



THE GENETICS OF HANDEDNESS AND DYSLEXIA

WILLIAM BRANDLER

Merton College

MRC Functional Genomics Unit

Department of Physiology, Anatomy, and Genetics

University of Oxford

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Trinity Term 2013

Abstract

The population level bias towards right-handedness in humans implies left-hemisphere dominance for fine motor control. Left-handedness and reduced cerebral asymmetry have been linked to neurodevelopmental disorders such as dyslexia. Understanding the biology of these traits at a genetic level is crucial for understanding the relationship between handedness and neurodevelopmental disorders.

Here I present genome-wide association study (GWAS) meta-analyses for both relative hand skill (handedness, $n = 728$) and reading-related traits ($n = 548$) in individuals with dyslexia. I uncovered a genome-wide significant association in an intron of *PCSK6* associated with relative hand skill. *PCSK6* is a protease that cleaves NODAL proprotein into an active form, and NODAL determines the development of left/right (LR) asymmetry in bilaterians. I performed pathway analyses of the GWAS data that revealed handedness is determined in part by the mechanisms that establish left/right (LR) asymmetry early in development, such as NODAL signalling and ciliogenesis. This finding replicated in a general population cohort unaffected with neurodevelopmental disorders ($n = 2,666$).

A key stage in LR asymmetry development is the rotation of cilia that creates a leftward flow of NODAL. Candidate genes for dyslexia are involved in both neuronal migration and ciliogenesis. Ciliopathies can cause not only LR body asymmetry phenotypes, but also cerebral midline phenotypes such as an absent corpus callosum. Furthermore, I identified a genome-wide significant association with non-word reading located in an intron of *MAP1B*, a gene involved in neuronal migration that causes an absent corpus callosum when disrupted in mice. However,

this finding did not replicate in two independent cohorts with dyslexia ($n = 156$ & 199), or in the general population cohort ($n = 2,359$). Though these cohorts had inadequate reading measures and poorly matched ascertainment for dyslexia.

I also performed copy number variation (CNV) pathway and burden analyses of 920 individuals with dyslexia and 1,366 unselected controls, but did not find that rare CNVs play a major role in the etiology of dyslexia.

Based on these results I propose that common variants in genes responsible for cilogenesis and corpus callosum development influence traits such as handedness and reading ability.

word count: 36,500

Acknowledgements

I'd like to thank Tony Monaco for answering my email on Christmas Eve 2008 saying he would be interested in considering me for a PhD. The Nuffield Department of Medicine Studentship for funding me these four years, and the panel of interviewers who offered me for the PhD scholarship in spite of the lack of direct clinical application of my research. Caleb Webber, who took me on to join his lab at a time after Tony had left Oxford and I was lacking support and at a loss about where my PhD project was going. Andrew Morris, my other official supervisor who taught me how not to crash the servers (alongside how to perform GWAS). Silvia Paracchini, my unofficial supervisor who played a key role in supporting most aspects of my research. Simon Fisher for offering constructive criticism and financial support for the project. Dorothy Bishop for critically evaluating my research papers. Also everyone in the Monaco and Webber groups; in particular Tom Scerri, Julia Steinberg, Tallulah Andrews, Stephen Meader, and Frank Honti.

Further back in time I'd like to thank Paul Ridd, my biology teacher at school who ignited my interest in the subject. David Summers, my director of studies at Cambridge who showed me that even the most seemingly boring subjects (like *E. coli* plasmid replication) could be made interesting if taught in the right way. Also Tasos Pavlopoulos, my undergraduate project supervisor who taught me the value of doing the best possible job you can, especially when others are relying on you.

Finally, I'd like to thank my parents, without their support I wouldn't have come close to what I have achieved.

Contents

List of Figures	6
List of Tables	10
1 Introduction	14
1.1 Literature review - handedness	14
1.1.1 Is handedness unique to humans?	15
1.1.2 When did handedness evolve in humans?	20
1.1.3 Why did handedness evolve in humans?	21
1.1.4 Linking left-handedness with human neurodevelopmental disorders	23
1.1.5 When does handedness become apparent during human development?	28
1.1.6 The genetic architecture of handedness: non-genetic, monogenic, or polygenic?	29
1.2 Literature review - the genetics of dyslexia	31
1.2.1 Core deficit in dyslexia	31
1.2.2 The relationship between dyslexia, neurodevelopmental disorders, and handedness	32
1.2.3 Heritability of dyslexia and linkage studies	33

1.2.4	Dyslexia candidate genes, assessment of the evidence	33
1.3	Thesis structure	37
2	Dyslexia ascertainment criteria and phenotyping of study participants	40
2.1	Study participant ascertainment criteria	40
2.2	How to measure handedness	45
2.2.1	Relative hand skill phenotype	47
3	Relative hand skill GWAS	51
3.1	Abstract	51
3.2	Introduction	52
3.2.1	Genome-wide association studies	52
3.3	Methods	58
3.3.1	Genotyping and quality control	58
3.3.2	Imputation	59
3.3.3	Statistical analysis	60
3.3.4	Fine mapping	60
3.4	GWAS results	62
3.4.1	Replication	62
3.4.2	Fine mapping of <i>PCSK6</i> locus	71
3.5	Discussion	76
3.5.1	Future perspectives	77
4	Genotyping of dyslexia cases on the Illumina OmniExpress SNP array	80
4.1	Preparation of DNA	82
4.2	Per-individual quality control	83
4.2.1	Identification of individuals with discordant sex information	84
4.2.2	Genotyping success rate	84
4.2.3	Identification of duplicated or related individuals	86

4.2.4	Identification of individuals with divergent ancestry	87
4.3	Per-SNP quality control	90
4.3.1	Identification of SNPs with a significantly different genotyping call rates between cases and controls	91
4.3.2	Identification of SNPs with elevated missing genotype rates .	91
4.3.3	Identification of SNPs with a significant deviation from Hardy-Weinberg Equilibrium	91
4.3.4	Removal of SNPs with a low minor allele frequency	93
5	Genome-wide association meta-analysis for relative hand skill	95
5.1	Abstract	95
5.2	Introduction	96
5.3	Methods	96
5.3.1	Genotyping and quality control	96
5.3.2	Imputation	97
5.3.3	Statistical analysis	98
5.3.4	Gene-set enrichment analysis	99
5.3.5	Genome-wide Complex Trait Analysis (GCTA)	100
5.4	Results	102
5.4.1	Relative hand skill phenotype	103
5.4.2	Relative hand skill GWA meta-analysis in individuals with reading disability	104
5.4.3	Relative hand skill GWAS in the ALSPAC general population	110
5.4.4	Gene-set enrichment analysis	113
5.4.5	Genome-wide complex trait analysis (GCTA)	120
5.5	Discussion	120
6	Genome-wide association study for dyslexia and reading-related traits	127

6.1	Abstract	127
6.2	Introduction	128
6.3	Methods	129
6.3.1	Cohorts	129
6.3.2	Genotyping, imputation, and statistical analysis	129
6.3.3	Principle component analysis	130
6.3.4	Reclustering of SNPs for cases and controls on the Illumina OmniExpress array	131
6.4	Results	134
6.4.1	Illumina 550K	134
6.4.2	Illumina OmniExpress case/control GWAS	134
6.5	Imputation	137
6.6	Meta-analysis	138
6.7	Quantitative reading-related traits GWAS	140
6.7.1	GWAS replication	149
6.7.2	Principle component analysis	150
6.7.3	Multivariate analysis	156
6.7.4	GSEA	159
6.8	Discussion	162
6.8.1	Case / control analysis	162
6.8.2	Nonword reading GWAS	163
6.8.3	Principle component analysis	165
6.8.4	Multivariate GWAS analysis	166
6.8.5	GSEA	166
7	CNV calling and quality control	168
7.1	Introduction	168
7.2	Methods	170

7.2.1	Cohorts and genotyping	170
7.3	Comparison of individuals genotyped on more than one SNP array .	173
7.4	Comparison of the CNV calling algorithms	179
7.5	CNV calling criteria	182
8	CNV analysis	184
8.1	Abstract	184
8.2	Introduction	185
8.3	Methods & Results	187
8.3.1	Case / control burden	187
8.3.2	Linear regression	190
8.3.3	CNVs that intersect known dyslexia genes	192
8.3.4	Pathway analysis using mouse phenotypes	194
8.4	Discussion	198
8.4.1	Limitations of the study design	199
8.4.2	Conclusions	200
9	Concluding remarks & future perspectives	202
10	Bibliography	208
11	Appendix	229
11.1	Fine mapping of <i>PCSK6</i> locus	229
11.2	Samples that failed QC	232
11.3	Most highly associated markers in each GWAS	235
11.4	Published papers	247

List of Figures

1.1	Flow chart outlining study design	39
2.1	PegQ distribution	49
2.2	Histograms and QQ plots of PegQ	50
3.1	Manhattan and Locus Zoom plot for relative hand skill GWAS of cohort 1	63
3.2	Forest plot for relative hand skill GWA meta-analysis	65
3.3	ECDF of PegQ in ALSPAC cohorts	67
3.4	QQ plot for relative hand skill GWAS of cohort 1	67
3.5	UCSC genome browser view of <i>PCSK6</i> ENCODE regulatory regions	74
3.6	UCSC genome browser view of secondary promoter of <i>PCSK6</i>	75
3.7	NODAL signalling pathway	78
4.1	Genotype failure rate versus heterozygosity	85
4.2	Example B allele frequency plots	86
4.3	Histogram of pairwise IBD metric	88
4.4	Principal component analysis of the GWAS data	89
4.5	SNP call rate histogram	92
5.1	Description of Meta-Analysis Gene-set Enrichment of variaNT Associations (MAGENTA) method	101

5.2	QQ plot for RD relative hand skill GWA meta-analysis	105
5.3	Manhattan and Locus Zoom plot for RD relative hand skill GWA meta-analysis	106
5.4	Venn diagram comparing GSEA for reading disabled cohorts and general population	115
6.1	rs4802078 genotype clusters	133
6.2	QQ plots of Illumina OmniExpress dyslexia case-control GWAS . . .	134
6.3	Principal component analysis of cases and controls	135
6.4	QQ and Manhattan plot of Illumina 550K dyslexia case-control GWAS	136
6.5	Manhattan plot of Illumina OmniExpress dyslexia case-control GWAS	137
6.6	QQ and Manhattan plot for the dyslexia case-control GWA meta- analysis	139
6.7	Histograms of reading-related traits	142
6.8	QQ plots of reading-related traits	143
6.9	QQ plots of the quantitative reading GWAS	144
6.10	Manhattan plots of the quantitative reading GWAS	145
6.11	<i>MAP1B</i> is associated with non-word reading in individuals with dyslexia	146
6.12	Conditioning on rs16875959 reveals there is only one variant driving association with non-word reading	147
6.13	Matrix scatter plot showing correlation between reading-related traits	152
6.14	Scatter plot of first two principle components of reading-related traits	153
6.15	QQ and Manhattan plot of first principle component of reading- related traits	154
7.1	Visualisation of high confidence CNV calls	172

7.2	Comparison of CNV calling in individuals genotyped on more than one array	175
7.3	Venn diagram of CNV calls in individuals genotyped on both the 550K and 300K arrays	176
7.4	Venn diagram of CNV calls in individuals genotyped on both the OmniExpress and 550K arrays	176
7.5	Comparison of size of overlapping CNV calls in individuals genotyped on both the 550K and 300K arrays	177
7.6	Comparison of size of overlapping CNV calls in individuals genotyped on both the OmniExpress and 550K arrays	178
7.7	Comparison of PennCNV and QuantiSNP CNV calls	180
7.8	Venn diagram of CNV calls using PennCNV and QuantiSNP	181
7.9	Comparison of size of overlapping CNV calls from PennCNV and QuantiSNP	182
8.1	Visualisation of the case-control burden analysis on the 550K array .	189
8.2	Visualisation of the case-control burden analysis on the OmniExpress array	189
8.3	The burden of CNVs in the OmniExpress cases is driven entirely by duplications	190
8.4	Correlation between reading ability and both CNV length and number of genes disrupted in individuals with dyslexia - at a CNV confidence level of 30	191
8.5	Correlation between reading ability and both CNV length and number of genes disrupted in individuals with dyslexia - at a CNV confidence level of 10	191
9.1	Establishment of LR asymmetry during development	206

9.2	Cilia and the biology of handedness and dyslexia	207
-----	--	-----

List of Tables

1.1	Handedness in other species	20
1.2	Summary of cohorts utilised in the project	38
2.1	Correlation between reading related measures, IQ, and PegQ	43
2.2	Correlation between quantitative measures of relative hand skill and dexterity	46
2.3	Skewness and kurtosis of PegQ scores	48
3.1	Association statistics for the most highly associated SNP in <i>PCSK6</i>	66
3.2	Most strongly associated SNPs in cohort 1 relative hand skill GWAS	68
3.3	<i>PCSK6</i> is not associated with traits other than PegQ	70
3.4	<i>PCSK6</i> VNTR is not associated with PegQ	73
4.1	Summary of individuals genotyped on the Illumina OmniExpress chip	82
4.2	Number of individuals that failed genotyping QC	90
4.3	SNPs that failed genotyping QC	94
5.1	GWAS design	102
5.2	Summary of ALSPAC PegQ and motor skills scores	104
5.3	Association of rs7182874 with relative hand skill	107

5.4	Top SNPs in PegQ GWA meta-analysis for individuals with reading disability	108
5.5	Top SNPs in ALSPAC general population PegQ GWA meta-analysis	111
5.6	LR asymmetry gene-set enrichment analysis of the relative hand skill GWA meta-analysis of individuals with reading disability	116
5.7	SNPs in LR asymmetry genes are associated with relative hand skill .	117
5.8	After removing <i>PCSK6</i> , the LR asymmetry GSEA of the RD relative hand skill GWAS is still significant	118
5.9	Three of the four significant LR phenotypes from the GSEA replicate in the the general population relative hand skill GWAS	119
5.10	GSEA for ciliopathy associated brain phenotypes in RD cohorts . . .	119
5.11	GSEA for ciliopathy associated brain phenotypes in general population cohort	119
6.1	Summary of λ for the quantitative reading measures GWAS	130
6.2	False positive associations from the Illumina OmniExpress dyslexia GWAS	132
6.3	Summary of λ for the dyslexia case-control GWAS	138
6.4	Summary of quantitative reading scores	141
6.5	Skewness and kurtosis of reading-related scores	141
6.6	Top SNPs from CCN GWA meta-analysis	148
6.7	Association of the SNP rs16875959 in <i>MAP1B</i> with phenotypes other than non-word reading	149
6.8	rs16875959 – associated with non-word reading, does not replicate in follow up cohorts	150
6.9	rs11594922 – associated with non-word reading, does not replicate in follow up cohorts	151
6.10	PCA of reading-related traits	151

6.11	Top SNPs from first principle component of reading-related traits	
	GWA meta-analysis	155
6.12	Reading-related traits multivariate GWAS analysis results	157
6.13	Abnormal ciliary body morphology is associated in the GSEA of the	
	irregular word reading (CCI) GWAS	160
6.14	Abnormal axon guidance is associated in the GSEA of the single-	
	word reading GWAS	161
7.1	Summary of CNV Cohorts	170
7.2	Descriptive statistics for individual CNV quality control	173
8.1	CNV burden analysis for the cases and controls genotyped on the	
	550K array	188
8.2	CNV burden analysis for the cases and controls genotyped on the	
	OmniExpress array	188
8.3	CNVs that intersect dyslexia candidate genes	193
8.4	Enrichment of genes affected by CNVs in individuals with dyslexia	
	whose mouse orthologues, when disrupted, yield an increased startle	
	reflex phenotype	196
8.5	Genes intersected by CNVs that are annotated to the phenotype in-	
	creased startle reflex	196
11.1	<i>PCSK6</i> coding SNPs identified by sequencing	229
11.2	<i>PCSK6</i> secondary promoter SNPs identified by sequencing	230
11.3	Cell line genotyping of <i>PCSK6</i> SNPs	231
11.4	List of 550K cases that failed quality control	232
11.5	List of WTCCC controls that failed quality control	233
11.6	List of OmniExpress cases and siblings that failed quality control	234
11.7	List of POBI OmniExpress controls that failed quality control	235

11.8	Top SNPs from the 550K dyslexia GWAS	236
11.9	Top SNPs from the Illumina OmniExpress dyslexia GWAS	238
11.10	Top 25 SNPs from the 550K case-control GWAS after imputation . . .	241
11.11	Top 25 SNPs from the OmniExpress case-control GWAS after imputation	242
11.12	Top SNPs from the dyslexia case-control GWA meta-analysis	243
11.13	Top SNPs from CCI GWA meta-analysis	244
11.14	Top SNPs from reading GWA meta-analysis	245
11.15	Top SNPs from spelling GWA meta-analysis	245

Chapter 1

Introduction

1.1 Literature review - handedness

Humans display such striking structural and functional cerebral asymmetries that we have been labeled, 'The Lopsided Ape' (Corballis, 1993). The population majority level of right-handedness across all human societies measured implies left-hemisphere dominance for controlling hand preference (Marchant and McGrew, 1998; Raymond and Pontier, 2004). This bias suggests there is some advantage to being right-handed, however if that is the case then why isn't everyone? On the other hand, given that our hands are symmetrical, why isn't the ratio closer to 50:50, such as it is in chimpanzees (subsection **section 1.1.1**)?

Studies of preliterate societies have shown that for most manual tasks the prevalence of right-handedness is only 55:45, but when we consider precision-grip tool or fine motor skill use right-handedness predominates (Marchant and McGrew, 1998). In traditional societies, right-handedness for precision-grip tool use varies in one study from 73.1–96.7% (Faurie et al., 2005). Similarly, the geographical varia-

tion of right-handedness for throwing and hammering varies from 72.5–95% (Raymond and Pontier, 2004). On average men are $1.23\times$ more likely to be left-handed than women (Papadatou-Pastou et al., 2008), and there are no known societies in which left-handedness predominates.

Handedness correlates with both footedness and eyedness. Right-footedness when playing soccer has a prevalence of 75-80%, with 89% of right-handers preferring their right foot, and 66% of left-handers preferring their left-foot (Carey et al., 2009). 66% of individuals have a right-eye preference, with 65% of right-handers and 43% of left-handers preferring their right eye (Bourassa et al., 1996). Lateralisation is therefore less strong for eye preference than it is for footedness or handedness tasks involving fine motor control use.

We want to understand why most, but not all, humans are right-handed for fine motor control. In order to begin to get a handle on this problem we first need to understand; the prevalence of handedness across the animal kingdom (**subsection 1.1.1**), when handedness evolved (**subsection 1.1.2**), why handedness evolved (**subsection 1.1.3**), if there are any disadvantages associated with being left-handed (**subsection 1.1.4**), at what stage during human development handedness becomes apparent (**subsection 1.1.5**), and finally what the genetic basis of handedness is (**subsection 1.1.6, chapter 3 & chapter 5**).

1.1.1 Is handedness unique to humans?

If we know when handedness evolved and what other species we share it with, then we can begin to narrow down the cause of its evolution. In a review of lateralised behaviour in published studies of 119 different species, 61 (51%) show ev-

idence for population-level asymmetries, 20 (17%) show evidence for individual-level asymmetries and 38 (32%) show no evidence for asymmetry (Strockens et al., 2012). Lateralisation appears widespread, seen in arthropods, fish, amphibians, reptiles, birds, and mammals. However, this review has collated all published studies without considering the quality of the evidence. For example, many of the studies that find a population level asymmetry tested fewer than 50 animals and have not been independently replicated (for examples see **Table 1.1**). Furthermore, publication bias will inflate the number of species that show asymmetry relative to those that do not (see **Box 1** on page 26).

However, in spite of these limitations there are some clear examples of behavioural lateralisation in both extinct and extant species. Examination of healed injuries from predators on Cambrian and post-Cambrian fossil trilobites finds that the majority are located on the right side (Babcock, 1993; Babcock and Robison, 1989), suggesting behavioural lateralisation of either trilobites, or their predators, dates back 500 million years. Furthermore spitting spiders (*Scytodes globula*) collected in the field that have lost one leg show a greater loss of left legs than right legs, which they often lose when attacked by prey (Ades and Ramires, 2002). When tested under laboratory conditions, the spiders prefer to touch prey with left leg I than right leg I (Ades and Ramires, 2002). Bees respond better to odours when presented to their right antenna over their left (Letzkus et al., 2006). Similarly bees learn visual stimuli better when presented to the right eye (Letzkus et al., 2008). The examples suggest that behavioural lateralisation is ancient and used for a number of different purposes in arthropods.

Turning to vertebrates, at least three species of the teleost poeciliid fish show pref-

erential turning to the left when avoiding predators and approaching members of the opposite sex, while two species prefers to turn right (Bisazza et al., 1997, 1998; Cantalupo et al., 1995). The heritability of predator approach is ≥ 0.5 in these fish (Bisazza et al., 2000). Zebrafish prefer to approach from the right (Miklosi and Andrew, 1999). In the zebrafish frequent situs inversus mutant (*fsi*, which causes a complete reversal of organ asymmetry), this pattern of behaviour switches (although other lateralised behaviours do not), suggesting lateralisation of behaviour is at least partially under genetic control in fish (Barth et al., 2005).

Rogers and Workman have hypothesised that lateralised behaviour evolved in land animals as a specialisation for object manipulation. They compared foot preference in domestic chicks, which use their feet to rake the ground when feeding, to budgerigars, which do not use their feet in feeding, or to manipulate any object (Rogers and Workman, 1993). Individuals from the two species were given the task of removing a piece of tape from their beaks. Chicks show a preference for the right foot, whereas budgerigars show equal use of the left and right foot (Rogers and Workman, 1993). Another study of chicks confirms this (Tommasi and Vallortigara, 1999). Similarly in mammals, wallabies show left hand preference during bipedal feeding, and a right hand preference for supporting the body in a tripedal stance, possibly to keep the left hand free for feeding or manipulating objects (Giljov et al., 2012). Male dogs preferentially use their left paw, whereas females prefer their right paw, and this finding replicates in a study assessing the paw dogs use for removing adhesive paper from their snouts (Quaranta et al., 2004; Wells, 2003).

An alternative hypothesis to object manipulation is that handedness evolved be-

cause compartmentalisation of the brain was a necessary precursor for language evolution (Corballis, 1993). In support of this the prevalence of left-footedness when reaching for objects among parrots that are kept as pets is 64% (Snyder and Harris, 1997). Although parrots do not have language like humans, they are excellent at imitating sounds such as human speech, and it may be this attribute that necessitated lateralisation of the brain. However, the African grey parrots from this study ($n = 70$), which have the largest lexicon of any species tested, display an equal prevalence of left- and right-footedness (Snyder and Harris, 1997). Furthermore, rats in laboratory conditions also show a right paw preference for reaching (Guven et al., 2003; Pence, 2002). However, when rats are exposed to novel environments, there is a leftward shift in handedness such that the prevalence is much closer to 50:50, suggesting the bias is artificially caused by being reared under unstimulating laboratory conditions (Tang and Verstynen, 2002), a similar process may be occurring parrots.

A population level bias in limb preference or turning direction occurs throughout the animal kingdom. However, it is not common and no species is as strongly lateralised as humans (**Table 1.1**). The wide variety of behaviours for which lateralised behaviour is present suggests it has evolved independently many times for many purposes. Furthermore these species have largely been tested in captivity, are distantly related to humans, and do not engage in precision-grip tool use.

Handedness in the great apes

Great apes, on the other hand, are closely related to humans, use tools, and have been studied both in captivity and in the wild.

Hopkins performed a meta-analysis of hand preference from 96 studies of both wild and captive great apes including; 1,044 chimpanzees, 97 bonobos, 280 gorillas, and 103 orangutans ($n = 1,524$). Hand preference is measured in different ways across the studies, from simple reaching to tool use, although many studies collected more than one measure. He finds a significant population level bias in right-handedness for chimpanzees and bonobos, but not gorillas or orang-utans (Hopkins, 2006). Overall, the frequency of right-handedness in chimpanzees is 54%. There is no difference observed between the sexes. Captive apes are significantly more right-handed than wild apes, suggesting the results may be inflated by captive apes exposed to right-handed human handlers.

Either way, the ratio of 54:46 of right-handedness among chimpanzees, though significant, is trivial when compared to the 90:10 ratio observed in humans. Furthermore the results may be influenced in part by publication bias, interaction with right-handed humans, and observational bias of humans. Recent reviews conclude that a right-sided bias in behaviour or bone morphology of chimpanzees is very weak or absent (Cashmore et al., 2008; Fitch and Braccini, 2013; Marchant and McGrew, 2013; Stock et al., 2013). Humans are significantly more right-handed than apes, and although the population bias may have started before the last common ancestor of humans and chimpanzees, handedness has largely evolved along the human lineage.

Table 1.1: **Selected examples of handedness in other species.** Only species that show a significant population level bias and have data on preferred side are included.

<i>species</i>	<i>measure</i>	<i>n</i>	Preferred side / %			<i>citation(s)</i>
			<i>left</i>	<i>right</i>	<i>none</i>	
trilobites	healed predation injury	81	27	69	4	(Babcock and Robison, 1989)
spitting spiders	limb loss	36	75	25	-	(Ades and Ramires, 2002)
spitting spiders	prey touching	22	60	40	-	(Ades and Ramires, 2002)
poeciliid fish	escape direction from predator	29	72	28	-	(Cantalupo et al., 1995)
green toads	feeding	11	60	40	-	(Sovrano, 2007)
common toads	removing object from snout	70	20	57	23	(Bisazza et al., 1996)
leatherback turtles	flipperedness	361	46	54	-	(Sieg et al., 2010)
tortoises	righting behaviour	34	18	53	29	(Stancher et al., 2006)
chicks	removing object from beak	38	33	67	-	(Rogers and Workman, 1993)
chicks	ground scratching	24	40	60	-	(Tommasi and Vallortigara, 1999)
parrots	foot preference	524	36	64	-	(Snyder and Harris, 1997)
wallabies	bipedal feeding	27	74	7	19	(Giljov et al., 2012)
wallabies	tripedal stance	27	11	52	37	(Giljov et al., 2012)
rats	foot preference	258	20	72	8	(Guven et al., 2003; Pence, 2002)
bonobos	hand preference (various)	97	42	58	-	(Hopkins, 2006)
chimpanzees	hand preference (various)	1,044	46	54	-	(Hopkins, 2006)

1.1.2 When did handedness evolve in humans?

Archaeological data can help us understand the question of when handedness evolved along the human lineage. In the caves of Western Europe there are examples of humans printing negative hands on the walls by blowing pigments onto their own hand applied on the rock, through a tube held in the other hand. These images date back to the Upper Palaeolithic ($\sim 35,000$ -10,000 years before present, for an example photograph of this practice, see <http://bit.ly/15e4M2W>). Most images are of left hands because the individuals use their right hands to hold the tube (Faurie et al., 2005). By comparing a replication experiment using French university students to that of the caves, no difference is detected between the two

proportions of left-handers, separated by over 10,000 years (Faurie et al., 2005). These results suggest the prevalence of left-handedness has not changed substantially since the Upper Palaeolithic. What about further back in time though?

Studies of fossil hominins suggest that handedness is an ancient trait. Hominins often used their teeth to clench materials while they cut them with a stone tool, analogous to how Inuit do so today (Faurie et al., 2005). The orientation of scratches on the enamel that are occasionally produced during this process can be used to assess handedness. Studies of *Homo neanderthalensis* from 32,000 (Frayer et al., 2010) and 130,000 years ago (Fox and Frayer, 1997), and *Homo heidelbergensis* from 530,000 years ago (Lozano et al., 2009), show patterns of scratches consistent with right-handedness at a level similar to modern human populations, suggesting the origins of handedness date back at least half a million years in humans (Frayer et al., 2012).

1.1.3 Why did handedness evolve in humans?

Since handedness is unique and universal to humans, it is a highly derived trait like language and bipedality are (McGrew et al., 2013). Language is also a highly lateralised trait, with 96% of aphasics have left hemisphere damage (Lovell, 1960). This led Geschwind to propose that language is one of the most striking examples of *cerebral dominance*, a trait observed in no other mammals apart from man (Geschwind, 1970). Since the left hemisphere controls the right side of the body and vice versa, humans are also left-hemisphere dominant for handedness.

The most prominent theory for the evolution of handedness is that it evolved as a consequence of the evolution of language. In particular, Corballis argues that

right-handedness could have evolved from the synchronisation of manual and facial gestures with a lateralised system of vocal production (Corballis, 2003). The hypothesis assumes that language gradually arose from a system of manual gestures, which themselves were facilitated by the advent of bipedality. However as plausible as the theory is, the argument reads like a Just-so story, an unfalsifiable narrative explanation of a biological trait.

There is correlation between handedness and cerebral dominance for language; one study used functional transcranial Doppler sonography to show that 96% of strong right-handers, compared to 73% of strong left-handers, show left-hemisphere dominance for language when performing a word-generation task ($n = 326$, $r = 0.284$, odds ratio (OR) = 8.88, (Knecht et al., 2000)). However, the compartmentalizing of these two functions in the same hemisphere could be a mere coincidence. fMRI studies are revealing that language processing involves a complex network of regions distributed throughout the brain (Demonet et al., 2005; Fisher and Marcus, 2006). For example bilateral activity is observed in certain brain regions for speech perception and semantic retrieval (Price, 2010). Furthermore the right hemisphere plays roles in appreciating the context of statements or interpreting figurative meaning (Bookheimer, 2002).

On balance although handedness and language lateralisation are correlated, a causal relationship between the two has not yet been found (Fitch and Braccini, 2013). Precisely why handedness evolved remains a mystery.

1.1.4 Linking left-handedness with human neurodevelopmental disorders

Although it is not yet clear why handedness evolved, the long evolutionary history of right-handedness among hominins suggests there is an advantage to being right-handed, but that begs the question of why isn't everyone right-handed? It is traditionally assumed that there is something wrong with being left-handed. This bias is ingrained in our language, a review of 50 Indo-European and non-Indo-European languages concludes that in almost all languages the right is thought of positively, and the left has negative connotations (Schiefenhovel, 2013). Researchers too typically hypothesize that left-handedness is associated with negative outcomes. Consequently left-handedness has been linked to all manner of disorders such as; alcoholism (Bakan, 1973), allergies and auto-immune disorders (Geschwind and Behan, 1982), autism (Colby and Parkison, 1977), cerebral palsy (Arnould et al., 2004; Lin et al., 2012; Vargha-Khadem et al., 1985), dyslexia (Geschwind and Behan, 1982; Satz and Fletcher, 1987), migraines (Geschwind and Behan, 1982), and schizophrenia (Dragovic and Hammond, 2005).

Since hand-writing preference is so easy to measure, being a simple tick-box on questionnaire (**section 2.2**), it is often included in clinical or epidemiological studies, but results are typically only published if they are significant (see **Box 1** on page 26). Consequently, many associations between handedness and disorders or traits appear to be due to publication bias, where initial small studies have shown associations that have not replicated in larger follow up studies or meta-analyses; such as with alcoholism (Poikolainen, 2000), allergies or auto-immune disorders (Bryden et al., 1994; Segalowitz et al., 1994), autism (Preslar et al., 2013), dyslexia (Bishop, 1990b; Francks et al., 2003b), and migraines (Hering, 1995; Lipscombe and

Prior, 2002; Messinger et al., 1988). The only systematic review of the relationship between left-handedness and human disorders was performed 23 years ago, and found no evidence to suggest handedness was associated with any developmental disorder (Bishop, 1990b).

Similarly Halpern and Coren famously first reported in *Nature* a study of baseball players that found left-handers lived on average 8 months shorter than right-handers (Halpern and Coren, 1988). They followed this up in a letter to the *New England Journal of Medicine* that stated that left-handers in fact have a life expectancy 9 years shorter than their dextral counterparts (Halpern and Coren, 1991), an order of magnitude greater than their previous estimate. The latter finding suggests being left-handed is almost as detrimental to your health as smoking, which reduces life expectancy by 10 years (Doll et al., 2004). However, a more comprehensive study found no association (Hicks et al., 1994). Furthermore, relaxed cultural pressure to conform to right-handedness throughout the 20th century means that left-handers are on average younger than right-handers, therefore those who have died will also be on average younger, confounding the above observation (Harris, 1993).

There are however, two disorders that are consistently associated with handedness in large studies; cerebral palsy and schizophrenia.

Paralysis or weakness of one-side of the body is common in cerebral palsy, forcing the other side to be dominant for handedness. Since it is random which side is affected, handedness is close to 50% in these patients (Arnould et al., 2004; Lin et al., 2012; Vargha-Khadem et al., 1985). This is an example of pathological handedness,

where handedness is influenced by an environmental insult, such as trauma during birth in the case of cerebral palsy, rather than genetic factors.

For schizophrenia, on the other hand, it seems more likely that its relationship with handedness is genetic, since schizophrenia is a highly heritable neurodevelopmental disorder (McGuffin et al., 1994). A meta-analysis of 3,175 individuals with schizophrenia shows that it is associated with an increased prevalence of left-handedness (odds ratio = 1.81, (Dragovic and Hammond, 2005)).

Box 1. Publication bias and the file-drawer problem in handedness research.

Handedness is very cheap and easy to measure and so is often included in studies as a 'bonus factor', even though it may have very little scientific merit for the study in question (Bishop, 1990b; Fairweather, 1976). Any significant result leads to a bonus publication, whereas non-significant results are often forgotten, gathering dust in researchers' file-drawers. This publication bias is known as the file-drawer problem (Rosenthal, 1979).

A recent example in molecular medicine was a study that linked an isoform in the Alzheimer's associated gene, Apolipoprotein E (APOE), and handedness (Bloss et al., 2010). There are two medically relevant common protein coding variants in the gene that exist in three combinations (or isoforms) - $\epsilon 2$, $\epsilon 3$, and $\epsilon 4$. Carriers of the $\epsilon 4$ isoform are more likely to develop Alzheimer's disease (Coon et al., 2007; Raber et al., 2004); carriers of the $\epsilon 2$ isoform are protected against Alzheimer's, whereas the wild-type $\epsilon 3$ carriers show typical odds. Bloss et al. found in a cohort of 147 individuals that $\epsilon 2$ carriers showed a higher prevalence of left-handedness (29.2%) relative to $\epsilon 3$ homozygotes (8.9%) and $\epsilon 4$ carriers (6.1%).

The odds of being left-handed in $\epsilon 2$ carriers compared to non-carriers was 3.88, unusually large for a genetic association study of common variants; which typically have odds ratios of 1.1-1.5 (Manolio et al., 2009). Indeed, out of 1,570 published genome-wide association studies, only 55 (3.5%) report odds ratios that can better this (Hindorff et al.), including the strong association of $\epsilon 4$ with Alzheimer's (OR = 4.1, (Coon et al., 2007)).

However, a meta-analysis of 19,540 individuals across 11 different cohorts (including the original study) fails to find any association between $\epsilon 2$ and left-handedness (Piper et al., 2013). All ten replication cohorts are larger than the original study, and none individually show a significant association. A further study of 4,438 individuals also fails to report any association between *APOE* isoforms and either left handedness or higher left-hand grip strength (Hubacek et al., 2012). This example highlights the need for replication of genetic associations, particularly where the initial study cohort is small (i.e. less than 500), in order to reduce publication of false-positives.

Handedness, cerebral lateralisation, and neurodevelopmental disorders

Atypical cerebral asymmetries have been associated with neurodevelopmental disorders; in particular dyslexia and schizophrenia. Structural MRI studies consistently show reduced gray matter volume in dyslexic readers in the right superior temporal gyrus and left superior temporal sulcus (Richlan et al., 2012). Functional MRI studies confirm hypoactivity in the right superior temporal gyrus, however in general activity appears to be greater in the right- and less in the left-hemisphere in individuals with dyslexia (Maisog et al., 2008), and functional transcranial Doppler ultrasound confirms this finding (Illingworth and Bishop, 2009). In schizophrenia, a structural MRI study does not find atypical cerebral asymmetry among 375 individuals with schizophrenia (Deep-Soboslay et al., 2010). Although global asymmetries may not be atypical, specific areas such as the planum temporale are less asymmetric in individuals with schizophrenia (Clark et al., 2010; Shapleske et al., 1999; Sommer et al., 2001), as they are in left-handers.

1.1.5 When does handedness become apparent during human development?

Although handedness and cerebral asymmetry has been correlated with various traits and disorders, it is difficult to determine cause and effect. Does weak cerebral laterality cause the neurodevelopmental disorder or does the disorder cause the weak laterality, or do genetic influences underlie both weak laterality and neurodevelopmental disorders (pleiotropy) (Bishop, 2013)? In order to help answer this question we first need to understand the development of handedness.

The right hand is more developed than the left at seven weeks (O’Rahilly and Muller, 2010). Using ultrasound, Hepper et al. found that 90% of fetuses suck their right thumb, and 10% their left, at 15 weeks of gestation (Hepper et al., 1991). This prenatal thumb sucking is related to postnatal handedness as shown by a follow up study of 75 individuals in which 60 of the fetuses who preferred their right thumb are indeed all right handed as teenagers, whereas of the 15 who preferred their left thumb, 10 became left handed and the other 5 were right handed (Hepper et al., 2005). In the original study, 274 / 282 fetuses were observed to suck their thumbs, however another study that followed ten fetuses throughout development only observes thumb sucking on 4 / 77 registrations in three fetuses. They also observe finger or hand sucking slightly more frequently, 11 and eight times respectively, with no obvious bias to the right (de Vries et al., 2001). It appears then that thumb sucking is a relatively rare observation. More common are arm movements, Hepper found that 85% of fetuses move their right arm more than their left at 10 weeks (Hepper et al., 1998). However, de Vries finds no lateralisation of hand–head arm movements throughout development from 12–38 weeks (de Vries et al., 2001). Furthermore, another study assessing arm movements from 33 weeks

until birth finds no difference between the left and right arms (Fong et al., 2005a). The Hepper results are so remarkable that independent verification of the findings is necessary.

While prenatal studies remain controversial, it is clear that newborns tend to develop a preference for the right very early in life. First, newborns prefer to turn their heads to the right for the first three months of life (Hopkins et al., 1987, 1990; Ronnqvist and Hopkins, 1998), and this preference is not apparent *in utero* until $\sim$ 38 weeks (Fong et al., 2005b; Ververs et al., 1994). This head position preference is correlated with hand preference for making hand–mouth contacts (Hopkins et al., 1987). However, after three months reaching movements are frequently bimanual (White et al., 1964). Then after $\sim$ six months, a right-sided bias for grasping emerges, though not as strong as it is for writing in adults (for review see (Fagard, 2013)). On balance, it appears that the development of lateralised motor behaviour begins to appear early in postnatal development, and this initial bias becomes stronger as humans get older.

1.1.6 The genetic architecture of handedness: non-genetic, mono-genic, or polygenic?

Given that handedness is apparent early in human development it is likely that both *in utero* environment and genetics are important in the development of handedness. Laland argues that humans have a universal pre-disposition towards right-handedness that derives from a series of selective sweeps throughout evolution that ratcheted up the propensity towards right-handedness (Laland, 2008). His theoretical model suggests that our genes load the die to favour the right, and any variation between individuals derives purely from environmental influences,

such as cultural pressure to conform (Laland et al., 1995). Conversely, Annett, Klar, and McManus have proposed single gene models that can explain the variation in handedness (Annett, 1985; Klar, 1996; McManus, 2004).

A study of over 25,000 twin pairs shows that preferred hand for writing or drawing is a weak genetic trait with a heritability of 24% (Medland et al., 2009), which appears to rule out exclusively non-genetic arguments. However, even though single gene theories fit data on the prevalence of handedness, linkage studies do not identify this single-gene locus, pointing instead to different regions of genome including; 2p12-q11 (Francks et al., 2002, 2003a), 10q26 (Van Agtmael et al., 2002), 12q21-23 (Warren et al., 2006), and Xq21 (Laval et al., 1998). Only two specific genes have been suggested as candidates for handedness. The imprinted gene *LRRTM1* in 2p12-q11 has been associated with handedness and schizophrenia (Francks et al., 2007), and a candidate gene approach identified the X-linked androgen receptor (*AR*) (Medland et al., 2005). The same polymorphism in *AR* was found to be associated with 'mixed-handedness' in an independent cohort (Hampson and Sankar, 2012). Furthermore published case/control genome-wide association studies (GWASs) for handedness have found no statistically significant associations, despite adequate sample sizes to detect a single locus with a strong effect size (Armour et al., 2013; Eriksson et al., 2010). The largest case/control study, mentioned in an abstract for the American Society of Human Genetics 2009 meeting, assessed 2,350 left-handed 'cases' and 21,093 right-handed controls, and found no genome-wide significant associations (<http://www.ashg.org/2009meeting/abstracts/fulltext/f21141.htm>).

McManus et al. conclude that handedness cannot be controlled by a single ge-

netic locus. Instead they estimate that at least 40 loci underlie the variation in this trait (McManus et al., 2013). Given the universality of right-handedness among humans, it seems likely that it has been selected for during evolution as Laland suggests. However both chance, cultural, environmental pressures, coupled with genetic variants then modulate this bias, i.e. it is a polygenic trait.

1.2 Literature review - the genetics of dyslexia

Developmental dyslexia is a specific reading disability that cannot be accounted for by a lack of intelligence, sensory problems, or inadequate opportunity to learn, which affects approximately 4-8% of the population (Butterworth and Kovas, 2013). The American Psychiatric Association recommends a significant discrepancy between IQ and a reading ability test to diagnose dyslexia. This makes diagnosis more probable for someone with high intelligence and excludes the possibility of someone with intellectual disability from being diagnosed (Butterworth and Kovas, 2013).

1.2.1 Core deficit in dyslexia

In most children with dyslexia, the core deficit is in phonological processing, i.e. categorising speech sounds and relating them to written language (Ramus and Szenkovits, 2008; Snowling, 1987, 2001). The deficit is a combination of poor phonological awareness, poor verbal short term memory, and slow retrieval of lexical phonological representations from long term memory (assessed by rapid naming tasks) (Wagner and Torgesen, 1987). Ramus argues that this deficit is not in the

phonological representation of sounds themselves, but in the retrieval of these representations for purposes such as reading (Ramus and Szenkovits, 2008).

1.2.2 The relationship between dyslexia, neurodevelopmental disorders, and handedness

Dyslexia shows significant comorbidity with other neurodevelopmental disorders; 23-49% have dyscalculia (Landerl and Moll, 2010), ~35% have specific language impairment (SLI) (Catts et al., 2002, 2008), and 33-45% of individuals with attention deficit hyperactivity disorder (ADHD) also have developmental dyslexia (Sexton et al., 2012). Furthermore, there is a broad literature suggesting individuals with dyslexia have motor coordination and balancing difficulties (Fawcett and Nicolson, 1999; Getchell et al., 2007; Iversen et al., 2005; Ramus et al., 2003; Stoodley and Stein, 2006; Stoodley et al., 2005). This has led some researchers to propose that cerebellar impairment underlies dyslexia (Nicolson and Fawcett, 2005; Stoodley and Stein, 2011). However, individuals with dyslexia do not show deficits in other tasks that require the use of the cerebellum, such as time estimation, casting doubt on a broad-scale cerebellar dysfunction (Ramus et al., 2003).

Bishop tested for association between handedness and dyslexia by performing a mega-analysis, combining the cases and controls across 25 studies, and found no significant relationship between the two traits (Bishop, 1990b). However, Eglinton and Annett reanalysed the data by performing a meta-analysis, which treats each study separately thus controlling for study heterogeneity, and found a significant increase in the prevalence of left-handedness among individuals with dyslexia (Eglinton and Annett, 1994). Four out of the 25 studies are individually significant,

two of which date from the 1950s (Eglinton and Annett, 1994). There have been no large studies or meta-analyses published in the literature since, and how strong the effect is.

1.2.3 Heritability of dyslexia and linkage studies

Understanding the genetic basis of dyslexia can help us understand the molecular basis of phonological deficits and help resolve the etiology of the disorder. Dyslexia is highly heritable, recent large-scale twin studies estimate a heritability of reading ability and dyslexia of 0.53-0.82 (Byrne et al., 2009; Hensler et al., 2010; Petrill et al., 2007). Linkage studies have implicated nine loci with dyslexia, however each locus has more negative studies published than positive ones (for review see (Scerri and Schulte-Korne, 2010)). While linkage studies have been successful for Mendelian disorders, where a single variant of large effect causes the disorder, true linkage has proved difficult to find for complex genetic disorders (Altmuller et al., 2001). It is likely that all nine DYX loci identified through linkage studies are false positives.

1.2.4 Dyslexia candidate genes, assessment of the evidence

Only a handful of candidate genes for dyslexia have been identified, initially either through high throughput genotyping of variants in genes located under linkage peaks (*CMIP*, *DCDC2*, & *KIAA0319*), expression correlating with risk haplotypes identified by linkage studies (*MRPL19* / *C2ORF3*), or identification of reciprocal chromosomal translocations segregating with the disorder (*DYX1C1* & *ROBO1*). *DCDC2*, *DYX1C1*, *KIAA0319*, and *ROBO1* are involved in neuronal mi-

gration or axon guidance (Galaburda et al., 2006). Below I've detailed the evidence for each gene's association with dyslexia.

CMIP

c-maf-inducing protein (*CMIP*) is located in the specific language impairment 1 (SLI1) linkage locus, and the variant rs6564903 was first found to be associated with nonword repetition in individuals with language impairment (Newbury et al., 2009). This variant has subsequently been found to be associated with reading and spelling (but not non-word repetition) abilities across the range of abilities in the longitudinal cohort ALSPAC (i.e. cohorts 3 & 4 combined (Scerri et al., 2011)).

DYX1C1

Dyslexia susceptibility 1 candidate 1 (*DYX1C1*), located in the *DYX1* locus on chromosome 15q21, was found to be disrupted by a reciprocal translocation with chromosome 2 [t(2;15)(q11;q21)] that segregated with dyslexia in one Finnish family (Taipale et al., 2003). Sequencing of the exons in unrelated individuals identified four SNPs coding for amino acid substitutions and three in the 5' untranslated region (UTR) of *DYX1C1*. Case / control association analysis then identified that one coding variant and one 5' UTR variant were associated with dyslexia. Ten different replication studies have been performed, one study provides support for these variants (Marino et al., 2007), three support the *opposite* alleles (Brkanac et al., 2007; Dahdouh et al., 2009; Wigg et al., 2004), two support different variants within the gene (Bates et al., 2010; Wigg et al., 2004), and four do not find any association (Bellini et al., 2005; Cope et al., 2005; Meng et al., 2005; Scerri et al., 2004). Furthermore, one study claims to replicate the association, but actually only shows a

non-significant trend for both variants that hasn't been corrected for multiple testing (Paracchini et al., 2011). The different ethnic background and ascertainment of the participants may explain the conflicting results (Paracchini et al., 2007). For example, the 5' UTR variant rs3743205 has a minor allele frequency of 0.002 in one study (Bates et al., 2010).

Aside from neuronal migration, *DYX1C1* is also involved in ciliogenesis (Chandrasekar et al., 2013; Ivliev et al., 2012), and individuals homozygous recessive for loss of function mutations have left/right asymmetry defects such as situs inversus, mirroring the phenotype observed when the gene is knocked out in mice (Tarkar et al., 2013). None of these individuals identified ($n = 12$) have dyslexia, casting doubt that the disruption of *DYX1C1* by a reciprocal translocation was causative of dyslexia in the initial study. Coding variants in *DYX1C1* are unlikely to be responsible for association with dyslexia unless they cause a gain of function.

DCDC2

A number of different variants in the 212kb gene doublecortin domain containing 2 (*DCDC2*) have been associated with dyslexia (for review see (Scerri et al., 2010)). Three variants have been associated in more than one study; rs793862 (4/6 studies), a haplotype of rs793862 & rs807701 (2/5 studies), and a polymorphic deletion (3/6 studies) (Scerri et al., 2010). The variants rs793862 & rs807701 were subsequently found to be associated with reading and spelling ability in cohort 3 (ALSPAC reading disabled), but not in cohort 4 (ALSPAC general population) (Scerri et al., 2011). Furthermore the risk alleles had a higher frequency in a case / control analysis of cohorts 3 & 4 (Scerri et al., 2011).

KIAA0319

Association of *KIAA0319* was first identified with a microsatellite marker (JA04) in exon 1 of the gene that is within the DYX2 linkage locus (Kaplan et al., 2002). This association has never been replicated. However, a specific risk-haplotype located upstream of the gene is associated with dyslexia in three independent samples from two studies (Cope et al., 2005; Francks et al., 2004). This haplotype has also been associated with reading and spelling ability across the distribution of abilities in ALSPAC (Paracchini et al., 2008; Scerri et al., 2011). Furthermore, the variant with the strongest association on this haplotype (rs9461045) reduces expression of *KIAA0319* (Dennis et al., 2009). Although one study found the risk haplotype was in fact associated with better reading scores (Luciano et al., 2007).

MRPL19 & C2ORF3

Fine-scale linkage disequilibrium mapping of DYX3 (2p11-p15) refined the location of the linkage to an intergenic region of 2p12 in two independent cohorts (Anthoni et al., 2007). The risk haplotype correlated with expression of nearby genes *MRPL19* & *C2ORF3*. However, no specific coding variants genotyped in these genes were associated with dyslexia. Furthermore, two independent studies failed to find evidence to support association of any variants at this locus (Paracchini et al., 2011; Scerri et al., 2011)

ROBO1

The evidence for roundabout axon guidance receptor homolog 1 (*ROBO1*) playing a role in the etiology of dyslexia comes from a reciprocal translocation with a breakpoint between exons 1 & 2 of *ROBO1* in a single individual (Hannula-Jouppi et al., 2005).

