Gene Fork Head Encodes a Nuclear-Protein and Is Expressed in the Terminal Regions of the Drosophila Embryo. *Cell* *57*, 645–658.
- Wernicke, C. (1874). The aphasic symptom complex: a psychological study on a neurological basis.
- Whalley, H.C., Pappmeyer, M., Romaniuk, L., Sprooten, E., Johnstone, E.C., Hall, J., Lawrie, S.M., Evans, K.L., Blumberg, H.P., Sussmann, J.E., et al. (2012). Impact of a microRNA MIR137 Susceptibility Variant on Brain Function in People at High Genetic Risk of Schizophrenia or Bipolar Disorder. *37*, 2720–2729.
- White, R.E., and Giffard, R.G. (2012). MicroRNA-320 induces neurite outgrowth by targeting ARPP-1. *Neuroreport* *23*, 590–595.
- White, S.A. (2010). Brain & Language. *Brain Lang* *115*, 21–28.
- Whitehouse, A.J.O., Bishop, D.V.M., Ang, Q.W., Pennell, C.E., and Fisher, S.E. (2011). CNTNAP2 variants affect early language development in the general population. *Genes, Brain and Behavior* *10*, 451–456.
- Wibbrand, K., Panja, D., Tiron, A., Ofte, M.L., Skaftnesmo, K.-O., Lee, C.S., Pena, J.T., Tuschl, T., and Bramham, C.R. (2010). Differential regulation of mature and precursor microRNA expression by NMDA and metabotropic glutamate receptor activation during LTP in the adult dentate gyrus in vivo. *European Journal of Neuroscience* *31*, 636–645.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* 75, 855–862.

Willemsen, M.H., Vallès, A., Kirkels, L.A.M.H., Mastebroek, M., Olde Loohuis, N., Kos, A., Wissink-Lindhout, W.M., de Brouwer, A.P.M., Nillesen, W.M., Pfundt, R., et al. (2011). Chromosome 1p21.3 microdeletions comprising *DPYD* and *MIR137* are associated with intellectual disability. *Journal of Medical Genetics* 48, 810–818.

Wilson, S.M., Saygin, A.P., Sereno, M.I., and Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Publishing Group* 7, 701–702.

Wirgenes, K.V., nderby, I.E.S.O., Haukvik, U.K., Mattingdal, M., Tesli, M., Athanasiu, L., Sundet, K., ssberg, J.I.R.O., Dale, A.M., Brown, A.A., et al. (2012). *TCF4* sequence variants and mRNA levels are associated with neurodevelopmental characteristics in psychotic disorders. *Translational Psychiatry* 2, e112–e119.

Wisniewski, W., Hunter, J.V., Hanchard, N.A., Willer, J.R., Shaw, C., Tian, Q., Illner, A., Wang, X., Cheung, S.W., Patel, A., et al. (2013). *TM4SF20* Ancestral Deletion and Susceptibility to a Pediatric Disorder of Early Language Delay and Cerebral White Matter Hyperintensities. *The American Journal of Human Genetics* 93, 197–210.

Woodcock, R.W. (1987). *Woodcock Reading Mastery Tests* (American Guidance Service).

Woods, K., Thomson, J.M., and Hammond, S.M. (2007). Direct Regulation of an Oncogenic Micro-RNA Cluster by E2F Transcription Factors. *Journal of Biological Chemistry* 282, 2130–2134.

World Health Organization (2010). *The ICD-10 classification of mental and behavioural disorders : diagnostic criteria for research* (Geneva: Geneva : World Health Organization).

Worthey, E.A., Raca, G., Laffin, J.J., Wilk, B.M., Harris, J.M., Jakielski, K.J., Dimmock, D.P., Strand, E.A., and Shriberg, L.D. (2013). Whole-exome sequencing supports genetic heterogeneity in childhood apraxia of speech. *J Neurodevelop Disord* 5, 29–29.

Wu, L., Fan, J., and Belasco, J.G. (2006a). MicroRNAs direct rapid deadenylation of mRNA. *Proc Natl Acad Sci USA* 103, 4034–4039.

Wu, Y., Borde, M., Heissmeyer, V., Feuerer, M., Lapan, A.D., Stroud, J.C., Bates, D.L., Guo, L., Han, A., Ziegler, S.F., et al. (2006b). *FOXP3* Controls Regulatory T Cell Function through Cooperation with NFAT. *Cell* 126, 375–387.

Yang, J.-H., Li, J.-H., Jiang, S., Zhou, H., and Qu, L.-H. (2013). ChIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data. *Nucleic Acids Res* 41, D177–D187.

Yeom, K.-H., Lee, Y., Han, J., Suh, M.R., and Kim, V.N. (2005). Characterization of DGCR8/Pasha, the essential cofactor for Drosha in primary miRNA processing. *Nucleic Acids Res* 34, 4622–4629.

Yi, R., Qin, Y., Macara, I.G., and Cullen, B.R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes & Development* 17, 3011–3016.