On balance, there is very little evidence for *DYX1C1*, *ROBO1*, *MRPL19*, or *C2ORF3* in the etiology of dyslexia. However, there is suggestive evidence for variants in *CMIP*, *KIAA0319*, and *DCDC2*.

1.3 Thesis structure

Understanding the molecular genetic basis of both handedness and dyslexia may contribute to answering questions about the evolution and development of handedness and cerebral asymmetry, and offer clues to the link with neurodevelopmental disorders such as dyslexia.

The aim of my research project therefore is to identify genetic variants associated with both relative hand skill (handedness) and dyslexia (reading disability), in order to better understand the biology of these traits. To achieve this I perform both a Genome-Wide Association Study (GWAS) and Copy Number Variant (CNV) analyses utilising a series of data sets that are summarised in **Table 1.2**.

Table 1.2: Summary of cohorts utilised in the project.

Cohort	Name	<i>n</i>	Year	Genotyping Array	Data			
					GWAS	CNV	Reading	Handedness
1	NeuroDYS (Oxford)	197	2007	Illumina 550K	✓	✓	✓	✓
2	Oxford Dyslexia	374	2012	Illumina OmniExpress	✓	✓	✓	✓
3	ALSPAC Dyslexia	169	2011	Illumina 550K	✓	✗	✓	✓
4	ALSPAC General Population	2,666	2011	Illumina 550K	✓	✗	✓	✓
5	NeuroDYS (Bonn)	200	2007	Illumina 300K	✗	✓	✓	✗
6	NeuroDYS (Cardiff)	208	2007	Illumina 300K	✗	✓	✗	✗

I start the project with handedness phenotype data and genotype data from the 197 samples genotyped on the Illumina 550K single nucleotide polymorphism (SNP) genotyping array (cohort 1). I then perform a GWAS for relative hand skill in this cohort, the most strongly associated SNP is located in intron 15 of *PCSK6*, a gene that causes left/right asymmetry phenotypes when disrupted in mice (Constam and Robertson, 2000). I then genotype this SNP in cohorts 2 and 3, reaching genome-wide significance in the combined meta-analysis (**chapter 3**). I then follow up this finding by searching for the etiological variant in *PCSK6* through fine-mapping of the region (**subsection 3.4.2**). In collaboration with Silvia Paracchini (University of St. Andrews), and Simon Fisher (Max Planck Institute for Psycholinguistics), I then genotype a total of 854 samples on the Illumina OmniExpress SNP array. This includes unrelated individuals with dyslexia (cohort 2, $n = 374$ after quality control), siblings, triplicates, and replicates. My role is to prepare the DNA for genotyping (**chapter 4**), perform a GWA meta-analysis for relative hand skill in cohorts 1-3 (**chapter 5**), a GWA meta-analysis of the quantitative reading traits in cohorts 1 and 2 (**chapter 6**), and finally CNV analysis in cohorts 1, 2, 5, and 6 (**chapter 7**). A flow chart of the work done in this project is summarised in **Figure 1.1**.

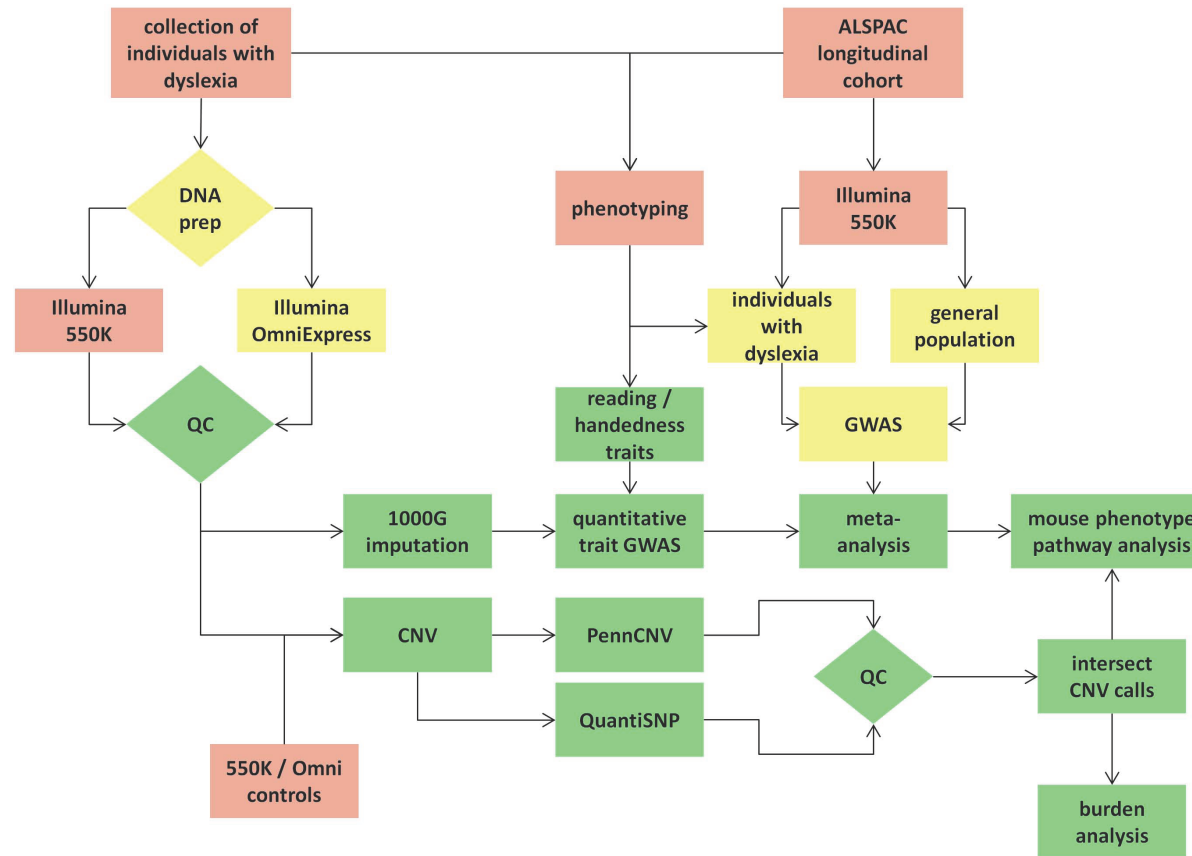


Figure 1.1: **Flow chart outlining the study design for my thesis.** The colour coding indicates work that I performed on my own, that was in collaboration with others, and that was done by others. CNV = copy number variant, GWAS = genome-wide association study, QC = quality control, 550K / OmniExpress = single nucleotide polymorphism genotyping arrays, PennCNV / QuantiSNP = CNV calling algorithms.

Chapter 2

Dyslexia ascertainment criteria and phenotyping of study participants

Below I have detailed the methods by which we ascertained individuals for dyslexia and phenotyped them for reading and relative hand skill. For statistical analyses and genotyping methods, please refer to the relevant methods sections for each chapter.

2.1 Study participant ascertainment criteria

The participants for this study were collected and phenotyped over a number of years

- **Cohort 1:** NeuroDYS (Oxford). Individuals were initially selected by Thomas Scerri from a collection of families with at least one child with reading disability (RD) (Fisher et al., 2002) for severity of phenotype to undergo a GWAS for reading disability. Approximately two thirds of the families were ascertained if the proband had a British Abilities Scales (BAS) single-word read-

ing score at least two standard deviations (SD) below that predicted by their intelligence quotient (IQ) derived from their verbal and non-verbal reasoning scores and if at least one other sibling had a record of reading disability. These criteria identified some probands with exceptionally high IQ scores and BAS reading scores that were above average. Therefore, for the remaining families the criteria were changed such that the subject's difference in their BAS single-word reading score had to be ≥ 1 SD below the population mean for their age-band, along with an IQ ≥ 90 . Proband exclusion criteria included other disorders such as specific language impairment (SLI), autism, or attention deficit-hyperactivity disorder (ADHD).

A battery of six psychometric tests were administered to all the subjects, we age-adjusted and standardised their scores against a normative control data set as described elsewhere (Fisher et al., 2002; Marlow et al., 2001). These included a series of quantitative reading measures (Marlow et al., 2001; Paracchini, 2009):

1. **single-word reading** - the standardised British Ability Scales (BAS) test of single-word reading was applied for individuals up to the age of 14.5 years (Elliot et al., 1983). For older individuals the Wide Range Achievement Test (WRAT-R) was used (Jastak and Wilkinson, 1984). (READ).
2. **spelling ability** - depending on the age of the participant either the BAS or WRAT-R test for single-word spelling was applied (SPELL).
3. **phonological decoding ability** - reading a set of 30 non-words, e.g. torlep, dorkit, tegwop (CCN).
4. **phonological awareness** - assessing an individual's ability to manipulate phonemes in words presented orally. The test consists of three sets of 10 phoneme manipulations, the first is simple phoneme deletion and

substitution (e.g. replace the first sound in 'dog' with 'l' to make 'log'); the second is complex phoneme deletion and substitution (e.g. replace the first sound in 'lip' with the first sound in 'pig' to make 'pip'); and finally spoonerisms (e.g. swap the first sounds of spoon/dog, which become doon/spog) (SPOON).

5. **orthographic coding** - reading a set of 30 irregular words, e.g. colonel, yacht, salmon (CCI).
6. **orthographic choice** - discriminating a set of 80 real words from pseudohomophones, e.g. rain versus rane (OLSON).

Individuals selected for cohort 1 are within the normal range of IQ (≥ 85), but have scores ≥ 1.5 SD below the mean for at least one of the 6 reading phenotypes. The subjects have a mean age of 11.4 years (SD 3.8, range: 6.3-25.4 years) and consist of 139 males and 58 females.

- **Cohort 2:** Oxford Dyslexia. These individuals were recruited from either the Dyslexia Research Centre clinics in Oxford and Reading or the Aston Dyslexia and Development Clinic in Birmingham (Scerri et al., 2010). The majority are between 8 and 18 years old. READ, SPELL, CCI, and CCN were collected for most individuals, as well as relative hand skill phenotypes and non-verbal performance IQ (MAT) and verbal IQ (SIM). I selected the individuals for inclusion in the GWA and CNV studies if they had a performance IQ ≥ 85 , at least one of the four reading scores ≥ 1.5 SD below that predicted by IQ, plus a mean of the four reading scores below the population average (to exclude those with exceptionally high IQ, but normal or even above average reading scores, consistent with cohort 1). The correlation between the various reading-related measures, IQ, and PegQ are given in **Table 2.1**.

Table 2.1: **Pearson’s correlation between quantitative reading-related measures, IQ, and PegQ** in cohorts 1 & 2 combined ($n = 558$).

	<i>READ</i>	<i>SPELL</i>	<i>CCN</i>	<i>CCI</i>	<i>MAT IQ</i>	<i>SIM IQ</i>	<i>PegQ</i>
<i>READ</i>	1	0.63	0.51	0.69	0.29	0.48	-0.01
<i>SPELL</i>	0.63	1	0.36	0.53	0.33	0.30	0.00
<i>CCN</i>	0.69	0.53	1	0.48	0.25	0.28	-0.03
<i>CCI</i>	0.67	0.47	0.48	1	0.21	0.35	0.02
<i>MAT IQ</i>	0.27	0.31	0.25	0.21	1	0.21	0.02
<i>SIM IQ</i>	0.45	0.28	0.28	0.35	0.21	1	0.00
<i>PegQ</i>	0.04	0.04	-0.03	0.02	0.02	0.00	1

- **ALSPAC:** The Avon Longitudinal Study of Parents and Children (ALSPAC) started collecting data on around 14,000 pregnant women in the Bristol area with a then expected delivery date between 1 April 1991 and 31 December 1992 (Golding et al., 2001). This cohort is described in detail on the website (<http://www.alspac.bris.ac.uk>) and elsewhere (Boyd et al., 2012). From age 7, all children were invited annually for assessments on a wide range of physical, behavioral, and neuropsychological traits, including reading and laterality-related measures. DNA is available from approximately 11,000 ALPSAC children. For the present study Thomas Scerri, Silvia Paracchini, and Dorothy Bishop assigned individuals from the ALSPAC cohort into a sub-group with reading disability, and a sub-group representing the general population without a neurodevelopmental disorder. In both cases, subjects were excluded based on both a self-reported ethnicity of ‘non-white’ and a performance $IQ \leq 85$ at age of 8.5.

- **Cohort 3:** Reading disabled (RD) sub-group. An assignment of RD was based on an age-adjusted single-word reading score 1 SD below the mean for both a 40-word reading test at 7.5 years and a 12-word reading test at 9.5 years, plus a performance IQ score ≥ 85 . Individuals missing any of these data points were excluded.
- **Cohort 4:** General population sub-group. Individuals were excluded if they met criteria for one of the following neurodevelopmental disorders:
 1. Attention Deficit Hyperactivity Disorder (ADHD): Defined based on a DAWBA DSM-IV clinical diagnosis at 7.5 years; individuals missing this data point were excluded
 2. Autistic features: a composite score of 7 measures from the Children's Communication Checklist (CCC) pragmatic aspects of communication ≤ -3 SD at 7.5 years
 3. Specific language impairment: If an individual met at least two of the following four criteria:
 - a composite score of 7 measures from the Children's Communication Checklist (CCC) pragmatic aspects of communication ≥ -3 SD and ≤ -1 SD at age 7.5 years
 - an age adjusted non-word repetition score ≤ -1 SD at 8.5 years
 - an age adjusted WOLD comprehension score ≤ -1 SD at 8.5 years
 - a questionnaire asking if the individual has ever had speech/language therapy given at 9.5 years; some flexibility for missing data points was allowable in assigning SLI

4. Reading disability (as defined above for the RD sub-group)
5. Performance IQ ≤ 85

2.2 How to measure handedness

Although most people do prefer one hand over the other, handedness is a matter of degree, not kind. There are three different questionnaires commonly used to assess handedness, the Edinburgh handedness inventory (Oldfield, 1971), the Annett handedness questionnaire (Annett, 1970), and the Crovitz-Zener score (Crovitz and Zener, 1962).

The Edinburgh handedness inventory is a 20-point questionnaire on preferred hand (left or right) for a number of tasks from writing to threading a needle (Oldfield, 1971). The Annett and Crovitz-Zener scores ask 12 and 14 questions, respectively, both of which have ten questions that overlap with the Edinburgh inventory (Annett, 1970; Crovitz and Zener, 1962). From these can be derived a laterality quotient that shows a J-shaped distribution, i.e. most people use their right-hand for most tasks, a sizeable minority use their left hand for most tasks, and an even spread of low numbers of individuals that lie somewhere in between.

Participants are typically subsequently categorised in terms of direction of handedness; either left-, mixed-, or right-handed, or in terms of degree of handedness; either consistently preferring one hand, or inconsistent. Since there are no standard cut-offs for these distinctions, this can potentially lead to post-hoc classifications. Simulations have shown that by varying the criteria used for classifying subjects as 'left' or 'right' handed, under certain circumstances the probability of obtaining

a significant result can be as high as 40% (Bishop, 1990a).

Furthermore, the three questionnaires do not assess fine motor performance, or dexterity. The epidemiological Avon Longitudinal Study of Parents and Children (ALSPAC) study measures these skills by assessing grip strength, sorting matches, marking squares, and the peg-board task. These measures are weakly correlated with each other (r^2 between 0.07 and 0.28, **Table 2.2**), suggesting that each measures a different aspect of laterality or dexterity.

The peg-board task is a particularly useful phenotype since it is quantitative, easy and quick to measure, and offers more information about dexterity than a questionnaire based assessment of preferred hand. The task measures the time taken by the subjects to move a row pegs from one location to another with the left hand (L) and right hand (R) separately (Annett, 1985). This provides a measure of relative hand skill (PegQ) that is normally distributed with a mean shifted to the right, since most individuals are faster at this task with their right hand. This measure was used throughout this research project to assess handedness, see **subsection 2.2.1**.

Table 2.2: Pearson’s correlation between quantitative measures of relative hand skill and dexterity measured in ALSPAC ($n = 4,282$).

	<i>Peg-Board</i>	<i>Grip Strength</i>	<i>Sorting Matches</i>	<i>Marking Squares</i>
<i>Peg-Board</i>	1	0.14	0.15	0.28
<i>Grip Strength</i>	0.14	1	0.07	0.19
<i>Sorting Matches</i>	0.15	0.07	1	0.17
<i>Marking Squares</i>	0.28	0.19	0.17	1

2.2.1 Relative hand skill phenotype

The peg-board task (Annett, 1985) was administered to the individuals who were analyzed in cohorts 1 and 2. The test involved the measurement of the time taken by the subjects to move, with each hand, a row of ten pegs on a board from one location to another. This test was repeated 5 times for each hand in all the subjects in cohorts 1 & 2. From these data I derived the relative hand skill measure (PegQ) for each subject, which was calculated as the difference between the average times for the left hand (L) and the right hand (R), divided by the average time for both hands combined (Francks et al., 2002):

$$PegQ = \frac{L - R}{(L + R)/2}$$

The average time take with both hands to complete the peg board task is referred to as hand motor skill throughout the thesis. It is a measure of overall skill as opposed to relative differences between the two hands measured by PegQ. Although overall hand motor skill improves with age (i.e. the peg-board task is performed in a shorter time with both hands), relative hand proficiency measured in this way does not change significantly between the ages of 3 and 15 (Annett, 1970). Therefore, although there is a large range of ages in our study (6-25 years), this should not significantly affect the results. The distribution of PegQ in cohort 1 has a mean of 0.080 (SD = 0.099), and in cohort 2 it is 0.061 (SD = 0.091). The PegQ distribution is unimodal, continuous, and approximately normal, making it suitable as a quantitative phenotype for a GWAS (**Figure 2.1** & **Figure 2.2**). PegQ has a heritability estimated between 0.23 and 0.36 (Francks et al., 2002, 2003b; Laval et al., 1998). Furthermore, PegQ is highly correlated with hand preference, for every 1 standard deviation shift to the right of the PegQ distribution; individuals are 13.4

times more likely to be right handed (95% confidence intervals: 6.11-36.49).

A peg-board task similar to that used in cohorts 1 and 2 was performed in the ALSPAC sample as part of a battery of manual dexterity tests known as Movement ABC (Henderson and Sugden, 1992). In this case there were 12 pegs on the table that the child picked up one-at-a-time and placed in a peg-board. After initial practice, children performed the test once (due to time constraints) with each hand. The mean PegQ score was slightly different in cohort 3 at 0.11 (SD = 0.199). However, given the slight difference in test this was to be expected; the distribution remains unimodal, continuous and approximately normal. One outlier was removed from further analysis (PegQ = 1.50). The PegQ distribution is highly correlated with both hand preference and dexterity (**subsection 5.4.1**). In cohorts 1, 2, 3 & 4, there is no evidence for strong skewness or kurtosis (**Table 2.3**). The visualisation of the distribution confirms this (**Figure 2.2**). I normalised all peg scores to a mean of 0 and SD of 1 for further analyses.

Table 2.3: **A summary of the skewness and kurtosis of the PegQ scores.** A skewness score ≤ -1 or ≥ 1 indicates the data is significantly skewed. A kurtosis score of 3 indicates a mesokurtic, normal, distribution. A score ≥ 3 indicates a leptokurtic distribution, with values concentrated around the mean and thicker tails. A score ≤ 3 indicates a platykurtic distribution, a flatter distribution with a wider spread of values around the mean.

<i>Cohort</i>	<i>Skewness</i>	<i>Kurtosis</i>
1 & 2	-0.55	3.5
3	0.01	2.96
4	-0.15	3.15

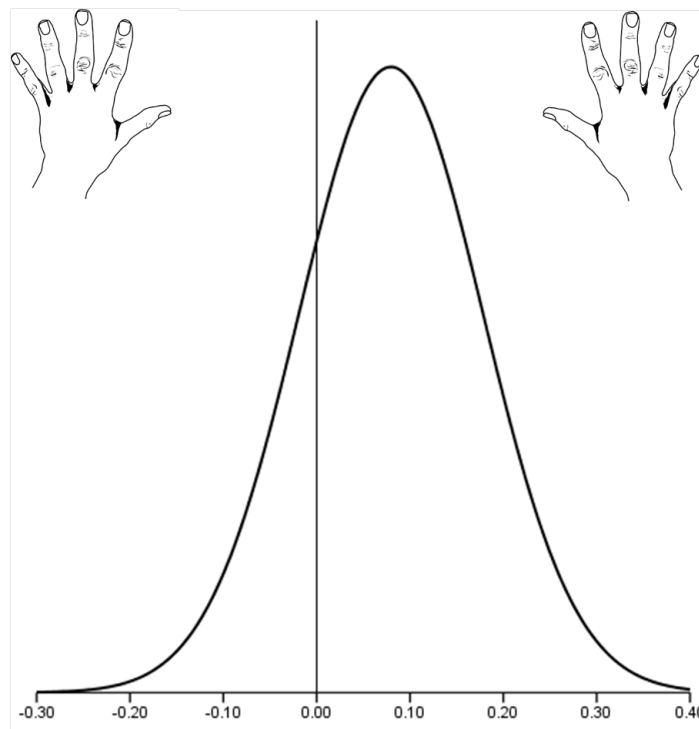
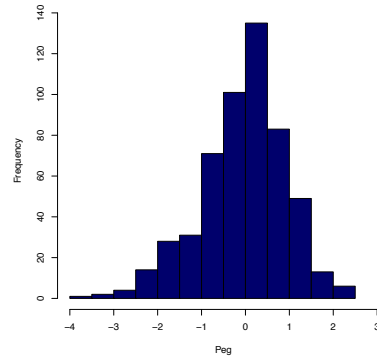
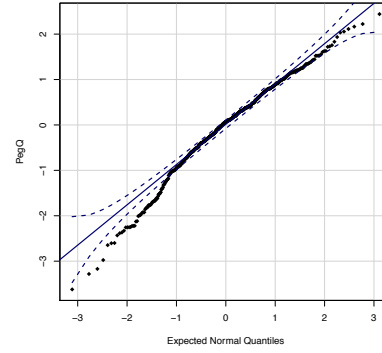


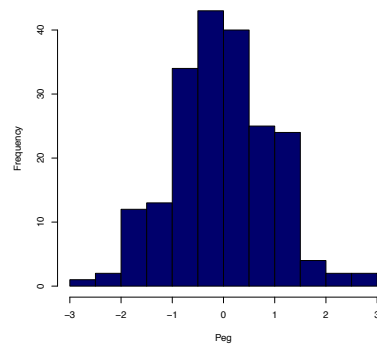
Figure 2.1: **Schematic of the PegQ distribution.** A positive PegQ indicates superior relative right-hand skill, and a negative PegQ indicates superior relative left-hand skill.



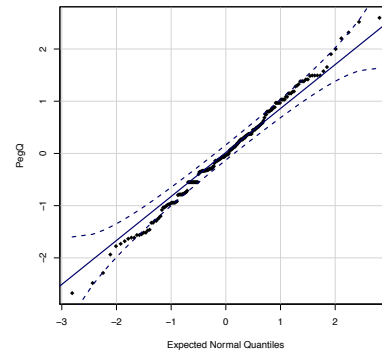
(a)



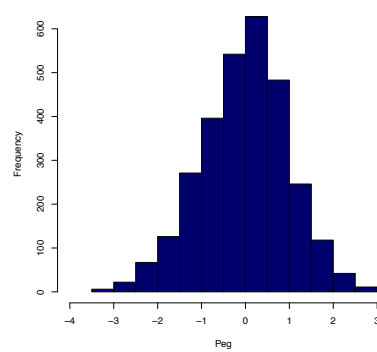
(b)



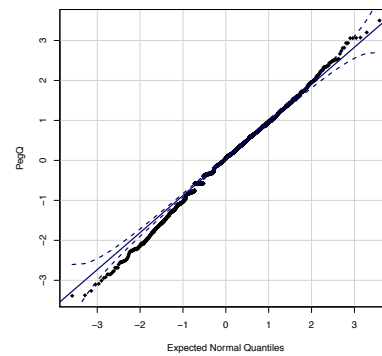
(c)



(d)



(e)



(f)

Figure 2.2: **Histograms and QQ plots of PegQ.** Scores are normalised to a mean of 0 and a standard deviation of 1. (a) Histogram & (b) QQ plot - cohorts 1 & 2. (c) & (d) - cohort 3. (e) & (f) - cohort 4. The 95% confidence intervals are given by the - - - lines.

Chapter 3

Relative hand skill GWAS

3.1 Abstract

I report a genome-wide association study (GWAS) for a quantitative measure of relative hand skill in individuals with dyslexia [reading disability (RD)]. The most highly associated marker, rs11855415 ($P = 4.7 \times 10^{-7}$), is located within *PCSK6*. I genotyped this marker in two independent cohorts with RD, and they both show the same trend, with the minor allele conferring greater relative right-hand skill. Meta-analysis of all three RD samples is genome-wide significant ($n = 744$, $P = 2.0 \times 10^{-8}$). Conversely, in the general population ($n = 2,666$), I observe a trend towards reduced laterality of hand skill for the minor allele ($P = 0.0020$). *PCSK6* is a protease that cleaves the left-right axis determining protein NODAL. Fine-mapping of the *PCSK6* locus fails to identify an etiological variant, but identifies a variable number tandem repeat located in a putative secondary promoter of the gene. This research is now being followed up by functional studies of *PCSK6* by collaborators at St. Andrew's University that promises insights into mechanisms underlying cerebral lateralisation and dyslexia.

3.2 Introduction

For a literature review on handedness, see **section 1.1**. In order to help determine the genetic basis of handedness, here I perform a GWAS for a quantitative measure of relative hand skill (peg-board task) in a cohort with RD. The most highly associated SNP is in an intron of *PCSK6*, a gene involved in left/right axis determination (Constam and Robertson, 2000). I then replicate this association in two other independent samples with RD reaching genome-wide significance in the combined meta-analysis. Fine-mapping of this region does not reveal the etiological variant.

3.2.1 Genome-wide association studies

GWASs compare the allele frequency of common variants across the genome in cases and controls. For quantitative traits, correlation between alleles of each variant and the trait is tested by linear regression using population-based cohorts. Since hundreds of thousands of variants across the whole genome are tested for association, a stringent P value threshold for significance is applied (see Box 2 on page 57).

GWASs have been successful in helping to identify the genetic basis of complex traits for two key reasons:

1. Collaborative efforts such as HapMap and the 1000 genomes projects that have allowed us to understand patterns of common human genetic variation across ethnic groups.
2. Improvements in cost efficiency of genotyping technology means that we can routinely assay hundreds of thousands of variants in large sample sizes.

Cataloguing common genetic variation in human populations began with the International HapMap project. The first phase of HapMap involved random shotgun sequencing of 270 individuals drawn from four populations, with an aim of identifying one common SNP every 5kb across the whole genome (The International HapMap Consortium, 2003). This has been followed up by the 1000 Genomes project, which has sequenced 1,092 individuals drawn from 14 geographically distinct populations at $2\text{--}6\times$ coverage (Abecasis et al., 2012). Thanks to these efforts dbSNP build 138 (April 25, 2013) contains 44,278,189 validated SNPs (http://www.ncbi.nlm.nih.gov/SNP/snp_summary.cgi). This includes almost all of the ~ 7.1 million SNPs with minor allele frequencies $\geq 5\%$ predicted to exist in human populations (Kruglyak and Nickerson, 2001).

GWASs are based on the premise that a causal variant is located on a haplotype, and therefore an allele in linkage disequilibrium (LD) with the causal variant will show association with a trait of interest, due to correlation between alleles in LD with each other (Stranger et al., 2011). The LD between SNPs also means that not all 7.1 million variants need to be genotyped in order to detect a true association. Instead a subset of variants are genotyped, and those that are not directly genotyped can be inferred through known patterns of LD between variants. This process is known as imputation, which predicts unobserved genotypes using haplotype patterns from a reference panel of individuals that have been densely genotyped (e.g. HapMap or the 1000 genomes project) (Howie et al., 2011). The imputed genotypes expand the set of SNPs that can be tested for association, increasing power and resolution to detect true association signals and etiological variants. Furthermore imputation facilitates meta-analysis of studies genotyped on different SNP arrays (as in **chapter 5**).

GWAS designs are based on the Common Disease / Common Variant (CD/CV) paradigm (Chakravarti, 1999; Lander, 1996), which proposes that variants with a high frequency in the population ($\geq 5\%$) contribute significantly to explaining the heritability of complex traits or disorders. Each variant has a modest effect size and explains several per cent of the genetic variance. This is contrasted with competing models; the infinitesimal model, broad-sense heritability model, and rare allele model (reviewed in (Gibson, 2011)). In the infinitesimal model many variants of small effect each contribute less than one percent of the variance and have odds ratios ≤ 1.1 , i.e. cases are at most $1.1 \times$ more likely to carry the variant. Ultimately every gene contributes to the trait in question, however the effect sizes become so small that a sample size greater than the entire population of the species is necessary to detect association. In the broad-sense heritability model there are interactions between genes and the environment ($G \times E$), genotypes (epistasis, $G \times G$), and epigenetic effects (such as parent-of-origin effects). The result is that some associations are only seen in certain environments or genetic backgrounds, and in a mixed population fewer associations will be found with a smaller proportion of the variance explained by each. Finally, the rare allele model proposes that many highly penetrant rare alleles ($f \leq 1\%$), each affecting only a handful of individuals or families, contribute to explaining the genetic variance. Since GWAS SNP arrays only genotype common variants (**section 3.2.1**), whole genome or exome sequencing is necessary to uncover rare variants. The models are not mutually exclusive and each contributes in varying amounts depending on the heritability and epidemiology of the trait in question (discussed in **chapter 9**).

With sufficiently large sample sizes and deep phenotyping, GWASs can also contribute to uncovering variants with small effect (infinitesimal model), or that only act on certain environmental or genetic backgrounds (broad-sense heritability model).

Furthermore GWASs are hypothesis free, testing for association across the entire genome, and they are *hypothesis generating*, highlighting candidate genes and genomic regions to be prioritised for further investigation (Stranger et al., 2011). This makes GWASs a powerful method for understanding genetic variation in common traits or diseases.

SNP Genotyping

GWASs rely on the ability to simultaneously genotype hundreds of thousands of variants in one affordable assay, this is achieved using DNA microarrays (LaFramboise, 2009).

Illumina SNP arrays were used in this project (**Table 1.2**). The Illumina whole-genome genotyping microarray uses a two colour, single-base extension (SBE) assay (Steemers et al., 2006). Hundreds of thousands of 50-mer oligonucleotide primer sequences are attached to beads on a BeadArray (Gunderson et al., 2004). These primers are complementary to the sequence directly upstream (5') of the position of the SNPs. DNA is amplified by $1,000\text{--}3,000\times$ and fragmented to 300-600bp to drive hybridisation to the primers. The primer is then enzymatically extended at the 3' end by a polymerase that incorporates a single labeled dideoxynucleotide, depending on the complementarity to the genotype of the SNP. The dideoxynucleotides ddCTP and ddGTP are labeled with biotin, ddATP and ddUTP are labeled with 2,4-dinitrophenol (DNP). The labeled dideoxynucleotides are then detected by staining with antibodies, resulting in a red signal if the SNP allele is **A** or **T**, and a green signal if the SNP allele is **G** or **C**. Therefore the only SNPs that cannot be detected in this assay are A/T and G/C SNPs, which account for 16.5% of SNPs in the genome (Steemers et al., 2006). However the genotypes of these can

be detected using an allele-specific primer extension (ASPE) assay that has two probes per SNP instead of one (this can be incorporated onto the same BeadArray as the SBE assays) (Pastinen et al., 2000). Alternatively they can be inferred through imputation.

Copy number variants (CNVs) can also be detected using this technology by combining information about allele frequency and signal intensity from runs of adjacent SNPs in the genome (methods detailed in **chapter 7 & Figure 7.1**).

Box 2. Significance thresholds in genome-wide association studies

GWASs involve testing hundreds of thousands, or millions, of variants for association with a specific trait or disorder. The accepted threshold value for significance in GWASs is $P \leq 5 \times 10^{-8}$, this is a one in 50 million probability of observing the association by chance. This threshold is determined by performing a Bonferroni correction assuming one million independent tests, $0.05 / 1 \times 10^6 = 5 \times 10^{-8}$ (Hirschhorn and Daly, 2005).

An alternative method for determining a significance threshold depends not only on the number of tests performed, but the a priori probability that there is a true association at any specified location in the genome (The Wellcome Trust Case-Control Consortium, 2007). This is given by the following equation:

$$\text{posterior odds for true association} = \text{prior odds} \times \text{power} / \text{significance threshold}$$

Calculating this is further complicated by the fact that the SNPs genotyped are not independent of each other due to linkage disequilibrium. However, if we assume there are one million independent regions of the genome, and ten detectable associations, then the prior odds are $10 / 1 \times 10^6 = 1 \times 10^{-5}$. If the power for detecting association is 0.5 and the threshold for significance is 5×10^{-7} , then the posterior odds for true association are $(1 \times 10^{-5} \times (0.5 / 5 \times 10^{-7})) = 10$. This means that the posterior odds in favour of a 'hit' being a true association is 10:1. This method relies on assumptions on the setting of parameters in the equation, on balance a conservative threshold of 5×10^{-8} will significantly reduce the risk of false positives (posterior odds 100:1) at the cost of true negatives that do not meet the threshold for significance.

3.3 Methods

For a description of the study participant ascertainment and the relative hand skill phenotype, see **section 2.1 & subsection 2.2.1**.

3.3.1 Genotyping and quality control

Individuals from cohort 1 ($n = 197$) were genotyped on the Illumina[®] 550k Platform according to the manufacturer's instructions. One individual was excluded based on having a call rate $\leq 95\%$. Andrew Morris filtered out SNPs with either; minor allele frequency (MAF) ≤ 0.05 (due to the small sample size), or successful genotype call rate ≤ 0.95 , or violation of Hardy-Weinberg equilibrium ($P \leq 5.7 \times 10^{-7}$; in autosomes). In total, 477,347 SNPs passed quality control. Principal Component Analysis (PCA) to test for individuals with divergent ancestry compared to CEU HapMap3 samples was performed as detailed previously (Anderson et al., 2010), and as a result four further individuals were excluded.

In order to investigate population structure within the sample, I repeated our principal components analysis without the HapMap3 samples and tested for association between PegQ and the first two principal components (PC, adjusting for age and sex) in a linear regression model. There was no evidence of association with either component ($P = 0.244$ for PC1 and $P = 0.184$ for PC2). I thus did not adjust for these components in the subsequent GWAS analysis.

Cohort 2 ($n = 376$) were genotyped for two SNPs, rs11855415 and rs9806256 using iPLEX assays from Sequenom[®] (San Diego, California, US), according to the manufacturer's instructions. I also re-genotyped the original 197 individuals (co-

hort 1) with these assays, and found 99% concordance for the directly genotyped rs9806256 and 98% concordance for the imputed SNP rs11855415. The Hardy Weinberg Equilibrium P values for the SNPs were; rs11855415 $P = 0.24$, rs9806256 $P = 0.82$. The Integrated Genotyping System (IGS) (Fiddy et al., 2006) was used to store genotypes.

The same two SNPs were genotyped in the ALSPAC children cohort by KBiosciences, Hertfordshire, UK (www.kbioscience.co.uk), using a fluorescence-based competitive allele-specific PCR (KASPar) assay. The Hardy Weinberg P values for the two SNPs were; rs11855415 $P = 0.018$, rs9806256 $P = 0.248$.

3.3.2 Imputation

To estimate the genotypes of SNPs not directly genotyped on the Illumina 550K platform, I used IMPUTE [v2.0.5 for autosomes, (Howie et al., 2009) and v1.0.0 for the X chromosome, (Marchini et al., 2007)]. Inferring the missing genotypes involved combining information from;

- i) Directly genotyped SNPs.
- ii) A densely genotyped reference panel (60 CEPH founders from HapMap Phase II, build 35).
- iii) A fine-scale recombination map.

Imputed SNPs were filtered out if they had an $MAF \leq 0.05$ or $proper\ info \leq 0.5$ (a measure of the quality of the imputation), allowing me to test a further 1,709,209 SNPs for association.

3.3.3 Statistical analysis

Directly genotyped and imputed SNPs were tested for association with the quantitative measure PegQ under an additive model using SNPTEST (v1.1.4, mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html) simultaneously specifying sex and age as covariates. X chromosome analysis in cohort 1 was conducted by running the association test separately for the two sexes in SNPTEST and combining the results using GWAMA [v1.4, (Magi and Morris, 2010)], which performs fixed effects meta-analysis. The results for the three cohorts were then meta-analyzed using GWAMA. The values of PegQ were standardised in all cohorts separately to give a mean of 0, and a SD of 1. For the two SNPs genotyped in ALSPAC in the general population, the absolute value of the standardised PegQ score was also tested as a quantitative phenotype (i.e. comparing scores close to the mean against scores at both tails of the PegQ distribution).

To rule out any underlying confounding effects I tested the two SNPs taken forward for replication, rs11855415 and rs9806256, for association with measures of hand motor skill, IQ and reading ability (CCI, CCN, OLSON, READ, SPELL, SPOON, defined in **section 2.1**), in each of the three cohorts as well as in the general population cohort 4 whenever data were available, and found no significant association (**Table 3.3**).

3.3.4 Fine mapping

16 Individuals from the original GWAS were resequenced over exons 14-23 plus the three 3' UTRs of alternatively spliced variants in the most strongly associated gene, *PCSK6*. A further 24 individuals, plus DNA from 12 neuronal cell lines were

resequenced over the putative secondary promoter in *PCSK6*. Primers were designed using primer3 (<http://frodo.wi.mit.edu/primer3/>). PCR reactions were performed using Bioline BIOTAQ kits. Typically, 50ng of DNA was used for each reaction. The PCR programme typically used was a variant of Touchdown PCR (Don et al., 1991), which uses high annealing temperatures in the early cycles to ensure high specificity between the primers and the target sequences:

- 1) Hot Start 95°C 5 minutes
- 2) Denaturation 95°C 30 seconds
- 3) Annealing 64°C 30 seconds -0.5°C per cycle
- 4) Elongation 72°C 30 seconds

Go to step 2) 20 times

- 5) Denaturation 95°C 30 seconds
- 6) Annealing 54.5°C 30 seconds
- 7) Elongation 72°C 30 seconds

Go to step 5) 15 times

- 8) Elongation 72°C 5 minutes
- 9) Hold 12°C

Following amplification by PCR, the sequencing reactions were performed using BigDye v3.1 Chemistry (Applied Biosystems). The sequencing was carried out

on an ABI 3730xl sequencer (Applied Biosystems) using an automated sequencing system based on the Sanger dideoxy-chain termination method (Sanger et al., 1977). Analyses of the sequence data were performed using Phred software (Ewing et al., 1998) and visualised using Consed (Gordon et al., 1998).

3.4 GWAS results

After the initial filtering and quality control cohorts, I tested 192 individuals for association with PegQ. There was no evidence of population structure and the genomic control statistic (λ) is 1.0031. No SNPs gave P values below 5×10^{-8} in cohort 1 (see **Table 3.2** & **Figure 3.1** & **Figure 3.4**). The strongest association signal for directly genotyped SNPs is rs9806256 ($P = 1.1 \times 10^{-6}$), while for imputed SNPs peak association was observed with rs11855415 ($P = 4.7 \times 10^{-7}$). Both SNPs are located within a cluster of 5 highly correlated SNPs, spanning 12kb from introns 14-18 of *PCSK6* (Proprotein convertase subtilisin/kexin type 6, also known as *PACE4*) on chromosome 15q26. Individuals with the minor (derived) allele have significantly greater relative right-hand skill compared with those carrying the major (ancestral) allele. The mean effect size of each copy of the minor allele is 0.60 standard deviations (SD) to the positive (right-handed) end of the PegQ distribution (β' in **Table 3.1**).

3.4.1 Replication

PCSK6 is known to play a key role in regulating left-right axis specification (Constam and Robertson, 2000), making it an attractive candidate gene for involvement in handedness. I therefore followed up the initial association by genotyping rs11855415 and rs9806256 in an independent sample of individuals with RD who

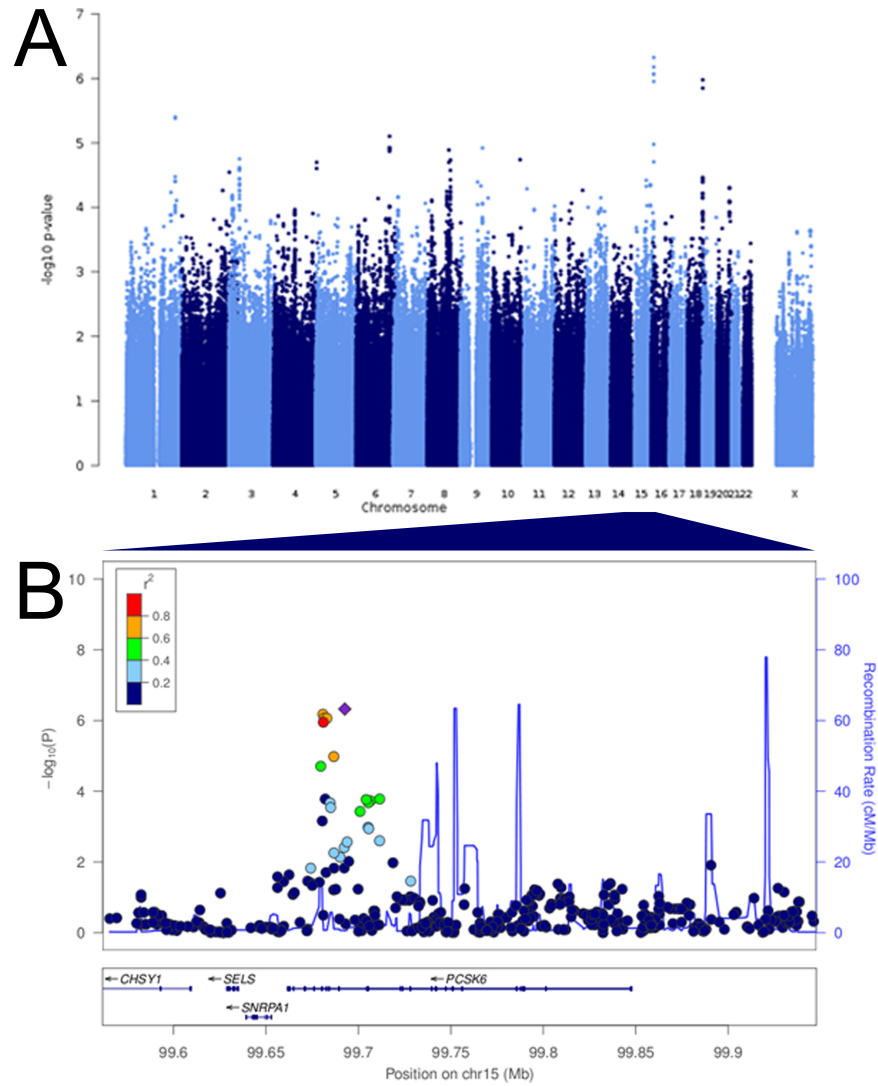


Figure 3.1: *PCSK6* is associated with relative hand skill in individuals with reading disability in the GWAS of cohort 1. (A) Manhattan plot of the P values for the meta-analysis of the RD Peg relative hand skill GWAS data-sets. The observed negative $\log_{10}$ transformed P values are plotted against chromosome and position along the genome. (B) Negative $\log_{10}$ of the P values for all genotyped and imputed SNPs in the *PCSK6* locus plotted in LocusZoom (Pruim et al., 2010). Linkage disequilibrium (r^2) with the most highly associated SNP, rs11855415 (♦), is shown by the colour of the SNPs. The recombination rate is shown by the blue line and the locations of genes in this locus are shown in the panel below the plot.

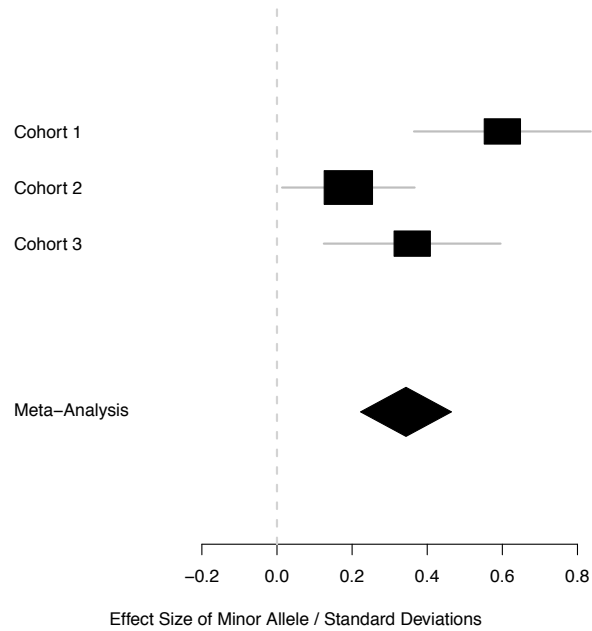
had also performed the same peg-board task (cohort 2, $n = 376$). The variants showed the same trend for increased right-hand skill in carriers of the minor allele; and the association was nominally significant with rs11855415 ($P = 0.033$, $\beta = 0.19$, **Table 3.1**), but did not meet significance for rs9806256.

Next, I genotyped these same two SNPs in individuals from the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort who have performed a similar peg-board task, and have been characterized by similar reading measures to the two previous cohorts. I selected individuals (cohort 3, $n = 197$) to closely match the ascertainment criteria for RD used in cohorts 1 and 2. The analyses show a significant increase in relative right-hand skill associated with the minor allele of both rs11855415 ($P = 0.0025$, $\beta = 0.36$, **Table 3.1**) and rs9806256 ($P = 0.00067$, $\beta = 0.40$).

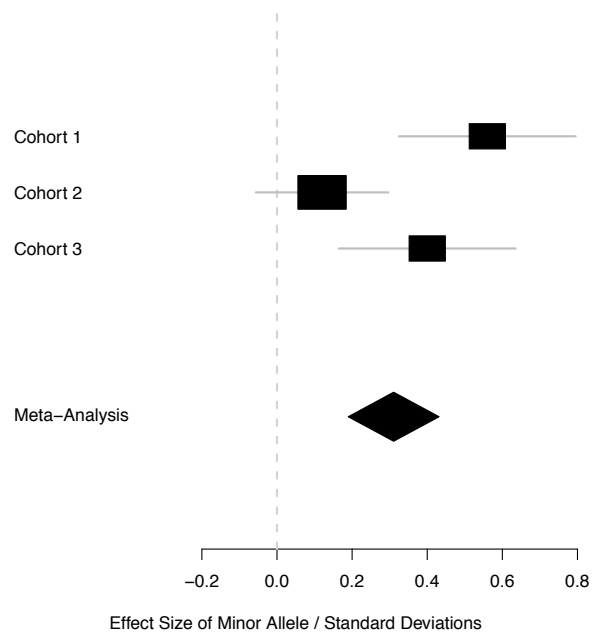
Meta-analysis of all 3 cohorts yielded a P value of 2.0×10^{-8} (**Table 3.1**) for rs11855415, exceeding accepted genome-wide thresholds for significance. For rs9806256, the meta-analysis yielded a P value of 2.34×10^{-7} . In the meta-analysis of RD individuals the average effect size of the minor allele is 0.35 SD towards the positive end of the PegQ distribution (**Figure 3.2**).

General population cohort

I tested the two SNPs for association in all the available ALSPAC children with a PegQ score ($n = 2,666$), excluding those with neurodevelopmental disorders and performance IQ ≤ 85 (cohort 4, **section 2.1**). Compared to the ALSPAC children with RD in cohort 3, the PegQ distribution shows neither a significant difference in the mean ($P = 0.54$), consistent with a previous study (Francks et al., 2003b),



(a)



(b)

Figure 3.2: Forest plot for relative hand skill GWA meta-analysis of *PCSK6* SNPs in individuals with dyslexia. (a) rs11855415 and (b) rs9806256.

Table 3.1: **PCSK6 is associated with relative hand skill in individuals with reading disability.** Association statistics for the most highly associated SNP, rs11855415, are given for each cohort of the study.

<i>Cohort</i>	<i>n</i>	<i>P value</i>	β	<i>SE</i>
1	185	4.70×10^{-7}	0.60	0.12
2	374	0.033	0.19	0.09
3	169	0.0025	0.35	0.12
Meta-analysis	728	1.99×10^{-8}	0.35	0.06
General Population (4)	2,666	0.25	-0.038	0.03

nor a difference in the likelihood of being left- or right-handed ($P = 0.69$). Furthermore, the distribution of PegQ is not significantly different between the two groups (**Figure 3.3**). In this general population sample I did not detect significant association between rs11855415 and relative hand skill ($P = 0.25$, $\beta = -0.038$). I then considered whether the variant may affect degree of handedness instead of direction. To test this I used the absolute value of the standardised PegQ score as a quantitative trait to test rs11855415 for association. This compared scores closer to the mean against scores at both tails of the PegQ distribution. This analysis shows a nominally significant trend towards the mean of the PegQ distribution for carriers of the minor allele ($P = 0.0020$).

Finally, I found no significant association between the SNPs and the reading, hand motor skill, and cognitive traits that were collected (**Table 3.3**).

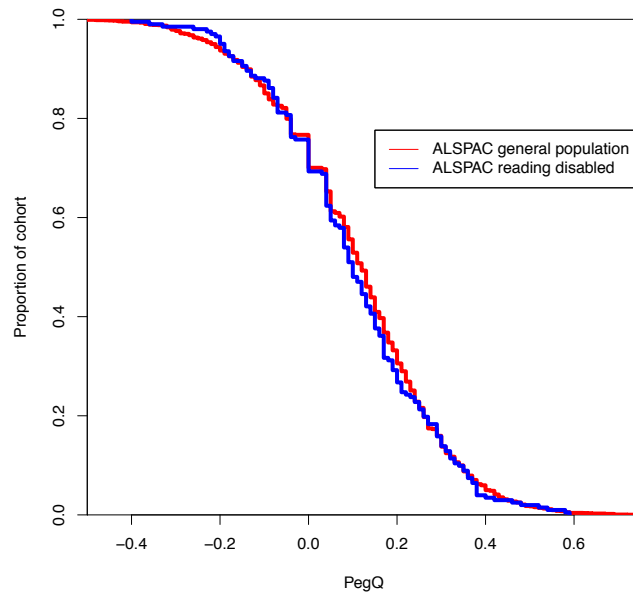


Figure 3.3: Empirical cumulative distribution function comparing PegQ in the **ALSPAC reading disabled** cohort 3 and the **ALSPAC general population** cohort 4. The two distributions are not significantly different from each other (two-sample Kolmogorov-Smirnov test $P = 0.71$).

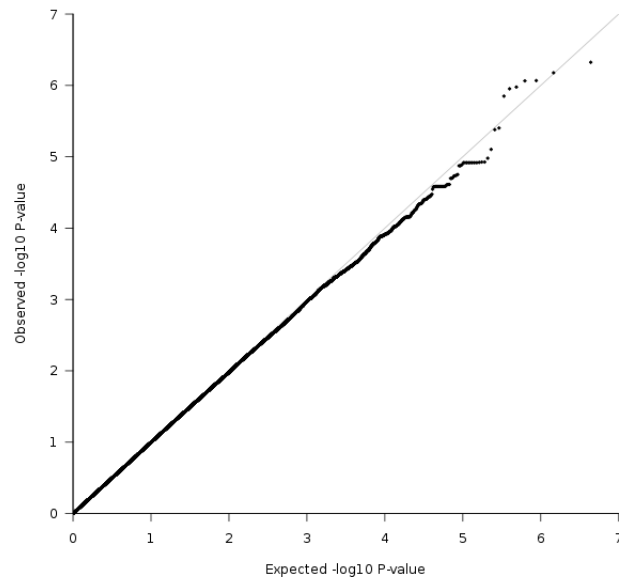


Figure 3.4: QQ plot for relative hand skill GWAS of cohort 1. The observed P values are plotted against the expected P values, a deviation from the null (diagonal line) indicates an enrichment of genuine associations, population stratification or an error in genotype calling.

Table 3.2: **A list of the 40 most strongly associated SNPs from the cohort 1 relative hand skill GWAS.** The genes to the left or right of the SNP are listed. MAF = Minor Allele Frequency, Info = quality of imputation, β = effect size in SD, SE = Standard Error, HWE = Hardy Weinberg Equilibrium, _i = imputed SNP.

<i>Rank</i>	<i>Chr</i>	<i>Position</i>	<i>Marker</i>	<i>MAF</i>	<i>Info</i>	β	<i>SE</i>	<i>HWE P</i>	<i>P</i>	<i>Gene</i>	<i>Feature</i>	<i>Left Gene</i>	<i>Right Gene</i>
1	15_i	99692646	rs11855415	0.25	0.95	-0.6	0.12	1	4.73×10^{-7}	<i>PCSK6</i>	intron	<i>SNRPA1</i>	<i>TM2D3</i>
2	15_i	99680812	rs1871978	0.28	0.98	0.58	0.12	0.59	6.65×10^{-7}	<i>PCSK6</i>	intron	<i>SNRPA1</i>	<i>TM2D3</i>
3	15_i	99681512	rs4965830	0.33	0.97	0.56	0.11	0.25	8.55×10^{-7}	<i>PCSK6</i>	intron	<i>SNRPA1</i>	<i>TM2D3</i>
4	15_i	99682995	rs7182874	0.33	0.97	0.56	0.11	0.25	8.64×10^{-7}	<i>PCSK6</i>	intron	<i>SNRPA1</i>	<i>TM2D3</i>
5	18_i	64725259	rs7238095	0.46	0.96	-0.52	0.11	0.47	1.05×10^{-6}	<i>CCDC102B</i>	intron	<i>TXNDC10</i>	<i>DOK6</i>
6	15	99681096	rs9806256	0.28	1	-0.56	0.12	0.59	1.11×10^{-6}	<i>PCSK6</i>	intron	<i>SNRPA1</i>	<i>TM2D3</i>
7	18_i	64710365	rs2051294	0.46	0.97	-0.51	0.11	0.47	1.42×10^{-6}	<i>CCDC102B</i>	intron	<i>TXNDC10</i>	<i>DOK6</i>
8	1_i	209459736	rs1284855	0.3	1	-0.47	0.1	0.03	3.96×10^{-6}	-	-	<i>FLVCR1</i>	<i>VASH2</i>
9	1_i	209465840	rs6702876	0.31	0.99	-0.47	0.1	0.03	4.17×10^{-6}	-	-	<i>FLVCR1</i>	<i>VASH2</i>
10	6_i	142405419	rs12206319	0.23	0.98	-0.53	0.12	0.68	7.89×10^{-6}	-	-	<i>LOC729076</i>	<i>LOC645503</i>
11	15	99686708	rs752028	0.33	1	0.5	0.11	0.25	1.05×10^{-5}	<i>PCSK6</i>	intron	<i>SNRPA1</i>	<i>TM2D3</i>
12	6_i	142401465	rs12212141	0.24	0.98	-0.51	0.12	0.55	1.19×10^{-5}	-	-	<i>LOC729076</i>	<i>LOC645503</i>
13	6_i	142402402	rs4541765	0.24	0.98	-0.51	0.12	0.55	1.19×10^{-5}	-	-	<i>LOC729076</i>	<i>LOC645503</i>
14	9	94425992	rs10739976	0.46	1	0.46	0.11	0.25	1.20×10^{-5}	<i>FBP2</i>	intron	<i>HIATL1</i>	<i>FBP1</i>
15	6_i	142389959	rs11155233	0.24	0.97	-0.51	0.12	0.54	1.21×10^{-5}	-	-	<i>LOC729076</i>	<i>LOC645503</i>
16	6_i	142386917	rs13192891	0.24	0.97	-0.51	0.12	0.54	1.21×10^{-5}	-	-	<i>LOC729076</i>	<i>LOC645503</i>
17	6_i	142397679	rs4128460	0.24	0.97	-0.51	0.12	0.54	1.21×10^{-5}	-	-	<i>LOC729076</i>	<i>LOC645503</i>
18	6_i	142391771	rs9496232	0.24	0.97	-0.51	0.12	0.54	1.21×10^{-5}	-	-	<i>LOC729076</i>	<i>LOC645503</i>
19	6_i	142387385	rs11968035	0.24	0.97	0.51	0.12	0.54	1.21×10^{-5}	-	-	<i>LOC729076</i>	<i>LOC645503</i>
20	6_i	142389596	rs12213080	0.24	0.97	0.51	0.12	0.54	1.21×10^{-5}	-	-	<i>LOC729076</i>	<i>LOC645503</i>
21	6_i	142388438	rs7764878	0.24	0.97	0.51	0.12	0.54	1.21×10^{-5}	-	-	<i>LOC729076</i>	<i>LOC645503</i>
22	6_i	142391691	rs9496231	0.24	0.97	0.51	0.12	0.54	1.21×10^{-5}	-	-	<i>LOC729076</i>	<i>LOC645503</i>
23	8_i	90092828	rs6469727	0.44	0.99	0.46	0.11	0.47	1.29×10^{-5}	-	-	<i>LOC100129100</i>	<i>RIPK2</i>

24	6_i	142450400	rs11155234	0.23	0.98	-0.51	0.12	0.42	1.34×10^{-5}	NMBR	intron	LOC645503	VTA1
25	6_i	142430974	rs12210427	0.23	0.98	-0.51	0.12	0.42	1.34×10^{-5}	-	-	LOC645503	NMBR
26	3_i	45015686	rs4683019	0.43	0.98	0.42	0.1	0.1	1.77×10^{-5}	EXOSC7	intron	ZDHHC3	CLEC3B
27	10_i	119161661	rs7922735	0.13	0.51	-1.18	0.13	1	1.82×10^{-5}	-	-	PDZD8	EMX2OS
28	8_i	97042119	rs1155562	0.13	0.99	-0.67	0.16	0.74	1.84×10^{-5}	-	-	C8orf37	LOC643228
29	8_i	97044921	rs1484629	0.13	0.98	0.67	0.16	0.74	1.86×10^{-5}	-	-	C8orf37	LOC643228
30	15	99679686	rs3825921	0.4	1	0.45	0.11	0.45	1.97×10^{-5}	PCSK6	intron	SNRPA1	TM2D3
31	8	90067358	rs1586475	0.44	1	-0.45	0.11	0.66	2.01×10^{-5}	-	-	LOC100129100	RIPK2
32	4	184067228	rs17330237	0.13	1	-0.62	0.15	0.21	2.01×10^{-5}	ODZ3	intron	MGC45800	DCTD
33	3_i	45059406	rs4683031	0.43	0.98	0.41	0.1	0.14	2.44×10^{-5}	-	-	CLEC3B	CDCP1
34	3_i	45062035	rs6790660	0.43	0.98	-0.41	0.1	0.14	2.44×10^{-5}	-	-	CLEC3B	CDCP1
35	3_i	45067810	rs4323037	0.43	0.98	0.41	0.1	0.14	2.45×10^{-5}	-	-	CLEC3B	CDCP1
36	4_i	184067015	rs17330216	0.13	0.91	-0.63	0.15	0.2	2.49×10^{-5}	ODZ3	intron	MGC45800	DCTD
37	3_i	45032054	rs9852384	0.43	1	-0.41	0.1	0.14	2.59×10^{-5}	-	-	EXOSC7	CLEC3B
38	3_i	45055574	rs1520079	0.43	1	0.41	0.1	0.14	2.61×10^{-5}	-	-	CLEC3B	CDCP1
39	3_i	45053186	rs9880792	0.43	1	0.41	0.1	0.14	2.61×10^{-5}	-	-	CLEC3B	CDCP1
40	3_i	45042421	rs1004877	0.43	1	0.41	0.1	0.14	2.61×10^{-5}	CLEC3B	near-gene-5	EXOSC7	CLEC3B

Table 3.3: **There is no association with reading, hand motor skill, and cognitive traits for the SNPs in PCSK6 associated with relative hand skill** (Bonferroni correction $P = 0.0018$). For the correlation between PegQ, the reading-related phenotypes, and IQ measures, see Table 2.1. A dash (-) indicates that no phenotype was available to test in that cohort.

<i>Trait</i>	<i>Marker</i>	<i>Cohorts</i>			
		1	2	3	4
Hand motor skill	rs11855415	0.28	0.46	0.75	0.34
	rs9806256	0.02	0.90	0.81	0.52
CCI	rs11855415	0.36	0.79	-	-
	rs9806256	0.30	0.91	-	-
CCN	rs11855415	0.24	0.28	-	-
	rs9806256	0.06	0.24	-	-
OLSON	rs11855415	0.20	-	-	-
	rs9806256	0.16	-	-	-
READ	rs11855415	0.98	0.28	0.41	0.31
	rs9806256	0.28	0.45	0.45	0.41
SPELL	rs11855415	0.90	0.26	-	-
	rs9806256	0.76	0.37	-	-
SPOON	rs11855415	0.86	-	-	-
	rs9806256	0.70	-	-	-
Performance IQ	rs11855415	0.91	0.65	0.81	0.48
	rs9806256	0.72	0.49	0.98	0.31
Verbal IQ	rs11855415	0.53	0.82	0.62	0.02
	rs9806256	0.20	0.53	0.76	0.05
Total IQ	rs11855415	0.67	0.94	0.86	0.11
	rs9806256	0.53	0.97	0.85	0.11

3.4.2 Fine mapping of *PCSK6* locus

The region of peak association (**Figure 3.1**) is located between introns 14-18 towards the 3' end of *PCSK6*. Therefore, in order to fine map the region to help identify the etiologic variant, first I sequenced exons 14-23 (plus flanking intronic sequence) and the three 3' UTRs of alternatively spliced variants of *PCSK6* that are located in this locus. This includes all the exons within the recombination hotspots that flank the region of peak association (**Figure 3.1**). 16 individuals from the original GWAS were chosen, representing the full distribution of both genotypes for rs11855415 and PegQ phenotype scores. Six variants were identified, four in the 3' UTRs and two intronic variants close to exons (**Table 11.1**). All the variants were previously known and of these, rs1030 and rs34318939 are in linkage disequilibrium (LD) with rs11855415. Rs1030 is located in a 3' UTR, however it was imputed in the original GWAS ($P = 0.0002$) and owing to its weaker association it is unlikely to represent the etiologic variant. Next, rs34318939 is located 48bp upstream of exon 16 and is a single nucleotide deletion of an adenine base. Its proximity to an exon suggests it may be a branch point for splicing, however the majority of branch points for splicing in humans are located 21-34 nucleotides upstream of the 3' end of an intron with a consensus sequence different to that observed around this SNP (Gao et al., 2008).

The sequencing of the exons and 3' UTRs suggest it is unlikely that the etiologic variant is a coding change in *PCSK6*. Next I turned my attention to the possibility that the etiologic variant may represent a regulatory change in an intron of the gene. To search for putative regulatory regions I utilised the ENCODE (Encyclopedia Of DNA Elements) regulatory tracks on the UCSC genome browser (<http://genome.ucsc.edu/>) detailing; enrichment of histone modifications

suggestive of enhancer and promoter activity, DNase clusters indicating open chromatin, regions of transcription factor binding, and transcription levels. The most highly associated SNP, rs11855415, is located within a region that shows evidence of all of these indicators of gene regulation (**Figure 3.5 & Figure 3.6**). Also, there are two spliced expressed sequence tags (ESTs) with exons originating in this locus, one in a sense direction to the gene, and the other antisense (**Figure 3.6**). Taken in combination, these data provide evidence that the most highly associated SNP is located in a novel secondary promoter of *PCSK6*.

I therefore sequenced the entire promoter region using overlapping primers in 24 individuals from the original GWAS. The sequencing identified two SNPs other than rs11855415; rs12900794 and rs12916087 (**Table 11.2** in the appendix). Both are located in transcription factor binding sites, and they are in weak linkage disequilibrium with rs11855415. However both were imputed in the original GWAS with *P* values of 0.004 and 0.02 respectively. Therefore assuming the imputation was accurate, they are unlikely to be causative.

There is also a variable number tandem repeat (VNTR) in the promoter region located 522bp downstream of rs11855415, which is flanked by the two DNase hypersensitivity clusters (**Figure 3.6**). The VNTR consists of 33bp repeated either nine times (ancestral) or six times (i.e. a derived 99bp deletion – rs10523972). This repeat is unique to humans and chimpanzees at this locus, and is not found in gorillas or other vertebrates. However, this could be an artefact of genome assembly rather than a genuinely novel sequence that has evolved since the last common ancestor of humans, chimpanzees and gorillas. The gorilla genome was sequenced by 2x coverage of ABI capillary-sequenced whole-genome shotgun read pairs with

Table 3.4: **The VNTR located in the secondary promoter of PCSK6 is not associated with relative hand skill (PegQ).** See Figure 3.6 for a visualisation of the genomic region around this VNTR.

<i>Cohort</i>	<i>n</i>	<i>MAF</i>	<i>P</i>	β	<i>SE</i>
1	170	0.20	0.44	-0.10	0.13
2	321	0.18	0.13	-0.16	0.10
Meta-analysis	491	0.18	0.10	-0.13	0.08

an average read length of 546bp, and then $\sim 50\times$ coverage of Illumina short read pairs, mostly 35-37 bp in length (Scally et al., 2012). The low coverage of long reads coupled with the repeat unit length of the VNTR relative to the short reads (33bp versus 35bp) means that the precise copy number of this VNTR may not have been accurately resolved by this sequencing method in gorilla.

I genotyped the VNTR through PCR amplification and gel electrophoresis. Given the large size of the polymorphism (99bp), it was possible to resolve the deletion on a gel without Sanger sequencing. The six copy repeat allele has a frequency of 0.18 in our cohorts, comparable to the associated variant rs11855415 ($MAF = 0.24$), however they are not in linkage disequilibrium ($r^2 = 0.050$) and the VNTR is not associated with PegQ (**Table 3.4**).

I also genotyped a panel of 12 neuronal cell lines for both rs11855415 and the VNTR, none of the cell lines were homozygote for the minor allele of rs11855415 (**Table 11.3** in the appendix).

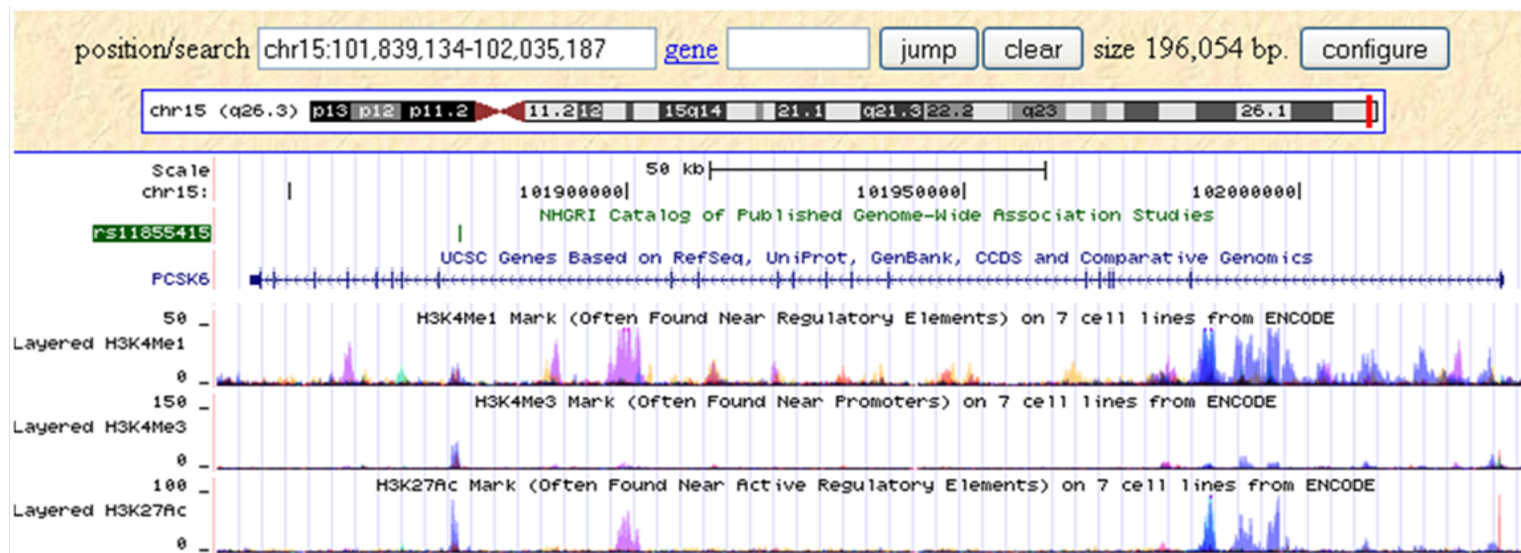


Figure 3.5: UCSC genome browser view of the ENCODE promoter and enhancer regions in *PCSK6*. The exons in the gene track are shown by vertical bars. The green bar represents the genome-wide significant SNP rs11855415. Levels of three different histone modifications are shown for seven cell lines: GM12878 lymphoblastoid, H1-hESC stem, HSMM skeletal muscle myoblast, HUVEC human umbilical vein endothelial, K562 immortalised myelogenous leukemia, NHEK keratinocyte, and NHLF normal human lung fibroblast.

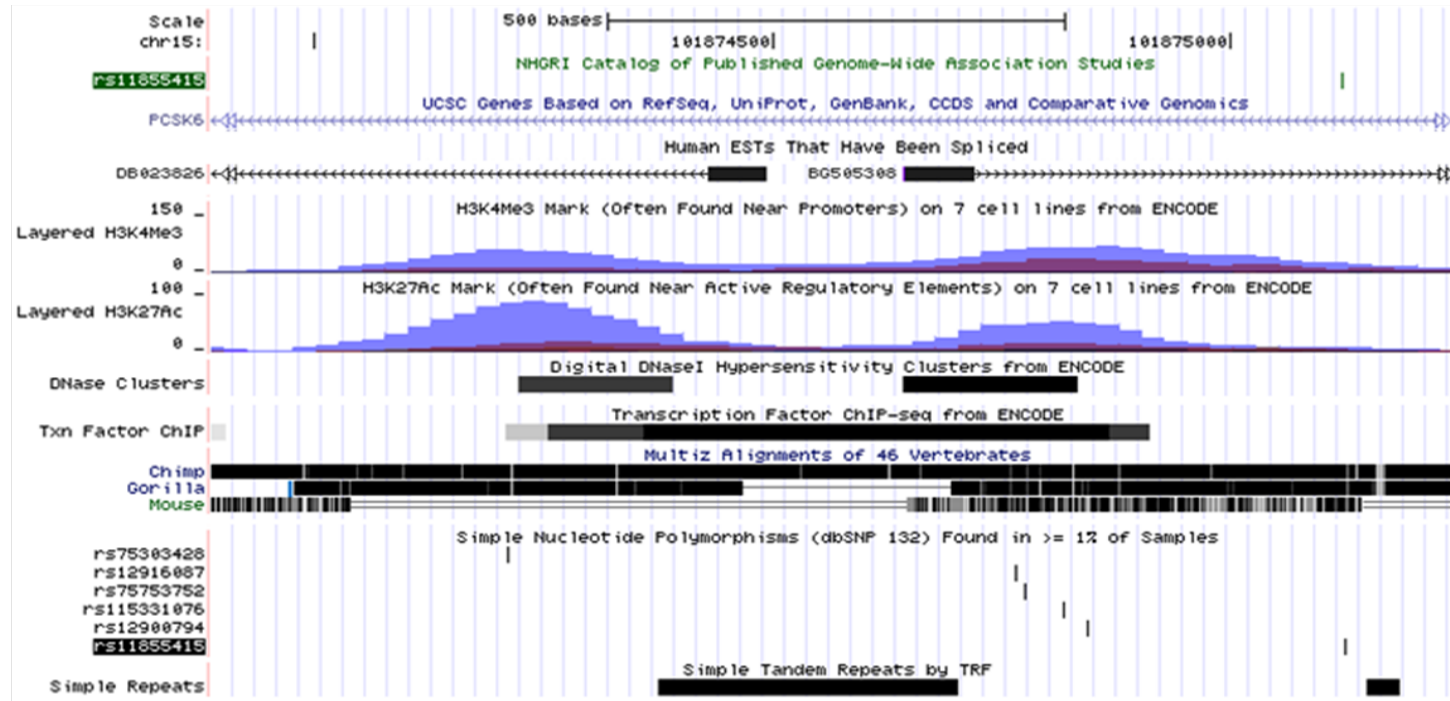


Figure 3.6: **UCSC genome browser view of the secondary promoter in *PCSK6*.** From top to bottom after the *PCSK6* gene track: 1) The black bars represent the first exons of two ESTs that originate in this locus, and the arrows represent the direction of transcription. 2) Tri methylation of lysine 4 in histone H3 (H3K4Me3) is an indicator of promoter activity. The peaks represent the **K562 immortalised myelogenous leukemia** and the **NHEK keratinocyte** cell lines, 3) Acetylation of lysine 27 on histone H3 (H3K27Ac) is an indicator of enhancer activity, 4) The grey bars represent DNaseI hypersensitive clusters; the degree of shading is proportional to the maximum signal strength observed, 5) The grey bar represents transcription factor binding sites, the degree of shading is proportional to the maximum signal strength observed, 6) Alignments of selected vertebrates to the human reference sequence at this locus, 7) SNPs, 8) simple repeats, including the VNTR.

3.5 Discussion

I have performed the first GWAS for a quantitative measure of relative hand skill, as assessed by the peg-board task (Annett, 1985). The most highly associated SNP, rs11855415, is within an intron of *PCSK6*. The minor allele of this SNP confers increased relative right-hand skill. This association replicated in two independent cohorts with RD and is the first SNP for handedness to be identified at a genome-wide significant level. This effect was specific for individuals with RD, while the minor allele of rs11855415 showed a significant trend towards reduced laterality of hand skill in the general population. Finally, fine-mapping of this locus did not reveal a clear candidate for the causative mutation.

PCSK6 is a proprotein convertase that processes proteins such as NODAL from their latent precursors into biologically active products (**Figure 3.7**, (Kiefer et al., 1991)). NODAL is a TGF β -related protein that specifies both the anteroposterior and left-right axes (Beck et al., 2002). Artificially reversing the direction of 'nodal flow' in mice causes asymmetry phenotypes such as situs inversus (Nonaka et al., 2002). In a proportion of mouse embryos null for *Pcsk6* (given in parentheses), the normally asymmetrically expressed *Nodal* (71%), *Lefty* (15%), and *Pitx2* (21%) mRNAs are bilaterally expressed (Constam and Robertson, 2000). 25% of embryos die prenatally, almost all with severe laterality defects such as heterotaxia (abnormal positioning of body organs), and / or heart asymmetry defects such as double outlet right ventricle and dextrocardia (Constam and Robertson, 2000). *PCSK6* is widely expressed in humans, with particularly high expression in the liver, corpus callosum (the band of axon fibres that connect the left and right hemispheres) and spinal cord (Johnson et al., 2003). Mutations in other NODAL pathway members have been linked to lateralisation defects such as holoprosencephaly, which is an

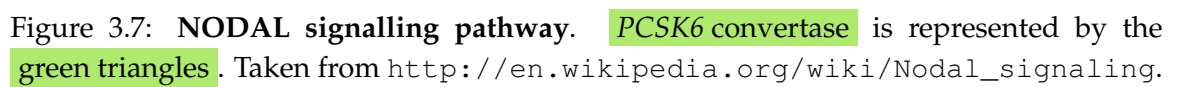
incomplete separation of the forebrain into discrete hemispheres (Mohapatra et al., 2009; Roessler et al., 2008).

The observation that the increase in relative right-hand skill associated with rs11855415 in *PCSK6* is specific to RD suggests there may be epistatic interaction between *PCSK6* and RD susceptibility genes. There have been a number of candidate genes proposed for RD, including *DCDC2* (Meng et al., 2005; Schumacher et al., 2006), *DYX1C1* (Taipale et al., 2003), *KIAA0319* (Cope et al., 2005; Francks et al., 2004), and *ROBO1* (Hannula-Jouppi et al., 2005). These genes are all involved in neuronal migration or axon guidance (Galaburda et al., 2006). It may be that the functional change in *PCSK6* subtly alters the initial left-right patterning of the early embryo, and this has a downstream effect during neuronal migration on the development of cerebral asymmetry, although we can only speculate at this stage. Future studies will seek to explain the difference between RD and general population subjects (also see **chapter 5 & chapter 9**).

The genome-wide significance of SNP association in *PCSK6*, coupled with the role of this gene in a known biological pathway for left-right asymmetry, make it a strong candidate for involvement in cerebral asymmetry and thus handedness.

3.5.1 Future perspectives

The results are consistent with the emerging view that handedness is determined by multiple interacting genetic and environmental factors. *PCSK6* has been found to be a direct target of the transcription factor FOXP2 (Vernes et al., 2007), whilst *PCSK6* may interact with *SRPX2* (Royer-Zemmour et al., 2008). Mutation in either *FOXP2* or *SRPX2* can lead to severe forms of language or speech impairment. It



would therefore be interesting to investigate the effect of *PCSK6* on handedness in individuals with other neurodevelopmental disorders such as SLI. Understanding the genes that modify the effects of *PCSK6* in RD individuals will offer us insight into the etiology of both cerebral asymmetry and dyslexia.

The functional work to identify the etiological variant and the effect it has on the regulation and transcription of *PCSK6* is now being undertaken by Robert Shore at the University of St. Andrews under the supervision of Silvia Paracchini. Thanks to funding from both Silvia Paracchini and Simon Fisher, I now turn my attention towards genotyping cohort 2 genome-wide, and performing GWA meta-analyses for both handedness and dyslexia, plus CNV analysis for dyslexia.

Chapter 4

Genotyping of dyslexia cases on the Illumina OmniExpress SNP array

In the previous study detailed above, we only had access to genome-wide high-density genotype data from cohort 1 (**chapter 3**). Below I describe the genotyping of the remaining Oxford UK dyslexia samples (cohort 2). Sibling and singleton DNA (but not parents) were genotyped at the Wellcome Trust Centre for Human Genetics (March 2012). In total 854 Human DNA samples were processed on the Illumina Infinium Human OmniExpress-12v1.0 beadchips (730,525 SNPs) using the Infinium HD Ultra protocol.

My role first was to prepare the DNA for genotyping. Then after quality control detailed below, I perform a GWAS using relative hand skill (PegQ) as a quantitative trait in the individual dyslexia cases (cohort 2, not including the siblings). The results of this are meta-analysed with the PegQ data from cohort 1 plus the dyslexia cases identified in ALSPAC (cohort 3, **chapter 5**). Then I perform both a dyslexia case/control and quantitative reading-related traits GWAS (**chapter 6**).

Finally I call and analyse CNVs in the dyslexia cohorts 1, 2, 5, and 6 (**chapter 7**). For a summary of all the cohorts, please refer to **Table 1.2**.

The genotyping is funded by Silvia Paracchini of the University of St. Andrews and Simon Fisher from the Max Planck Institute for Psycholinguistics in Nijmegen (MPI). The analysis will be performed in collaboration with Clyde Francks and Alessandro Gialluisi at the MPI (who will perform sibling-based analyses as part of a larger meta-analysis for dyslexia).

Controls will be used for a case / control dyslexia GWAS (**chapter 6**) and CNV analysis (**chapter 7**). We are collaborating with Bruce Winney and Sir Walter Bodmer, who have genotyped 335 British individuals on the Illumina OmniExpress chip at the WTCHG, as part of the People of the British Isles (POBI) study (Winney et al., 2012).

Furthermore, we have access to 1,167 controls (from the Wellcome Trust Case Control Consortium) that were genotyped on the Illumina 550K chip.

Both cohorts are unselected or phenotyped for reading ability, and therefore ~5-10% of controls will have dyslexia. However, the loss of power this causes is approximately the same as that due to a reduction of the sample size by 10% (Colhoun et al., 2003)

4.1 Preparation of DNA

To prepare the DNA for genotyping on the Illumina OmniExpress chip, I normalized it to $\sim 50\text{ng}/\mu\text{l}$ ($\pm 20\text{ng}/\mu\text{l}$).

First I quantified the DNA using the Quant-iTTM PicoGreen[®] kit (Invitrogen), an ultra sensitive fluorescent nucleic acid stain for quantifying double-stranded DNA (dsDNA) in solution, following the standard protocol.

For samples where the DNA concentration was below $40\text{ng}/\mu\text{l}$, vacuum centrifugation at 43°C was performed to concentrate the samples, followed by re-quantification using PicoGreen. Samples which had a DNA concentration less than $30\text{ng}/\mu\text{l}$ after vacuum centrifugation were excluded from genotyping. $10\mu\text{l}$ was aliquoted into 96-well skirted plates for samples with a DNA concentration of between $30\text{ng}/\mu\text{l}$ – $70\text{ng}/\mu\text{l}$. For those with a DNA concentration $\geq 70\text{ng}/\mu\text{l}$, the samples were diluted down to $50\text{ng}/\mu\text{l}$ and $10\mu\text{l}$ was aliquoted into 96-well skirted plates. Siblings were included in the same genotyping batch to minimize batch effects.

In total, there were 854 samples genotyped. 830 samples were independent, and a further 12 samples were done in triplicate. This is in order to assess the consistency in genotyping quality between samples.

A summary of the 854 samples is shown in **Table 4.1**.

Table 4.1: Summary of the numbers of cases and siblings genotyped on the Illumina OmniExpress chip. The 98 siblings are related to 80 of the 446 unrelated probands, and the other 264 siblings are related to 187 of the original cases genotyped on the Illumina 550K chip (cohort 1). The replicate samples are repeats of probands genotyped in cohort 1.

<i>Category</i>	<i>Number of Individuals</i>
Unrelated probands	446
Siblings	98
Siblings of 550K probands	264
Replicate samples	22
Triplicate samples	12
Total	854

4.2 Per-individual quality control

I followed the protocol laid out in Anderson et al., 2010 (Anderson et al., 2010). I also perform the quality control for cohort 1 again in parallel with this analysis (in **chapter 3**, QC was done by Andrew Morris).

The first stage of the quality control aimed to remove individuals who may introduce false positives or noise into the data because they:

- have discordant sex information.
- have a low genotyping success rate.
- are duplicated or related to other individuals.
- show divergent ancestry.

Samples that failed quality control and the reasons for their exclusion are summarized in **Table 4.2**. A more detailed explanation can be found in **Table 11.6** for cases

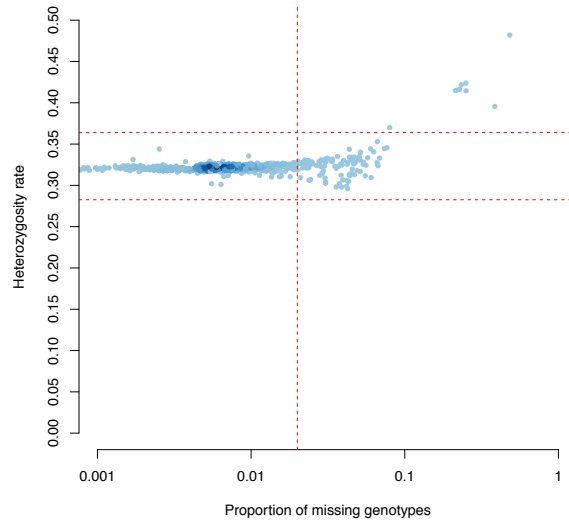
and siblings, and **Table 11.7** for controls.

4.2.1 Identification of individuals with discordant sex information

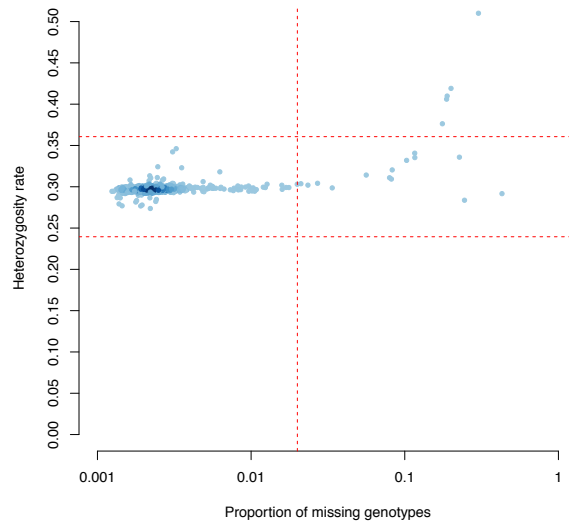
The sex of an individual can be determined by checking the heterozygosity of SNP calls on the X chromosome. Since males only have one copy of the X chromosome, they should be homozygous for all SNPs tested. If an individual shows a discordant sex based on genetic data compared to our phenotype data, this suggests either the individual has been labelled incorrectly at some stage of the data collection, or that the sample has been swapped or contaminated along the way. Therefore I removed all individuals with discordant sex information as a first QC step (**Table 4.2**).

4.2.2 Genotyping success rate

I plotted the distribution of the genotyping call rate against the heterozygosity rates in order to identify outlying individuals in **Figure 4.1**. 11 cases and 13 controls have call rates below the cutoff of 98%, and/or a heterozygosity rate ± 2 standard deviations from the mean. Individuals with low call rates indicate the sample is of low quality or is contaminated, this is highlighted by example plots of B allele frequency for chromosome 1 shown in **Figure 4.2**. The overall success rate was 91.4% for the 232 samples in the 550K dyslexia cases (cohort 1, plus 32 extra samples accidentally included from cohort 6), 91.5% for the 1,167 WTCCC 550K controls, 98.6% for the 830 samples in the OmniExpress dyslexia cases (cohort 2 plus siblings), and 96.1% for the 335 POBI OmniExpress controls (**Table 4.2**).



(a)



(b)

Figure 4.1: **Genotype failure rate versus heterozygosity** across cases, their siblings, and controls. Shading indicates sample density and dashed lines denote quality control thresholds. (a) Samples genotyped on the Illumina 550K chip. (b) Samples genotyped on the Illumina OmniExpress chip.

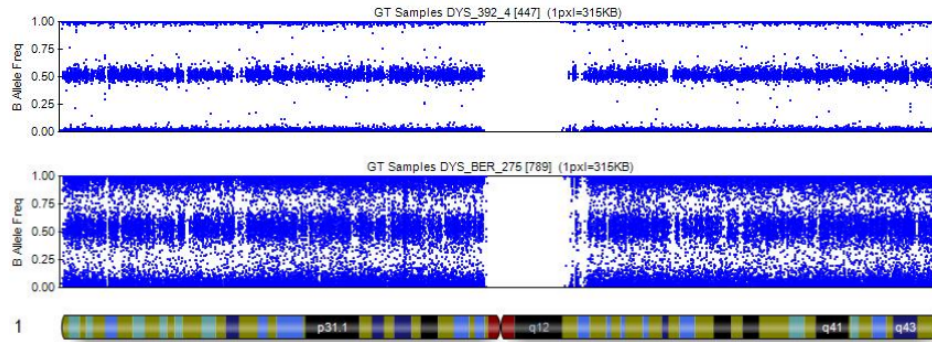


Figure 4.2: The B allele frequency (BAF) over chromosome 1 for two individuals, **DYS_392_4** has a genotyping success rate of 99.998%, compared to **DYS_BER_275**, which has an 89.02% success rate. BAF refers to the proportion of the genotyping signal from the SNP array that comes from a single allele. Since diploid organisms have two alleles at each locus, the expected BAF is either: 0, 0.5, or 1. Significant departures from these values suggest genotyping error. I chose a genotyping success rate quality control threshold of 98% to take forward for the GWAS.

4.2.3 Identification of duplicated or related individuals

If duplicate or related individuals are included in association studies, then the sample may no longer be representative of the allele frequencies observed in the entire population because genotypes within certain families will be overrepresented among the sample. To avoid such biases, I calculated the identity by state (IBS) metric for each pair of individuals in order to identify the average proportion of alleles shared in common at genotyped SNPs. In order to accurately measure this proportion, regions of extended linkage disequilibrium (LD) (such as the human leukocyte antigen (HLA) region) are removed from the data set and remaining regions are pruned to remove SNPs in LD with each other (defined as $r^2 \geq 0.2$).

The degree of shared ancestry for each pair (identity by descent, IBD) was calculated using PLINK (Purcell et al., 2007). $IBD = 1$ indicates duplicates or monozygotic twins, $IBD = 0.5$ indicates the pair are siblings. I chose an $IBD \geq 0.1875$ as the threshold to remove individuals who weren't already known to be siblings,

which lies halfway between second- (0.25) and third-degree relatives (0.125). I plotted three histograms, the first two for individuals who are expected to be unrelated from the Illumina 550K and OmniExpress projects (**Figure 4.3a** & **Figure 4.3b**), and the second for individuals from the same family which were genotyped (**Figure 4.3c**). If unrelated individuals had an $IBD \geq 0.1875$ or siblings had an $IBD \leq 0.35$ or ≥ 0.65 , they were removed from further analyses. Individuals who previously failed genotyping (**subsection 4.2.2**) were not included for this analysis.

4.2.4 Identification of individuals with divergent ancestry

Allele frequency differences between cases and controls can be due to different population origins rather than disease state. In order to control for population stratification, I identified individuals with divergent ancestry using principal component analysis (PCA), after removing individuals who failed genotyping (**subsection 4.2.2**). I excluded individuals with a second principal component score less than 0.072 (**Table 4.2**), indicating significant deviation from European ancestry based on the CEU HapMap3 reference samples. Furthermore, I used principal components as covariates in downstream analyses. The PCA plot is shown in **Figure 4.4**.

A summary of the numbers of individuals that failed QC is shown in **Table 4.2**. Finally, the full list of individuals that failed quality control, along with the reasons, are listed in the appendix (**section 11.2**) in **Table 11.4** for 550K cases, **Table 11.5** for the WTCCC controls, **Table 11.6** for OmniExpress cases and siblings, and **Table 11.7** for the POBI controls.

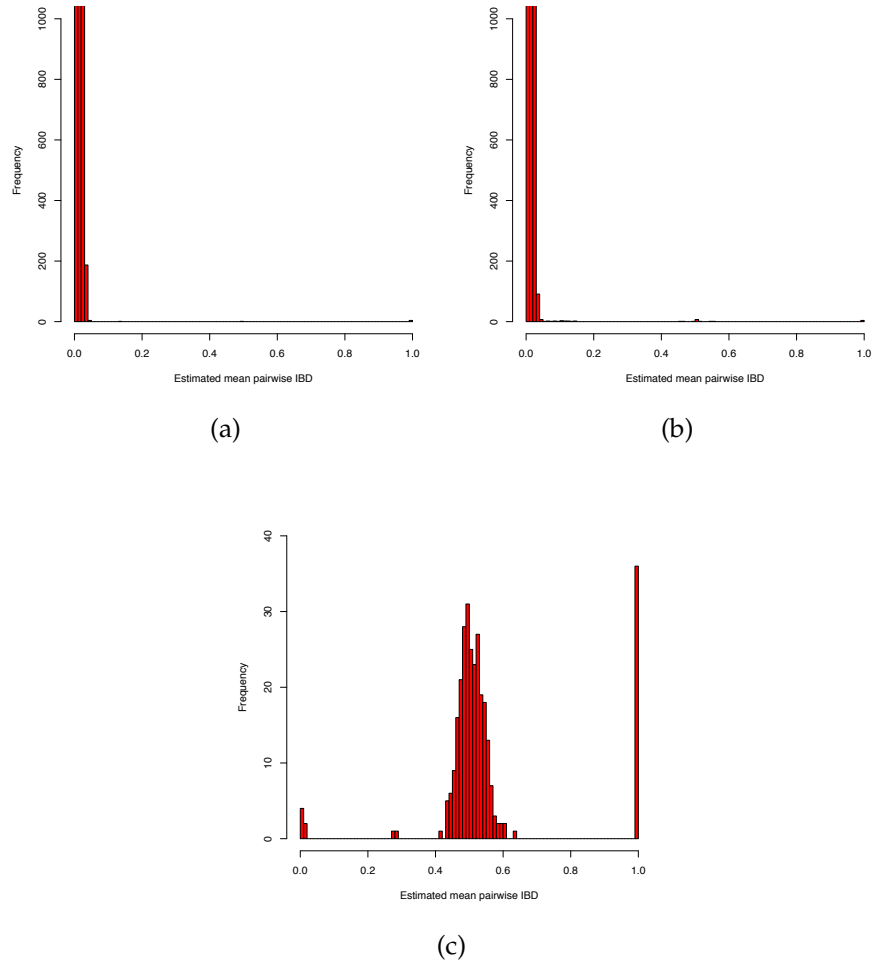


Figure 4.3: **Histogram of the pairwise IBD metric for all cases and controls.** (a) Cases and controls genotyped on the Illumina 550K chip. (b) Unrelated cases and controls on the Illumina OmniExpress chip. Individuals with an $IBD \geq 0.1875$ were removed from further analyses. (c) Siblings. Individuals with an $IBD \leq 0.35$ or ≥ 0.65 were removed from further analyses.

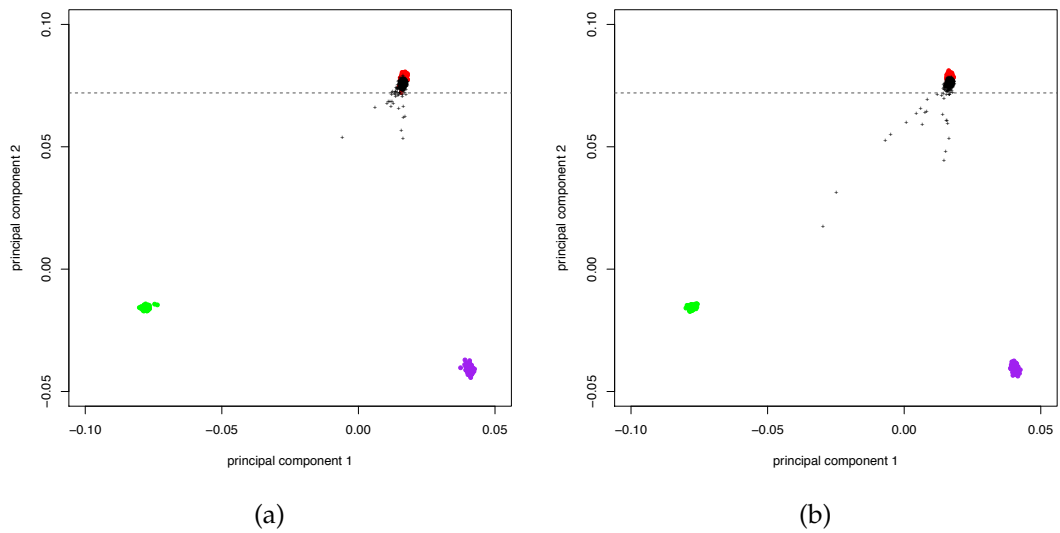


Figure 4.4: **Ancestry clustering based on principal component analysis of the GWAS data** for (a) Illumina 550K and (b) Illumina OmniExpress data. HapMap3 reference samples are **CEU** (European), **CHB** & **JPT** (Han Chinese and Japanese) and **YRI** (Yoruban). Genome-wide association samples are shown as black +s. Individuals with a second principal component score less than 0.072 (black dashed line) were removed from the study.

Table 4.2: **Summary of the numbers of cases, siblings and controls genotyped on the Illumina OmniExpress chip that failed at each stage of the per-individual QC.**

<i>Reason for exclusion</i>	<i>550K</i>		<i>OmniExpress</i>	
	<i>Cases</i>	<i>Controls</i>	<i>Cases</i>	<i>Controls</i>
Total Number	232	1,167	854	335
Discordant Sex	4	2	4	10
No sex info	-	-	1	3
Call rate	20	99	11	13
Duplicated / Related	-	5	32	10
Divergent Ancestry	4	20	25	1
Total Remaining	204	1,041	781	298

4.3 Per-SNP quality control

After removing all individuals that failed the previous QC steps in **section 4.2**, I then removed markers that could introduce false positives because they:

- have significantly different missing genotype rates between cases and controls at $P \leq 1 \times 10^{-5}$.
- have an excessive missing genotype rate (call rate ≤ 0.98).
- show a significant deviation from Hardy-Weinberg Equilibrium ($P \leq 1 \times 10^{-5}$).
- have a minor allele frequency (MAF) ≤ 0.01 .

4.3.1 Identification of SNPs with a significantly different genotyping call rates between cases and controls

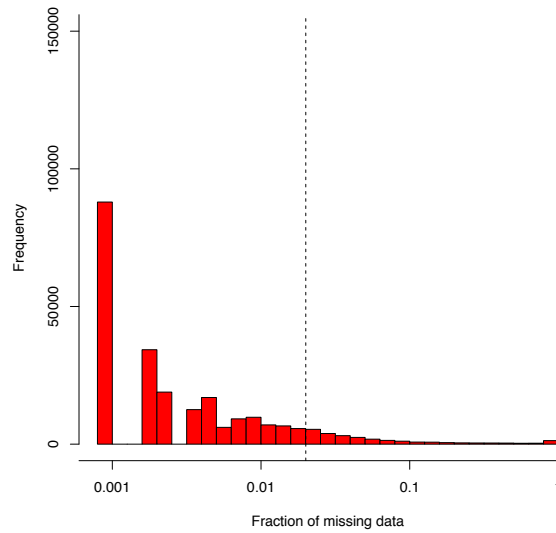
Differences in allele frequency between cases and controls are more likely to be due to genotyping errors than disease state. In total, 13,694 (550K) and 1,340 (OmniExpress) SNPs show significant differences in call rate between cases and controls (at $P \leq 1 \times 10^{-5}$, a conservative P value cut-off given the number of SNPs tested).

4.3.2 Identification of SNPs with elevated missing genotype rates

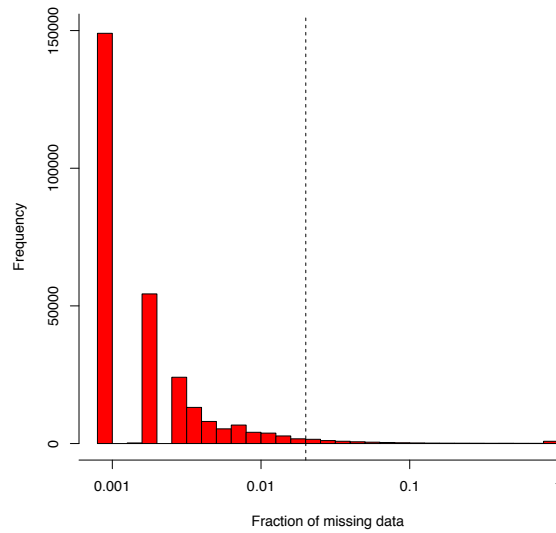
Markers with a low call rate may introduce false positives, however a balance must be struck as each SNP excluded represents a potential missed opportunity to identify a true disease associated variant. Though this can be negated as many variants will be recaptured by imputation of other markers in linkage disequilibrium by using data from the 1000 genomes project (Howie et al., 2011). After plotting a histogram of the call rate, I chose a call rate threshold of 98%, which excluded 15,343 (550K) and 5,808 (OmniExpress) SNPs – after first removing SNPs that showed significant differences in call rate between cases and controls (**Figure 4.5**).