Zeesman, S., Nowaczyk, M.J.M., Teshima, I., Roberts, W., Cardy, J.O., Brian, J., Senman, L., Feuk, L., Osborne, L.R., and Scherer, S.W. (2006). Speech and language impairment and oromotor dyspraxia due to deletion of 7q31 that involves FOXP2. *Am. J. Med. Genet.* 140A, 509–514.

Zeng, Y., Yi, R., and Cullen, B.R. (2003). MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc Natl Acad Sci USA* 100, 9779–9784.

Zhang, H.M., Kuang, S., Xiong, X., Gao, T., Liu, C., and Guo, A.Y. (2013a). Transcription factor and microRNA co-regulatory loops: important regulatory motifs in biological processes and diseases. *Briefings in Bioinformatics*.

Zhang, H., Kolb, F.A., Brondani, V., Billy, E., and Filipowicz, W. (2002a). Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *Embo J* 21, 5875–5885.

Zhang, J., Webb, D.M., and Podlaha, O. (2002b). Accelerated protein evolution and origins of human-specific features: Foxp2 as an example. *Genetics* 162, 1825–1835.

Zhang, Y., Ueno, Y., Liu, X.S., Buller, B., Wang, X., Chopp, M., and Zhang, Z.G. (2013b). The MicroRNA-17-92 Cluster Enhances Axonal Outgrowth in Embryonic Cortical Neurons. *J Neurosci* 33, 6885–6894.

Zhao, C., Sun, G., Li, S., and Shi, Y. (2009). A feedback regulatory loop involving microRNA-9 and nuclear receptor TLX in neural stem cell fate determination. *Nat Struct Mol Biol* 16, 365–371.

Zhao, C., Sun, G., Ye, P., Li, S., and Shi, Y. (2013). MicroRNA let-7d regulates the TLX/microRNA-9 cascade to control neural cell fate and neurogenesis. *Sci. Rep.* 3, 1329–1329.

Zhou, B., Zhong, Q., Minoo, P., Li, C., Ann, D.K., Frenkel, B., Morrissey, E.E., Crandall, E.D., and Borok, Z. (2008). Foxp2 Inhibits Nkx2.1-Mediated Transcription of SP-C via Interactions with the Nkx2.1 Homeodomain. *Am J*

Respir Cell Mol Biol 38, 750–758.

Zhou, P., Xu, W., Peng, X., Luo, Z., Xing, Q., Chen, X., Hou, C., Liang, W., Zhou, J., Wu, X., et al. (2013). Large-scale screens of miRNA-mRNA interactions unveiled that the 3'UTR of a gene is targeted by multiple miRNAs. *PLoS ONE* 8, e68204.

Zhou, R., Yuan, P., Wang, Y., Hunsberger, J.G., Elkahoul, A., Wei, Y., Damschroder-Williams, P., Du, J., Chen, G., and Manji, H.K. (2009). Evidence for selective microRNAs and their effectors as common long-term targets for the actions of mood stabilizers. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology* 34, 1395–1405.

Zhu, P.J., Huang, W., Kalikulov, D., Yoo, J.W., Placzek, A.N., Stoica, L., Zhou, H., Bell, J.C., Friedlander, M.J., Krnjević, K., et al. (2011). Suppression of PKR promotes network excitability and enhanced cognition by interferon-γ-mediated disinhibition. *Cell* 147, 1384–1396.

Zhu, Y., Kalbfleisch, T., Brennan, M.D., and Li, Y. (2009). A MicroRNA gene is hosted in an intron of a schizophrenia-susceptibility gene. *Schizophrenia Research* 109, 86–89.

Zongaro, S., Hukema, R., D'Antoni, S., Davidovic, L., Barbry, P., Catania, M.V., Willemsen, R., Mari, B., and Bardoni, B. (2013). The 3' UTR of FMR1 mRNA is a target of miR-101, miR-129-5p and miR-221: implications for the molecular pathology of FXTAS at the synapse. *Hum Mol Genet* 22, 1971–1982.

Zorc, M., Skok, D.J., Godnic, I., Calin, G.A., Horvat, S., Jiang, Z., Dovc, P., and Kunej, T. (2012). Catalog of microRNA seed polymorphisms in vertebrates. *PLoS ONE* 7, e30737.

Zöllner, S., and Pritchard, J.K. (2007). Overcoming the Winner's Curse: Estimating Penetrance Parameters from Case-Control Data. *The American Journal of Human Genetics* 80, 605–615.

Zweier, C., Jong, E.K. de, Zweier, M., Orrico, A., Ousager, L.B., Collins, A.L., Bijlsma, E.K., Oortveld, M.A.W., Ekici, A.B., Reis, A., et al. (2009). CNTNAP2 and NRXN1 Are Mutated in Autosomal-Recessive Pitt-Hopkins-like Mental Retardation and Determine the Level of a Common Synaptic Protein in *Drosophila*. *The American Journal of Human Genetics* 85, 655–666.