4.3.3 Identification of SNPs with a significant deviation from Hardy-Weinberg Equilibrium

The Hardy-Weinberg principle states that the allele and genotype frequencies in a population remain constant, i.e. in equilibrium, with each other from generation to generation. Given a SNP with two alleles, **A** and **a**, whose frequencies are denoted p and q , the expectation is that if mating is random, under the null hypothesis the frequency of the genotypes in the next generation of a population will be $\text{freq}(\text{AA})$



(a)



(b)

Figure 4.5: **Histogram of the fraction of missing data for each SNP across all individuals that passed the 'per-individual' quality control steps in section 4.2.** The dashed vertical line represents the threshold (2%) at which SNPs were removed from subsequent analysis because of an excess failure rate. **(a)** Illumina 550K. **(b)** Illumina OmniExpress.

$= p^2$, $\text{freq}(\mathbf{Aa}) = 2pq$, $\text{freq}(\mathbf{aa}) = q^2$. A significant deviation from these frequencies could suggest the alleles are under selection, however, it is more likely that the deviation is due to a genotyping calling errors. Furthermore, given that up to 730,525 SNPs are being tested – a large number are expected to deviate from the equilibrium by chance. Therefore, I chose $P \leq 1 \times 10^{-5}$ for exclusion of SNPs to reflect a balance between the number of SNPs tested, the likelihood that disease causing variants may be under selection, and the need to eliminate SNPs that may generate false positives due to failed genotyping. In total 1,260 (550K) and 911 (OmniExpress) markers failed on the Hardy-Weinberg Equilibrium test performed in PLINK, **Table 4.3**.

4.3.4 Removal of SNPs with a low minor allele frequency

Given the small scale of our study, our power to detect genuine associations at low minor allele frequencies (MAF) is reduced. I chose an MAF cut-off of 0.01 in the first instance, which removed 21,751 (550K) and 80,386 (OmniExpress) SNPs.

A summary of the number of SNPs that failed at each QC stage is shown in **Table 4.3**. After these per-SNP QC steps, 505,608 (550K) and 643,789 (OmniExpress) SNPs were taken forward for the GWAS analyses.

Table 4.3: **Summary of the numbers of SNPs genotyped on the Illumina OmniExpress and 550K chip that failed at each stage of the per-SNP QC.** SNPs were removed if they; showed a significant difference in call rate between cases and controls at $P \leq 1 \times 10^{-5}$, have an excessive missing genotype rate (≥ 0.02), show a significant deviation from Hardy-Weinberg Equilibrium ($P \leq 1 \times 10^{-5}$), or have a minor allele frequency (MAF) ≤ 0.01 (n.b. some SNPs failed on multiple categories).

<i>Reason for exclusion</i>	<i>550K</i>	<i>OmniExpress</i>
Total number of SNPs	555,352	730,525
Different call rate between cases / controls	13,694	1,340
No founder genotypes	-	93
Call rate	15,343	5,808
HWE	1,260	911
MAF	21,751	80,386
Total number remaining	505,608	643,789

Chapter 5

Genome-wide association meta-analysis for relative hand skill

5.1 Abstract

I report a genome-wide association study meta-analysis for a quantitative measure of relative hand skill in cohorts 1-3, all individuals with dyslexia [reading disability (RD)] ($n = 728$). The most strongly associated variant, rs7182874 ($P = 8.68 \times 10^{-9}$), is located in *PCSK6*, confirming the association reported in **chapter 3**. As *PCSK6* is known to regulate NODAL in the development of left/right (LR) asymmetry in mice, I hypothesized that there may be an over-representation of highly associated variants within the orthologues of genes whose disruption in mice yields LR asymmetry phenotypes. In order to test this I developed a novel approach to GWAS pathway analysis, coupling gene-set enrichment analysis with mouse phenotype data. Four out of 15 LR asymmetry phenotypes show an over-representation ($\text{FDR} \leq 5\%$). I replicated three of these phenotypes; situs inversus, heterotaxia, and double outlet right ventricle, in an independent general popula-

tion cohort ($n = 2,666$, $\text{FDR} \leq 5\%$). These findings lead me to propose that handedness is a polygenic trait controlled in part by the molecular mechanisms that establish LR body asymmetry early in development.

5.2 Introduction

For a literature review on handedness, see **section 1.1**. For an explanation of GWAS, see **subsection 3.2.1**. Here I build on the study in **section 3.2** by genotyping the replication cohort 2 genome-wide, and obtaining access to genotype data from ALSPAC (cohorts 3 & 4). I then perform a combined GWA meta-analysis after imputation (**subsection 5.3.2**) (Howie et al., 2009), and a hypothesis led gene set enrichment analysis using the program MAGENTA (**subsection 5.3.4**) (Segre et al., 2010).

5.3 Methods

For a description of the cohorts and their ascertainment see **section 2.1**, for the relative hand skill phenotype see **subsection 2.2.1**.

5.3.1 Genotyping and quality control

Genotypes for cohort 1 were generated on the Illumina[®] 550K SNP array (**subsection 3.3.1**). Genotypes for cohort 2 were generated for the present study on the Illumina[®] OmniExpress SNP array (730,525 markers), for a description of the QC see **section 4.2**. After QC, 458,236 (cohort 1) and 589,768 (cohort 2) single nucleotide polymorphisms (SNPs) with a minor allele frequency $\geq 5\%$ were taken forward for

imputation (**subsection 5.3.2**).

Genotyping for ALSPAC (cohorts 3 and 4) was carried out using the Illumina 550K array. Individuals were excluded by our ALSPAC collaborators on the basis of the following: sex mismatches, minimal or excessive heterozygosity, disproportionate levels of individual missingness, cryptic relatedness measured as proportion of identity by descent, and insufficient sample replication. The remaining individuals were assessed for evidence of population stratification by multidimensional scaling analysis and compared with Hapmap II (release 22) European descent (CEU), Han Chinese, Japanese and Yoruba reference populations; all individuals with non-European ancestry were removed in the ALSPAC dataset. SNPs with a minor allele frequency of $\leq 5\%$, a call rate of $\leq 95\%$ or evidence for violations of Hardy-Weinberg equilibrium were removed.

5.3.2 Imputation

In cohorts 1 and 2, I inferred haplotypes prior to imputation using the Segmented HAPlotype Estimation and Imputation Tool (shapeit v2 - <http://www.shapeit.fr/> (Delaneau et al., 2012)). I used IMPUTE (Howie et al., 2009) (v2.2.2 - http://mathgen.stats.ox.ac.uk/impute/impute_v2.html) to estimate the genotypes of SNPs not directly genotyped. Imputed SNPs were filtered out if they had an MAF ≤ 0.05 or imputation info ≤ 0.8 . I then used the 1000 Genomes Phase I integrated variant set (March 2012, build 37) as the reference panel for imputation, testing 5,829,682 SNPs for association.

In ALSPAC autosomal genotypic data are imputed by ALSPAC using Markov Chain Haplotyping software (MACH v.1.0.16) and phased haplotype data from

CEU individuals (Hapmap release 22, Phase II NCBI B36, dbSNP 126) based on a cleaned dataset of 8,365 individuals and 464,311 autosomal SNPs. After imputation, all poorly imputed SNPs ($r^2 \leq 0.30$) and those with low minor allele frequency ($MAF \leq 0.05$) are excluded from analyses.

5.3.3 Statistical analysis

Association with PegQ was tested under an additive model using SNPTEST (v2.4.1 -https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html) simultaneously specifying sex and age as covariates. The genomic inflation factor (λ) was calculated for directly genotyped and imputed SNPs combined, and shows no evidence of population structure, λ is 1.01 for cohort 1, 1.00 for cohort 2, 1.01 for cohort 3, and 1.00 for cohort 4. The values of PegQ were standardised in all cohorts separately to give a mean of 0, and a SD of 1. The results for the three cohorts were then meta-analyzed under a fixed-effects model using GWAMA (Magi and Morris, 2010). a fixed effects meta-analysis applied in the presence of heterogeneity will tend to exaggerate the P value, and a standard random effects meta-analysis applied when the effects are actually due to a common genetic mechanism will be slightly conservative (DerSimonian and Laird, 1986). GWAMA performs tests of heterogeneity of effects across studies, and can then perform random-effects meta-analysis for each SNP by calculating the random-effects variance component (Magi and Morris, 2010). No heterogeneity was observed between the studies. In total, 2,095,209 SNPs that were either directly genotyped or imputed and overlapped in all three studies and passed quality control were tested for association. Finally, in order to compare relative hand skill and hand motor skill in cohorts 3 and 4, I used a one-tailed Student's T-Test, not assuming equal variance.

5.3.4 Gene-set enrichment analysis

I performed pathway analysis using Meta-Analysis Gene-set Enrichment of variant Associations (MAGENTA (Segre et al., 2010)). MAGENTA first maps SNPs to genes (defined as from 20kb upstream of the 5' UTR to 10kb downstream of the 3' UTR), and then selects the most strongly associated SNP, after correcting for confounding factors such as gene size, the number of SNPs, linkage disequilibrium, number of recombination hotspots, in order to calculate a P value for each gene. It then ranks all gene P values and tests if there is an overrepresentation of a given set of genes above a user-specified cutoff of all genes (see **Figure 5.1** for an explanation of the method). I chose the 75th percentile as a cutoff because it yields the best power for detecting enrichment in traits that are highly polygenic with many associations with weak effect (Segre et al., 2010).

I wanted to test the hypothesis that genes involved in the determination of LR asymmetry are enriched for SNPs with low P values in the GWAS data. In order to do this I utilized mouse knockout phenotype data for the pathway analysis because;

1. The effect of disruptions of genes involved in the development of LR asymmetry have been better studied in mouse models compared to humans. Asymmetry phenotypes are extremely rare in humans, for example situs inversus has a prevalence of ~ 1 in 10,000 (Torgersen, 1950).
2. Phenotypes from mouse knock-out experiments are typically very similar to observations of human patients with mutations in orthologous genes (Webber et al., 2009).

I obtained this data by mining the Mammalian Phenotype Ontology (MPO) from the Mouse Genome Informatics database (MGI; <http://www.informatics.jax.org>) (Eppig et al., 2007). The MPO is a hierarchical set of phenotypes consisting of 30 overarching terms, such as *Nervous System* (MP:0003631) and *Behaviour / Neurological* (MP:0005386), and hundreds of sub-phenotypes. Each gene may be assigned to multiple phenotypes, and if it is assigned to a sub-phenotype, it is automatically assigned to any overarching terms in the hierarchy. For example *Pcsk6* is assigned to embryogenesis (MP:0005380) → abnormal embryogenesis/ development (MP:0001672) → abnormal developmental patterning (MP:0002084) → abnormal left-right axis patterning (MP:0001706, plus many other phenotypes).

There are 6,573 genes in the MGI database (August 2012, <http://www.informatics.jax.org/>, (Blake et al., 2009)). Phenotypic annotations were mapped to the human orthologs of these 6,573 genes, using human:mouse gene orthology relationships defined by the MGI resource. The median size of human orthologs is 30.5kb and the mean is 75.9kb. I tested all pathways in the mouse genome informatics database (MGI, $n = 15$) that cause asymmetry defects when knocked out in mice and contain more than 10 and less than 100 genes (a trade-off between power and specificity), against a background of all 6,573 genes in the MGI database.

5.3.5 Genome-wide Complex Trait Analysis (GCTA)

Another alternative to individually testing SNPs for association is to estimate the phenotypic variance explained by all SNPs together (Yang et al., 2010). To test this I used GCTA (Genome-wide complex trait analysis), which fits a linear model to the phenotypic data and then uses restricted maximum likelihood (REML) to estimate the variance explained by all the SNPs (Yang et al., 2010). GCTA takes binary PED

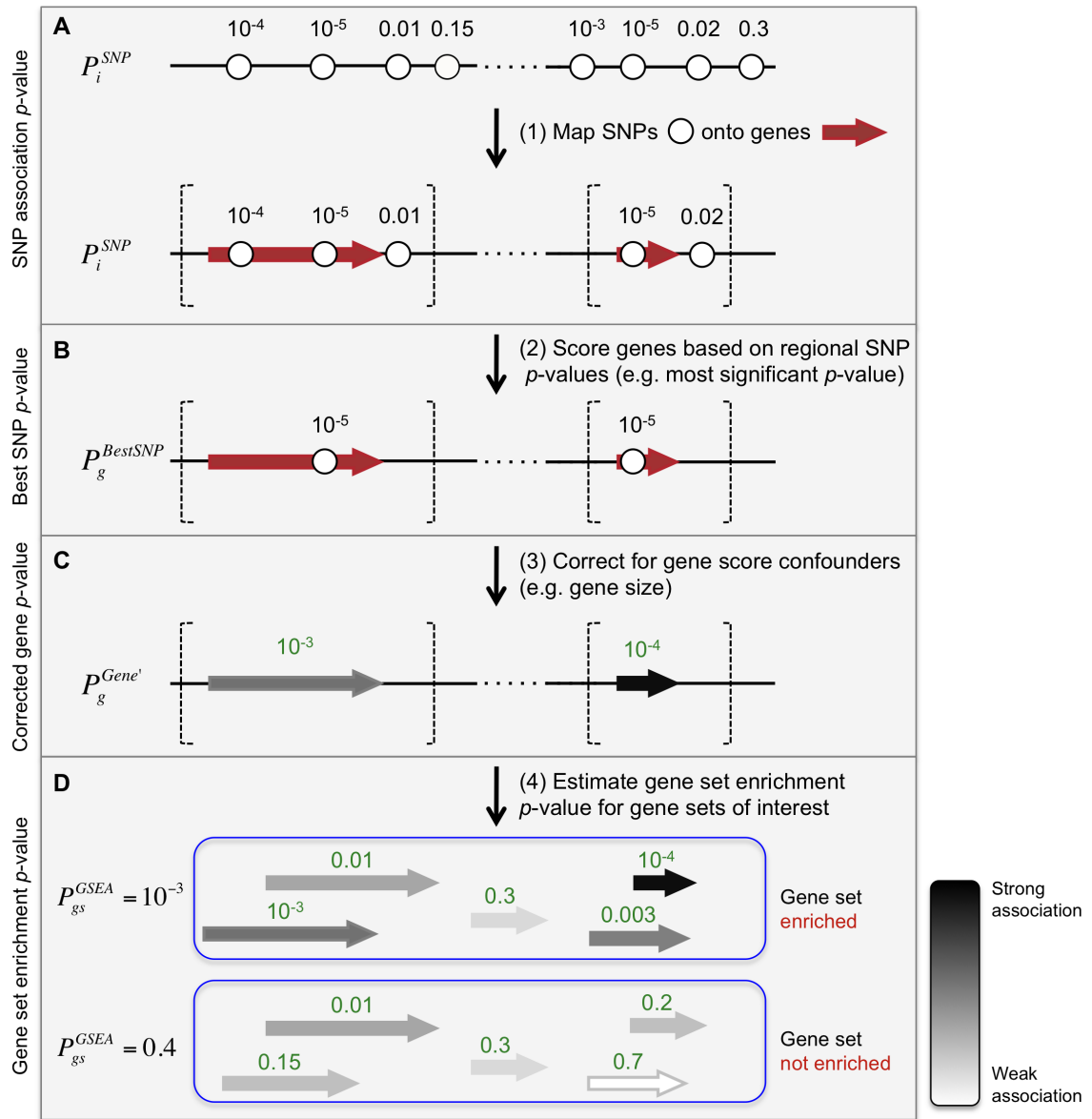


Figure 5.1: Description of Meta-Analysis Gene-set Enrichment of variaNT Associations (MAGENTA) method. (A) Genetic variants and their association scores are mapped onto genes. Gene boundaries (vertical dashed lines) were defined as 20kb upstream and 10kb downstream to the most extreme transcript start and end sites of the gene (red arrow), respectively. (B) The most significant SNP that lies within the extended gene boundaries is assigned to each gene in the genome. (C) Step-wise multivariate linear regression analysis applied to regress out confounding effects such as gene size, the number of SNPs, linkage disequilibrium, and number of recombination hotspots. (D) Non-parametric statistical test applied to test if genes in a gene set are enriched for highly ranked gene scores more than expected by chance, compared to randomly sampled gene sets of identical size from the genome (adapted from (Segre et al., 2010)).

(BED) files as input. I converted the SNPtest formatted GWAS data into PED files using GTOOL. I then converted the data into BED files and merged cohorts 1 & 2 using PLINK (Purcell et al., 2007).

5.4 Results

The study participants are derived from four cohorts (**Table 5.1**), three with RD (cohorts 1-3, $n = 185, 374$, and 169 respectively), and a general population cohort (cohort 4, $n = 2,666$). The samples from cohorts 3 and 4 are part of the Avon Longitudinal Study of Parents and Children (ALSPAC). We divided them into individuals with RD (cohort 3), and those unaffected with RD (cohort 4), based on their IQ and reading scores (for ascertainment criteria, see **section 2.1**). All have had their relative hand skill assessed.

Table 5.1: **GWAS design.** Three cohorts with reading disability and a further general population cohort were genotyped and tested after imputation for association with relative hand skill.

<i>cohort</i>	<i>n</i>	<i>Genotyping Array</i>	<i>Imputation Reference Panel</i>	<i>Imputation Software</i>	<i>Reading Disabled</i>
1	185	Illumina 550K	1000 Genomes March 2012	Impute	✓
2	374	Illumina OmniExpress	1000 Genomes March 2012	Impute	✓
3	169	Illumina 550K	HapMap Phase II	MACH	✓
4	2,666	Illumina 550K	HapMap Phase II	MACH	✗

After QC, I performed a quantitative GWAS for relative hand skill (handedness, assessed using the Peg board test) in four separate cohorts, shown in **Table 5.1**. I analysed each cohort separately and then meta-analysed cohorts 1-3 (the reading disabled groups). In each cohort individually, no SNP was genome-wide significant at $P \leq 5 \times 10^{-8}$.

5.4.1 Relative hand skill phenotype

For a description of the relative hand skill peg-board task see **subsection 2.2.1**. I investigated the correlation between hand preference for writing and relative hand skill using logistic regression in all the individuals in our cohorts for which we have data. The PegQ phenotype is highly correlated with the preferred hand of the participants. In cohorts 1 & 2, for every 1 standard deviation shift to the right of the Peg distribution, individuals are 13.4 times more likely to be right handed (95% Confidence intervals: 6.11–36.49). In ALSPAC (cohorts 3 & 4), they are 3.23 times more likely to be right handed (95% Confidence intervals: 2.83–3.70). The weaker correlation in ALSPAC is probably due to the fact that the peg-board task was performed only once, as opposed to five times in cohorts 1 and 2 (**subsection 2.2.1**).

Furthermore the PegQ phenotype also correlates with dexterity, measured as the time taken to perform the task with either the right hand (cohorts 1 & 2, Pearson's $r = -0.42$, cohorts 3 & 4 $r = -0.50$) or the left hand (cohorts 1 & 2, Pearson's $r = 0.27$, cohorts 3 & 4 $r = 0.59$).

Individuals with RD (cohort 3) are not significantly more likely to be left-handed ($f = 11.4\%$) than those in the general population (cohort 4, $f = 10.9\%$, two-tailed Fisher's exact test $P = 0.82$). Furthermore, they are not more likely to be shifted toward the left-handed distribution of the peg-board task, however their average time for both hands is slower, indicating lower overall motor skill (**Table 5.2**).

Table 5.2: **Individuals with reading disability are slower than controls at performing the peg-board test.** Comparison of the individuals in general population to those with reading disability in ALSPAC for both relative hand skill (PegQ) and the average time taken to complete the peg-board test with both hands (Motor Skill). One-tailed Student's T-Test, not assuming equal variance.

<i>Phenotype</i>	<i>General Population Ave</i>	<i>General Population SD</i>	<i>Dyslexia Ave</i>	<i>Dyslexia SD</i>	<i>P</i>
PegQ	0.11	0.19	0.10	0.18	0.32
Motor Skill	23.57	3.18	24.39	3.65	0.0011

5.4.2 Relative hand skill GWA meta-analysis in individuals with reading disability

Cohorts 1 & 2 have been genotyped on the Illumina 550K & OmniExpress SNP arrays, respectively, and I imputed genotypes up to the 1000 Genomes reference panel (Phase 1, March 2012 release, **Table 5.1**). The samples from cohorts 3 & 4 come from the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort. The participants have performed a similar peg-board task (**section 5.3**), been characterized by similar reading measures to cohorts 1 & 2, been genotyped on the Illumina 550K SNP array, and had genotypes imputed up to the HapMap II CEU reference panel (**Table 5.1**). We selected individuals (cohort 3) to closely match the ascertainment criteria for RD used in the other two cohorts. Furthermore we selected a general population cohort, excluding individuals showing phenotypes compatible with a neurodevelopmental disorder diagnosis (**section 2.1**).

I tested 2,095,209 SNPs for association with relative hand skill in a meta-analysis of the three RD cohorts. One locus is genome-wide significant, the intronic marker rs7182874 within the 186kb gene proprotein convertase subtilisin/kexin type 6 (*PCSK6*) on chromosome 15 ($P = 8.68 \times 10^{-9}$, $\beta = 0.37$, **Table 5.3**). Each copy of

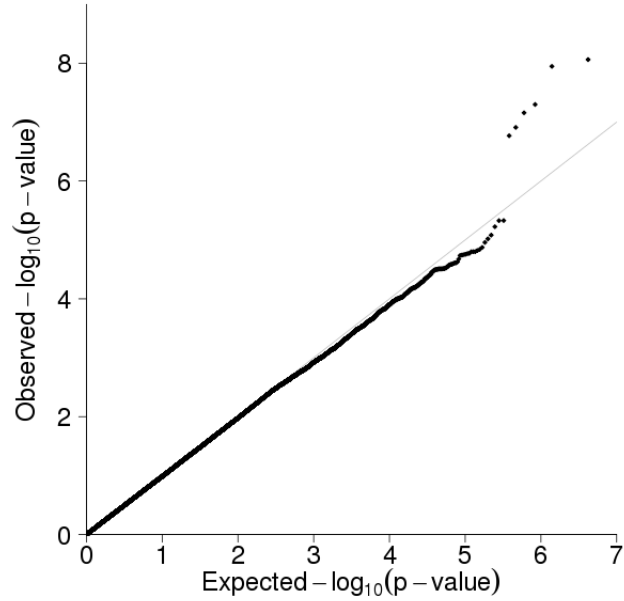


Figure 5.2: **QQ plot for the RD relative hand skill GWA meta-analysis.** The observed P values are plotted against the expected P values, a deviation from the null (diagonal line) indicates an enrichment of genuine associations, population stratification or an error in genotype calling.

the minor (derived) allele is associated with a shift of 0.37 standard deviations to the right of the relative hand skill distribution. See **Table 3.1** for a comparison with the original GWAS and replication study of the SNP rs11855415 in *PCSK6*. The two SNPs are in high linkage disequilibrium, the P value for rs11855415 is 6.96×10^{-8} . The most highly associated SNPs in the GWA meta-analysis are given in **Table 5.4**. The QQ plot is shown in **Figure 5.2**. A manhattan plot, and zoomed in plot of the *PCSK6* locus are shown in **Figure 5.3**.

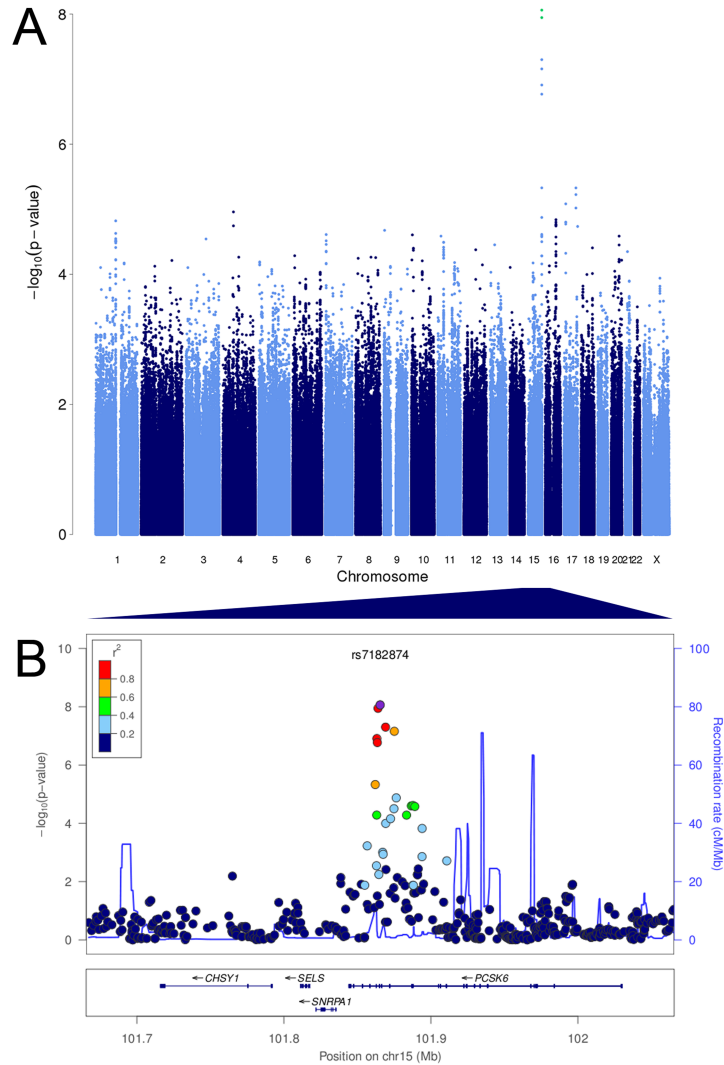


Figure 5.3: ***PCSK6* is associated with relative hand skill in individuals with reading disability.** Manhattan and Locus Zoom plot for RD relative hand skill GWA meta-analysis. (A) Manhattan plot of the P values for the meta-analysis of the RD PegQ relative hand skill GWAS data-sets (cohorts 1–3). The observed negative log₁₀ transformed P values are plotted against chromosome and position along the genome. The two ●s represent Markers that are genome-wide significant at $P \leq 5 \times 10^{-8}$. (B) Negative log₁₀ of the P values for all genotyped and imputed SNPs around *PCSK6* plotted in LocusZoom (Pruim et al., 2010). Linkage disequilibrium (r^2) with the most highly associated SNP, rs7182874 (●), is shown by the colour of the SNPs. The recombination rate is shown by the line and the locations of genes in this locus are shown in the panel below the plot.

Table 5.3: **Association of the most highly associated SNP, rs7182874** (ancestral allele = C; derived, minor allele = T), with relative hand skill in individuals with RD in each cohort of the study. For comparison to the original GWAS, see **Table 3.1**.

<i>Cohort</i>	<i>n</i>	<i>P</i>	β	<i>SE</i>
1	185	2.30×10^{-6}	0.55	0.12
2	374	0.0099	0.26	0.10
3	169	0.0049	0.34	0.12
Meta-analysis	728	8.69×10^{-9}	0.37	0.064

Table 5.4: A list of the 40 most strongly associated SNPs from the meta-analysis of the **relative hand skill GWA meta-analysis of individuals with reading disability**. All the genes within 100kb to the left or right of the SNP are listed. +++ indicates that the effect allele is associated with an increased relative right hand skill in both studies, and --- indicates increased relative left hand skill. EAF = Effect Allele Frequency, β = effect size of each copy of the effect allele, in standard deviations (95% Lower and Upper confidence intervals given), SE = Standard Error.

Chr	Position	Variant	EAF	β	β 95L	β 95U	SE	P value	Effects	Gene	Left Genes	Right Genes
15	101865472	rs7182874	0.71	-0.37	-0.5	-0.24	0.064	8.68×10^{-9}	---	PCSK6	VIMP,LOC100507472,CHSY1,SNRPA1	-
15	101863989	rs4965830	0.71	-0.37	-0.49	-0.24	0.064	1.13×10^{-8}	---	PCSK6	VIMP,LOC100507472,CHSY1,SNRPA1	-
15	101869185	rs752028	0.7	-0.34	-0.47	-0.22	0.063	5.02×10^{-8}	---	PCSK6	VIMP,LOC100507472,CHSY1,SNRPA1	-
15	101875123	rs11855415	0.77	-0.37	-0.5	-0.24	0.068	6.96×10^{-8}	---	PCSK6	VIMP,LOC100507472,CHSY1,SNRPA1	-
15	101863289	rs1871978	0.75	-0.35	-0.48	-0.22	0.067	1.23×10^{-7}	---	PCSK6	VIMP,LOC100507472,CHSY1,SNRPA1	-
15	101863573	rs9806256	0.74	-0.35	-0.48	-0.22	0.067	1.7×10^{-7}	---	PCSK6	VIMP,LOC100507472,CHSY1,SNRPA1	-
15	101862163	rs3825921	0.62	-0.27	-0.39	-0.16	0.06	4.69×10^{-6}	---	PCSK6	VIMP,LOC100507472,CHSY1,SNRPA1	-
17	70584265	rs718073	0.84	0.35	0.2	0.5	0.077	4.7×10^{-6}	+++	LOC100499467	-	SLC39A11,LINC00511
17	70584327	rs718072	0.84	0.35	0.2	0.5	0.077	5.96×10^{-6}	+++	LOC100499467	-	SLC39A11,LINC00511
17	10567483	rs169313	0.44	0.25	0.14	0.35	0.055	8.28×10^{-6}	+++	-	MYH3	MAGOH2,SCO1,TMEM220,C17orf48
17	70587320	rs2537985	0.84	0.34	0.19	0.49	0.077	9.57×10^{-6}	+++	LOC100499467	-	SLC39A11,LINC00511
4	61966273	rs4599440	0.8	-0.31	-0.44	-0.17	0.069	1.1×10^{-5}	---	-	-	-
15	101876416	rs2086678	0.54	-0.25	-0.36	-0.14	0.057	1.34×10^{-5}	---	PCSK6	VIMP,LOC100507472,CHSY1,SNRPA1	-
16	62849577	rs2962457	0.78	0.29	0.16	0.42	0.066	1.45×10^{-5}	+++	-	-	-
1	119150184	rs4659111	0.51	-0.24	-0.36	-0.13	0.057	1.51×10^{-5}	---	-	-	-
16	62815924	rs2962469	0.79	0.29	0.16	0.42	0.067	1.58×10^{-5}	+++	-	-	-
17	10565791	rs201608	0.44	0.24	0.13	0.34	0.055	1.58×10^{-5}	+++	-	MYH3	MAGOH2,SCO1,TMEM220,C17orf48
17	10563050	rs201611	0.44	0.24	0.13	0.35	0.055	1.59×10^{-5}	+++	-	MYH3	MAGOH2,SCO1,TMEM220,C17orf48
17	10562314	rs201612	0.44	0.24	0.13	0.35	0.056	1.69×10^{-5}	+++	-	MYH3	MAGOH2,SCO1,TMEM220,C17orf48
16	62840744	rs2408330	0.78	0.29	0.16	0.42	0.067	1.7×10^{-5}	+++	-	-	-
16	62794811	rs2052871	0.78	0.28	0.15	0.41	0.066	1.75×10^{-5}	+++	-	-	-
16	62796921	rs2962461	0.78	0.28	0.15	0.41	0.066	1.76×10^{-5}	+++	-	-	-
4	61967025	rs7439943	0.79	-0.3	-0.44	-0.17	0.071	1.8×10^{-5}	---	-	-	-
16	62798852	rs2962462	0.78	0.28	0.15	0.41	0.066	1.81×10^{-5}	+++	-	-	-
17	80563659	rs11078005	0.58	-0.25	-0.36	-0.13	0.058	1.84×10^{-5}	---	-	FO XK2	WDR45L,RAB40B
9	4309006	rs4567095	0.31	-0.24	-0.36	-0.13	0.057	2.11×10^{-5}	---	-	GLIS3	-
1	119117474	rs10754383	0.5	-0.24	-0.35	-0.13	0.057	2.35×10^{-5}	---	-	-	-
15	101887794	rs1030	0.67	-0.25	-0.37	-0.14	0.06	2.44×10^{-5}	---	PCSK6	VIMP,LOC100507472,CHSY1,SNRPA1	-
7	5528883	rs4724701	0.45	0.23	0.12	0.33	0.054	2.45×10^{-5}	+++	FBXL18	TNRC18	ACTB,MIR589

10	6470994	rs17299670	0.1	0.43	0.23	0.62	0.1	2.48×10^{-5}	+++	PRKCQ	LOC399715	-
15	101886755	rs7175368	0.67	-0.25	-0.37	-0.13	0.06	2.54×10^{-5}	---	PCSK6	VIMP,LOC100507472,CHSY1,SNRPA1	-
11	19953089	rs17612287	0.12	-0.39	-0.57	-0.21	0.093	2.59×10^{-5}	---	NAV2	-	-
20	46794500	rs707510	0.75	0.29	0.15	0.42	0.068	2.59×10^{-5}	+++	-	-	-
15	101889075	rs10152965	0.68	-0.26	-0.37	-0.14	0.061	2.65×10^{-5}	---	PCSK6	VIMP,LOC100507472,CHSY1,SNRPA1	-
16	62816545	rs2962470	0.78	0.28	0.15	0.41	0.066	2.68×10^{-5}	+++	-	-	-
1	119157211	rs1742854	0.51	-0.24	-0.35	-0.13	0.056	2.84×10^{-5}	---	-	-	-
3	118408200	rs6765745	0.082	0.43	0.23	0.63	0.1	2.87×10^{-5}	+++	-	-	-
1	119146884	rs12137431	0.51	-0.24	-0.35	-0.13	0.056	2.98×10^{-5}	---	-	-	-
1	119153953	rs10157240	0.51	-0.24	-0.35	-0.13	0.057	2.99×10^{-5}	---	-	-	-
1	119152617	rs10923630	0.51	-0.24	-0.35	-0.12	0.057	3.06×10^{-5}	---	-	-	-

5.4.3 Relative hand skill GWAS in the ALSPAC general population

No SNP is genome-wide significant at $P \leq 5 \times 10^{-8}$ in the ALSPAC general population (cohort 4). The most strongly associated SNP is rs7883190, located ~ 6 kb upstream of *GPC3* (**Table 5.5**), a gene that causes heart and lung asymmetry defects when disrupted in mice (Ng et al., 2009).

Table 5.5: A list of the 40 most strongly associated SNPs from the ALSPAC general population Peg GWAS. $n = 2,666$.

<i>Chr</i>	<i>Position</i>	<i>Variant</i>	<i>EA</i>	β	<i>SE</i>	<i>P value</i>	<i>Gene</i>	<i>Left Genes</i>	<i>Right Genes</i>
23	132851311	rs7883190	0.62	0.12	0.025	2.08×10^{-6}	GPC3	-	-
23	132848355	rs5933366	0.62	0.12	0.025	2.14×10^{-6}	GPC3	-	-
23	132828231	rs2267527	0.61	0.11	0.024	2.25×10^{-6}	GPC3	-	-
23	132870184	rs11796475	0.61	0.12	0.025	2.37×10^{-6}	GPC3	-	-
23	132823205	rs2267524	0.61	0.11	0.024	2.49×10^{-6}	GPC3	-	-
23	132823202	rs2267523	0.61	0.11	0.024	2.8×10^{-6}	GPC3	-	-
20	11157522	rs6040449	0.63	0.14	0.03	3.92×10^{-6}	-	-	LOC339593
20	11157476	rs6040448	0.63	0.14	0.03	3.93×10^{-6}	-	-	LOC339593
20	52844426	rs13037755	0.71	0.26	0.057	4.87×10^{-6}	-	-	-
20	11179353	rs2207105	0.37	-0.14	0.03	5.73×10^{-6}	-	-	LOC339593
20	11150285	rs6040439	0.66	0.14	0.03	6.47×10^{-6}	-	-	LOC339593
4	68102080	rs7700004	0.19	0.16	0.036	9.64×10^{-6}	-	CENPC1	UBA6,STAP1
4	68097974	rs7660528	0.81	-0.16	0.036	9.67×10^{-6}	-	CENPC1	UBA6,STAP1
4	68105033	rs7672027	0.81	-0.16	0.036	9.7×10^{-6}	-	CENPC1	UBA6,STAP1
23	132876368	rs12558772	0.61	0.12	0.026	9.83×10^{-6}	GPC3	-	-
5	176072947	rs11741729	0.59	-0.13	0.03	1.01×10^{-5}	-	MIR4281,SNCB,EIF4E1B,TSPAN17	UNC5A
20	11175697	rs726867	0.66	0.13	0.03	1.03×10^{-5}	-	-	LOC339593
20	11157119	rs6074263	0.65	0.13	0.03	1.08×10^{-5}	-	-	LOC339593
20	11179875	rs6040476	0.62	0.13	0.03	1.14×10^{-5}	-	-	LOC339593
4	68108419	rs10013617	0.19	0.16	0.037	1.2×10^{-5}	STAP1	CENPC1	UBA6
9	18766840	rs1549986	0.084	-0.23	0.051	1.23×10^{-5}	ADAMTSL1	ADAMTSL1	-
22	45461009	rs801719	0.67	-0.13	0.031	1.26×10^{-5}	CERK	GRAMD4	TBC1D22A
4	67924140	rs6419917	0.78	-0.15	0.035	1.34×10^{-5}	-	-	CENPC1
6	45895446	rs2894651	0.51	-0.15	0.036	1.38×10^{-5}	-	-	CLIC5
6	31321790	rs3130409	0.46	-0.13	0.029	1.39×10^{-5}	-	HCG27,TCF19,CCHCR1,POU5F1,PSORS1C3	HLA-C

20	11184424	rs1883952	0.66	0.13	0.03	1.47×10^{-5}	-	-	LOC339593
20	11147629	rs6033061	0.33	-0.14	0.032	1.53×10^{-5}	-	-	LOC339593
20	11152883	rs2223853	0.35	-0.13	0.03	1.62×10^{-5}	-	-	LOC339593
4	68112718	rs2046599	0.19	0.16	0.037	1.67×10^{-5}	STAP1	CENPC1	UBA6
20	11186791	rs6040484	0.34	-0.13	0.03	1.68×10^{-5}	-	-	LOC339593
3	53253281	rs12493802	0.43	0.12	0.029	1.81×10^{-5}	TKT	PRKCD	DCP1A
6	45885772	rs4714875	0.39	-0.14	0.032	1.81×10^{-5}	-	-	CLIC5
6	105955716	rs9486077	0.95	0.28	0.066	1.85×10^{-5}	PREP	-	-
23	112181619	rs968859	0.43	0.1	0.024	1.95×10^{-5}	-	-	-
23	112187659	rs981430	0.43	0.1	0.024	1.95×10^{-5}	-	-	-
23	112179757	rs1033439	0.43	0.1	0.024	1.96×10^{-5}	-	-	-
23	112189725	rs4829544	0.43	0.1	0.024	1.96×10^{-5}	-	-	-
23	132805699	rs1264382	0.72	0.11	0.027	1.96×10^{-5}	GPC3	-	-
23	112177928	rs880355	0.43	0.1	0.024	1.97×10^{-5}	-	-	-
23	112199687	rs11797987	0.43	0.1	0.024	1.97×10^{-5}	-	-	-

5.4.4 Gene-set enrichment analysis

PCSK6 causes LR asymmetry defects when knocked out in mice (Constam and Robertson, 2000; Zhou et al., 2012). I therefore tested for an enrichment of highly associated (though not genome-wide significant) SNPs within genes annotated to mouse genome informatics (MGI) LR asymmetry phenotypes using the program MAGENTA (Segre et al., 2010).

Four phenotypes out of the 15 tested are significant at $FDR \leq 5\%$ in the combined RD meta-analysis using an enrichment cut-off that equals the 75th percentile of all gene *P* values; **abnormal right lung morphology** (MP:0009570), **double outlet right ventricle** (MP:0000284), **heterotaxia** (MP:0004133), and **situs inversus** (MP:0002766) (Table 5.6). The genes that comprise the LR asymmetry phenotype sets overlap; thus FDR correction for multiple testing is conservative. A list of the ten most highly associated genes in the RD cohort is given in Table 5.7. After removing the known gene *PCSK6* from the analysis, three phenotypes remain significant at $FDR \leq 5\%$ (Table 5.8). Therefore the association in *PCSK6* is not solely driving the results.

I took forward the four phenotypes that show an over-representation of genes with low *P* value SNPs in the RD cohorts for testing in the general population (cohort 4). Three out of the four phenotypes show enrichment at ($FDR \leq 5\%$), **double outlet right ventricle**, **heterotaxia**, and **situs inversus** (Table 5.9). I have therefore shown that our findings from the RD cohort hold true for the general population.

There are 116 unique genes whose disruption in their mouse orthologue is annotated with these three phenotypes. 47 genes in the RD cohorts, and 42 in the

general population cohort, show P values above the 75th percentile of all gene P values (the expected number is one quarter of 116, i.e. 29). Of these, 16 overlap between the RD meta-analysis and the general population cohort (**Figure 5.4**, one-tailed Fisher's exact test, $P = 0.42$). Although the same phenotypes show enrichments, the associated SNPs and genes are different, i.e. there is both allelic and locus heterogeneity between the cohorts.

Four out of the five most highly associated genes from the mouse phenotypes tested are involved in ciliogenesis (*RFX3*, *MNS1*, *GLI3*, and *PKD2*, **Table 5.7**). Ciliopathies cause cerebral phenotypes such as hydroencephaly, and agenesis of the corpus callosum and cerebellar vermis; two brain midline structures that connect the two hemispheres of the cerebrum and cerebellum, respectively. I therefore also tested these three phenotypes in both the meta-analysis of the individuals with RD (cohorts 1-3), and separately in the general population (cohort 4). The absent corpus callosum (MP:0002196) phenotype showed an over-representation in the RD cohort (adjusted $P = 0.029$, **Table 5.10**), but not in the general population (cohort 4, **Table 5.11**). Of the associated genes from the absent corpus callosum phenotype, only *GLI3* overlaps with the associated genes in **Table 5.7**. Hydroencephaly (MP:0001891) and abnormal cerebellar vermis (MP:0000864) did not show an over-representation in either cohort.

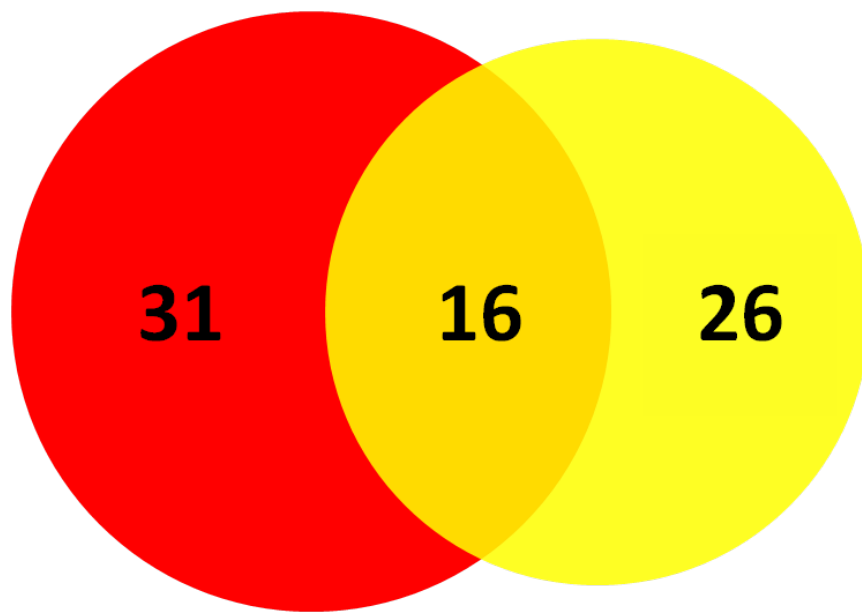


Figure 5.4: There is locus heterogeneity between the **RD** and **general population** cohorts in the GSEA. Venn diagram showing the number of genes that are associated with the three phenotypes; double outlet right ventricle, heterotaxia, and situs inversus, in the GSEA in both cohorts. 16 genes **overlap** between the two cohorts (one-tailed Fisher's exact test, $P = 0.72$). In total, 116 genes are annotated to these three phenotypes.

Table 5.6: In the relative hand skill GWA meta-analysis of individuals with reading disability, there is an over-representation of highly associated SNPs within human orthologues of genes involved in causing LR asymmetry phenotypes when knocked out in mice. 15 phenotypes that contain between 10 and 100 genes, and cause asymmetry defects when knocked out in mice, were tested using MAGENTA (Segre et al., 2010). The observed number of genes were compared to the expected using a cut-off that equals the 75th percentile of all gene P values, which were then adjusted for multiple testing using an FDR correction.

<i>Mouse Phenotype ID</i>	<i>Phenotype</i>	<i>Genes</i>	<i>Median Size / KB</i>	<i>Mean Size</i>	<i>Expected</i>	<i>Observed</i>	<i>Adjusted P</i>
MP:0010429	abnormal heart left ventricle outflow tract morphology	47	42	83	12	13	0.40
MP:0003922	abnormal heart right atrium morphology	20	50	87	5	4	0.74
MP:0009569	abnormal left lung morphology	22	51	89	6	8	0.20
MP:0001706	abnormal left-right axis patterning	62	48	73	16	20	0.17
MP:0009570	abnormal right lung morphology	32	68	120	8	16	0.0087
MP:0000644	dextrocardia	26	64	113	7	11	0.070
MP:0000284	double outlet right ventricle	87	47	92	22	35	0.016
MP:0002625	heart left ventricle hypertrophy	48	29	107	12	12	0.60
MP:0000276	heart right ventricle hypertrophy	25	18	96	6	3	0.93
MP:0004133	heterotaxia	42	46	80	11	18	0.025
MP:0000542	left-sided isomerism	17	64	92	4	8	0.068
MP:0004158	right aortic arch	36	27	80	9	14	0.07
MP:0000531	right pulmonary isomerism	15	39	77	4	5	0.35
MP:0000508	right-sided isomerism	16	46	88	4	5	0.37
MP:0002766	situs inversus	24	63	103	6	11	0.046

Table 5.7: **SNPs in LR asymmetry genes are associated with relative hand skill.** I've listed the ten lowest gene P values that are also within one of the four enriched phenotypes from **Table 5.6** in the RD meta-analysis. MAF = minor allele frequency, β = effect size of each copy of the minor allele, in standard deviations.

<i>Gene</i>	<i>Chromosome</i>	<i>Gene Size / kb</i>	<i>Gene P</i>	<i>Best SNP</i>	<i>SNP P</i>	<i>MAF</i>	β
<i>PCSK6</i>	15	186	3.94×10^{-8}	rs7182874	8.68×10^{-9}	0.29	0.37
<i>PKD2</i>	4	70	0.00034	rs2728105	0.00015	0.46	0.22
<i>MNS1</i>	15	36	0.00087	rs6493858	0.00032	0.44	-0.21
<i>GLI3</i>	7	276	0.0056	rs2282920	0.00016	0.41	0.23
<i>RFX3</i>	9	301	0.0079	rs2991306	0.00032	0.11	-0.33
<i>ACVR2B</i>	3	39	0.017	rs7645704	0.0038	0.07	0.34
<i>MGRN1</i>	16	66	0.019	rs2279408	0.0036	0.20	0.20
<i>LAMA5</i>	20	58	0.019	rs477859	0.0032	0.22	0.20
<i>RARA</i>	17	48	0.025	rs2715553	0.0085	0.47	-0.15
<i>FZD1</i>	7	4	0.027	rs6977978	0.0055	0.14	-0.20

Table 5.8: After removing *PCSK6* from the GSEA analysis, there is still an over-representation of highly associated SNPs within genes that cause LR asymmetry phenotypes in the RD relative hand skill GWAS. See Table 5.6 for an explanation

<i>Mouse Phenotype ID</i>	<i>Phenotype</i>	<i>Genes</i>	<i>Median Gene Size / KB</i>	<i>Mean Gene Size / KB</i>	<i>Expected</i>	<i>Observed</i>	<i>Adjusted P</i>
MP:0010429	abnormal heart left ventricle outflow tract morphology	47	42	83	12	13	0.41
MP:0003922	abnormal heart right atrium morphology	20	50	87	5	4	0.73
MP:0009569	abnormal left lung morphology	22	51	89	6	8	0.22
MP:0001706	abnormal left-right axis patterning	61	46	71	15	19	0.24
MP:0009570	abnormal right lung morphology	31	66	117	8	15	0.022
MP:0000644	dextrocardia	25	62	110	6	10	0.10
MP:0000284	double outlet right ventricle	86	46	91	22	34	0.022
MP:0002625	heart left ventricle hypertrophy	48	29	107	12	12	0.57
MP:0000276	heart right ventricle hypertrophy	25	18	96	6	3	0.94
MP:0004133	heterotaxia	41	45	77	10	17	0.052
MP:0000542	left-sided isomerism	16	60	86	4	7	0.094
MP:0004158	right aortic arch	36	27	80	9	14	0.091
MP:0000531	right pulmonary isomerism	15	39	77	4	5	0.35
MP:0000508	right-sided isomerism	16	46	88	4	5	0.39
MP:0002766	situs inversus	24	63	103	6	11	0.044

Table 5.9: Three of the four significant phenotypes from **Table 5.6** replicate in the GSEA of the **general population** relative hand skill GWAS (see **Table 5.6** for an explanation).

<i>Mouse Phenotype ID</i>	<i>Phenotype</i>	<i>No. Genes</i>	<i>Median Gene Size / KB</i>	<i>Mean Gene Size / KB</i>	<i>Expected</i>	<i>Observed</i>	<i>Adjusted P</i>
MP:0009570	abnormal right lung morphology	31	70	123	8	9	0.32
MP:0000284	double outlet right ventricle	85	47	90	21	32	0.019
MP:0004133	heterotaxia	40	50	82	10	16	0.026
MP:0002766	situs inversus	23	64	106	6	11	0.023

Table 5.10: **There is an over-representation of highly associated SNPs within genes involved in causing the ciliopathy related phenotype absent corpus callosum in the reading disabled cohorts.** I performed GSEA for both absent corpus callosum, hydrocephaly, and abnormal cerebellar vermis development, which typically present comorbid with LR asymmetry phenotypes associated with ciliopathies, **Table 5.6** for an explanation of the methods.

<i>Mouse Phenotype ID</i>	<i>Phenotype</i>	<i>Genes</i>	<i>Median Gene Size / KB</i>	<i>Mean Gene Size / KB</i>	<i>Expected</i>	<i>Observed</i>	<i>Adjusted P</i>
MP:0002196	absent corpus callosum	37	102	162	9	16	0.029
MP:0001891	hydrocephaly	100	56	92	25	21	1
MP:0000864	abnormal cerebellum vermis morphology	42	96	143	11	6	0.96

Table 5.11: **There is not an over-representation of highly associated SNPs within genes involved in causing ciliopathy related brain phenotypes in the general population cohorts.** I performed GSEA for both absent corpus callosum, hydrocephaly, and abnormal cerebellar vermis development, which typically present comorbid with LR asymmetry phenotypes associated with ciliopathies, **Table 5.6** for an explanation of the methods.

<i>Mouse Phenotype ID</i>	<i>Phenotype</i>	<i>Genes</i>	<i>Median Gene Size / KB</i>	<i>Mean Gene Size / KB</i>	<i>Expected</i>	<i>Observed</i>	<i>Adjusted P</i>
MP:0002196	absent corpus callosum	38	103	160	10	8	0.74
MP:0001891	hydrocephaly	96	55	90	24	31	0.20
MP:0000864	abnormal cerebellum vermis morphology	41	98	144	10	10	0.81

5.4.5 Genome-wide complex trait analysis (GCTA)

I used GCTA to estimate the phenotypic variance explained by all SNPs together (Yang et al., 2010). Although this method does not inform about the biology of the trait, it allows for an estimate to be made of the contribution of common genetic variation to its heritability. The estimate of the proportion of variation explained by directly genotyped autosomal SNPs in cohorts 1 & 2 is 0.00 (SE = 0.566). For genotyped and imputed autosomal SNPs combined the estimate is 0.00 (SE = 0.514).

5.5 Discussion

I have performed a genome-wide association study (GWAS) meta-analysis for a quantitative measure of handedness and found association with variants that implicate genes involved in the determination of left/right body asymmetry. To achieve this, I developed a novel approach to GWAS pathway analysis, coupling gene-set enrichment analysis (GSEA) with mouse phenotype data. In the RD meta-analysis, there was an over-representation of variants with low P values within human orthologues of genes that cause LR asymmetry phenotypes when knocked-out in mice and I replicated the same over-representation in the general population. I also confirmed that the gene *PCSK6* was associated with relative hand skill at genome-wide significance (RD meta-analysis, cohorts 1-3, rs7182874 $P = 8.68 \times 10^{-9}$).

The genes identified by GSEA in the meta-analysis of cohorts with RD are involved in the determination of LR asymmetry in mice (**Table 5.7**). *Mns1* and the transcription factor *Rfx3* both play key roles in ciliogenesis (El Zein et al., 2009; Zhou et al., 2012). The transcription factor *Gli3* is expressed at primary cilia, localizing to both

the distal tip of the cilia as well as the nucleus, and is involved in Sonic hedgehog signal transduction (Haycraft et al., 2005). The unidirectional fluid flow created by clockwise rotation of cilia that breaks asymmetry is detected via the Ca^{2+} channel PKD2 (Yoshida et al., 2012). This flow concentrates the morphogen NODAL on the left, which is cleaved into an active form by the protease PCSK6 (Constam and Robertson, 1999). Artificially reversing the direction of 'nodal flow' in mice causes asymmetry phenotypes such as situs inversus (Nonaka et al., 2002). The NODAL signal is then transduced through the transmembrane receptor ACVR2B (Yeo and Whitman, 2001). Retinoic acid acts through receptors such as RARA to affect expression of *Nodal*, *Pitx2*, and *Lefty* in mice (Chazaud et al., 1999). FZD1 also acts upstream of NODAL, it is a receptor for WNT3A (Gazit et al., 1999), which activates the Delta/Notch pathway to regulate expression of *Nodal* in mice (Nakaya et al., 2005). Similarly, in a significant proportion of mouse embryos where *Mgrn1* is disrupted, *Pitx2*, *Lefty1*, and *Lefty2* are abnormally expressed (Cota et al., 2006). LAMA5 is a component of the extracellular matrix, its precise role in the development of LR asymmetry is unknown. However, mice in which *Lama5* has been disrupted develop Polycystic Kidney Disease (PKD) (Shannon et al., 2006), a phenotype comorbid with LR asymmetry defects caused by ciliopathies (Badano et al., 2006).

I have shown that our findings from the RD cohorts hold true for the general population, three of the four phenotype gene sets tested are significant at $\text{FDR} \leq 5\%$ (heterotaxia, situs inversus, and double outlet right ventricle). However, the associated variants and genes are different. For example, the most highly associated marker in the general population GWAS (rs7883190, $P = 2.08 \times 10^{-6}$, **Table 5.5**) is located 6kb upstream of *GPC3* on the X chromosome, a gene that causes the double outlet right ventricle phenotype when disrupted in mice (Ng et al., 2009).

If genes that are critical to the breaking of LR body symmetry influence handedness and brain asymmetry, individuals with disorders that cause asymmetry defects such as situs inversus (SI) should have a greater prevalence of left-handedness. However, studies dating back as far as 1836 have shown that individuals with SI present with a prevalence of left-handedness equal to the rest of the population (Cockayne, 1938; Torgersen, 1950; Watson, 1836). A more recent study of individuals with primary ciliary dyskinesia (a ciliopathy that causes SI) has confirmed this finding (McManus, 2004). However, two brain imaging studies of individuals with SI have shown a significant reversal of the typical pattern of right-frontal and left-occipital petalia asymmetry (Ihara et al., 2010; Kennedy et al., 1999), of which one study also showed a significant reversal of language dominance (Ihara et al., 2010). However, each study only assessed three individuals with SI, and given its rarity (~ 1 in 10,000 (Torgersen, 1950)), large-scale studies have yet to be performed to uncover the precise relationship between body asymmetry, brain asymmetry and handedness.

To help explain this relationship, I hypothesize that the same molecular mechanism for visceral LR asymmetry is reused later during cerebral development. In support of this, RFX3 regulates the expression of *Gli3* in the telencephalon in mice, which in turn regulates the distribution of guidepost neurons necessary for corpus callosum formation (Benadiba et al., 2012). Alongside asymmetry defects, *Rfx3* deficient mice show corpus callosum agenesis and disorganization of guidepost neurons necessary for axon pathfinding across the brain's midline (Benadiba et al., 2012). Humans with ciliopathies commonly present with two specific structural changes within the brain; hydrocephaly and an absent corpus callosum (Badano

et al., 2006). Because four of the five most strongly associated genes shown in **Table 5.7** are involved in ciliogenesis, I therefore performed GSEA on these two phenotypes as well. Enrichment was detected in absent corpus callosum, but not in hydroencephaly. The only gene that overlaps with the LR asymmetry genes in **Table 5.7** in the absent corpus callosum phenotype is *GLI3*. However, *Rfx3* has not yet been annotated to this phenotype in the mouse phenotype ontology (Benadiba et al., 2012), while *PCSK6* is also known to have high expression in the corpus callosum (Johnson et al., 2003). However, the evidence for a relationship between handedness and corpus callosum size is mixed (for review see (Corballis et al., 2012)). Furthermore, a study of 12 infants with agenesis of the corpus callosum show similar bias in right-handedness and performance on unimanual hand skill tasks to controls (Sacco et al., 2006).

The relationship between these genes and both corpus callosum development and brain asymmetry are not well understood in humans. However, brain asymmetries have been better studied in zebrafish and have revealed that these processes are under the control of Nodal-related genes, suggesting a conserved pathway controlling both visceral and cerebral asymmetries (Concha et al., 2000). A particular line of zebrafish, frequent-situs-inversus (*fsi*), displays reversal of both body and neuronal asymmetries, as well as of some characteristic behavioral phenotypes, such as eye preference and direction of prey approach (Barth et al., 2005). Our study implicates Nodal-related genes in the development of handedness and therefore brain asymmetry in humans.

This study has three key limitations; both genotyping and phenotyping study heterogeneity, and low power due to a small sample size. Firstly, genotypes from

ALSPAC were imputed with HapMap phase II instead of the 1000 genomes reference panel. This means I was restricted to analyzing only ~ 2 million SNPs instead of over 6 million common SNPs in the 1000 genomes project. The imputation accuracy is also lower with HapMap II, however I restricted analyses to common variants ($MAF \geq 5\%$), for which imputation accuracy is high (box 4 of (Marchini and Howie, 2010)). Secondly, not only was the relative hand skill phenotype slightly different in ALSPAC (**section 2.1**), it was only measured once instead of five times. Test-retest reliability of the pegboard task is lower when the test is performed once (correlation = 0.37-0.7) versus three times (correlation ≥ 0.8) (Buddenberg and Davis, 2000). Furthermore the phenotype is only a proxy for one specific aspect of handedness, namely manipulating objects quickly with each hand separately. It is strongly correlated with both hand preference and dexterity (**subsection 5.4.1**), and it is not clear whether the associations observed are due to handedness or dexterity or a combination of the two. Thirdly, although I interrogated a quantitative phenotype, which is more powerful than binary case-control GWASs that have been performed (Armour et al., 2013; Eriksson et al., 2010), the numbers of participants were relatively small (discovery cohorts: $n = 744$ and replication: $n = 2,666$). The zero estimate of genetic contribution to phenotypic variance and high standard error when performing GCTA in cohorts 1 & 2 further suggests that the GWAS is underpowered. Therefore I was unable to estimate the proportion of phenotypic variance explained by all the SNPs together, the true contribution of common genetic variation is swamped by noise. If handedness is indeed controlled by many genes, much larger cohorts will be needed to unravel its full biological complexity.

Based on my results I propose that the same biological mechanism underlying the development of LR asymmetry in the body plays a role in determining handed-

ness. However, although I replicated the same LR asymmetry phenotypes in the general population as the RD cohorts, different SNPs located in different genes within these same pathways were responsible for driving the association. Furthermore, the genome-wide significant SNP in *PCSK6* does not even show the same trend in the general population (**Table 5.3**). This suggests there is both allelic and locus heterogeneity between the RD and general population cohorts. In addition, an independent study found that the variable number tandem repeat (VNTR) located 522bp from the genome-wide significant associated SNP in *PCSK6* is associated with degree of handedness in a cohort not selected for dyslexia, further supporting this hypothesis (Arning et al., 2013).

This heterogeneity may derive from epistasis between genes involved in dyslexia and those involved in handedness. In support of this a cilia-related coexpression module derived from publically available microarray datasets finds that the dyslexia associated genes, *DCDC2*, *DYX1C1*, and *KIAA0319* are coexpressed in cilia, suggesting they interact with genes involved in the determination of LR asymmetry (Ivliev et al., 2012). *Dyx1c1* is upregulated during ciliogenesis and localizes to centrioles and basal bodies of cilia in multiciliated tracheal epithelial cells in mice (Hoh et al., 2012). Disrupting *Dyx1c1* in mice causes chronic airway disease, laterality defects, and male infertility, resembling primary ciliary dyskinesia (PCD) (Tarkar et al., 2013). Similarly, inhibition of *dyx1c1* in zebrafish reduces the length of cilia, and produces asymmetry phenotypes such as situs inversus (Chandrasekar et al., 2013; Tarkar et al., 2013). Furthermore, recessive loss of function mutations in *DYX1C1* have been identified in 12 patients with PCD (none of whom have dyslexia) (Tarkar et al., 2013).

No consistent association has been found between either hand preference or relative hand skill and dyslexia (Bishop, 1990b; Francks et al., 2003b), and my study confirms this. However, early work by Galaburda et al. reported eight consecutive post-mortem specimens with RD that had reduced planum temporale asymmetry (Galaburda, 1989; Galaburda et al., 1985). This suggests subtle structural asymmetries might be involved in dyslexia. Furthermore, reading performance is correlated with motor abilities on the peg-board test (Francks et al., 2003b; Stoodley and Stein, 2006). I replicated this finding by comparing cohorts 3 and 4 in the current study - demonstrating that individuals with RD are slower at the peg-board test than those in the general population ($P = 0.001$, **Table 5.2**). This is consistent with a broad literature showing motor skill difficulties in children with dyslexia (Fawcett et al., 1996; Getchell et al., 2007; Iversen et al., 2005; Nicolson and Fawcett, 1990; Ramus et al., 2003).

The precise relationship between handedness, cerebral asymmetry, and neurodevelopmental disorders such as dyslexia remains to be determined. However, my results have implicated some of the same genes involved in body asymmetry in handedness, and therefore the development of cerebral asymmetry in both individuals with RD and the general population. Larger scale GWA or sequencing studies of handedness, coupled with a deeper understanding of the genetic basis of language-related disorders such as dyslexia will help determine whether handedness and language do indeed develop hand-in-hand.

Chapter 6

Genome-wide association study for dyslexia and reading-related traits

6.1 Abstract

I report a genome-wide association study meta-analysis for both a case-control dyslexia GWAS, and quantitative measures of reading ability in cohorts 1 & 2 (cases $n = 562$, controls $n = 1,339$). In the case-control analysis, no SNP was genome-wide significant. In the quantitative analysis, I tested four reading-related traits individually and combined in a multivariate analysis. There was one genome-wide significant locus associated with non-word reading. The most strongly associated variant, rs16875959 ($P = 4.89 \times 10^{-9}$), is located in *MAP1B*, which plays a key role in axonogenesis and neuronal migration. There was also a suggestive signal in rs11594922 ($P = 2.52 \times 10^{-6}$), located in the gene *CACNB2*, an L-type Ca^{2+} channel associated with five other neurodevelopmental disorders. However, these two signals do not replicate in either of the two reading disabled cohorts 3 & 5 ($n = 166$ & 199), or in the general population cohort 4 ($n = 2,365$), suggesting they are either

false positives or that larger, better ascertained, cohorts will be necessary in order to replicate the signal. Furthermore, gene-set enrichment analysis of the GWAS data finds no evidence to support theories that axonogenesis / neuronal migration, brain midline development, or ciliogenesis are important in the etiology of dyslexia.

6.2 Introduction

For an introduction to dyslexia and what is known about the genetics of the disorder, see **section 1.2**.

In order to identify new candidate genes for dyslexia, here I perform a genome-wide association case/control study. Furthermore within the dyslexia cohorts I perform a quantitative GWAS of a set of four correlated reading-related traits that assess; single-word reading, spelling, orthographic coding (CCI, irregular word reading), and phonological decoding ability (CCN, non-word reading, **section 2.1**). I also combine these four correlated phenotypes in a Bayesian multivariate linear model. I find one genome-wide significant locus associated with CCN in the gene *MAP1B*, and a borderline significant variant in *CACNB2*. I obtain access to genotype data for these variants in ALSPAC, dividing the cohort into individuals with reading disability and the general population (cohorts 3 and 4), as well as the NeuroDYS Bonn sample (cohort 5, **Table 1.2**), and find that these SNPs do not replicate in the independent cohorts. Furthermore, I perform gene-set enrichment analyses (like in **subsection 5.4.4**) and find no evidence for an enrichment of SNPs located in genes that cause axonogenesis / neuronal migration, brain midline, or ciliogenesis defects when disrupted in mice.

6.3 Methods

6.3.1 Cohorts

I performed a case-control GWAS for dyslexia in two cohorts. The first is 204 individuals (cohort 1) compared to 1,041 unselected controls from the 1958 Birth Cohort genotyped on the Illumina 550K SNP array as part of the Wellcome Trust Case Control Consortium (WTCCC, (The Wellcome Trust Case-Control Consortium, 2007)). The second is 358 unrelated individuals (cohort 2) compared to 298 unselected controls from the People of the British Isles Study genotyped on the Illumina OmniExpress SNP array (POBI, (Winney et al., 2012)). All cases and controls were genotyped by the Bioinformatics core at the Wellcome Trust Centre for Human Genetics in Oxford.

6.3.2 Genotyping, imputation, and statistical analysis

Genotyping was performed as described in **subsection 3.3.1** & **section 4.2**. Imputation was performed as in **subsection 5.3.2**, I only tested SNPs that had a MAF ≥ 0.01 , and an imputation info ≥ 0.8 (an estimation of the uncertainty in the genotype estimation). For the 550K dyslexia case-control GWAS, after imputation I tested 7,820,691 variants for association, and for the Illumina OmniExpress dyslexia case-control GWAS, I tested 8,077,363 variants for association. For numbers of SNPs tested for association with each quantitative reading trait, see **Table 6.1**. Statistical analysis was performed using SNPtest as described in **subsection 5.3.3**.

The four quantitative reading traits tested; READ, SPELL, phonological decoding ability (CCN), and the irregular word test (CCI), are described in **section 2.1**. Most

of the individuals performed below average for their age-group in these tests, a summary of their Z scores relative to the general population for each test in the two cohorts is given in **Table 6.4**. The four quantitative traits were combined in a Bayesian linear model and tested for association in SNPtest using sex, age, and non-verbal IQ as covariates. This test outputs a $\log_{10}$ Bayes Factor (BF), and a BF of 1,000 ($\log_{10}$ BF = 3) is suggestive of association. For a SNP then to be declared as associated, a combined BF of 1,000,000 in the discovery and replication cohorts is necessary (Jonathan Marchini, *personal communication*).

Table 6.1: **Summary of the genomic inflation factor, λ , for the quantitative reading measures GWAS.** λ was calculated for both directly genotyped and imputed markers in both studies.

Reading Measure	550K					OmniExpress				
	Genotyped λ	<i>n</i>	Imputed λ	<i>n</i>	Total <i>n</i>	Genotyped λ	<i>n</i>	Imputed λ	<i>n</i>	Total <i>n</i>
READ	1.01	482,050	1.00	7,446,337	7,928,387	1.03	636,184	1.02	7,457,072	8,093,256
SPELL	1.02	485,034	1.00	7,495,770	7,980,804	1.02	636,163	1.02	7,454,385	8,090,548
CCI	1.01	485,092	1.00	7,500,969	7,986,061	1.03	637,022	1.02	7,472,247	8,109,269
CCN	1.00	485,125	1.00	7,468,301	7,953,426	1.03	636,206	1.02	7,446,293	8,082,499

6.3.3 Principle component analysis

In the initial GWAS case/control analysis on the 550K SNP array, one SNP appeared to reach genome-wide significance at $P \leq 5 \times 10^{-8}$. However, given the small number of individuals tested this was a surprising result. I therefore considered whether there was any population stratification between cases and controls driving the association. I performed principle component analysis (PCA) as detailed previously (Anderson et al., 2010). A plot of the first two principal components (PC, **Figure 6.3**) suggested some stratification that I controlled for by taking the first ten PCs as covariates. After this, no SNP reaches genome-wide significance at $P \leq 5 \times 10^{-8}$, and the genomic inflation factor (λ), calculated as the median chi-squared value divided by 0.456, is 1.072.

6.3.4 Reclustering of SNPs for cases and controls on the Illumina OmniExpress array

Initially six SNPs reach genome-wide significance at $P \leq 5 \times 10^{-8}$ in the Illumina OmniExpress case/control GWAS, **Table 6.2**. However, given the power of the study (358 cases and 298 controls), how low some of the P values are (e.g. rs4802078 – $P = 2.38 \times 10^{-50}$), and the lack of further significant SNPs in linkage disequilibrium, they are false positives. An inspection of the genotype calling in *GenomeStudio* (Illumina®) reveals that these significant P values are due to errors in allele calling between the cases and controls that has led to two clusters being assigned the same genotype in cases, but not in controls. An example, rs4802078, is shown in **Figure 6.1**.

Furthermore, inspection of the QQ plot (**Figure 6.2**) reveals an early departure from the expected P value, which is confirmed by the genomic inflation factor ($\lambda = 1.078$). A possible explanation for this inflation of P values is population stratification. I had excluded individuals with non-European ancestry in **subsection 4.2.4**, however there may be more subtle population stratification generated because the controls originate from across the British Isles (POBI), whereas the cases are from Southern England. Inflation could also be caused by differences in genotype calling between cases and controls. For example, ‘informative missingness’ is where all failed genotypes in the cases or the controls are either heterozygotes or homozygotes for a specific allele. Given that the cases and controls were genotyped separately, subtle differences in sample quality or batch effects may have led to informative missingness generating inflated P values (McCarthy et al., 2008).

In order to correct the artefacts generated by differences in allele calling between

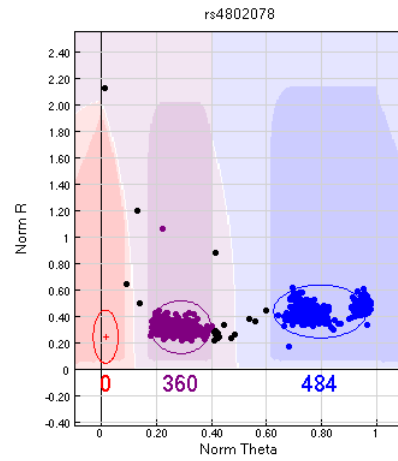
cases and controls, and control for population stratification (**Figure 6.3**) I decided to:

- Combine the cases and controls, and recluster the SNPs.
- Calculate the first ten principal components (PC), and use them as covariates in the association analyses.

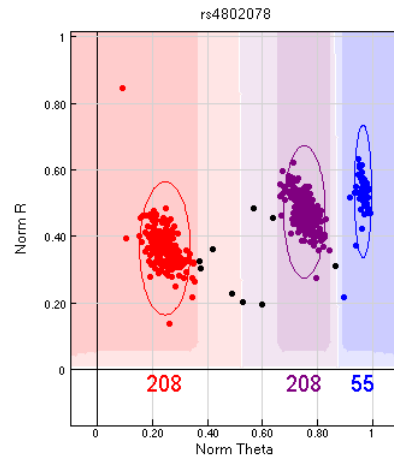
Reclustering of the SNPs removed the artefacts observed in **Table 6.2**, for example the SNP rs4802078 has now been clustered correctly (**Figure 6.1**) and the P value has changed from $P = 2.38 \times 10^{-50}$ to $P = 0.14$. However, the genomic inflation factor is still high ($\lambda = 1.07$). Taking the first ten principal components as covariates did not reduce the inflation factor ($\lambda = 1.08$), suggesting that the inflation may be due to either artefacts such as DNA quality that we cannot control for, batch effects, or is due to a genuine enrichment of SNPs with low P values. The QQ plot is shown in **Figure 6.2**. Had any signals reached genome-wide significance after this process we would also perform genomic control of summary statistics to correct for variation between studies (Devlin and Roeder, 1999).

Table 6.2: False positive associations from the Illumina OmniExpress dyslexia GWAS. Spurious associations arose from using the genotyping calling algorithm separately for cases and controls. Listed are SNPs with $P \leq 5 \times 10^{-8}$ from the dyslexia case-control GWAS. Inspection of the allele clusters in *GenomeStudio* suggests that these results are false positives generated by errors in allele calling between cases and controls, **Figure 6.1**.

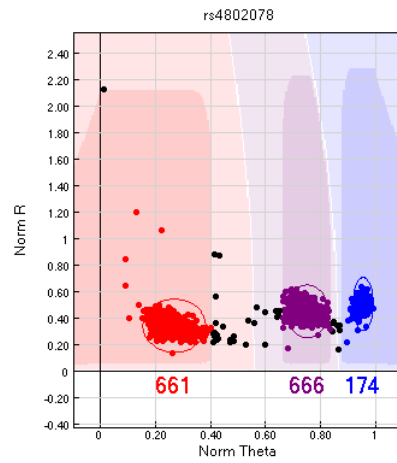
<i>Chr</i>	<i>rsnum</i>	<i>Position</i>	<i>Case f</i>	<i>Control f</i>	<i>P value</i>	<i>OR</i>	<i>Gene</i>
7	rs3757496	32861417	0.21	0.63	3.69×10^{-52}	0.16	-
19	rs4802078	40968842	0.24	0.65	2.38×10^{-50}	0.17	<i>BLVRB</i>
1	rs7519458	4007110	0.15	0.53	2.56×10^{-48}	0.16	<i>LOC728716</i>
14	rs12883530	58097226	0.16	0.54	2.45×10^{-47}	0.17	<i>SLC35F4</i>
2	rs6546667	71313262	0.59	0.36	6.12×10^{-17}	2.6	-
21	rs11910489	18046904	0.11	0.0084	2.11×10^{-14}	15	-



(a)



(b)



(c)

Figure 6.1: Plots of the clusters generated by Illumina's *GenomeStudio* program for the SNP rs4802078 in (a) cohort 2, (b) POBI control cohort, and (c) the two cohorts combined.

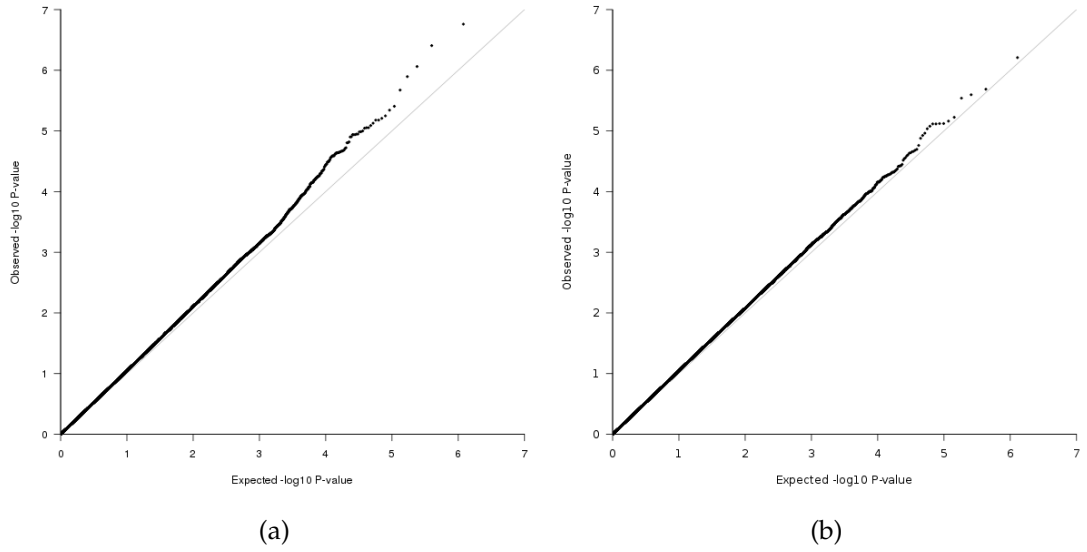


Figure 6.2: QQ plots of the P values for Illumina OmniExpress dataset, comparing dyslexia cases (cohort 2) to POBI controls, before and after controlling for allele calling artefacts and population stratification. (a) Initial plot. the deviation from the null (diagonal line) is likely due to both systematic error in genotype calling between cases and controls, exemplified in Figure 6.1, and population stratification. (b) Plot after controlling for artefacts and stratification.

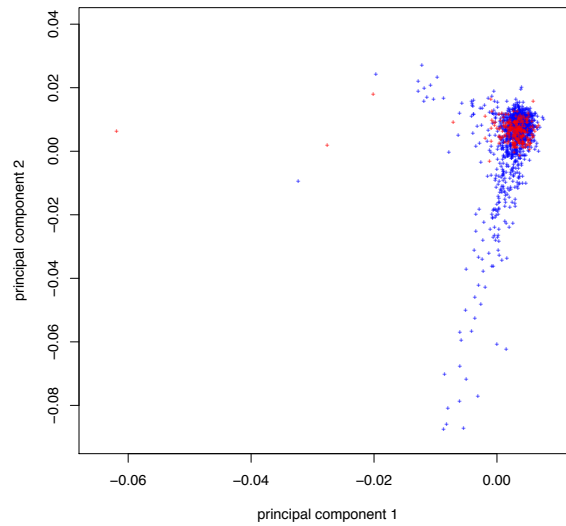
6.4 Results

6.4.1 Illumina 550K

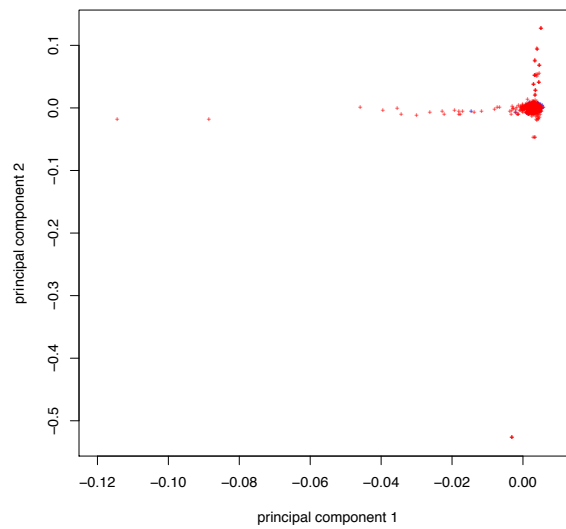
The most highly associated SNP is rs838600 ($P = 2.32 \times 10^{-6}$). It is located in an intron of *SLC9A9* on chromosome 3. A list of the SNPs with $P \leq 1 \times 10^{-4}$ is shown in Table 11.8. A quantile-quantile (QQ) and manhattan plot are shown in Figure 6.4.

6.4.2 Illumina OmniExpress case/control GWAS

The most highly associated SNP is rs8083477 ($P = 6.18 \times 10^{-7}$), located 91,588bp upstream of *RAB12* on chromosome 18. The Manhattan plot, and table of the most highly associated SNPs are shown in Figure 6.5 & Table 11.9.

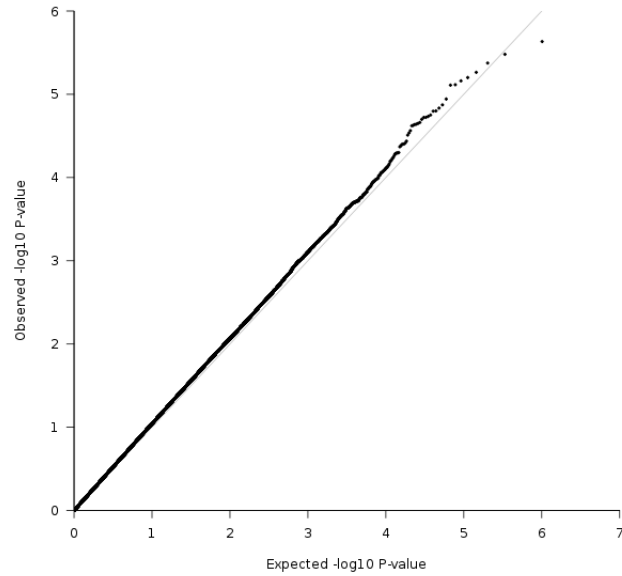


(a)

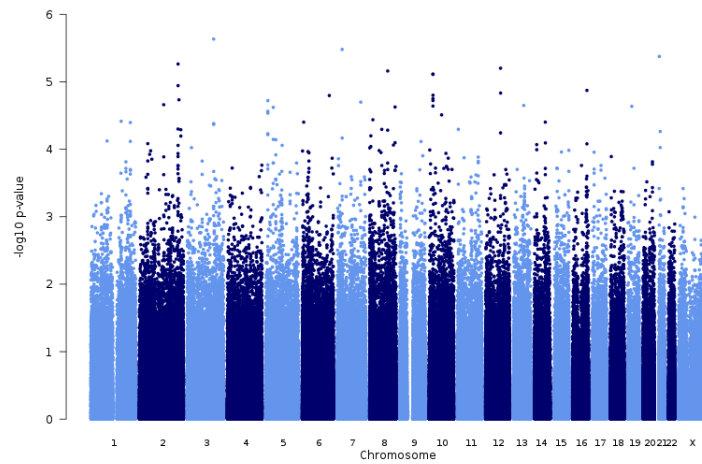


(b)

Figure 6.3: **Ancestry clustering based on principal component analysis of the cases and controls** genotyped on the (a) Illumina 550K and (b) Illumina OmniExpress chips. Controls are labelled as +, and cases +. The first ten principal components were used as covariates to control for population stratification in the GWAS.



(a)



(b)

Figure 6.4: **QQ and Manhattan plots of the P values for the Illumina 550K dyslexia case-control GWAS.** (a) QQ plot, the P values are plotted against the expected P values, a deviation from the null (diagonal line) indicates an enrichment of genuine associations, population stratification or an error in genotype calling. (b) Manhattan plot, the negative $\log_{10}$ transformed P values are plotted against chromosome and position along the genome.

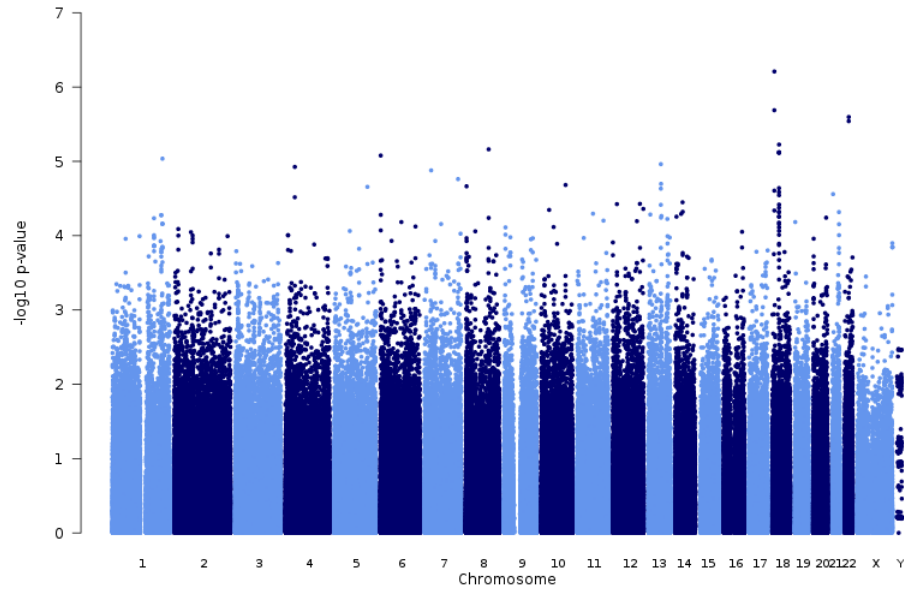


Figure 6.5: A Manhattan plot of the P values for the Illumina OmniExpress dataset, comparing dyslexia cases to POBI controls.

6.5 Imputation

For the 550K dyslexia case-control GWAS, after imputation I tested 7,820,691 variants for association, none of which were genome-wide significant at $P \leq 5 \times 10^{-8}$ (**Table 11.10**).

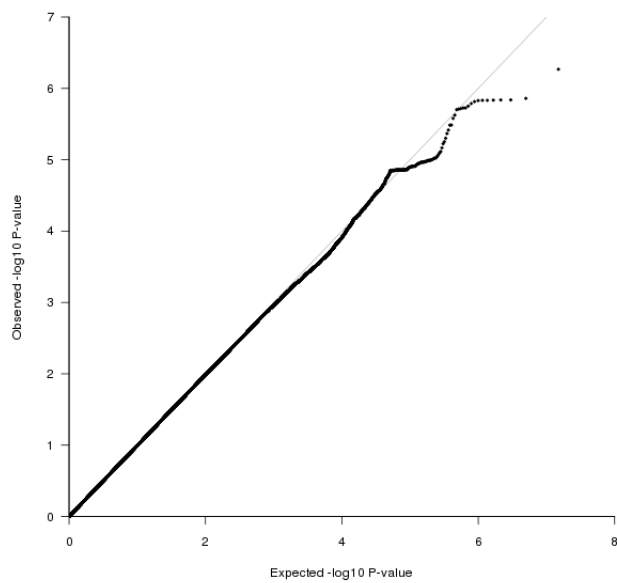
For the Illumina OmniExpress dyslexia case-control GWAS, after imputation I tested 8,077,363 variants for association, none of which are genome-wide significant at $P \leq 5 \times 10^{-8}$ (**Table 11.11**).

6.6 Meta-analysis

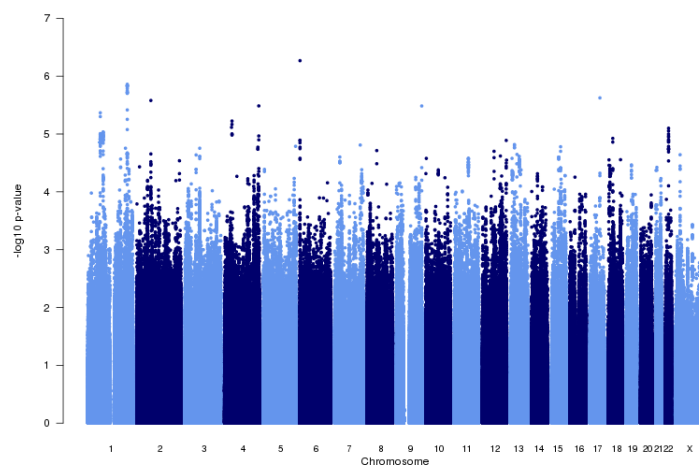
I meta-analysed the combined Illumina 550K and OmniExpress dyslexia case-control GWAS data-sets imputed using Phase I of the 1000 Genomes Project using the program GWAMA (Magi and Morris, 2010), which performs fixed-effects meta-analysis with inverse variance weighting of allelic effect estimates. In total, I tested 7,476,092 markers that were either directly genotyped or imputed in both data sets. No marker was genome-wide significant at $P \leq 5 \times 10^{-8}$ (**Table 11.12**). The most highly associated variant, rs9502966 ($P = 5.39 \times 10^{-7}$), is located 54.8kb upstream of the single exon gene *FOXC1* on chromosome 6. The genomic inflation statistic (λ) is shown for each data set in **Table 6.3**, and the QQ and Manhattan plots are shown in **Figure 6.6**.

Table 6.3: **Summary of λ for the dyslexia case-control GWAS.** λ was calculated for both directly genotyped and imputed markers in both studies.

<i>Cohort</i>	<i>Genotyped λ</i>	<i>Genotyped n</i>	<i>Imputed λ</i>	<i>Imputed n</i>	<i>Total n</i>
550K	1.072	492,969	1.07	7,327,722	7,820,691
OmniExpress	1.078	637,105	1.084	7,440,258	8,077,363
Combined					7,476,092



(a)



(b)

Figure 6.6: QQ and Manhattan plot of the P values for the meta-analysis of the dyslexia case-control GWA data-sets.

6.7 Quantitative reading-related traits GWAS

Quantitative reading-related traits were collected as part of the original ascertainment of individuals with dyslexia including; single-word reading ability (READ), spelling ability (SPELL), an irregular word test (CCI, which assesses orthographic coding ability), and phonological decoding ability (CCN, non-word reading, **section 2.1**). Most of the individuals performed below average for their age-group in these tests, a summary of their Z scores relative to the general population for each test in the two cohorts is given in **Table 6.4**. Visual inspection of the histograms and QQ plots for each phenotype, alongside kurtosis and skewness scores, reveals that each phenotype is roughly normally distributed (**Figure 6.7**, **Figure 6.8**, & **Table 6.5**). However CCI and CCN show slightly 'fat' tails, i.e. platykurtic distributions with a wider spread of values around the mean (kurtosis = 2.5 & 2.2 respectively). SPELL is slightly skewed to the left (skewness = -0.43).

I tested these quantitative traits in each cohort, using sex and age as covariates, and meta-analysed the results using GWAMA (Magi and Morris, 2010). The QQ and Manhattan plots are shown in **Figure 6.9** & **Figure 6.10** and a summary of λ and the number of SNPs tested is in **Table 6.1**). No marker is genome-wide significant for READ, SPELL, or CCI (the details of the most highly associated SNPs can be found in the appendix in **Table 11.13**, **Table 11.14**, and **Table 11.15**). However, three markers in one locus on chromosome 5 are genome-wide significant with CCN (**Table 6.6**). The most highly associated marker, rs16875959 ($P = 4.89 \times 10^{-9}$), is located in intron 4 of the six exon, 102,280bp, gene microtubule associated protein 1B (*MAP1B*) on chromosome 5 (**Figure 6.11**). The effect allele has a minor allele frequency of 0.11 and each copy is associated with an increased score of 0.46 standard deviations, i.e. the minor (derived) allele is protective against phono-

Table 6.4: **A summary of the quantitative reading-related scores in both cohorts.** Each score is standardised on a population level to a mean of zero and a standard deviation of one (Z scores). Since the cohort is composed of individuals with dyslexia, most score below zero for each reading measure, and the standard deviation (SD) for the measures in each cohort is smaller than one.

<i>Reading Measure</i>	<i>n</i>	<i>550K</i>			<i>n</i>	<i>OmniExpress</i>		
		<i>Mean Z Score</i>	<i>Range</i>	<i>SD</i>		<i>Mean Z Score</i>	<i>Range</i>	<i>SD</i>
<i>READ</i>	198	-1.25	-3.00 – 0.70	0.69	358	-1.22	-2.87 – 1.20	0.67
<i>SPELL</i>	190	-1.56	-3.60 – 0.30	0.8	357	-1.28	-3.00 – 0.13	0.61
<i>CCI</i>	195	-1.66	-4.27 – 0.29	0.85	350	-1.38	-3.51 – 1.22	0.87
<i>CCN</i>	197	-1.00	-2.87 – 1.06	0.83	358	-0.88	-2.64 – 0.66	0.77

Table 6.5: **A summary of the skewness and kurtosis of the reading-related phenotypes measured in cohorts 1 and 2 combined.** A skewness score ≤ -1 or ≥ 1 indicates the data is significantly skewed. A kurtosis score of 3 indicates a mesokurtic distribution, i.e. that the sample is normally distributed. A score greater than 3 indicates a leptokurtic distribution, with values concentrated around the mean and thicker tails. A score less than 3 indicates a platykurtic distribution, flatter than a normal distribution with a wider spread of values around the mean.

<i>Phenotype</i>	<i>Skewness</i>	<i>Kurtosis</i>
<i>READ</i>	0.16	3.1
<i>SPELL</i>	-0.43	3.1
<i>CCI</i>	-0.0087	2.5
<i>CCN</i>	0.0025	2.2

logical decoding difficulties. Conditioning on the most highly associated SNP, rs16875959, suggests that the association at this locus is driven by one etiological allele (**Figure 6.12**). rs16875959 is associated with the other reading phenotypes at $P \leq 0.05$, as well as SIM IQ (verbal reasoning), but not MAT IQ (non-verbal reasoning) or PegQ (**Table 6.7**). This is a reflection of the correlation between these traits (shown in **Table 2.1**).

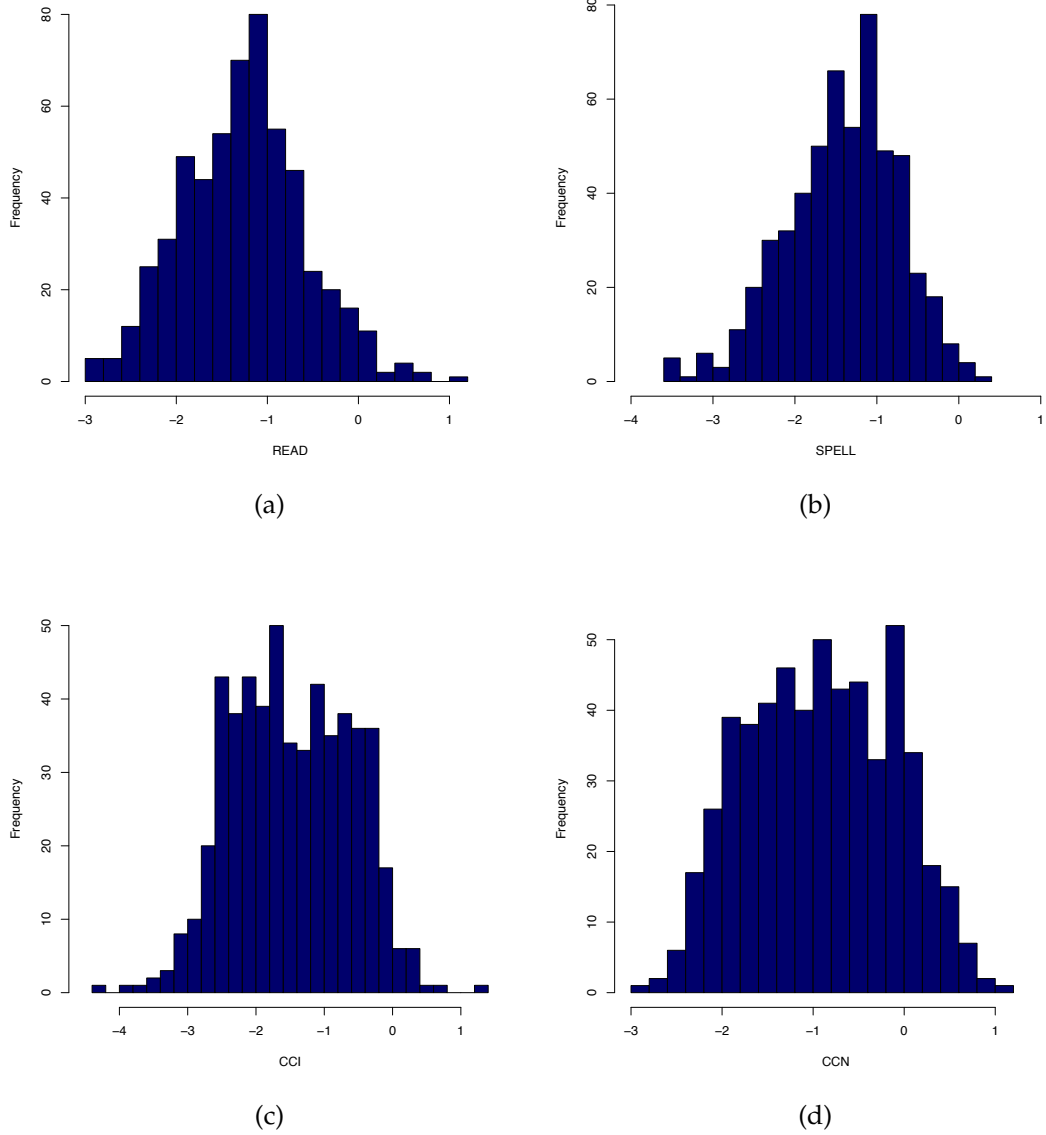
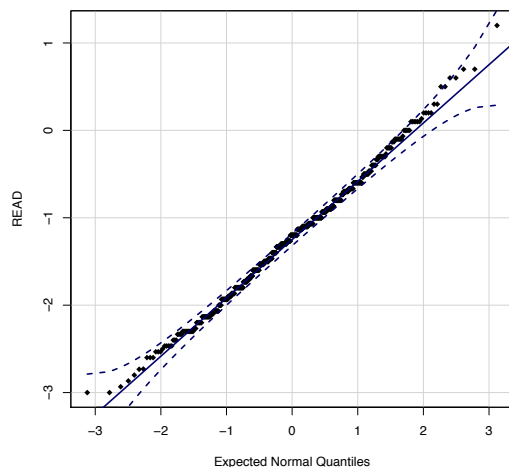
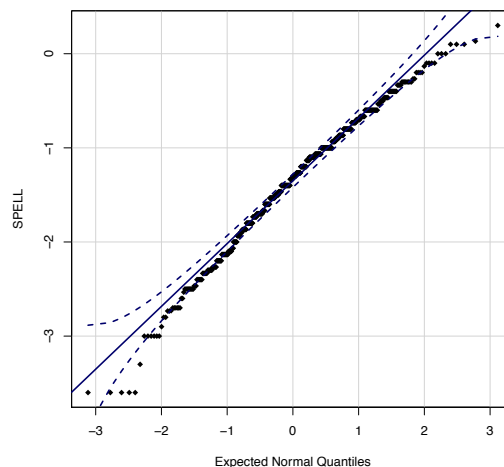


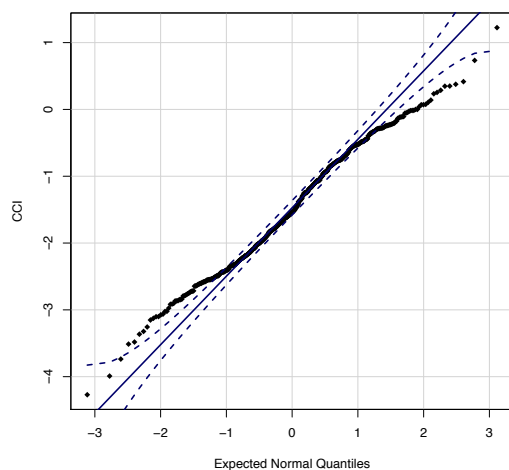
Figure 6.7: **Histograms of reading-related traits in cohorts 1 & 2;** (a) READ, (b) SPELL, (c) CCI and (d) CCN. Since the cohorts are all individuals with dyslexia, the reading Z scores are typically below the mean relative to a general population reference.



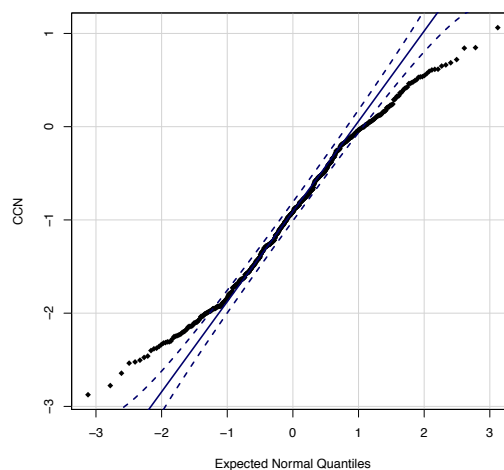
(a)



(b)



(c)



(d)

Figure 6.8: **QQ plots of reading-related traits; (a) READ, (b) SPELL, (c) CCI and (d) CCN.** 95% confidence intervals are given by the - - - lines.

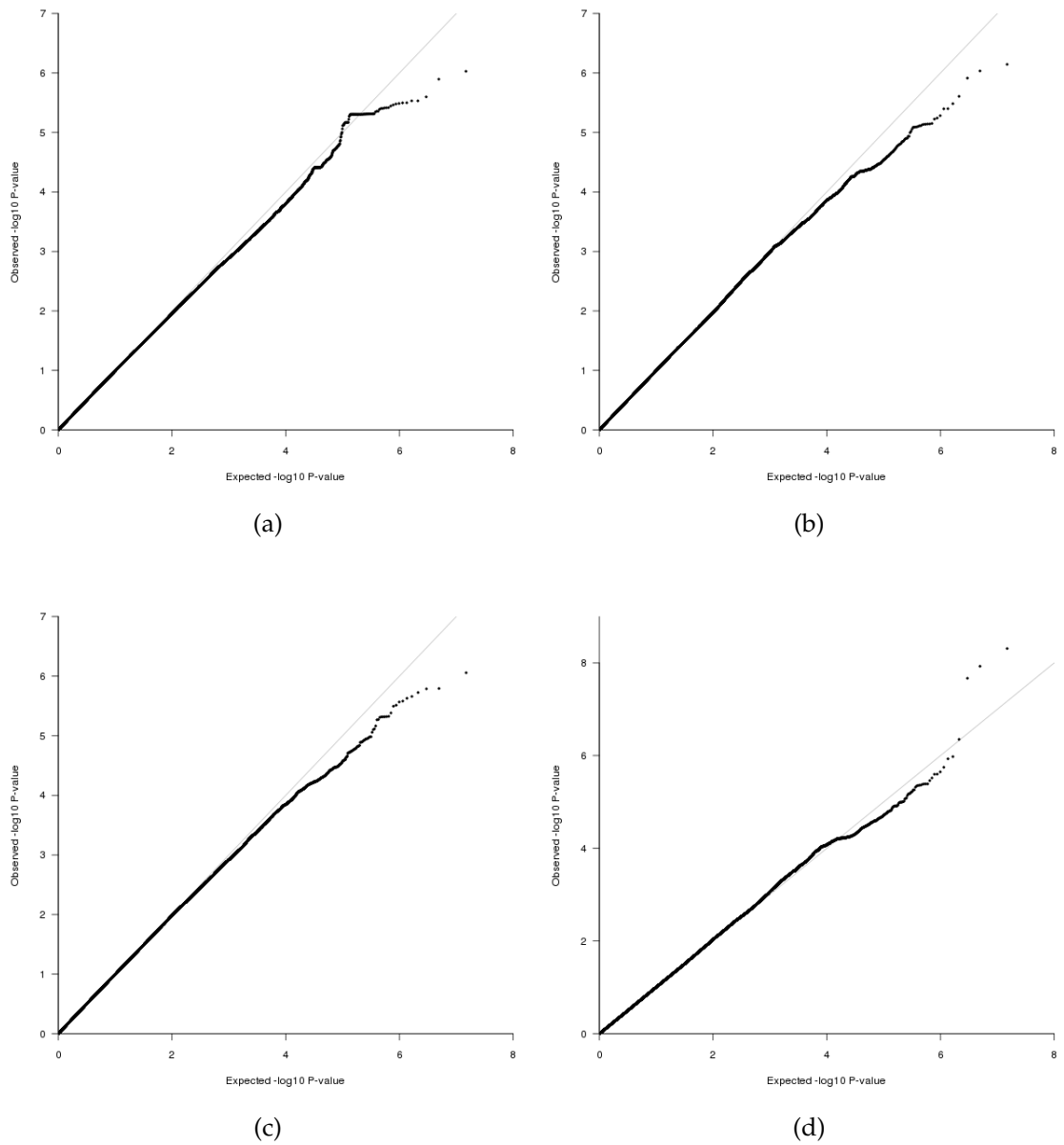


Figure 6.9: QQ plots of the P values for the quantitative reading scores in the combined dyslexia cohorts; (a) READ, (b) SPELL, (c) CCI and (d) CCN.

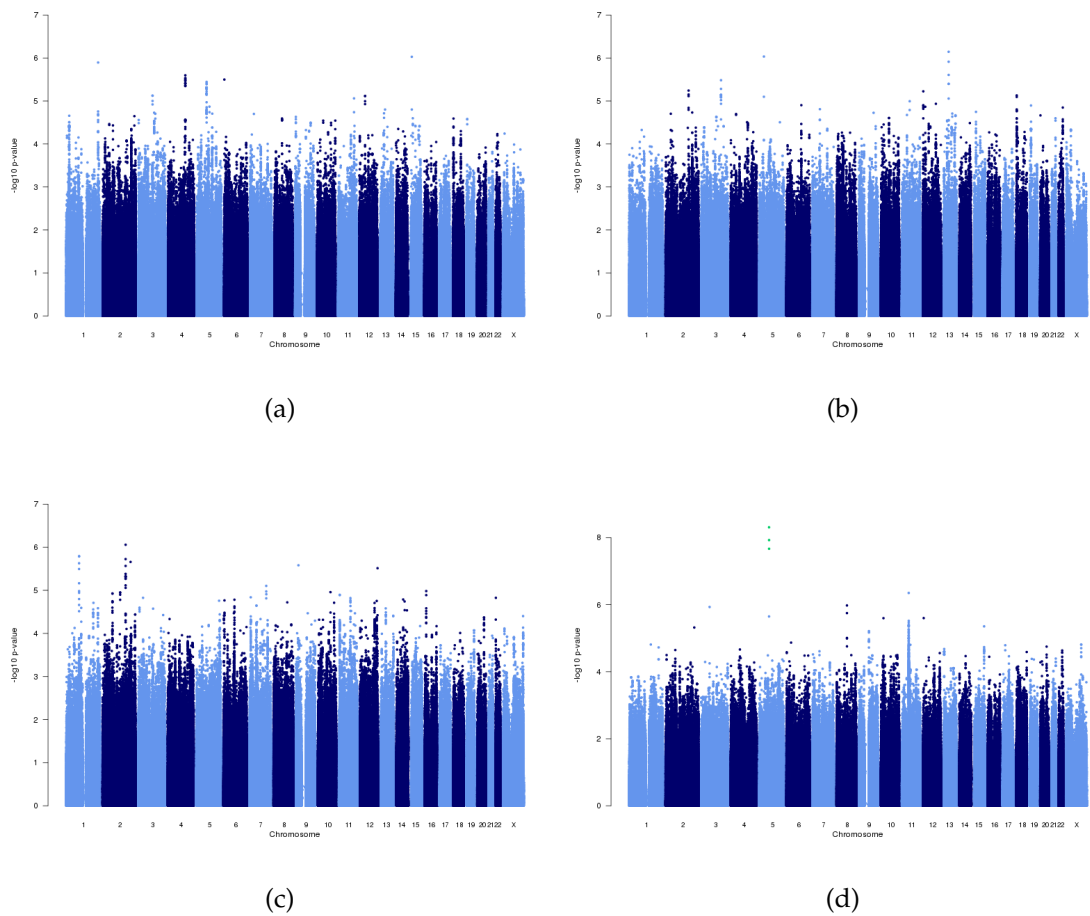


Figure 6.10: **Manhattan plots of the P values for the quantitative reading scores in the combined dyslexia cohorts; (a) READ, (b) SPELL, (c) CCI and (d) CCN. Genome-wide significant SNPs are shown as ●.**

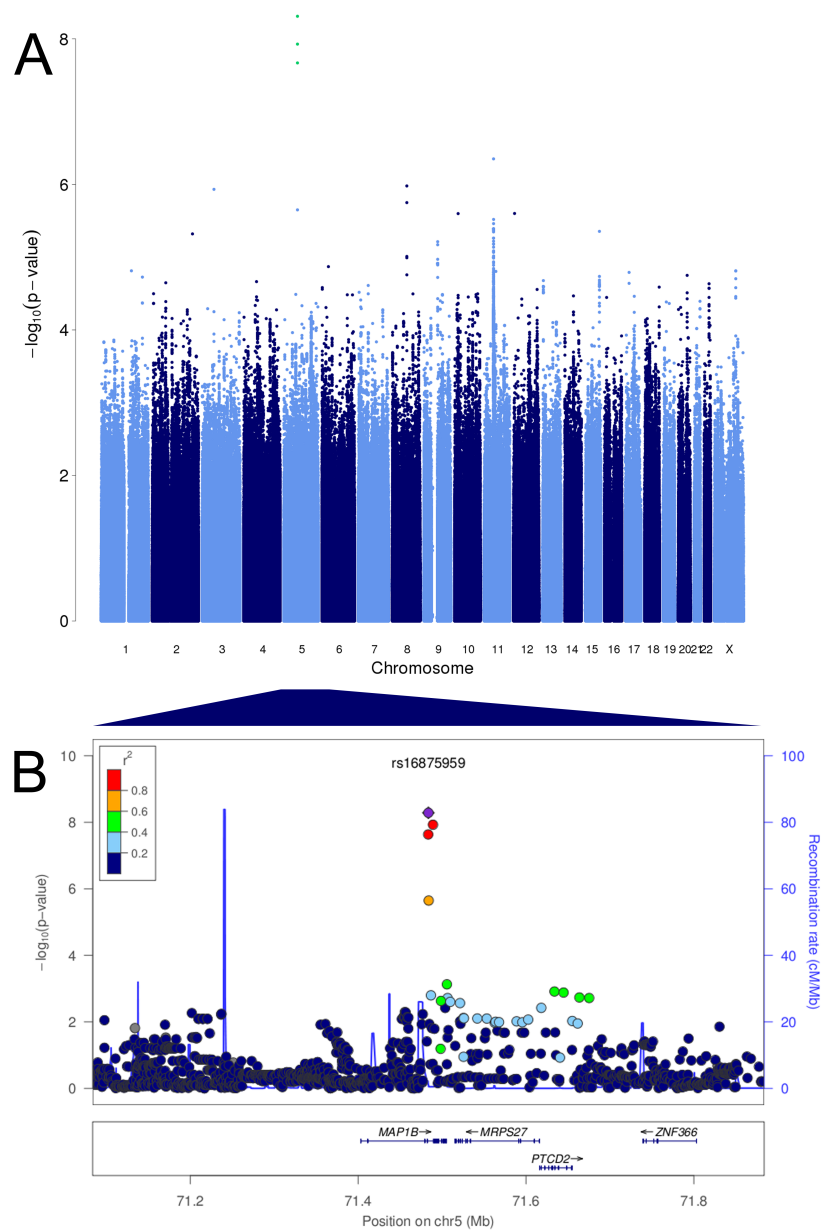


Figure 6.11: ***MAP1B* is associated with non-word reading in individuals with dyslexia.** **A)** Manhattan plots of the P values for the GWAS for non-word reading score in the combined dyslexia cohorts, the genome-wide significant SNPs are shown as ●. **B)** Visualisation of P values at the *MAP1B* locus. Negative $\log_{10}$ of the P values for all genotyped and imputed SNPs around *MAP1B* plotted in LocusZoom (Pruim et al., 2010) in the PegQ GWA meta-analysis. Linkage disequilibrium (r^2) with the most highly associated SNP, rs16875959 (◆), is calculated based on the November 2010 release of the 1000 Genomes European population and is shown by the colour of the SNPs. The recombination rate is shown by the line and the locations of genes in this locus are shown in the panel below the plot.

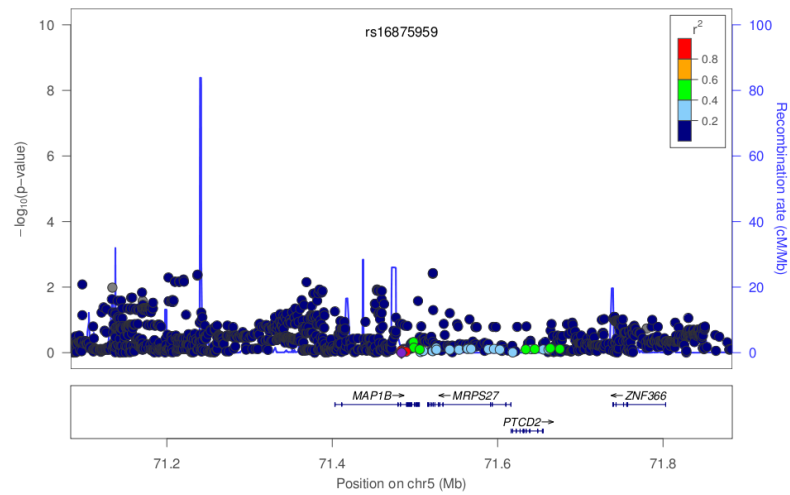


Figure 6.12: Conditioning on rs16875959 reveals there is only one variant driving association with non-word reading (CCN) at this locus.

Table 6.6: **A list of SNPs with $P \leq 5 \times 10^{-6}$ from the CCN GWA meta-analysis.** All the genes within 100kb to the left or right of the SNP are listed. ++ indicates that the effect allele is associated with an increased CCN score in both studies, and -- indicates the effect allele is associated with an decreased CCN score.

<i>Chr</i>	<i>Position</i>	<i>Variant</i>	<i>Effect Allele Frequency</i>	β	β 95L	β 95U	<i>SE</i>	<i>P value</i>	<i>Number of Samples</i>	<i>Effects</i>	<i>Gene</i>	<i>Left Genes</i>	<i>Right Genes</i>
5	71483741	rs16875959	0.11	0.46	0.31	0.61	0.078	4.89×10^{-9}	548	++	MAP1B	MIR4803	MRPS27
5	71489305	rs12516429	0.11	0.45	0.3	0.61	0.079	1.18×10^{-8}	548	++	MAP1B	MIR4803	MRPS27
5	71483451	rs59462818	0.11	0.43	0.28	0.58	0.077	2.14×10^{-8}	549	++	MAP1B	MIR4803	MRPS27
11	49690460	rs28665132	0.49	-0.26	-0.36	-0.16	0.051	4.46×10^{-7}	547	--	LOC440040	-	-
8	76028076	rs11991922	0.47	0.25	0.15	0.35	0.051	1.05×10^{-6}	547	++	-	CRISPLD1	-
3	61040359	rs60746884	0.085	0.44	0.26	0.62	0.09	1.17×10^{-6}	547	++	FHIT	-	-
8	76030974	rs1455797	0.47	0.24	0.14	0.34	0.051	1.78×10^{-6}	547	++	-	CRISPLD1	-
5	71484112	rs9885364	0.15	0.32	0.19	0.45	0.067	2.24×10^{-6}	548	++	MAP1B	MIR4803	MRPS27
12	5370301	rs975514	0.18	0.3	0.17	0.42	0.063	2.51×10^{-6}	549	++	-	-	-
10	18642730	rs11594922	0.18	-0.29	-0.41	-0.17	0.062	2.52×10^{-6}	549	--	CACNB2	-	CACNB2
11	49900218	rs1722012	0.57	-0.23	-0.33	-0.14	0.05	3.03×10^{-6}	547	--	-	LOC440040	OR4C13
11	49683158	rs11040462	0.56	-0.23	-0.32	-0.13	0.049	3.46×10^{-6}	548	--	LOC440040	-	-
11	49641518	rs11040445	0.54	-0.23	-0.33	-0.13	0.05	4.07×10^{-6}	547	--	LOC440040	-	-
11	49661351	rs7479914	0.54	-0.23	-0.33	-0.13	0.05	4.07×10^{-6}	547	--	LOC440040	-	-
11	49657228	rs10839317	0.56	-0.23	-0.32	-0.13	0.049	4.08×10^{-6}	547	--	LOC440040	-	-
11	49687202	rs10839326	0.54	-0.23	-0.33	-0.13	0.05	4.16×10^{-6}	547	--	LOC440040	-	-
11	49934876	rs9704447	0.49	-0.24	-0.34	-0.14	0.052	4.32×10^{-6}	547	--	-	-	OR4C13,OR4C12
11	49656382	rs9971508	0.54	-0.23	-0.33	-0.13	0.05	4.34×10^{-6}	547	--	LOC440040	-	-
15	95186945	chr15:95186945:1	0.061	0.48	0.28	0.69	0.1	4.42×10^{-6}	548	++	-	-	-
11	49690504	rs10769604	0.55	-0.23	-0.32	-0.13	0.049	4.5×10^{-6}	547	--	LOC440040	-	-
2	211151859	rs72994488	0.04	-0.62	-0.89	-0.35	0.14	4.79×10^{-6}	547	--	-	ACADL	MYL1

Table 6.7: Association of the most highly associated SNP with non-word reading (CCN), rs16875959, and the other quantitative phenotypes collected for cohorts 1 & 2.

<i>Phenotype</i>	β	β 95L	β 95U	<i>SE</i>	<i>P</i>
CCN	0.46	0.31	0.61	0.078	4.89×10^{-9}
READ	0.17	0.036	0.3	0.067	0.012
SPELL	0.12	0.003	0.25	0.062	0.045
CCI	0.26	0.098	0.43	0.085	0.0019
MAT IQ	0.042	-0.11	0.19	0.076	0.58
SIM IQ	0.26	0.10	0.42	0.08	0.0011
PegQ	0.094	-0.10	0.29	0.10	0.35

6.7.1 GWAS replication

I took forward two SNPs that were associated with non-word reading (CCN) for replication; rs16875959 in *MAP1B*, and rs11594922 in *CACNB2*. Although the latter is not genome-wide significant ($P = 2.52 \times 10^{-6}$), it was chosen because this gene has previously been associated with five neurodevelopmental disorders (Smoller et al., 2013). The two SNPs were tested in cohorts 3, 4, and 5 for non-word reading. In ALSPAC (cohorts 3 & 4), a non-word reading test consisting of ten items was applied at age of nine years old (e.g. douter, mishammer, loker). In cohort 5 (NeuroDYS Bonn, **Table 1.2**), a comparable non-word test was performed in German. Neither SNP was significant in any of the cohorts, and after meta-analysis, neither SNP is even nominally associated (at $P \leq 0.05$, **Table 6.8** & **Table 6.9**).

Power calculation

The failure to replicate signals may be due to a lack of power in the small replication cohorts. I tested the power for replication using *GWAPower* (Feng et al., 2011), assuming the phenotype measured was identical and that each SNP explains at

Table 6.8: The SNP rs16875959, located in *MAP1B* and associated with non-word reading, does not replicate in follow up cohorts.

<i>Cohort</i>	<i>n</i>	<i>P</i>	β	<i>SE</i>	<i>EAF</i>
1	196	3.79×10^{-5}	0.58	0.14	0.092
2	353	1.40×10^{-5}	0.40	0.093	0.12
1 & 2	549	4.89×10^{-9}	0.46	0.078	0.11
3	167	0.70	-0.03	0.16	0.12
5	199	0.91	0.026	0.22	0.12
1, 2, 3, & 5	915	0.76	0.044	0.15	0.11
general population (4)	2,366	0.95	-0.003	0.048	0.10

least 1% of the heritability of the trait. The power to detect association at $P = 0.05$ is 0.36 for cohort 3 ($n = 167$), and 0.40 for cohort 5 ($n = 199$). These are over-estimates because the phenotype tested was not identical to CCN, the SNPs are unlikely to explain such a large proportion of the heritability, and the SNP is not likely to be in complete linkage disequilibrium with the causative allele. However, making these assumptions, a sample size of at least 617 would be necessary to have 80% power to replicate the signal.

6.7.2 Principle component analysis

Since each reading-related phenotype is correlated (**Figure 6.13** & **Table 2.1**), an alternative to testing the phenotypes individually is to calculate the principle components of the four phenotypes and use the first principle component (PC) as a quantitative trait for GWAS. The data can be explained by four PCs, the first of which captures 65.4% of the variation in the phenotypes (**Table 6.10**). Furthermore, a scatterplot of the first two principle components shows there is no obvious subgrouping of the cohorts (**Figure 6.14**).

Table 6.9: The SNP rs11594922, located in *CACNB2* and associated with non-word reading, does not replicate in follow up cohorts.

<i>Cohort</i>	<i>n</i>	<i>P</i>	β	<i>SE</i>	<i>EAF</i>
1	196	0.0012	-0.35	0.11	0.19
2	353	0.00041	-0.26	0.075	0.17
1 & 2	549	2.52×10^{-6}	-0.29	0.062	0.18
3	167	0.11	-0.20	0.13	0.20
5	197	0.79	-0.05	0.12	0.20
1, 2, 3, & 5	913	0.39	-0.08	0.10	0.18
general population (4)	2,366	0.38	0.034	0.039	0.18

No SNPs were GW-significantly associated with the first PC of the reading-related traits (**Figure 6.15** & **Table 6.11**). The most strongly associated variant is rs10773664 ($P = 6.25 \times 10^{-7}$), located in *TMEM132D* on chromosome 12.

Table 6.10: Summary of principle component analysis (PCA) of reading-related traits. PC = principle component.

	<i>PC1</i>	<i>PC2</i>	<i>PC3</i>	<i>PC4</i>
Standard deviation	1.62	0.82	0.66	0.52
Proportion of variance explained	0.65	0.17	0.11	0.07
Cumulative proportion	0.65	0.82	0.93	1.0

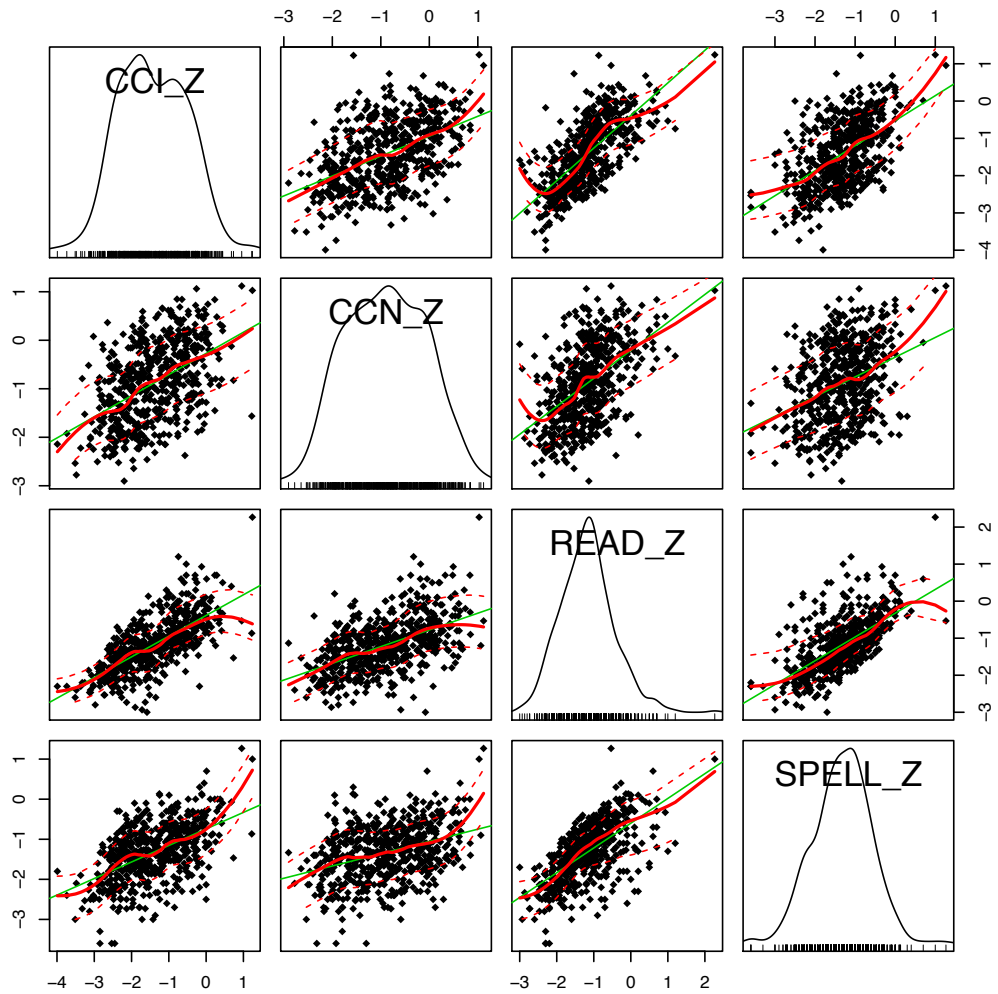


Figure 6.13: **Matrix scatter plot showing correlation between reading-related traits.** Histograms along the diagonal show the distribution of each trait. Each of the off-diagonal cells is a scatterplot of two of the four traits, for example, the second cell in the first row is a scatterplot of CCI (y-axis) against CCN (x-axis). The values for each trait are standardised against a population-level mean of 0 and standard deviation of 1.

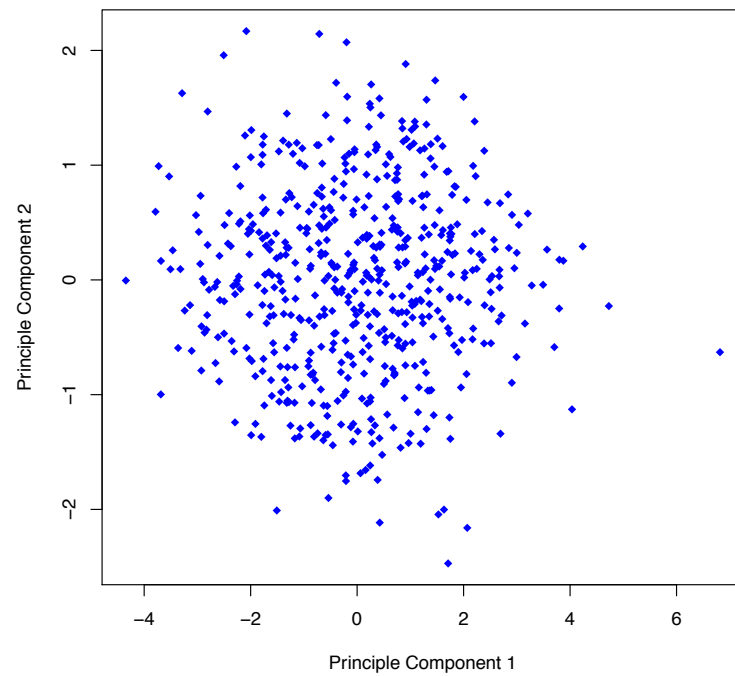


Figure 6.14: **Scatter plot of first two principle components of reading-related traits.**

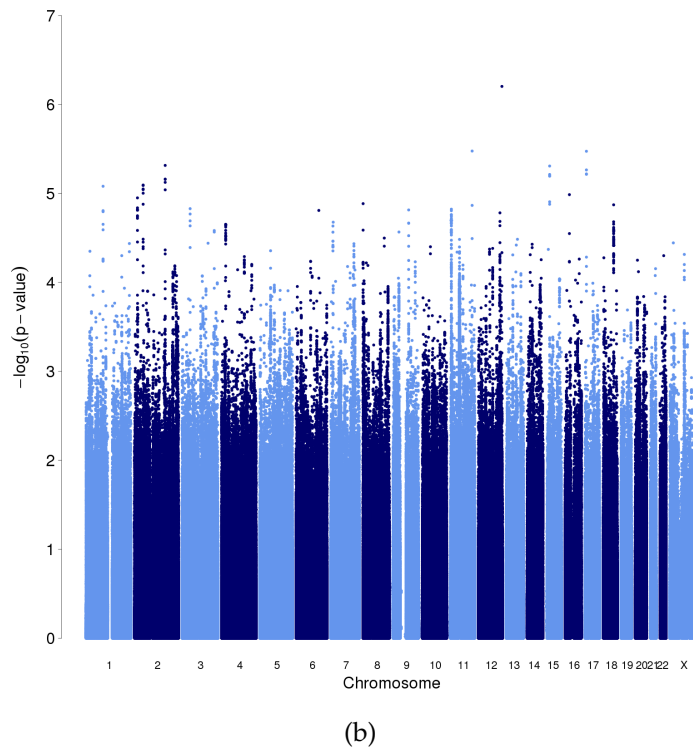
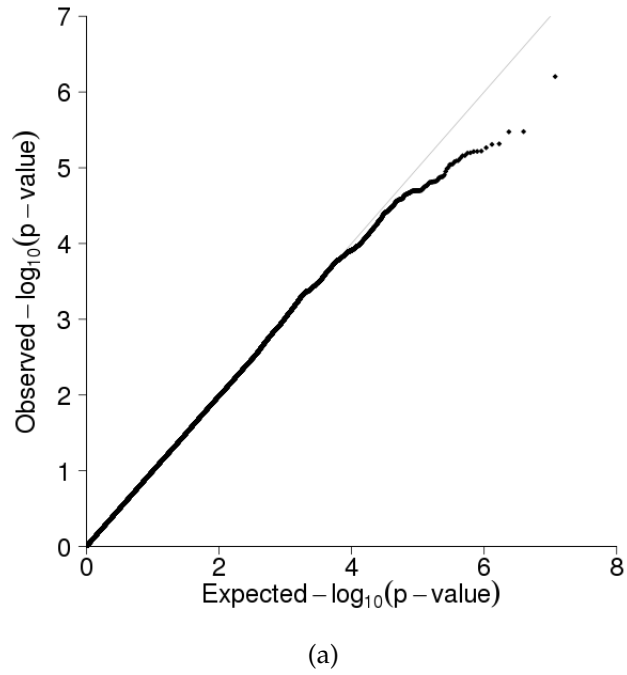


Figure 6.15: **QQ and Manhattan plots of the P values for the GWAS of the first principle component (PC) of reading-related traits.** (a) QQ plot, the P values are plotted against the expected P values, a deviation from the null (diagonal line) indicates an enrichment of genuine associations, population stratification or an error in genotype calling. (b) Manhattan plot, the negative $\log_{10}$ transformed P values are plotted against chromosome and position along the genome.

Table 6.11: A list of SNPs with $P \leq 1 \times 10^{-6}$ from the first PC of reading-related traits GWA meta-analysis. All the genes within 100kb to the left or right of the SNP are listed.

<i>Chr</i>	<i>Position</i>	<i>Variant</i>	<i>Effect Allele Frequency</i>	β	β 95L	β 95U	<i>SE</i>	<i>P value</i>	<i>Number of Samples</i>	<i>Effects</i>	<i>Gene</i>	<i>Left Genes</i>	<i>Right Genes</i>
12	129940481	rs10773664	0.45	0.48	0.29	0.66	0.095	6.25×10^{-07}	563	++	TMEM132D	-	-
11	120086918	rs499434	0.86	0.64	0.37	0.91	0.14	3.33×10^{-6}	561	++	OAF	TRIM29	POU2F3,LOC649133
17	7804405	rs62062524	0.16	-0.66	-0.93	-0.38	0.14	3.35×10^{-6}	561	-	CHD3	LSMD1,TMEM88,DNAH2,KDM6B,CYB5D1	SCARNA21,KCNAB3,+3 genes
2	169095090	rs10930316	0.29	-0.48	-0.68	-0.27	0.1	4.83×10^{-6}	562	-	STK39	-	-
15	37172064	rs8036656	0.36	-0.45	-0.64	-0.26	0.099	4.91×10^{-6}	563	-	LOC145845	C15orf41,CSNK1A1P1	MEIS2
17	7822732	rs11654506	0.16	-0.64	-0.91	-0.36	0.14	5.43×10^{-6}	561	-	-	LSMD1,TMEM88,DNAH2,+5 genes	KCNAB3,CNTROB,+2 genes
17	7812870	chr17:7812870:D	0.16	-0.63	-0.91	-0.36	0.14	6.02×10^{-6}	561	-	CHD3	LSMD1,TMEM88,DNAH2,+5 genes	KCNAB3,CNTROB,+3 genes
17	7823466	rs66593967	0.16	-0.63	-0.91	-0.36	0.14	6.08×10^{-6}	562	-	-	LSMD1,TMEM88,DNAH2,+5 genes	KCNAB3,+3 genes
15	37171230	rs7171195	0.36	-0.45	-0.64	-0.25	0.099	6.11×10^{-6}	561	-	LOC145845	C15orf41,CSNK1A1P1	MEIS2
15	37171459	rs11634677	0.36	-0.45	-0.64	-0.25	0.099	6.34×10^{-6}	562	-	LOC145845	C15orf41,CSNK1A1P1	MEIS2
15	37171699	rs8036711	0.36	-0.45	-0.64	-0.25	0.099	6.38×10^{-6}	561	-	LOC145845	C15orf41,CSNK1A1P1	MEIS2
2	169094179	rs72884536	0.28	-0.47	-0.68	-0.27	0.1	6.89×10^{-6}	562	-	STK39	-	-
2	169096848	rs1517324	0.29	-0.47	-0.67	-0.26	0.1	6.92×10^{-6}	562	-	STK39	-	-
2	169091327	chr2:169091327:D	0.29	-0.46	-0.66	-0.26	0.1	7.49×10^{-6}	562	-	STK39	-	-
2	47271430	rs11125113	0.25	0.5	0.28	0.73	0.11	8.03×10^{-6}	561	++	TTC7A	-	C2orf61
2	47261639	rs6717063	0.23	0.51	0.29	0.73	0.11	8.12×10^{-6}	563	++	TTC7A	MCFD2	C2orf61
1	95057013	rs1408341	0.4	0.43	0.24	0.62	0.096	8.29×10^{-6}	562	++	-	ABCD3,F3	-
2	47263232	rs4953447	0.23	0.51	0.28	0.73	0.11	8.91×10^{-6}	562	++	TTC7A	MCFD2	C2orf61
2	47264332	chr2:47264332:D	0.21	0.53	0.3	0.76	0.12	9.04×10^{-6}	561	++	TTC7A	MCFD2	C2orf61
2	169095296	rs1400635	0.29	-0.46	-0.66	-0.26	0.1	9.09×10^{-6}	561	-	STK39	-	-
2	47266784	rs4953448	0.25	0.5	0.28	0.72	0.11	9.92×10^{-6}	561	++	TTC7A	MCFD2	C2orf61

6.7.3 Multivariate analysis

An alternative to PCA is to combine the four quantitative reading-related traits in a Bayesian linear model using SNPtest. This test outputs a Bayes Factor (BF), which can be combined in an additive fashion to meta-analyse multiple cohorts (Jonathan Marchini, *personal communication*). For a SNP to be declared associated, a combined BF of $\geq 1,000,000$ is necessary. In the combined meta-analysis of cohorts 1 & 2, 155 SNPs showed a BF $\geq 1,000$, which is suggestive of association. Although suggestive, none of the BF are strong enough to declare a novel association (see **Table 6.12** for the 50 most strongly associated variants).

Table 6.12: **The 50 most strongly associated SNPs in the reading-related traits multivariate GWAS analysis.** BF = Bayes Factor, a score of $\geq 1,000,000$ is necessary to declare a new association.

<i>Chr</i>	<i>Variant</i>	<i>Position</i>	<i>MAF</i>	<i>log₁₀ BF cohort 1</i>	<i>log₁₀ BF cohort 2</i>	<i>Combined BF</i>	<i>Gene</i>	<i>Left Genes</i>	<i>Right Genes</i>
18_i	rs12605760	57194036	0.36	0.035	4.4	29,836	CCBE1	-	-
18_i	rs8095438	57193739	0.37	0.12	3.9	10,500	CCBE1	-	-
18_i	rs78972038	57192796	0.37	0.094	3.9	9,902	CCBE1	-	-
18_i	rs1944994	57192231	0.37	0.08	3.8	7,848	CCBE1	-	-
20	rs6038007	4855759	0.34	0.31	3.5	5,724	SLC23A2	RASSF2	-
20	rs1110277	4854682	0.34	0.31	3.4	5,192	SLC23A2	RASSF2	-
18_i	chr18:57206429:D	57206429	0.24	-0.025	3.7	5,149	CCBE1	-	-
11_i	rs9704447	49934876	0.48	0.031	3.7	4,929	-	-	OR4C13,OR4C12
1	rs2012813	237834701	0.41	-0.065	3.7	4,567	RYR2	-	-
17_i	rs11079803	46019901	0.46	-0.21	3.8	3,954	PNPO	SP6,SP2	COPZ2,PRR15L,MIR152,CDK5RAP3
20_i	rs8115867	4855678	0.32	-0.0053	3.6	3,777	SLC23A2	RASSF2	-
20	rs8122118	4855880	0.32	-0.031	3.6	3,656	SLC23A2	RASSF2	-
2_i	rs1579863	47071213	0.32	-0.089	3.6	3,383	LOC100134259	SOCS5,LOC388948	MCFD2,TTC7A
17_i	rs8066893	46009812	0.43	-0.034	3.5	2,798	-	LRR46,SCRN2,SP6,SP2	COPZ2,PRR15L,PNPO,CDK5RAP3
18_i	rs4940469	57205121	0.28	0.12	3.3	2,795	CCBE1	-	-
11_i	rs10839474	50028156	0.43	-0.2	3.6	2,737	-	OR4C13,OR4C12	-
11_i	rs1722022	49908279	0.48	-0.25	3.7	2,656	-	LOC440040	OR4C13,OR4C12
18_i	rs990507	57205384	0.28	0.13	3.3	2,653	CCBE1	-	-
18_i	chr18:57198279:D	57198279	0.28	0.13	3.3	2,592	CCBE1	-	-
18_i	rs12960222	57204688	0.28	0.13	3.3	2,589	CCBE1	-	-
18_i	rs55718166	57204636	0.28	0.13	3.3	2,588	CCBE1	-	-
18_i	rs4940468	57203025	0.28	0.13	3.3	2,588	CCBE1	-	-
18_i	rs4940467	57201802	0.28	0.13	3.3	2,588	CCBE1	-	-
18_i	rs4940466	57201534	0.28	0.13	3.3	2,588	CCBE1	-	-
18_i	rs1944992	57198342	0.28	0.13	3.3	2,588	CCBE1	-	-
18_i	rs1944991	57199949	0.28	0.13	3.3	2,588	CCBE1	-	-
18_i	rs1944989	57202242	0.28	0.13	3.3	2,588	CCBE1	-	-
18	rs4566280	57197873	0.28	0.13	3.3	2,588	CCBE1	-	-
6_i	rs6938154	135582073	0.35	0.39	3	2,582	-	MYB	AHI1
1_i	rs57749080	237833352	0.38	-0.14	3.5	2,560	RYR2	-	-
18_i	chr18:57206431:D	57206431	0.25	-0.0053	3.4	2,538	CCBE1	-	-
18_i	chr18:57206787:D	57206787	0.28	0.13	3.3	2,510	CCBE1	-	-
18_i	rs9957878	57207953	0.28	0.13	3.3	2,481	CCBE1	-	-

18_i	rs12970757	57206932	0.28	0.11	3.3	2,370	CCBE1	-	-
18_i	rs12970223	57206937	0.28	0.11	3.3	2,370	CCBE1	-	-
6_i	rs9494191	135579115	0.35	0.38	3	2,262	-	MYB	AHI1
18_i	rs28525136	57209265	0.28	0.13	3.2	2,259	CCBE1	-	-
18_i	chr18:57206416:D	57206416	0.24	0.039	3.3	2,162	CCBE1	-	-
18_i	rs35157814	57209962	0.28	0.13	3.2	2,151	CCBE1	-	-
11_i	rs9787774	49932670	0.5	-0.14	3.5	2,147	-	-	OR4C13,OR4C12
18_i	rs6567099	57210447	0.28	0.13	3.2	2,103	CCBE1	-	-
11_i	rs4128561	128415129	0.43	3.4	-0.097	2,083	ETS1	ETS1	-
18_i	rs6567100	57210748	0.28	0.13	3.2	2,049	CCBE1	-	-
18	rs1944993	57197019	0.28	0.022	3.3	2,032	CCBE1	-	-
11_i	rs1721990	49912154	0.48	-0.16	3.5	2,023	-	LOC440040	OR4C13,OR4C12
4_i	rs10029253	65307826	0.48	0.13	3.2	1,992	-	TECRL	-
18_i	rs4940892	57211277	0.27	0.13	3.2	1,946	CCBE1	-	-
18_i	chr18:57201340:I	57201340	0.28	0.12	3.2	1,933	CCBE1	-	-
18_i	rs12970742	57206888	0.3	-0.018	3.3	1,836	CCBE1	-	-
18_i	rs35443920	57196526	0.28	-0.015	3.3	1,791	CCBE1	-	-

6.7.4 GSEA

I used gene-set enrichment analysis (as in **subsection 5.4.4**), to test for an enrichment of SNPs associated with the four reading-related traits (**section 2.1**) located in genes that cause neuronal migration / axonogenesis, cilia morphology, or brain midline phenotypes when disrupted in mice.

The phenotype abnormal ciliary body morphology (MP:0005099) is associated with the irregular word test (CCI, **Table 6.13**), and abnormal axon guidance (MP:0002961) is associated with SPELL (**Table 6.14**). However, considering that four reading-related traits were tested for three different sets of mouse phenotypes, there is no evidence to suggest that there is an enrichment of SNPs with low P values located in genes that cause these phenotypes when disrupted in mice.

Table 6.13: In the irregular word (CCI) GWA meta-analysis of individuals with reading disability, there is an over-representation of highly associated SNPs within human orthologues of genes involved in causing abnormal ciliary body morphology when knocked out in mice. Three phenotypes that contain between 10 and 100 genes, and cause cilia defects when knocked out in mice, were tested using MAGENTA (Segre et al., 2010). The observed number of genes were compared to the expected using a cut-off that equals the 75th percentile of all gene *P* values, which were then adjusted for multiple testing using an FDR correction.

<i>Mouse Phenotype ID</i>	<i>Phenotype</i>	<i>Genes</i>	<i>Median Size / KB</i>	<i>Mean Size</i>	<i>Expected</i>	<i>Observed</i>	<i>Adjusted P</i>
MP:0004131	abnormal embryonic ciliary morphology	25	45	69	6	6	0.56
MP:0005099	abnormal ciliary morphology	17	33	76	4	9	0.026
MP:0004132	absent embryonic cilia	12	51	60	3	4	0.44

Table 6.14: In the single word spelling GWA meta-analysis of individuals with reading disability, there is an over-representation of highly associated SNPs within human orthologues of genes involved in causing abnormal axon guidance when knocked out in mice. Four phenotypes that contain between 10 and 200 genes, and cause axonogenesis or neuronal migration defects when knocked out in mice, were tested using MAGENTA (Segre et al., 2010). The observed number of genes were compared to the expected using a cut-off that equals the 75th percentile of all gene *P* values, which were then adjusted for multiple testing using an FDR correction.

<i>Mouse Phenotype ID</i>	<i>Phenotype</i>	<i>Genes</i>	<i>Median Size / KB</i>	<i>Mean Size</i>	<i>Expected</i>	<i>Observed</i>	<i>Adjusted P</i>
MP:0005404	abnormal axon morphology	159	36	71	40	50	0.068
MP:0003651	abnormal axon outgrowth	61	59	131	15	17	0.42
MP:0002961	abnormal axon guidance	60	45	157	15	23	0.046
MP:0006009	abnormal neuronal migration	112	57	133	28	27	0.58

6.8 Discussion

I have performed both a GWAS case / control meta-analysis for dyslexia, a quantitative reading-related trait meta-analysis, a multivariate GWAS analysis, and GSEA of the GWAS data.

6.8.1 Case / control analysis

I found no genome-wide significant variants in the GWAS case/control analysis for dyslexia in cohorts 1 & 2. However, in cohort 2 the most strongly associated variant was in *RAB12* (Ras oncogene family member 12, $P = 6.18 \times 10^{-7}$), which lies on chromosome 18p11.22. This is within the locus DYX6 (18p11.2, an 8.3Mb region of chromosome 18) that has previously been associated with dyslexia using a quantitative-trait locus (QTL) based genome-wide linkage screen (Fisher et al., 2002). A finer-scale linkage map of chromosome 18 found linkage across 18p11.2–18q12.2 (47Mb), replicating the previous study, although broadening out the linkage peak (Scerri et al., 2010). Association analysis of 3000 SNPs across this region identified *MC5R*, *DYM*, and *NEDD4L* as potential candidate genes (Scerri et al., 2010). However, seven further studies have failed to replicate chromosome 18 linkage (reviewed in (Scerri and Schulte-Korne, 2010)). Our samples partially overlap with those used for both these studies, and aside from the association in *RAB12*, the SNP rs1426310 on 18q12.1 lies under this linkage peak ($P = 5.95 \times 10^{-6}$). This is a missense variant in the final exon of the gene *DSG1*, which is involved in desmosome adhesion in the epithelial cells. Mutations in *DSG1* cause Pemphigus foliaceus, an autoimmune blistering disease of the skin (Ayed et al., 2002). Neither *RAB12* nor *DSG1* are promising candidate for dyslexia. Furthermore, the linkage

region is broad, containing at least 159 genes, so finding association in this locus by chance has a reasonable likelihood (<http://genome.ucsc.edu/>). The most strongly associated variant in the case/control meta-analysis (cohorts 1 & 2) is 54.8kb upstream of the single exon gene *FOXC1* on chromosome 6. Mutations in *FOXC1* cause both Axenfeld-Rieger syndrome (characterised by eye abnormalities that can lead to glaucoma) and an absent cerebellar vermis (Dandy-Walker malformation, (Aldinger et al., 2009)).

6.8.2 Nonword reading GWAS

In the nonword reading (CCN) GWAS, the genome-wide significant locus is located in *MAP1B* (rs16875959, $P = 4.89 \times 10^{-9}$), another gene involved in brain mid-line development (Meixner et al., 2000). More specifically, homozygous null mice for *Map1b* display an absent corpus callosum, defects in axon guidance, and a reduced number of large myelinated axons (Meixner et al., 2000). Knockdown of *Map1b* in rat embryonic cortical neurons impairs axonogenesis (Tymanskyj et al., 2012) and overexpression of *Map1b* induces both neurite extension and neuronal death in rats (Gomi and Uchida, 2012). Furthermore, the four dyslexia candidate genes; *DCDC2*, *DYX1C1*, *KIAA0319*, and *ROBO1* are, like *MAP1B*, involved in neuronal migration or axon guidance (Galaburda et al., 2006).

CCN is a particularly useful measure for interrogating the phonological deficit in dyslexia because it tests the ability to read non-words (e.g. dorkit, tegwop, **section 2.1**, (Paracchini, 2009)), which may explain why this phenotype had significant results initially. Since individuals taking the test have never come across these words, it accurately assesses their awareness of the sound structure of spo-

ken words (i.e. phonological decoding).

However, the genome-wide significant variant associated with non-word reading did not replicate in three independent cohorts. The lack of replication of *MAP1B* in the other cohorts suggests the finding is either; a false positive, due to underpowered replication cohorts (**subsection section 6.7.1**), or to heterogeneity in ascertainment between the cohorts. In support of the latter hypothesis, in cohorts 1 & 2 the set of 30 non-words tested are quite dissimilar to real words (e.g. torlep, dorkit). However, only ten non-words were tested in ALSPAC, and these are more similar to real words, often adding or changing a prefix or suffix to a real word (e.g. mishammer, amazive). The non-word reading deficit in dyslexia is particularly marked when the non-words are constructed to be dissimilar to real words (Snowling, 1980), suggesting that, coupled with fewer items tested, the measure in ALSPAC is a less accurate measure of reading disability. In cohort 5 the words were different too, this time in another language because the participants German, although the deficit in non-word reading in German dyslexics is similar to that in English dyslexics (Ziegler et al., 2003). The variant associated with *CACNB2* shows the same trend in the replication cohorts that have reading disability (**Table 6.9**), but the reading disabled cohorts are underpowered for the trend to be significant (**subsection section 6.7.1**). I estimated that a sample size of at least 617 ascertained identically to cohorts 1 & 2, will be necessary to have 80% power to replicate the signal. The effect size of GWAS signals are typically overestimated due to 'winner's curse' (McCarthy et al., 2008; Zollner and Pritchard, 2007). For reading ability there are likely to be hundreds of variants associated with the trait (see **subsection 3.2.1**). The one signal that does show up as genome-wide significant in our study does so because the effect size is overestimated. Therefore much larger cohorts are necessary to replicate the initial signal, however we are limited by the

small replication cohorts we have access to. The exception is the ALSPAC general population cohort 4, so assuming the nonword phenotype in ALSPAC is a good proxy for CCN, the lack of replication suggests that the associations, if real, are specific only to individuals who already are susceptible to reading difficulties. Finally the CCN measure shows a slightly platykurtic distribution, i.e. there is a wide spread of values around the mean (**Table 6.5 & Figure 6.8**). This may inflate effect sizes further because the association test assumes that the distribution is normal, however the QQ plot and lambda ($\lambda = 1.02$) suggest there is not a systematic inflation of P values (**Figure 6.9**).

6.8.3 Principle component analysis

The PCA GWAS tested the first PC of the reading-related traits, which captures 65.4% of the variation in the four phenotypes (**Table 6.10**). The most strongly associated variant is rs10773664 ($P = 6.25 \times 10^{-7}$), located in *TMEM132D* on chromosome 12 (**Table 6.11**).

A two-SNP haplotype in *TMEM132D* (rs11060369-rs7309727) has previously been associated with panic disorder (Erhardt et al., 2011), and both the haplotype and individual association with rs7309727 were replicated in a follow up study (Erhardt et al., 2012). The variant rs11060369 is not in linkage disequilibrium with rs10773664 ($r^2 = 0.096$) and is not associated with the first PC in our sample ($P = 0.32$). However, rs7309727 is in weak LD with rs10773664 ($r^2 = 0.266$) and is associated with the first PC of the reading related traits ($P = 0.0036$). Variants in *TMEM132D* may contribute to risk for multiple neurodevelopmental disorders, as has been observed by the Cross-Disorder group of the Psychiatric Genomics

Consortium, who found variants that contribute to risk of five neurodevelopmental disorders in a GWAS (autism, ADHD, bipolar, depression, and schizophrenia (Smoller et al., 2013)).

6.8.4 Multivariate GWAS analysis

The multivariate GWAS analysis combined information from all four phenotypes into one test and did not find any associated variants with a Bayes Factor great enough to suggest significance. The most strongly associated variant is located in the 266kb gene *CCBE1* on chromosome 18q21.32 (outside of the linkage peak). Mutations in *CCBE1* are associated with generalized lymph vessel dysplasia, a condition characterised by fluid build up in tissues that also causes mental retardation (Alders et al., 2009), which is not a promising candidate gene for dyslexia.

6.8.5 GSEA

Finally, the gene-set enrichment analysis provided no support for theories that neuronal migration (Galaburda et al., 2006), ciliogenesis (Chandrasekar et al., 2013), or brain midline development (this thesis) play a role in the etiology of dyslexia. However, the GWAS only assessed reading-related traits specifically in individuals with dyslexia (who are at the tail of the distribution for reading), and was underpowered for GWAS of a common trait ($n = 549$).

On balance, this study has not strongly implicated any genes or biological pathways in dyslexia. Although dyslexia is a highly heritable trait, it is also very common and has no apparent effect on fertility. It therefore is likely to fall un-

der the common disease common variant (CD/CV) paradigm (**subsection 3.2.1**, (Chakravarti, 1999; Lander, 1996)), suggesting that many variants with a high frequency in the population ($\geq 5\%$), each of small effect size, contribute to susceptibility for developing the disorder. Much larger sample sizes that are well ascertained and phenotyped will be necessary in order to understand the genetic etiology of the dyslexia.

Chapter 7

CNV calling and quality control

7.1 Introduction

Copy number variations (CNVs) derive from duplications, triplications, insertions, deletions, or translocations of segments of DNA ranging in size from 1 kilobase (kb) to many megabases (Mb) that are not identifiable under a light microscope by conventional chromosomal banding (Stankiewicz and Lupski, 2010). The mechanisms by which CNVs are generated include; non-allelic homologous recombination, non-homologous end joining, and fork stalling and template switching (Lupski and Stankiewicz, 2005; Stankiewicz and Lupski, 2010). Copy number variable regions (CNVRs) encompass at least 12% of the genome (Redon et al., 2006), and overlap 13.5% of genes, directly altering the coding sequence of at least 5.5% of genes (Conrad et al., 2010).

Previously, CNVs were detected using locus based approaches such as fluorescent in-situ hybridisation (FISH), a time-consuming, targeted method that requires prior knowledge of the chromosomal region of interest (Shinawi and Cheung,

2008). However, thanks to chromosome microarray technologies such as array comparative genomic hybridisation (array CGH) and SNP microarrays, the whole genome can now be interrogated at once, facilitating the discovery of disease-associated CNVs. Array CGH involves cohybridising differentially labelled genomic DNA of both a patient (Cyanine 3, Cy3) and control (Cyanine 5, Cy5) to an array containing DNA targets. The interrogating probes (targets) are pieces of human genomic DNA in the form of bacterial artificial chromosomes (BACs) clones (size of 75-200 kb), smaller insert clones such as cosmids or fosmids (30-50 kb), or oligonucleotides (25-85 mers). If there is an altered Cy3: Cy5 fluorescence ratio this indicates a loss or a gain of the patient DNA at that specific genomic region (Shinawi and Cheung, 2008). SNP arrays, on the other hand, do not compare patients to controls. Instead, CNVs are detected by comparing the total fluorescence of each probe to the rest of probes in the genome (log R ratio), and the intensity of fluorescence for each allele (B allele frequency, **Figure 7.1**). They offer higher resolution for detecting CNVs from individual patients, because each probe is a single nucleotide, and SNP arrays can have hundreds of thousands, or millions of probes. Genomic imbalances can subsequently be validated with other cytogenetic and molecular methods that include FISH and quantitative PCR.

Here I lay out the methodology used for copy number variation (CNV) calling and quality control, justifying the methods and thresholds employed when using CNV calling algorithms. For a detailed discussion of CNVs, their relationship to neurodevelopmental disorders such as dyslexia, burden & pathway analysis, see **chapter 8**.

7.2 Methods

7.2.1 Cohorts and genotyping

I had access to CNV data from four cohorts with reading disability; the Oxford dyslexia cohorts genotyped on the Illumina 550K and OmniExpress arrays plus those from Bonn and Cardiff genotyped on the Illumina 300K as part of the NeuroDYS project (cohorts 1, 2, 5, & 6, **Table 1.2**, **Table 7.1**). I also obtained access to control cohorts from two sources; the Wellcome Trust Case Control Consortium (WTCCC) 1958 Birth Cohort genotyped on the Illumina 550K array (The Wellcome Trust Case-Control Consortium, 2007), plus the People of the British Isles (POBI) study genotyped on the Illumina OmniExpress array (**chapter 4**, (Winney et al., 2012)). All cases and controls were genotyped by the Bioinformatics core at the Wellcome Trust Centre for Human Genetics in Oxford.

Table 7.1: **Summary of cohorts used for CNV analysis.**

<i>Cohort</i>	<i>Name</i>	<i>Case/Control</i>	<i>n</i>	<i>Genotyping Array</i>	<i>Reading Scores</i>
1	NeuroDYS (Oxford)	case	197	Illumina 550K	✓
2	Oxford Dyslexia	case	374	Illumina OmniExpress	✓
5	NeuroDYS (Bonn)	case	200	Illumina 300K	✓
6	NeuroDYS (Cardiff)	case	208	Illumina 300K	✗
WTCCC	Wellcome Trust Case Control Consortium	control	1,167	Illumina 550K	✗
POBI	People of the British Isles	control	336	Illumina OmniExpress	✗

First, SNPs were mapped to the hg19 assembly of the human genome using the batch coordinate conversion program *liftOver* (<http://genome.ucsc.edu/cgi-bin/hgLiftOver>) (Hinrichs et al., 2006). Copy Number Variants (CNVs) were called using two algorithms, PennCNV (adjusting for genomic waves (Diskin et al., 2008; Wang et al., 2007)) and QuantiSNP (Colella et al., 2007).

PennCNV and QuantiSNP both work by combining the following information in a hidden Markov model to call CNVs:

- Log R ratio – a measure of the magnitude of the combined fluorescent intensity signals from both sets of probes, which will be higher in duplications and lower in deletions relative to the expected copy number state of 2.
- B allele frequency – the relative ratio of the fluorescent signals from one allelic probe to the other.

Deletions are called by a shift downwards in the log R ratio, and loss of heterozygosity indicated by the disappearance of the heterozygous state (B allele frequency = 0.5). In duplications, the log R ratio is shifted upwards, and since the copy number is 3, there are four possible genotypes for each SNP (with alleles A & B); AAA, AAB, ABB, BBB, and correspondingly four B allele frequencies; 0, 1/3, 2/3, & 1 (**Figure 7.1**).

Individuals were removed if they had a Log R Ratio standard deviation ≥ 0.3 , drifting B allele frequency (BAF) ≥ 0.01 , and a waviness factor $\geq \pm 0.05$. These are the standard parameters for the PennCNV program. Further outliers were removed if they had an extreme number of CNV calls ($\geq 3\text{rd Quartile} + 3 \times \text{inter-quartile range}$). Individuals were also removed if the aggregate length of CNV calls was more than 5Mb, suggestive of karyotypic abnormalities. For example one control was identified with trisomy of chromosome 21 (Down's syndrome) and was subsequently removed. After these quality control steps, 920 cases, and 1,366 controls were taken forward for CNV analysis (**Table 7.2**).

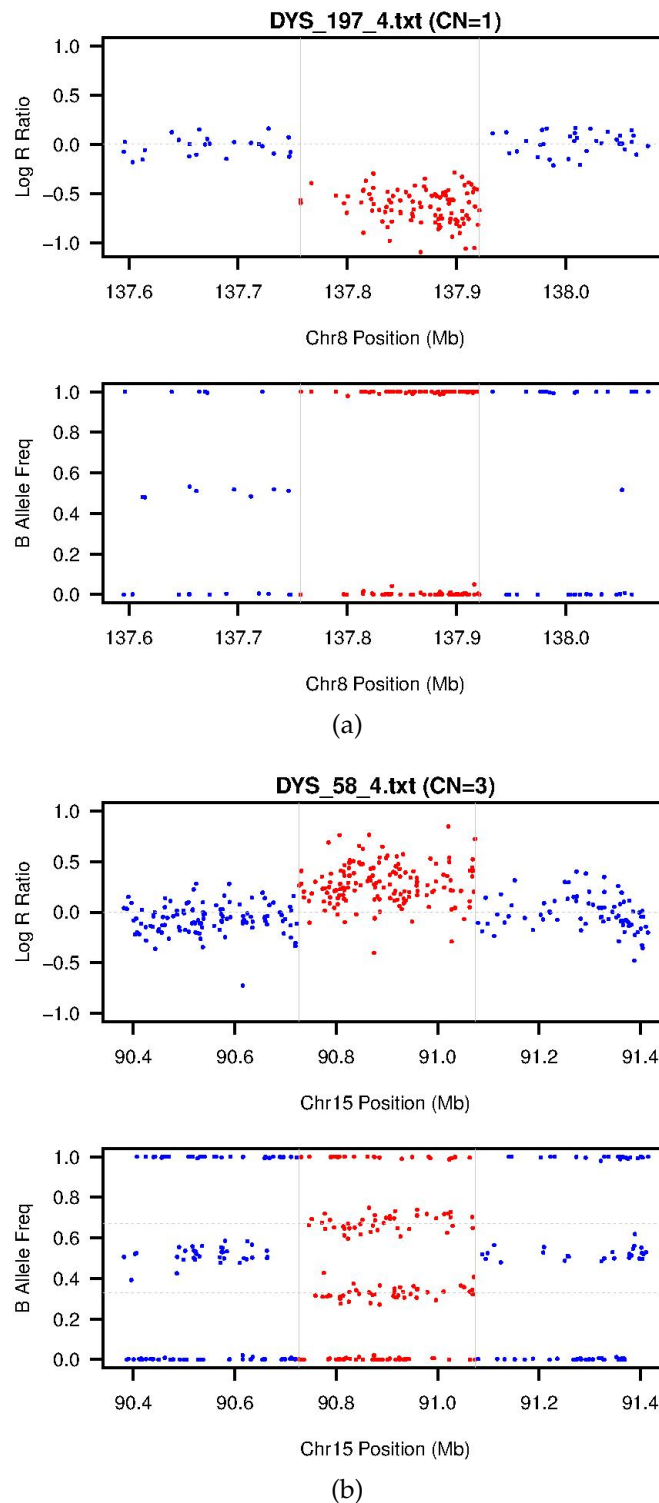


Figure 7.1: **Visualisation of high confidence CNV calls made by PennCNV.** SNPs that are called within the boundaries of CNVs are shown in **red**. (a) In **deletion** CNVs the log R ratio is shifted down and the B allele frequency is either 0 or 1. (b) In **duplication** CNVs the log R ratio is shifted up and the B allele frequency is either 0, 1/3, 2/3, or 1.

After the individual QC, I determined the quality control thresholds for each CNV called, by comparing the overlap of CNV calls for individuals genotyped on more than one SNP array (**section 7.3**), and I also compared the overlap of CNV calls between the two algorithms on each SNP array (**section 7.4**).

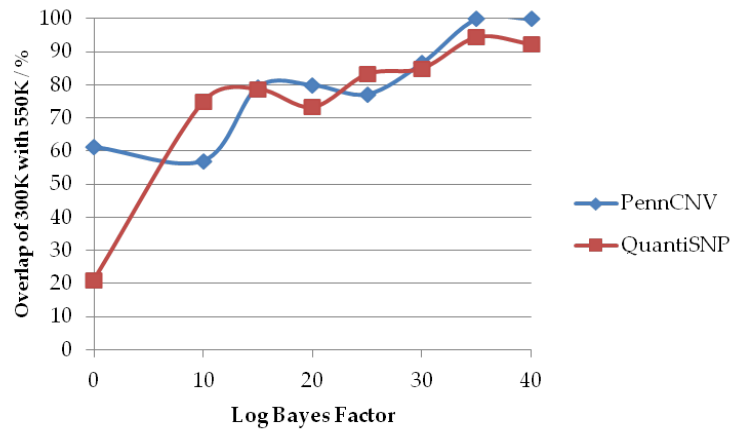
Table 7.2: **Descriptive statistics for individual CNV quality control.** Aside from the SNP genotyping QC detailed in **section 4.2**, SNPs were removed if the program *liftOver* failed to convert them to build hg19 of the human genome. Individuals were removed if they had a Log R Ratio Standard deviation ≥ 0.3 , drifting B allele frequency (BAF) ≥ 0.01 , and a waviness factor $\geq \pm 0.05$.

<i>Cohort</i>	<i>SNP array</i>	<i>Before QC</i>		<i>After QC</i>	
		<i>SNPs</i>	<i>n</i>	<i>SNPs</i>	<i>n</i>
1	550K	555,352	200	501,063	184
2	OmniExpress	730,525	446	643,227	384
5	300K	317,503	200	299,479	199
6	300K	317,503	208	290,558	153
WTCCC	550K	555,352	1,167	501,063	1,064
POBI	OmniExpress	730,525	336	643,227	302

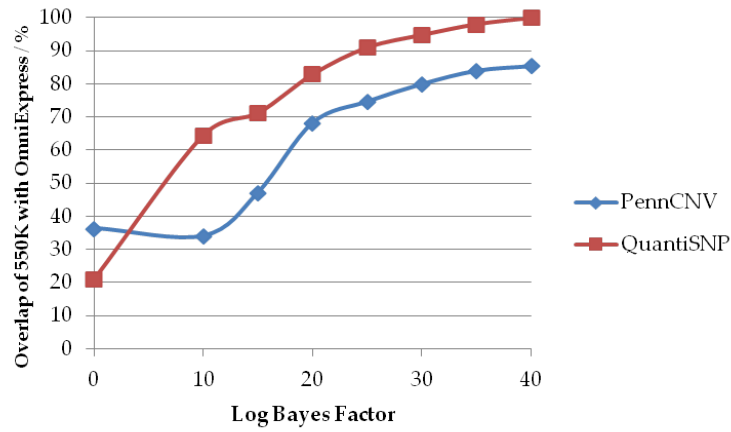
7.3 Comparison of individuals genotyped on more than one SNP array

I compared the overlap of CNV calls by PennCNV and QuantiSNP for individuals genotyped on more than one SNP array (Illumina 300K and Illumina 550K, $n = 11$; Illumina 550K and OmniExpress, $n = 27$). Since the OmniExpress SNP array has $\sim 180,000$ more SNPs than the 550K, and the 550K SNP array has $\sim 250,000$ more SNPs than the 300K array, more CNVs should be called with higher confidence

in the more densely typed arrays. In order to assess the degree of overlap and the concordance, I compared the CNV calls at a number of different confidence thresholds (log bayes factor, LBF). First I assessed the proportion of CNVs from the less densely genotyped array that were also called by the other array for each calling algorithm at a range of confidence thresholds (**Figure 7.2**). The proportion increased with confidence score and started to tail off around $\text{LBF} = 30$. Since the number of CNVs called decreases exponentially with confidence score (**Figure 7.7b**), the choice of threshold is therefore a trade off between the number of CNVs called, and the confidence that they are true positives. The overlaps at a confidence score of $\text{LBF} = 30$ are visualised in Venn diagrams, showing significantly more CNVs are called on the more densely typed arrays (**Figure 7.3 & Figure 7.4**). For the CNVs that did overlap, the size of the CNVs were compared to assess how well the algorithms were defining CNV boundaries on the SNP arrays (**Figure 7.5 & Figure 7.6**). There is no apparent bias in the size of CNVs called on more than one SNP array, suggesting that differences are not due to the density of SNPs on the arrays.



(a)



(b)

Figure 7.2: Comparison of CNV calling in individuals genotyped on more than one array. I assessed the proportion of CNVs from the less densely genotyped array that were also called by the other array for both **PennCNV** and **QuantiSNP** at a range of confidence thresholds. (a) 300K and 550K ($n = 11$). (b) OmniExpress and 550K ($n = 27$).

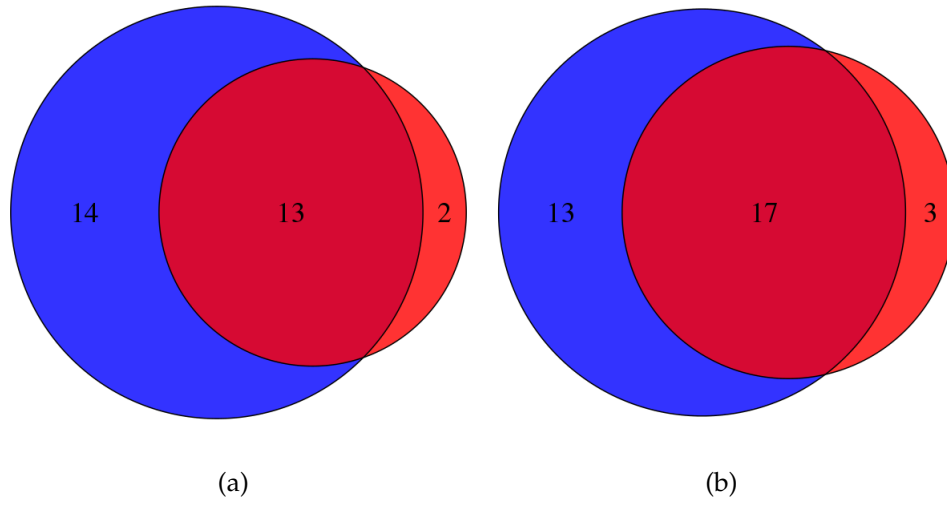


Figure 7.3: Venn diagram of CNV calls in individuals genotyped on the 550K and 300K SNP arrays ($n = 11$), CNV calling thresholds were ≥ 3 SNPs / CNV, CNV size ≥ 1 kb, and LBF ≥ 30 . (a) PennCNV and (b) QuantiSNP.

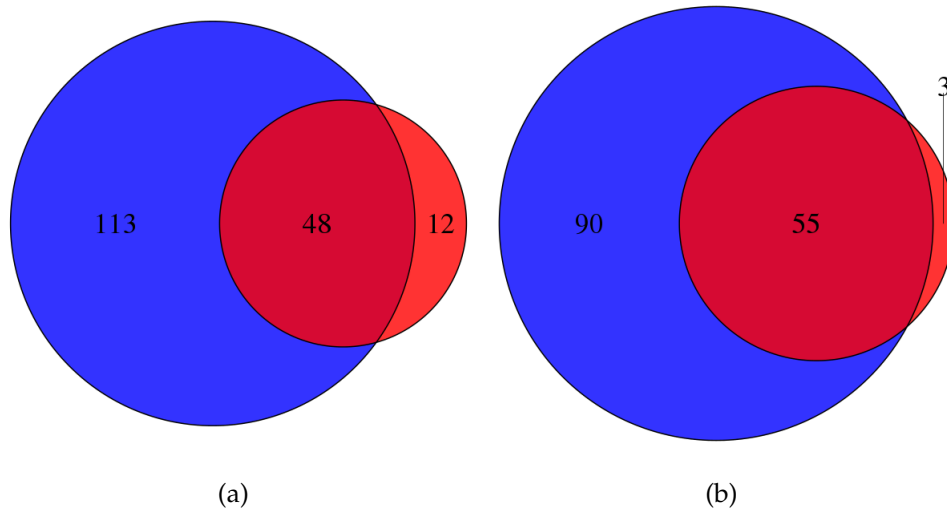
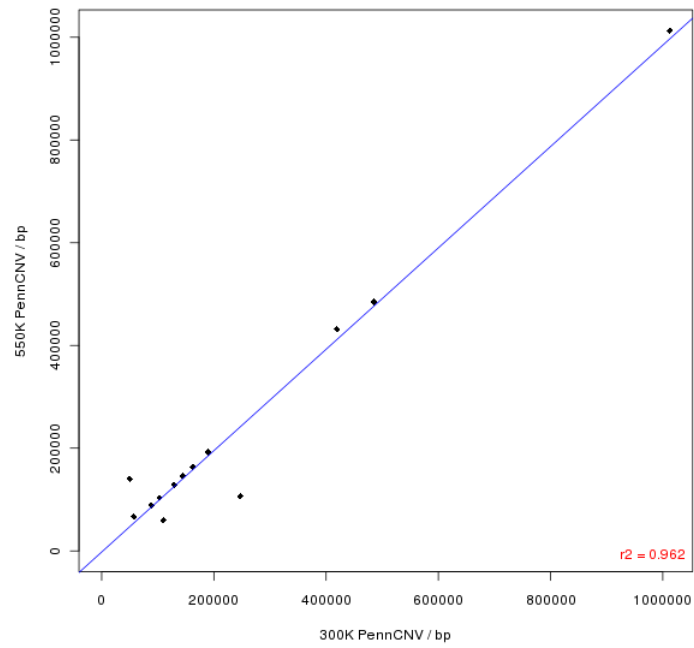
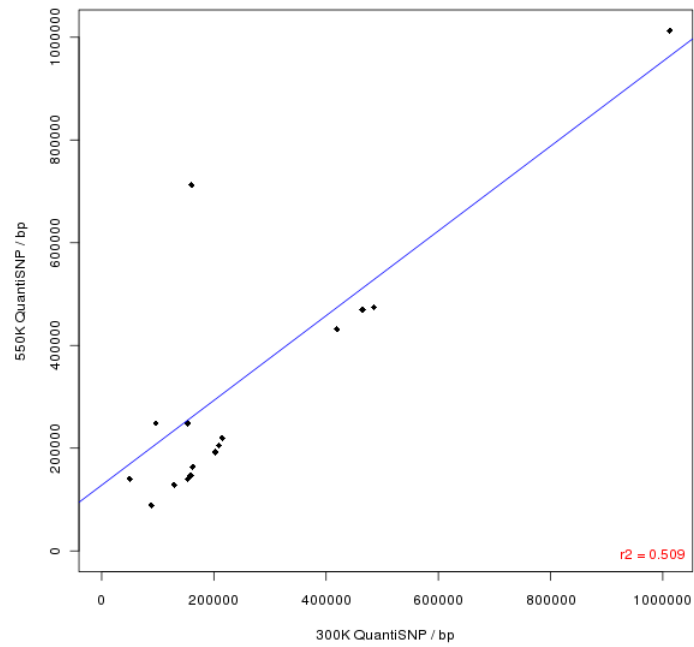


Figure 7.4: Venn diagram of CNV calls in individuals genotyped on the OmniExpress and 550K SNP arrays ($n = 27$), CNV calling thresholds were ≥ 3 SNPs / CNV, CNV size ≥ 1 kb, and LBF ≥ 30 . (a) PennCNV and (b) QuantiSNP.

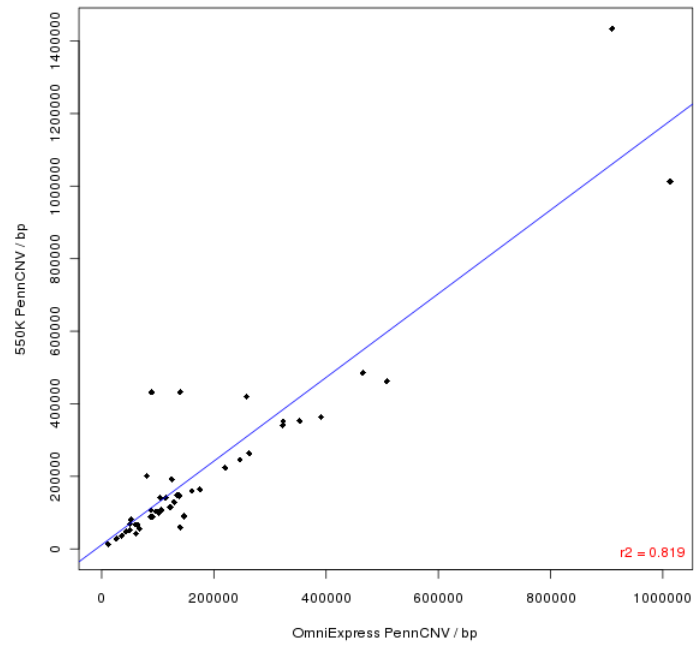


(a)

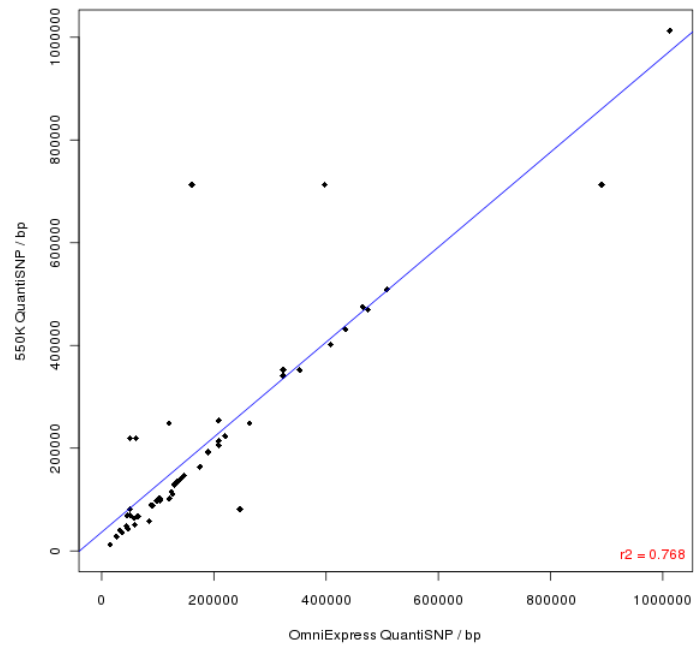


(b)

Figure 7.5: Comparison of size of overlapping CNV calls in individuals genotyped on both the 300K and 550K SNP arrays ($n = 11$). (a) **PennCNV** and (b) **QuantiSNP**.



(a)

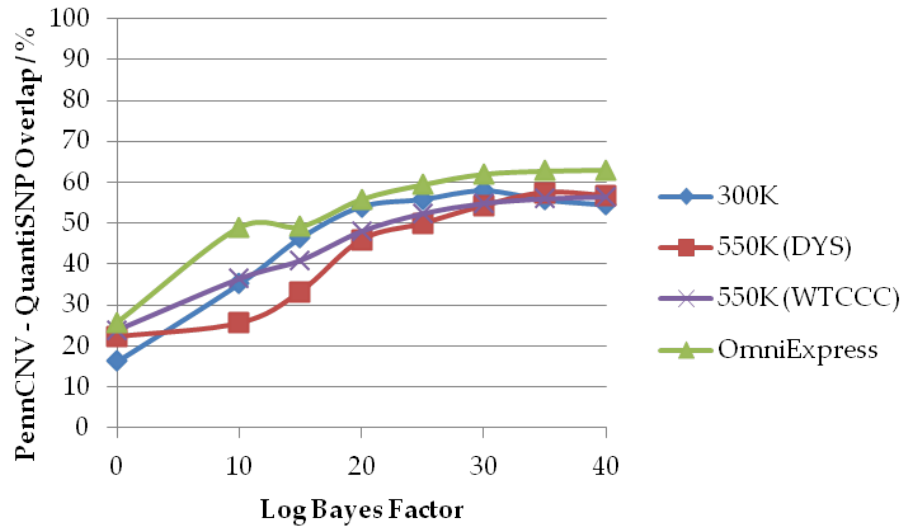


(b)

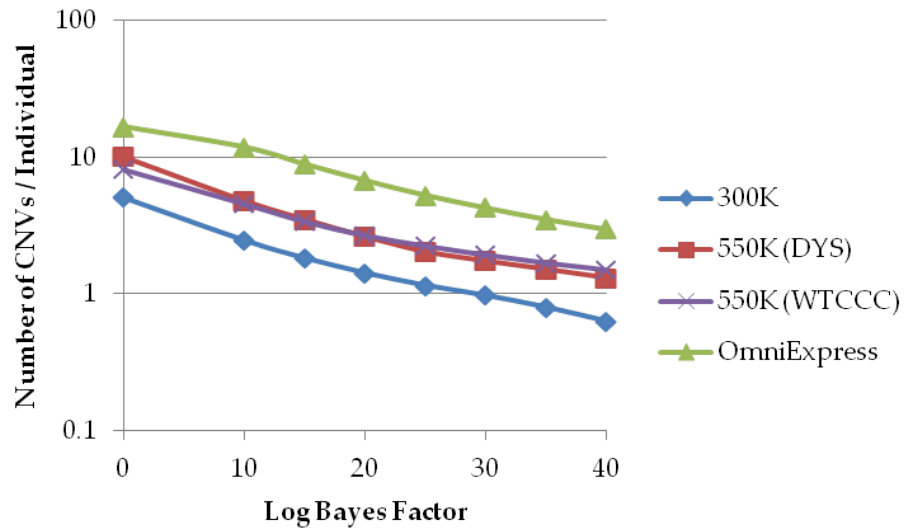
Figure 7.6: Comparison of size of overlapping CNV calls in individuals genotyped on both the **OmniExpress** and **550K** SNP arrays ($n = 27$). (a) **PennCNV** and (b) **QuantiSNP**.

7.4 Comparison of the CNV calling algorithms

I compared the overlap of CNV calls between the algorithms PennCNV and QuantiSNP for individuals genotyped on each array separately. The proportion of CNVs that were called by both algorithms was compared to the total number of CNVs called for a number of confidence thresholds (**Figure 7.7a**). As in **section 7.3**, the proportion of overlapping CNVs increased with confidence score and started to tail off around $\text{LBF} = 30$. QuantiSNP was more permissive than PennCNV across the arrays, calling more CNVs (**Figure 7.8**). The number of CNVs called also decreased exponentially with confidence score (**Figure 7.7b**). Based on these results and **section 7.3**, I chose a threshold of $\text{LBF} = 30$. For the CNVs that were called by both algorithms, the size of the CNVs was assessed to determine how well the algorithms were defining CNV boundaries on the SNP arrays (**Figure 7.9**). The correlation coefficient between overlapping CNVs from the same individuals called by PennCNV and QuantiSNP is $r^2 = 0.981$ for those genotyped on the 300K array, $r^2 = 0.810$ for 550K, and $r^2 = 0.823$ for OmniExpress. Some overlapping CNVs show large differences in size between the two algorithms, and there is no apparent bias in the size of CNVs called by PennCNV or QuantiSNP. I will therefore only take the intersection (as opposed to the union) of CNVs called by both algorithms forward for further analysis as a conservative measure.



(a)



(b)

Figure 7.7: Comparison of the overlap between PennCNV and QuantiSNP for the 300K dyslexia cases (cohorts 5 & 6), 550K dyslexia cases (cohort 1), 550K WTCCC controls and the OmniExpress dyslexia cases (cohort 2) & POBI controls. (a) The proportion of CNVs that overlap on both algorithms compared to the total number of CNVs called for a number of LBF confidence thresholds. (b) The number of CNV calls across difference LBF confidence thresholds decreases exponentially.

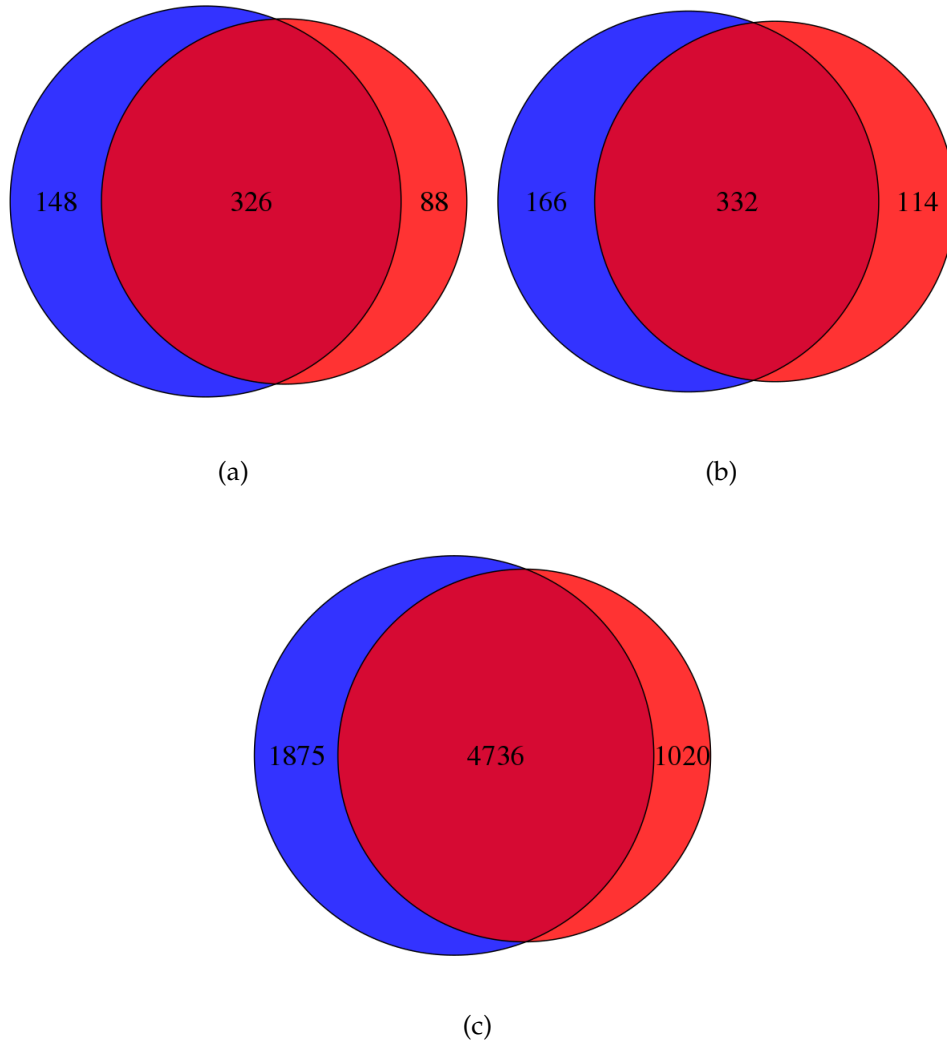


Figure 7.8: Venn diagram of CNV calls using **QuantiSNP** and **PennCNV** calling algorithms, CNV calling thresholds were ≥ 3 SNPs / CNV, CNV size ≥ 1 kb, and LBF ≥ 30 . (a) 300K, (b) 550K and (c) OmniExpress.

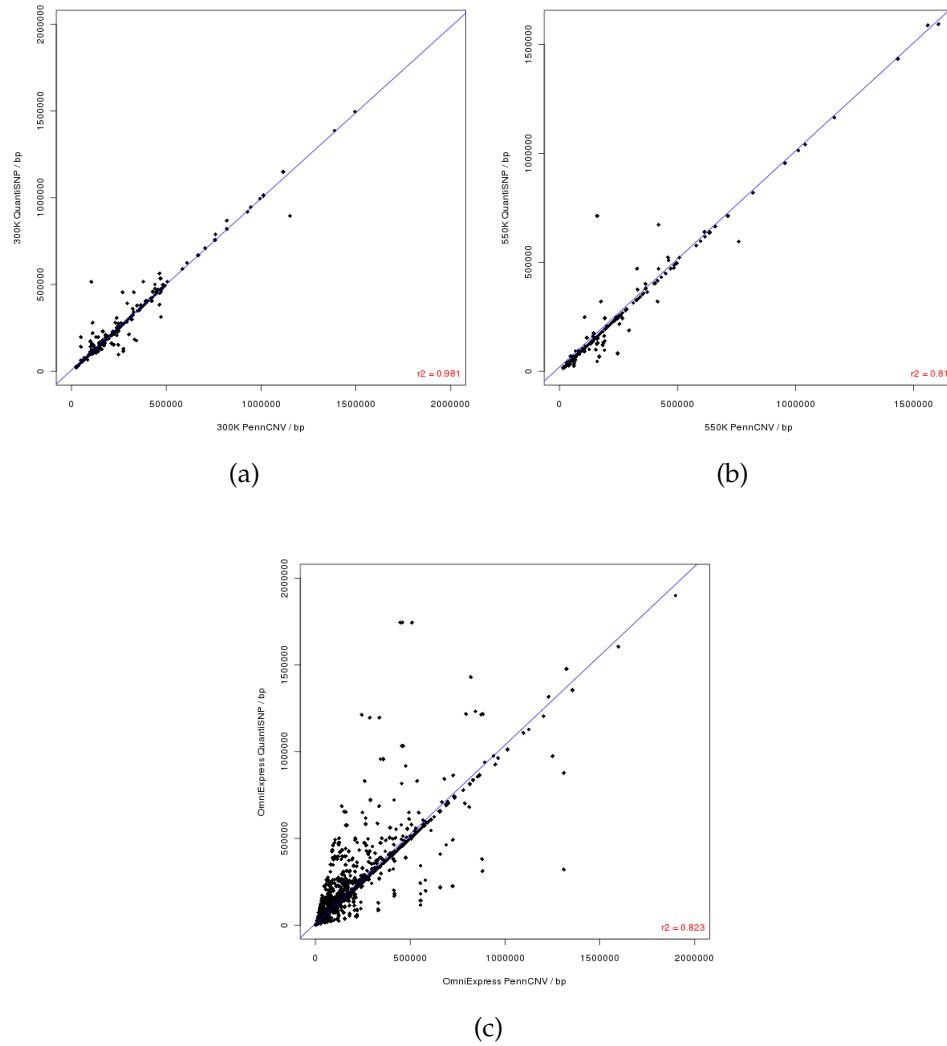


Figure 7.9: Comparison of size of overlapping CNV calls from **PennCNV** and **QuantiSNP**. (a) 300K, (b) 550K and (c) OmniExpress.

7.5 CNV calling criteria

As a result of the CNV quality control and calling methods detailed above, I decided to only analyse CNVs that met the following criteria:

- Minimum of 3 SNPs / CNV.
- Confidence score (LBF) ≥ 30 .

- The intersection of CNVs called by both PennCNV and QuantiSNP algorithms.

Furthermore, the QuantiSNP website recommends a confidence score of 30 to ensure the false positive rate is $\leq 1\%$ (<https://sites.google.com/site/quantisnp/howto/>). Previous work has shown that small CNVs (1-50kb) are called with the same accuracy as large CNVs (Pinto et al., 2011), so no size cutoff was chosen. These stringent criteria for calling CNVs should minimize the risk of false positives.

Chapter 8

CNV analysis

8.1 Abstract

I examined a novel set of copy number variants (CNVs) among a set of 920 unrelated individuals with dyslexia, identifying a significant enrichment of genes affected by duplication CNVs whose mouse orthologues, when disrupted, result in an increased startle reflex phenotype (4.3-fold enrichment; $P = 0.00028$; FDR adjusted $P = 0.047$). However, excluding CNVs commonly observed in two control cohorts unselected for dyslexia ($n = 1,366$), this enrichment is no longer significant and only contributes to 0.54% (5 / 920) of cases. There is a weak negative correlation between length of CNV / individual and reading scores, suggesting poorer readers may have a higher burden of CNVs. However, the combination of different SNP arrays used for genotyping cases and the way DNA was handled compared to controls make it difficult to draw firm conclusions from the data, and the findings above await independent replication. On balance, the combination of burden and mouse phenotype pathway analyses do not suggest that CNVs have a major contribution to the etiology of dyslexia.

8.2 Introduction

For a literature review on the genetics of dyslexia, see **section 6.2**, for an introduction to how CNVs arise and how they are detected, see **section 7.1**.

How do CNVs cause disease?

CNVs can cause disease through altering gene dosage; for example the dysmyelinating peripheral neuropathies Charcot-Marie-Tooth disease type 1A (CMT1A) and hereditary neuropathy with liability to pressure palsies (HNPP), are caused respectively by over-expression and haploinsufficiency of *PMP22* (Lupski et al., 1992). CNVs can unmask recessive alleles, e.g. Sotos syndrome can be caused by a deletion of one copy of *F12* coupled with a SNP that reduces expression of the other copy (Kurotaki et al., 2005). Alternatively, CNVs can disrupt regulatory sequences, as in the case of Liebenberg syndrome, a disorder characterised by a partial homeotic arm-to-leg transformation. Array CGH analyses in two families with Liebenberg syndrome revealed deletions located in *H2AFY*, a gene that has no role in limb development. However, the deletion also removes an insulator element and moves a cis-regulatory element within close proximity of *PITX1*. This 'enhancer-adoption' results in ectopic expression of *PITX1* in the forelimb that causes Liebenberg syndrome (Spielmann et al., 2012) (for review on CNVs in non-coding regions see (Spielmann and Klopocki, 2013)).

CNVs contribute to risk of developing a neurodevelopmental disorder

Beyond Mendelian disease, CNVs play key roles in the etiology of complex neurodevelopmental disorders, which are typically associated with phenotypic heterogeneity (Carvill and Mefford, 2013). For example, ~600kb microduplications of 16p11.2 have been associated with schizophrenia, autism, and bipolar disorder, with the reciprocal microdeletion being associated with developmental delay and autism (McCarthy et al., 2009). Recent studies highlight a role for additional low frequency variants, or 'second hits' to account for this variability (Carvill and Mefford, 2013). For example, individuals with developmental delay that have an inherited 520kb 16p12.1 microdeletion are six times more likely to have a second large CNV compared with controls (Girirajan et al., 2010). A significant burden of CNVs in cases relative to controls has been observed in attention deficit hyperactivity disorder (Williams et al., 2010), autism (Girirajan et al., 2011; Pinto et al., 2010; Sebat et al., 2007), & schizophrenia (The International Schizophrenia Consortium, 2008; Walsh et al., 2008). There has been one study that has analysed the burden of CNVs ($\geq 1\text{Mbp}$) in individuals with dyslexia ($n = 376$) compared to controls ($n = 337$), which finds that they are virtually indistinguishable (Girirajan et al., 2011).

Here I present CNV analysis of 920 individuals with dyslexia genotyped on three different SNP arrays compared to 1,366 controls. I compare the burden of CNVs in dyslexia cases and controls, perform linear regression to test for correlation between single word reading (READ) and CNV burden within the cases, assess overlap of CNVs with known dyslexia candidate genes, and perform functional enrichment analysis utilising mouse phenotype data. Consistent with Girirajan et al., this study shows no evidence for CNVs playing a major role in the etiology of dyslexia.

8.3 Methods & Results

8.3.1 Case / control burden

I assessed whether or not there is an increased burden of CNVs in cases versus controls using PLINK <http://pngu.mgh.harvard.edu//users/mert2481purcell/plink/> (Purcell et al., 2007). PLINK assesses different, but overlapping, aspects of burden using permutation of cases and controls (two-sided test over 100,000 iterations); from the number of CNVs / individual to the proportion of CNVs that intersect genes. I only compared cases and controls that were genotyped on the same SNP array, i.e. cohort 1 & the 1958 birth cohort WTCCC controls genotyped on the 550K array, and cohort 2 & the POBI controls genotyped on the OmniExpress array. There was no burden comparing cohort 1 and the WTCCC controls genotyped on the Illumina 550K SNP array (**Table 8.1**). However, I observed a significant burden of duplications when comparing cohort 2 to the POBI controls that were both genotyped on the Illumina OmniExpress array (**Table 8.2**). In order to visualise this burden, I plotted the proportion of individuals in each cohort against both total length of CNV and number of genes intersected by CNVs (**Figure 8.1 & Figure 8.2**) using the empirical cumulative distribution function in R (*ecdf*, <http://stat.ethz.ch/R-manual/R-devel/library/stats/html/ecdf.html>). Cases consistently have a longer total length of CNV than controls on the OmniExpress SNP array, rather than the burden being driven by a small proportion of cases with large CNVs, and this burden derives entirely from duplications (**Figure 8.3**).

Table 8.1: CNV burden analysis for the dyslexia cases (cohort 1) and the WTCCC controls genotyped on the Illumina 550K array. The burden analysis was performed in PLINK by 1,000,000 permutations of cases and controls, for all CNVs and then separately for deletion and duplication CNVs. Only nominally significant *P* values are shown. OR = Odds Ratio.

[illegible]

Table 8.2: Burden analysis for the dyslexia cases (cohort 2) and the POBI controls genotyped on the Illumina OmniExpress array.

[illegible]

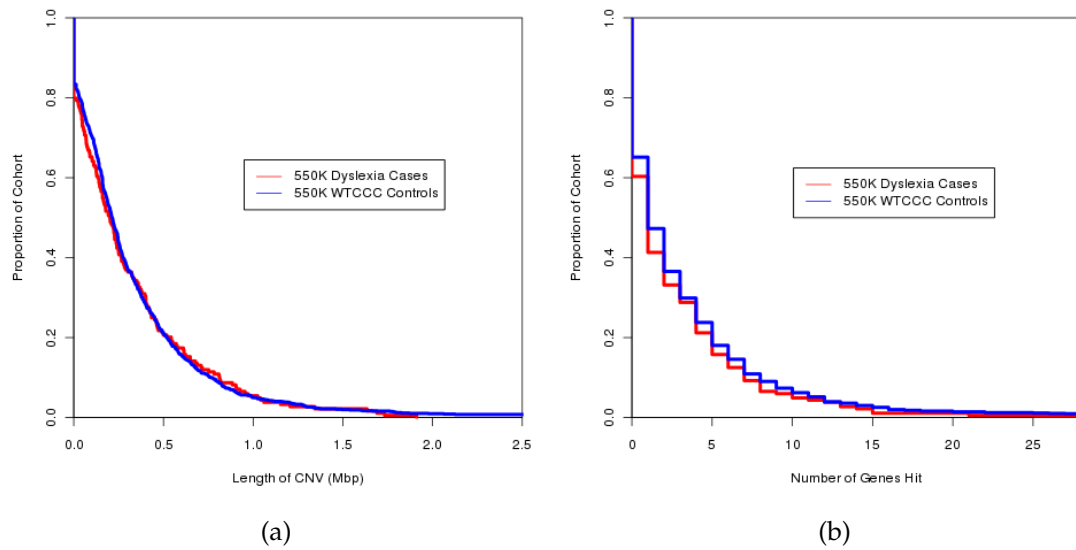


Figure 8.1: **There is no burden comparing cases and controls genotyped on the 550K array.** The proportion of individuals plotted against (a) total length of CNV / individual, and (b) number of genes intertested by CNVs / individual, for both **WTCCC controls** and the **dyslexia cases** (cohort 1).

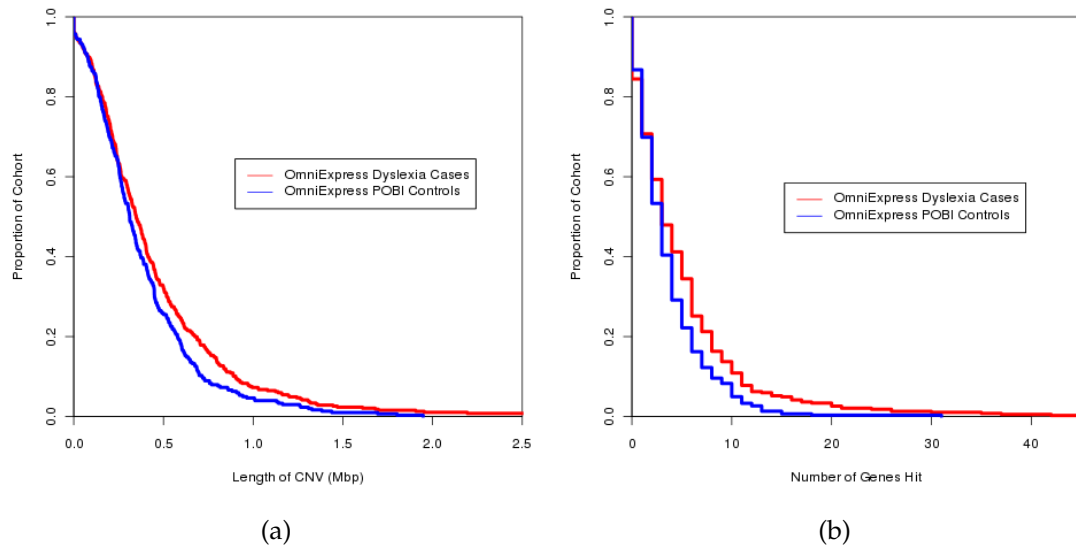


Figure 8.2: **There is a significant burden comparing cases and controls genotyped on the OmniExpress array.** The proportion of individuals plotted against (a) total length of CNV / individual, and (b) number of genes intertested by CNVs / individual, for both **POBI controls** and the **dyslexia cases** (cohort 2)

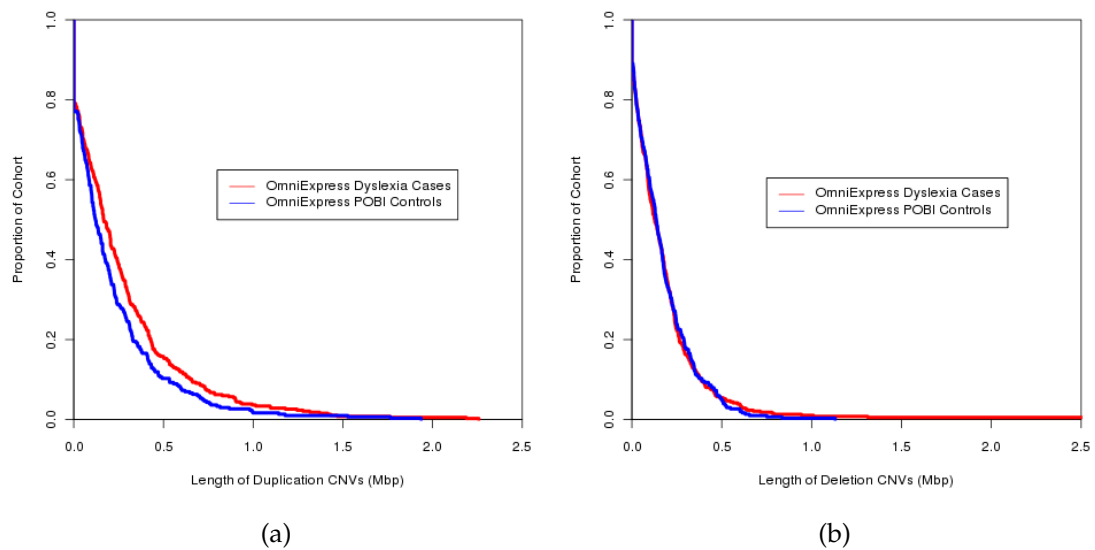


Figure 8.3: **The burden of CNVs in the OmniExpress cases is driven entirely by duplications.** The proportion of individuals plotted against (a) total length of **duplication** CNVs / individual and (b) total length of **deletion** CNVs / individual, for both **POBI controls** and the **dyslexia cases** (cohort 2).

8.3.2 Linear regression

I tested for the correlation between single word reading (READ) score ranks and the ranks of both the length of CNV per individual and number of genes disrupted per individual in a linear regression model, taking performance IQ as a covariate. Ranks are chosen because the CNV length in each individual was not normally distributed, and outliers could inflate the correlation. I then meta-analysed the results for each cohort, using a fixed-effects model in the *rmeta* package (<http://cran.r-project.org/web/packages/rmeta/index.html>). At a confidence score threshold of $\text{LBF} = 30$, reading scores are negatively correlated with both length of CNV and number of genes disrupted per individual. However, this correlation was weak, and only significant for length of CNV (**Figure 8.4**). With a more permissive threshold for calling CNVs of $\text{LBF} = 10$, this correlation improved

slightly and is significant for both length of CNV and number of genes intersected (Figure 8.5).

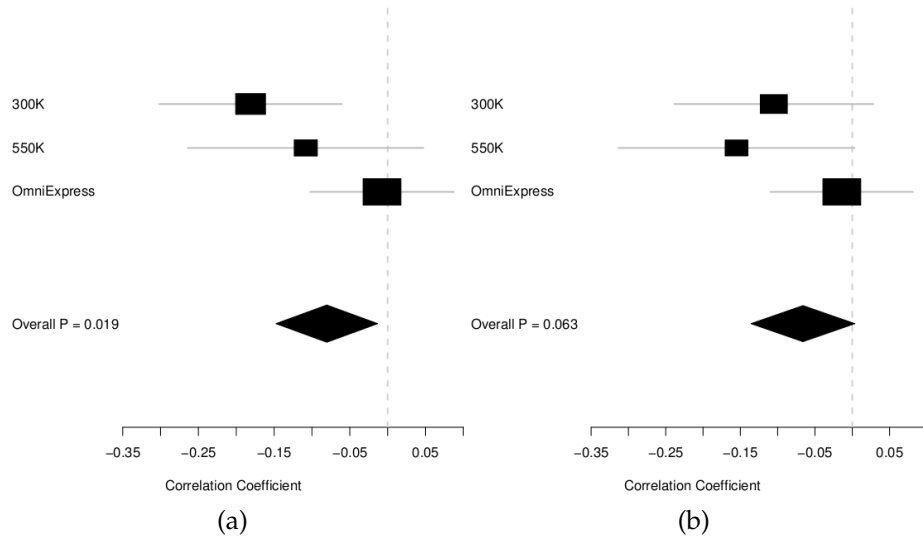


Figure 8.4: **Reading ability is correlated with both CNV length and number of genes disrupted in individuals with dyslexia at a confidence score of 30 for CNV calls.** Forest Plot of the correlation between single-word reading scores and (a) **length** of CNV (b) **genes** disrupted in each individual, correcting for IQ. The intersection of CNVs that had a confidence score of 30, and were called by both PennCNV and QuantiSNP are used for this analysis.

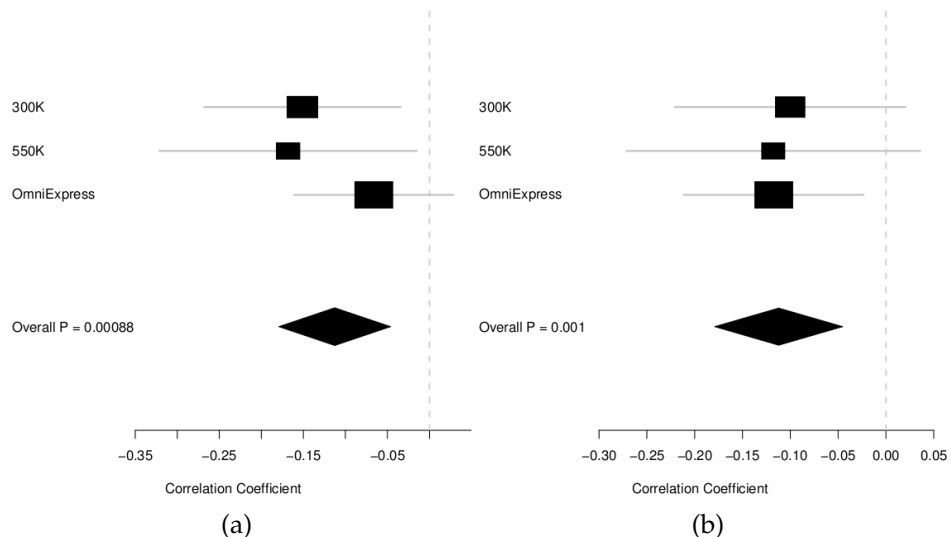


Figure 8.5: **Reading ability is correlated with both CNV length and number of genes disrupted in individuals with dyslexia at a confidence score of 10 for CNV calls.** See Figure 8.4 for an explanation of the plots.

8.3.3 CNVs that intersect known dyslexia genes

I looked within cases and controls to assess whether candidate genes for dyslexia are disrupted in individuals with dyslexia compared to the controls (*CMIP*, *DCDC2*, *DYX1C1*, *KIAA0319*, *MAP1B* & *ROBO1*). One case had a 584kb deletion that intersected *CMIP* (**Table 8.3**). One control had a deletion that intersected *DYX1C1*, and two cases had a CNV that intersected *ROBO2*, a paralogue of *ROBO1* (one duplication and one deletion). Therefore, qualitatively there is not an enrichment of CNVs intersecting these genes in our dyslexia cases.

Table 8.3: A list of CNVs that intersect known dyslexia candidate genes in the case / control CNV analysis.

<i>Individual</i>	<i>Status</i>	<i>Chr</i>	<i>Start</i>	<i>End</i>	<i>Copy Number</i>	<i>Confidence</i>	<i>SNPs</i>	<i>Size</i>	<i>Genes Intersected</i>
LMR-2103-001	Case	3	77,406,508	77,790,068	4	58.2	61	383,560	<i>ROBO2</i>
DYS_BER_064	Case	3	77,698,442	77,968,488	1	205.4	53	270,046	<i>ROBO2</i>
WTCCC66559	Control	15	55,788,832	55,809,660	1	11.8	4	20,828	<i>DYX1C1</i> , <i>DYX1C1-CCPG1</i>
LW-3055-001	Case	16	81,224,234	81,808,233	1	261.8	109	583,999	<i>PKD1L2</i> , <i>LOC100129617</i> , <i>CMIP</i> , <i>MIR4720</i> , <i>GAN</i> , <i>BCMO1</i>

8.3.4 Pathway analysis using mouse phenotypes

Although there was neither a consistent burden between cases and control, nor was there strong correlation between CNV burden and reading scores, the CNVs may intersect genes in the cases that fall into specific clusters or biological pathways. In order to interrogate this, I utilised mouse knockout phenotypic data (for the rationale behind using this data and explanation of the ontology see **subsection 5.3.4**). I first tested the 30 overarching terms for association, then I chose Nervous System ($n = 308$ phenotypes) and Behaviour / Neurological ($n = 166$ phenotypes) categories for further interrogation, because I hypothesized that they were most likely to be relevant for a neurodevelopmental disorder such as dyslexia. I considered only phenotypes that are associated with more than 1% of all genes within each overarching phenotype, so as to reduce uninformative results and to increase methodological power. To correct for hundreds of phenotype tests, False Discovery Rate (FDR) of less than 5% was applied (Benjamini and Hochbert, 1995). This is conservative given that the FDR correction assumes each test is independent, whereas the mouse phenotype gene sets overlap significantly. Reasoning that over-expression through duplication might have a different phenotype to reduction or loss of expression through deletion, I also examined the sets of deletion and duplication CNVs separately.

Human genes intersected by CNVs in cases and controls were assigned using the Ensembl Ensmart 70 database (<http://www.ensembl.org/index.html>). At least one protein-coding exon of every transcript associated with a gene must be intersected by a CNV for it to be considered to be disrupted by a CNV. This method is less prone to gene length biases, particularly important as brain specific expressed genes are typically longer than average (Webber, 2011). It also ensures that the

protein product of the gene is affected whichever transcript is expressed.

Hypergeometric test

First, I tested cases on their own (without controls) by employing the hypergeometric distribution. This method assesses the cases as a whole and tests the null hypothesis that CNVs intersect genes randomly, rather than clustering in a set of genes whose mouse orthologues cause specific phenotypes when disrupted. I considered all cases from the 300K array ($n = 352$), 550K array ($n = 184$), and Omni-Express ($n = 384$) combined ($n = 920$). CNVs are not clustered within genes that are associated with any of the 30 overarching terms, or the 308 Nervous System phenotypes, in all the CNVs, deletions, or duplications. Out of the 166 within the Behaviour / Neurological category, I identified a significant enrichment of genes affected by duplication CNVs in individuals with dyslexia whose mouse orthologues, when disrupted, yield an **increased startle reflex** phenotype (MP:0001488) (4.3-fold enrichment; $P = 0.00028$; FDR adjusted $P = 0.047$). This finding is borderline significant in the CNVs as a whole, including both duplications & deletions (3.5-fold enrichment; $P = 0.00031$, **Table 8.4**). Eight genes contribute to this enrichment; *ABCA2*, *PREX2*, *RLN3*, *SRR*, *TSC1*, *PRODH*, *SEPT5*, & *SLC2A3*, of which the latter three are observed disrupted in control CNVs (**Table 8.5**). The analysis was not significant when genes intersected by CNVs in the controls are removed from the analysis (3.1-fold enrichment; $P = 0.0071$, **Table 8.4**). There was one instance in both the cases and controls where a single large CNV ($\geq 1\text{Mb}$) intersects two genes (*SEPT5* & *PRODH*) that both cause the increased startle reflex when disrupted in mice.

Table 8.4: **There is a significant enrichment of genes affected by CNVs in individuals with dyslexia whose mouse orthologues, when disrupted, yield an increased startle reflex phenotype** (MP:0001488). This enrichment is no longer significant when genes that are also intersected by CNVs in the control sets are removed from the analysis (P value threshold = 0.0003).

<i>SNP array</i>	<i>P</i>	<i>Genes</i>	<i>P control genes removed</i>	<i>Genes</i>
300K	0.0017	4	0.26	1
550K	0.096	2	0.29	1
OmniExpress	0.0047	5	0.057	3
Combined	0.00031	8	0.0071	5

Table 8.5: **Genes intersected by CNVs that are annotated to the phenotype increased startle reflex** (MP:0001488). Number of CNVs that intersect these genes in the cases and controls are given.

<i>Gene</i>	<i>Chromosome</i>	<i>Gene size / kb</i>	<i>n cases</i>	<i>n controls</i>
<i>ABCA2</i>	9	21.1	1	0
<i>PREX2</i>	8	279.3	1	0
<i>PRODH</i>	22	23.8	5	23
<i>RLN3</i>	19	2.8	1	0
<i>SEPT5</i>	22	9.1	1	1
<i>SLC2A3</i>	12	17.1	14	49
<i>SRR</i>	17	21.3	1	0
<i>TSC1</i>	9	53.3	1	0

Fisher's exact case / control test

The hypergeometric test does not compare cases to controls, nor does it consider the frequency with which genes are intersected by CNVs. In order to then compare individuals with dyslexia to controls, I implemented a population frequency-based method for determining functional enrichments between case / control sets of CNVs using a Fisher's exact test. Our approach (developed by Stephen Meader & Caleb Webber) compares the frequency at which cases with CNVs affecting a particular functional association occur compared to that frequency observed for controls. This is contrasted with other approaches that compare a list of genes disrupted in cases to those in controls (Webber, 2011). While the latter approach counts genes, and thus individuals may contribute any number of genes but each gene may be counted only once, this approach counts individuals, which will still be counted even if their CNVs hit a gene already observed. Considering each mouse model phenotype in turn, the number of individuals with CNVs that affect at least one gene whose mouse ortholog's disruption yields that particular phenotype are counted. After repeating this separately for the control individuals, I tested whether genes associated with particular mouse phenotypes are affected more frequently in cases than controls using the Fisher's exact test.

I compared cohort 1 to the WTCCC controls (550K) and cohort 2 to the POBI controls (OmniExpress) for both Nervous System and Behaviour / Neurological phenotypes. No phenotypes are associated after correction for multiple testing. Furthermore, increased startle reflex is not even nominally significant ((at $P \leq 0.05$) when comparing cases to controls on either SNP array, because of the high frequency with which *PRODH* & *SLC2A3* are disrupted in both cases and controls (Table 8.5).

8.4 Discussion

I have compared the burden of CNVs in dyslexia cases and controls, performed linear regression to test for correlation between single word reading and CNV burden within the cases, assessed overlap of CNVs with known dyslexia candidate genes, and performed functional enrichment analysis utilising mouse phenotype data.

There was a burden of duplications in cohort 2 compared to the POBI controls, but this was not observed comparing cohort 1 to the WTCCC controls. This suggests that the burden is an artefact caused by differences in the way cases and controls were handled (**subsection 8.4.1**). Linear regression showed a slight negative correlation between reading scores and burden of CNVs (length of CNV per individual and number of genes disrupted per individual), this was significant but too weak to have any explanatory power for dyslexia. Even though there is correlation it does not imply causation, as for example dyslexia is comorbid with other disorders such as; dyscalculia (Landerl and Moll, 2010), specific language impairment (SLI) (Catts et al., 2002, 2008), and attention deficit hyperactivity disorder (ADHD) (Sexton et al., 2012), and these individuals may be overrepresented at the bottom tail of the distribution of reading ability (although efforts were made to remove individuals with other neurodevelopmental disorders from our study, **section 2.1**). Furthermore in the cases only one individual had a CNV that intersected a known dyslexia candidate gene, *CMIP*. Haploinsufficiency of this locus has been associated with autism (Van der Aa et al., 2012), suggesting a potential comorbidity with other neurodevelopmental phenotypes. There is therefore not strong evidence for a burden effect, either compared to controls, stratifying on severity within the cases, or intersecting known dyslexia candidate genes.

In the mouse phenotype pathway analysis, I found an enrichment of genes affected by CNVs in individuals with dyslexia whose mouse orthologues, when duplicated, yield an increased startle reflex phenotype. An increased startle reflex is observed in individuals with deficits in prepulse inhibition, which attenuates the startle response after a weaker preceding stimulus, or prepulse. Deficits in prepulse inhibition are associated with schizophrenia (Braff et al., 2001). Both decreased & abnormal prepulse inhibition were nominally associated with all CNVs in the dyslexia cohorts (MP:0009142 $P = 0.012$ & MP:0003088 $P = 0.018$). Neither prepulse inhibition nor startle reflex have been previously associated with dyslexia. Furthermore only seven genes annotated to increased startle reflex were affected by duplications, of which two were commonly disrupted in both cases and controls. Therefore, assuming the remaining five all contribute to susceptibility to dyslexia, they are relevant for only 0.54% of the cases (5/920). Also the association was only borderline significant after correcting for multiple testing ($P = 0.047$), and therefore awaits independent replication. Finally, startle reflex was not associated in the case / control analyses, nor was any other phenotype, however only cohorts 1 (550K) & 2 (OmniExpress) were analysed because there was no suitable control for the 300K cases.

8.4.1 Limitations of the study design

There were serious methodological issues that make it difficult to identify robust, true positive associations in the CNV data. First, the cases were genotyped on three different SNP arrays (300K, 550K and OmniExpress), while the controls were only available on the 550K and OmniExpress. The controls were not genotyped on the same plates or at the same time, leading to potential batch effects. DNA was col-

lected using different methods, most cases in cohorts 1 & 2 were collected by cheek swabs, whereas the POBI sample used blood samples (Winney et al., 2012), as did the 1958 birth cohort (the WTCCC controls: <http://www.b58cgene.sgul.ac.uk/collection.php>). Furthermore the DNA had been stored for ~15 years in the case of cohorts 1 & 2, whereas the DNA for the 1958 birth cohort was collected between 2002 & 2004, and the POBI study began in 2004 and collected samples over a number of years (<http://www.peopleofthebritishisles.org/>). These differences in the way DNA was collected, stored, and handled better may explain the systematic increase in duplications observed in cohort 2 compared to the POBI controls, particularly because the same burden was not observed comparing cohort 1 to the WTCCC controls. Furthermore, the burden of total length of CNV / individual (odds ratio = 1.21, **Table 8.2**) is greater than that for autism (OR = 1.19) or schizophrenia (OR = 1.15) (Pinto et al., 2010; The International Schizophrenia Consortium, 2008), suggestive of an artefact rather than a true positive. Finally, since the controls were not selected to exclude individuals with reading disability ~5% will have dyslexia, equivalent to a loss of power approximately equivalent to a reduction of the sample size by 10% (Colhoun et al., 2003).

8.4.2 Conclusions

On balance, consistent with Girirajan et al. (Girirajan et al., 2011), in this study it appears that there is not strong evidence for CNVs playing a major role in the etiology of dyslexia. However, with much larger sample sizes effects may become apparent. CNVs with phenotypic effects strong enough to be detected in a study of this scale are more likely to be relevant for disorders that increase mortality or reduce reproductive success. For example, autism has a very early onset, low fertility, increased mortality, high heritability, and a paternal age effect (Uher, 2009).

Individuals with autism typically do not pass on their genes to the next generation, so there will be strong selection against mutations that significantly increase the likelihood of developing the disorder (Uher, 2009). Disorders such as autism are more likely therefore to be associated with rare or de novo mutations of large effect, such as CNVs intersecting haploinsufficient genes. Dyslexia, on the other hand, is not associated with significant reproductive disadvantage or increased mortality, and although one study has linked it with advanced paternal age (Saha et al., 2009), it seems more likely that dyslexia falls under the common disease / common variants (CD/CV) paradigm (Chakravarti, 1999; Lander, 1996). Assessing epidemiological data is therefore crucial when designing GWA or CNV studies, and much larger sample sizes will be required to detect robust associations between SNPs or CNVs that explain a significant proportion of the heritability of the disorder.

Chapter 9

Concluding remarks & future perspectives

Before this project started, there were only a handful of candidate genes for handedness and dyslexia. The dyslexia candidate genes *DCDC2*, *DYX1C1*, *KIAA0319*, and *ROBO1* are known to play a role in neuronal migration (Galaburda et al., 2006). Coexpression microarray data, transcriptional profiling, and gene knock-down studies have now shown that *DYX1C1*, *KIAA0319*, and *DCDC2* also play roles in ciliogenesis (Chandrasekar et al., 2013; Hoh et al., 2012; Ivliev et al., 2012; Tarkar et al., 2013). Cilia create a leftward NODAL flow crucial for LR asymmetry determination (Hirokawa et al., 2006), and I've shown that genes involved in LR asymmetry determination are important in the development of handedness (chapter 5).

The biological mechanism that determines LR asymmetry in embryonic development involves the rotation of motile cilia that create a leftward flow in the node during gastrulation (**Figure 9.1A**, (Hirokawa et al., 2012)). This flow is detected

by a second type of primary, non-motile mechanosensory, cilia (Tabin and Vogan, 2003). The polycystic kidney disease 2 (*PKD2*) gene, expressed on the cilium is involved in transducing this signal into an increase of intracellular calcium ions on the left side of the node that act as a secondary messenger to trigger left-sided expression of genes such as NODAL (Takao et al., 2013). The expression of NODAL on the left edge of the node induces further expression of itself and other genes in a positive feedback loop that spreads expression to the lateral plate mesoderm and signals left-sided positional information to cells (**Figure 9.1B**, (Babu and Roy, 2013)). Cilia mediate many important functions in development and defective cilia cause many syndromes or disease, known as ciliopathies, which can cause asymmetry defects such as situs inversus (Badano et al., 2006; Fliegauf et al., 2007). Not all individuals with ciliopathies develop situs inversus because disrupting cilia rotation prevents leftward flow, but does not reverse it. LR asymmetry is still established; however dysregulation of flow direction creates a fluctuating asymmetry direction where up to 50% of embryos are born with situs inversus.

Four out of the five most strongly associated genes in the GSEA of the GWAS for relative hand skill in the dyslexia cohort are involved in ciliogenesis: meiosis-specific nuclear structural protein 1 (*MNS1*), regulatory factor X 3 (*RFX3*), GLI family zinc finger 3 (*GLI3*) as well as *PKD2* (**Figure 9.2, Table 5.7**, (Brandler et al., 2013)). Disruption of *Mns1*, *Rfx3*, or *Pkd2* in mice can cause situs inversus (Bonnafe et al., 2004; Pennekamp et al., 2002; Zhou et al., 2012).

Surprisingly, individuals with situs inversus do not show an increased likelihood of being left-handed (McManus, 2004), therefore it was previously thought that mechanisms which regulate body asymmetries were distinct from those that reg-

ulate brain asymmetry (Sun and Walsh, 2006). Yet I have shown that genes that cause asymmetry defects such as situs inversus are important in the development of handedness (**chapter 5**). It is possible therefore, that compensatory mechanisms allow for the normal development of handedness in individuals with situs inversus, suggesting the development of handedness is more complex than just involving early LR asymmetry determining genes.

Aside from LR asymmetry phenotypes; ciliopathies can cause cerebral phenotypes such as an absent corpus callosum or cerebellar vermis (Badano et al., 2006), structures that connect the two hemispheres of the cerebrum and cerebellum, respectively. Individuals with either structure missing in the brain display motor coordination problems (Economou and Katsetos, 2012; Mueller et al., 2009). Motor coordination and balance problems have been consistently observed in individuals with dyslexia, and it has been hypothesised that cerebellar dysfunction underlies both reading and coordination difficulties in dyslexia (Nicolson and Fawcett, 2005; Stoodley and Stein, 2011).

Intriguingly of the two genes I identify at genome-wide significance associated with relative hand skill and reading ability (**chapter 3 & chapter 6**); *PCSK6* is highly expressed in the corpus callosum (Johnson et al., 2003), and *MAP1B* is involved in neuronal migration and causes an absent corpus callosum phenotype when knocked out in mice (Meixner et al., 2000). Therefore I propose that the molecular mechanisms for establishing LR asymmetry in the body are reused later in the development of midline structures of the brain, which in turn influences traits such as handedness and reading ability.

This hypothesis is based on initial data and awaits confirmation. However, with larger cohorts, more detailed phenotyping, and increasingly affordable DNA genotyping and sequencing technologies, we will begin to unravel the relationship between handedness, cerebral asymmetry, and neurodevelopmental disorders such as dyslexia.

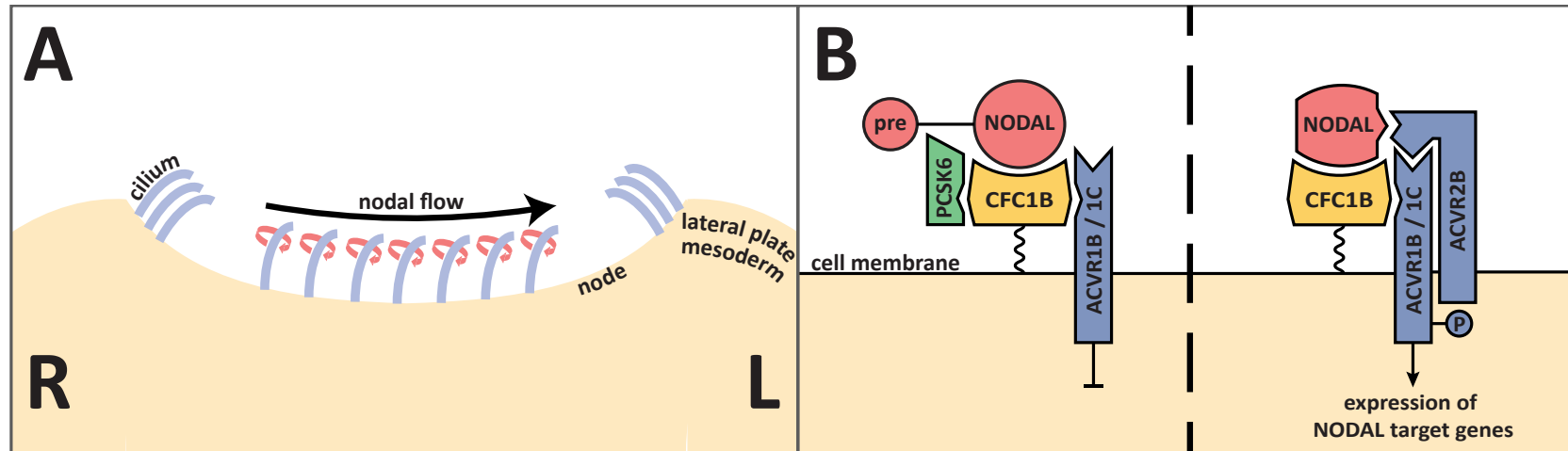


Figure 9.1: Establishment of LR asymmetry during development.

A) Cross-section of the developing embryo during gastrulation viewed from the posterior. The node is a pit that forms transiently at the midline during gastrulation and contains two types of primary **cilia**. Posteriorly angled clockwise rotating cilia create a leftward flow and stationary mechanosensory cilia detect this flow (Babu and Roy, 2013; Tabin and Vogan, 2003), which leads to an increase of intracellular calcium ions that act as a secondary messenger to trigger left-sided expression of genes such as NODAL (Takao et al., 2013).

B) Zoomed in representation of NODAL signalling at the surface of a cell on the left side of both the node and lateral plate mesoderm. Cryptic family protein 1B (CFC1B) is tethered to the membrane by a glycosylphosphatidylinositol (GPI, a glycolipid) anchor (Watanabe et al., 2007), and it recruits NODAL proprotein (pre-NODAL), proprotein convertase subtilisin/kexin type 6 (PCSK6), and activin type I receptors (ACVR1B / ACVR1C) (Blanchet et al., 2008a,b; Reissmann et al., 2001; Yeo and Whitman, 2001). PCSK6 then cleaves pre-NODAL into and active form, and a type II activin receptor (ACVR2B) forms a complex with the NODAL ligand, type I receptors and CFC1B (Reissmann et al., 2001). Type I, type II receptors, and NODAL exist as homodimers and the binding of the NODAL ligand causes the receptors to combine into a heterotetrameric complex (for simplicity proteins are shown as monomers) (Yamashita et al., 1994). Phosphorylation of type I receptors by ACVR2B then transmits the NODAL signal via a signal transduction pathway that activates expression of NODAL target genes, specifying that the cell is on the left side of the embryo (Wrana et al., 1994). Variants in both *PCSK6* and *ACVR2B* have been associated with relative hand skill in individuals with dyslexia (Table 5.7, (Brandler et al., 2013)).

Chapter 10

Bibliography

- G. Abecasis, P. K. Tam, C. D. Bustamante, E. A. Ostrander, S. W. Scherer, S. J. Chanock, P. Y. Kwok, A. J. Brookes, G. R. Abecasis, A. Auton, L. D. Brooks, M. A. DePristo, R. M. Durbin, R. E. Handsaker, H. M. Kang, G. T. Marth, and G. A. McVean. An integrated map of genetic variation from 1,092 human genomes. *Nat Genet*, 491(7422):56–65, 2012.
- C. Ades and E. N. Ramires. Asymmetry of leg use during prey handling in the spider scytodes globula (scytodidae). *Journal of Insect Behavior*, 15(4):563–570, 2002.
- M. Alders, B. M. Hogan, E. Gjini, F. Salehi, L. Al-Gazali, E. A. Hennekam, E. E. Holmberg, M. M. Mannens, M. F. Mulder, G. J. Offerhaus, T. E. Prescott, E. J. Schroor, J. B. Verheij, M. Witte, P. J. Zwijnenburg, M. Vikkula, S. Schulte-Merker, and R. C. Hennekam. Mutations in ccbe1 cause generalized lymph vessel dysplasia in humans. *Nat Genet*, 41(12):1272–4, 2009.
- K. A. Aldinger, O. J. Lehmann, L. Hudgins, V. V. Chizhikov, A. G. Bassuk, L. C. Ades, I. D. Krantz, W. B. Dobyns, and K. J. Millen. Foxc1 is required for normal cerebellar development and is a major contributor to chromosome 6p25.3 dandy-walker malformation. *Nat Genet*, 41(9):1037–42, 2009.
- J. Altmuller, L. J. Palmer, G. Fischer, H. Scherb, and M. Wjst. Genomewide scans of complex human diseases: true linkage is hard to find. *Am J Hum Genet*, 69(5):936–50, 2001.
- C. A. Anderson, F. H. Pettersson, G. M. Clarke, L. R. Cardon, A. P. Morris, and K. T. Zondervan. Data quality control in genetic case-control association studies. *Nat Protoc*, 5(9):1564–73, 2010.
- M. Annett. The growth of manual preference and speed. *Br J Psychol*, 61(4):545–58, 1970.
- M. Annett. *Left, Right, Hand and Brain: The Right Shift Theory*. Psychology Press, London, 1985. 978-0863770180.
- H. Anthoni, M. Zucchelli, H. Matsson, B. Muller-Myhsok, I. Fransson, J. Schumacher, S. Massinen, P. Onkamo, A. Warnke, H. Griesemann, P. Hoffmann, J. Nopola-Hemmi, H. Lyytinen, G. Schulte-Korne, J. Kere, M. M. Nothen, and M. Peyrard-Janvid. A locus on 2p12 containing the co-regulated mrpl19 and c2orf3 genes is associated to dyslexia. *Hum Mol Genet*, 16(6): 667–77, 2007.
- J. A. Armour, A. Davison, and I. C. McManus. Genome-wide association study of handedness excludes simple genetic models. *Heredity*, 2013.
- L. Arning, S. Ocklenburg, S. Schulz, V. Ness, W. M. Gerding, J. G. Hengstler, M. Falkenstein, J. T. Eppelen, O. Gunturkun, and C. Beste. Vntr polymorphism is associated with degree of handedness but not direction of handedness. *PLoS One*, 8(6):e67251, 2013.

- C. Arnould, M. Penta, A. Renders, and J. L. Thonnard. Abilhand-kids: a measure of manual ability in children with cerebral palsy. *Neurology*, 63(6):1045–52, 2004.
- M. B. Ayed, P. Martel, M. Zitouni, D. Gilbert, H. Turki, M. Mokni, A. B. Osman, M. R. Kamoun, A. Zahaf, S. Makni, H. Masmoudi, and F. Tron. Tunisian endemic pemphigus foliaceus is associated with desmoglein 1 gene polymorphism. *Genes Immun*, 3(6):378–9, 2002.
- L. E. Babcock. Trilobite malformations and the fossil record of behavioral asymmetry. *Journal of Paleontology*, 67(2):217–229, 1993.
- L. E. Babcock and R. A. Robison. Preferences of paleozoic predators. *Nature*, 337(6209):695–696, 1989.
- D. Babu and S. Roy. Left-right asymmetry: cilia stir up new surprises in the node. *Open Biol*, 3(5):130052, 2013.
- J. L. Badano, N. Mitsuma, P. L. Beales, and N. Katsanis. The ciliopathies: an emerging class of human genetic disorders. *Annu Rev Genomics Hum Genet*, 7:125–48, 2006.
- P. Bakan. Left-handedness and alcoholism. *Percept Mot Skills*, 36(2):514, 1973.
- K. A. Barth, A. Miklosi, J. Watkins, I. H. Bianco, S. W. Wilson, and R. J. Andrew. fsi zebrafish show concordant reversal of laterality of viscera, neuroanatomy, and a subset of behavioral responses. *Curr Biol*, 15(9):844–50, 2005.
- T. C. Bates, P. A. Lind, M. Luciano, G. W. Montgomery, N. G. Martin, and M. J. Wright. Dyslexia and *dyx1c1*: deficits in reading and spelling associated with a missense mutation. *Mol Psychiatry*, 15(12):1190–6, 2010.
- S. Beck, J. A. Le Good, M. Guzman, N. Ben Haim, K. Roy, F. Beermann, and D. B. Constam. Extraembryonic proteases regulate nodal signalling during gastrulation. *Nat Cell Biol*, 4(12):981–5, 2002.
- G. Bellini, C. Bravaccio, F. Calamoneri, M. Donatella Cocuzza, P. Fiorillo, A. Gagliano, D. Mazzone, E. M. del Giudice, G. Scuccimarra, R. Militerni, and A. Pascotto. No evidence for association between dyslexia and *dyx1c1* functional variants in a group of children and adolescents from southern Italy. *J Mol Neurosci*, 27(3):311–4, 2005.
- C. Benadiba, D. Magnani, M. Niquille, L. Morle, D. Valloton, H. Nawabi, A. Ait-Lounis, B. Otsmane, W. Reith, T. Theil, J. P. Hornung, C. Lebrand, and B. Durand. The ciliogenic transcription factor *rfx3* regulates early midline distribution of guidepost neurons required for corpus callosum development. *PLoS Genet*, 8(3):e1002606, 2012.
- Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, 57(1):289–300, 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing.
- A. Bisazza, C. Cantalupo, A. Robins, L. J. Rogers, and G. Vallortigara. Right-pawedness in toads. *Nature*, 379(6564):408–408, 1996.
- A. Bisazza, R. Pignatti, and G. Vallortigara. Laterality in detour behaviour: interspecific variation in poeciliid fish. *Animal Behaviour*, 54:1273–1281, 1997.
- A. Bisazza, L. Facchin, R. Pignatti, and G. Vallortigara. Lateralization of detour behaviour in poeciliid fish: The effect of species, gender and sexual motivation. *Behav Brain Res*, 91(1-2):157–164, 1998.
- A. Bisazza, L. Facchin, and G. Vallortigara. Heritability of lateralization in fish: concordance of right-left asymmetry between parents and offspring. *Neuropsychologia*, 38(7):907–912, 2000.

- D. V. Bishop. Cerebral asymmetry and language development: cause, correlate, or consequence? *Science*, 340(6138):1230531, 2013.
- D. V. M. Bishop. How to increase your chances of obtaining a significant association between handedness and disorder. *Journal of Clinical and Experimental Neuropsychology*, 12(5):812–816, 1990a.
- D.V.M. Bishop. *Handedness and developmental disorder*, volume 110. Cambridge University Press, 1990b. 0521411955.
- J. A. Blake, C. J. Bult, J. T. Eppig, J. A. Kadin, and J. E. Richardson. The mouse genome database genotypes::phenotypes. *Nucleic Acids Res*, 37(Database issue):D712–9, 2009.
- M. H. Blanchet, J. A. Le Good, D. Mesnard, V. Oorschot, S. Baflast, G. Minchiotti, J. Klumperman, and D. B. Constam. Cripto recruits furin and pace4 and controls nodal trafficking during proteolytic maturation. *EMBO J*, 27(19):2580–91, 2008a.
- M. H. Blanchet, J. A. Le Good, V. Oorschot, S. Baflast, G. Minchiotti, J. Klumperman, and D. B. Constam. Cripto localizes nodal at the limiting membrane of early endosomes. *Sci Signal*, 1(45):ra13, 2008b.
- C. S. Bloss, D. C. Delis, D. P. Salmon, and M. W. Bondi. Apoe genotype is associated with left-handedness and visuospatial skills in children. *Neurobiol Aging*, 31(5):787–95, 2010.
- E. Bonnafe, M. Touka, A. AitLounis, D. Baas, E. Barras, C. UCLA, A. Moreau, F. Flamant, R. Dubruille, P. Couble, J. Collignon, B. Durand, and W. Reith. The transcription factor rfx3 directs nodal cilium development and left-right asymmetry specification. *Mol Cell Biol*, 24(10):4417–27, 2004.
- S. Bookheimer. Functional mri of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, 25:151–188, 2002.
- D. C. Bourassa, I. C. McManus, and M. P. Bryden. Handedness and eye-dominance: a meta-analysis of their relationship. *Laterality*, 1(1):5–34, 1996.
- A. Boyd, J. Golding, J. Macleod, D. A. Lawlor, A. Fraser, J. Henderson, L. Molloy, A. Ness, S. Ring, and G. Davey Smith. Cohort profile: The ‘children of the 90s’—the index offspring of the avon longitudinal study of parents and children. *Int J Epidemiol*, 2012.
- D. L. Braff, M. A. Geyer, and N. R. Swerdlow. Human studies of prepulse inhibition of startle: normal subjects, patient groups, and pharmacological studies. *Psychopharmacology (Berl)*, 156(2-3):234–58, 2001.
- W. M. Brandler, A. P. Morris, D. M. Evans, T. S. Scerri, J. P. Kemp, N. J. Timpson, B. St Pourcain, G. D. Smith, S. M. Ring, J. Stein, A. P. Monaco, J. B. Talcott, S. E. Fisher, C. Webber, and S. Paracchini. Common variants in left/right asymmetry genes and pathways are associated with relative hand skill. *PLoS genetics*, 9(9):e1003751, 2013.
- Z. Brkanac, N. H. Chapman, M. M. Matsushita, L. Chun, K. Nielsen, E. Cochrane, V. W. Berninger, E. M. Wijsman, and W. H. Raskind. Evaluation of candidate genes for dyx1 and dyx2 in families with dyslexia. *Am J Med Genet B Neuropsychiatr Genet*, 144B(4):556–60, 2007.
- M. P. Bryden, I. C. McManus, and M. B. Bulman-Fleming. Evaluating the empirical support for the geschwind-behan-galaburda model of cerebral lateralization. *Brain Cogn*, 26(2):103–67, 1994.
- L. A. Buddenberg and C. Davis. Test-retest reliability of the purdue pegboard test. *Am J Occup Ther*, 54(5):555–8, 2000.
- B. Butterworth and Y. Kovas. Understanding neurocognitive developmental disorders can improve education for all. *Science*, 340(6130):300–305, 2013.

- B. Byrne, W. L. Coventry, R. K. Olson, S. Samuelsson, R. Corley, E. G. Willcutt, S. Wadsworth, and J. C. DeFries. Genetic and environmental influences on aspects of literacy and language in early childhood: Continuity and change from preschool to grade 2. *Journal of Neurolinguistics*, 22(3): 219–236, 2009.
- C. Cantalupo, A. Bisazza, and G. Vallortigara. Lateralization of predator-evasion response in a teleost fish (*girardinus falcatus*). *Neuropsychologia*, 33(12):1637–1646, 1995.
- D. P. Carey, D. T. Smith, D. Martin, G. Smith, J. Skriver, A. Rutland, and J. W. Shepherd. The bi-pedal ape: plasticity and asymmetry in footedness. *Cortex*, 45(5):650–61, 2009.
- G. L. Carvill and H. C. Mefford. Microdeletion syndromes. *Curr Opin Genet Dev*, 23(3):232–9, 2013.
- L. Cashmore, N. Uomini, and A. Chapelain. The evolution of handedness in humans and great apes: a review and current issues. *Journal of Anthropological Sciences*, 86:7–35, 2008.
- H. W. Catts, M. E. Fey, J. B. Tomblin, and X. Zhang. A longitudinal investigation of reading outcomes in children with language impairments. *J Speech Lang Hear Res*, 45(6):1142–57, 2002.
- H. W. Catts, M. S. Bridges, T. D. Little, and J. B. Tomblin. Reading achievement growth in children with language impairments. *J Speech Lang Hear Res*, 51(6):1569–79, 2008.
- A. Chakravarti. Population genetics—making sense out of sequence. *Nat Genet*, 21(1 Suppl):56–60, 1999.
- Gayathri Chandrasekar, Liselotte Vesterlund, Kjell Hultenby, Isabel Tapia-PÃ¡ez, and Juha Kere. The zebrafish orthologue of the dyslexia candidate gene *dyx1c1* is essential for cilia growth and function. *PLoS One*, 8(5):e63123, 2013. <http://dx.doi.org/10.1371>
- C. Chazaud, P. Chambon, and P. Dolle. Retinoic acid is required in the mouse embryo for left-right asymmetry determination and heart morphogenesis. *Development*, 126(12):2589–96, 1999.
- G. M. Clark, T. J. Crow, T. R. Barrick, S. L. Collinson, A. C. James, N. Roberts, and C. E. Mackay. Asymmetry loss is local rather than global in adolescent onset schizophrenia. *Schizophr Res*, 120(1-3):84–6, 2010.
- E. A. Cockayne. Congenital pyloric stenosis in first and second consins. *Arch Dis Child*, 13(75): 249–53, 1938.
- K. M. Colby and C. Parkison. Handedness in autistic children. *J Autism Child Schizophr*, 7(1):3–9, 1977.
- S. Colella, C. Yau, J. M. Taylor, G. Mirza, H. Butler, P. Clouston, A. S. Bassett, A. Seller, C. C. Holmes, and J. Ragoussis. Quantisnp: an objective bayes hidden-markov model to detect and accurately map copy number variation using snp genotyping data. *Nucleic Acids Res*, 35(6):2013–25, 2007.
- H. M. Colhoun, P. M. McKeigue, and G. Davey Smith. Problems of reporting genetic associations with complex outcomes. *Lancet*, 361(9360):865–72, 2003.
- M. L. Concha, R. D. Burdine, C. Russell, A. F. Schier, and S. W. Wilson. A nodal signaling pathway regulates the laterality of neuroanatomical asymmetries in the zebrafish forebrain. *Neuron*, 28(2):399–409, 2000.
- D. F. Conrad, D. Pinto, R. Redon, L. Feuk, O. Gokcumen, Y. Zhang, J. Aerts, T. D. Andrews, C. Barnes, P. Campbell, T. Fitzgerald, M. Hu, C. H. Ihm, K. Kristiansson, D. G. Macarthur, J. R. Macdonald, I. Onyiah, A. W. Pang, S. Robson, K. Stirrups, A. Valsesia, K. Walter, J. Wei, C. Tyler-Smith, N. P. Carter, C. Lee, S. W. Scherer, and M. E. Hurles. Origins and functional impact of copy number variation in the human genome. *Nature*, 464(7289):704–12, 2010.

- D. B. Constam and E. J. Robertson. Regulation of bone morphogenetic protein activity by pro domains and proprotein convertases. *J Cell Biol*, 144(1):139–49, 1999.
- D. B. Constam and E. J. Robertson. Spc4/pace4 regulates a tgfbeta signaling network during axis formation. *Genes Dev*, 14(9):1146–55, 2000.
- K. D. Coon, A. J. Myers, D. W. Craig, J. A. Webster, J. V. Pearson, D. H. Lince, V. L. Zismann, T. G. Beach, D. Leung, L. Bryden, R. F. Halperin, L. Marlowe, M. Kaleem, D. G. Walker, R. Ravid, C. B. Heward, J. Rogers, A. Papassotiropoulos, E. M. Reiman, J. Hardy, and D. A. Stephan. A high-density whole-genome association study reveals that apoe is the major susceptibility gene for sporadic late-onset alzheimer’s disease. *J Clin Psychiatry*, 68(4):613–8, 2007.
- N. Cope, D. Harold, G. Hill, V. Moskvina, J. Stevenson, P. Holmans, M. J. Owen, M. C. O’Donovan, and J. Williams. Strong evidence that kiaa0319 on chromosome 6p is a susceptibility gene for developmental dyslexia. *Am J Hum Genet*, 76(4):581–91, 2005.
- M. C. Corballis. From mouth to hand: gesture, speech, and the evolution of right-handedness. *Behav Brain Sci*, 26(2):199–208; discussion 208–60, 2003.
- M. C. Corballis, G. Badzakova-Trajkov, and I. S. Haberling. Right hand, left brain: genetic and evolutionary bases of cerebral asymmetries for language and manual action. *Wiley Interdisciplinary Reviews-Cognitive Science*, 3(1):1–17, 2012.
- M.C. Corballis. *The lopsided ape: Evolution of the generative mind*. Oxford University Press, USA, 1993. 0195083520.
- C. D. Cota, P. Bagher, P. Pelc, C. O. Smith, C. R. Bodner, and T. M. Gunn. Mice with mutations in mahogunin ring finger-1 (mgrn1) exhibit abnormal patterning of the left-right axis. *Dev Dyn*, 235(12):3438–47, 2006.
- H. F. Crovitz and K. Zener. A group-test for assessing hand- and eye-dominance. *Am J Psychol*, 75: 271–6, 1962.
- F. Dahdouh, H. Anthoni, I. Tapia-Paez, M. Peyrard-Janvid, G. Schulte-Korne, A. Warnke, H. Remschmidt, A. Ziegler, J. Kere, B. Muller-Myhsok, M. M. Nothen, J. Schumacher, and M. Zucchelli. Further evidence for dyx1c1 as a susceptibility factor for dyslexia. *Psychiatr Genet*, 19(2):59–63, 2009.
- J. I. de Vries, R. H. Wimmers, I. A. Ververs, B. Hopkins, G. J. Savelsbergh, and H. P. van Geijn. Fetal handedness and head position preference: a developmental study. *Dev Psychobiol*, 39(3):171–8, 2001.
- A. Deep-Soboslay, T. M. Hyde, J. P. Callicott, M. S. Lener, B. A. Verchinski, J. A. Apud, D. R. Weinberger, and B. Elvevag. Handedness, heritability, neurocognition and brain asymmetry in schizophrenia. *Brain*, 133(10):3113–22, 2010.
- O. Delaneau, J. Marchini, and J. F. Zagury. A linear complexity phasing method for thousands of genomes. *Nat Methods*, 9(2):179–81, 2012.
- J. F. Demonet, G. Thierry, and D. Cardebat. Renewal of the neurophysiology of language: functional neuroimaging. *Physiol Rev*, 85(1):49–95, 2005.
- M. Y. Dennis, S. Paracchini, T. S. Scerri, L. Prokunina-Olsson, J. C. Knight, R. Wade-Martins, P. Coghill, S. Beck, E. D. Green, and A. P. Monaco. A common variant associated with dyslexia reduces expression of the kiaa0319 gene. *PLoS Genet*, 5(3):e1000436, 2009.
- R. DerSimonian and N. Laird. Meta-analysis in clinical trials. *Control Clin Trials*, 7(3):177–88, 1986.
- B. Devlin and K. Roeder. Genomic control for association studies. *Biometrics*, 55(4):997–1004, 1999.

- S. J. Diskin, M. Li, C. Hou, S. Yang, J. Glessner, H. Hakonarson, M. Bucan, J. M. Maris, and K. Wang. Adjustment of genomic waves in signal intensities from whole-genome snp genotyping platforms. *Nucleic Acids Res*, 36(19):e126, 2008.
- R. Doll, R. Peto, J. Boreham, and I. Sutherland. Mortality in relation to smoking: 50 years' observations on male british doctors. *BMJ*, 328(7455):1519, 2004.
- R. H. Don, P. T. Cox, B. J. Wainwright, K. Baker, and J. S. Mattick. 'touchdown' pcr to circumvent spurious priming during gene amplification. *Nucleic Acids Res*, 19(14):4008, 1991.
- M. Dragovic and G. Hammond. Handedness in schizophrenia: a quantitative review of evidence. *Acta Psychiatr Scand*, 111(6):410–9, 2005.
- A. Economou and C. D. Katsetos. Patterns of cognitive and fine motor deficits in a case of dandy-walker continuum. *J Child Neurol*, 27(7):930–7, 2012.
- E. Eglinton and M. Annett. Handedness and dyslexia: a meta-analysis. *Percept Mot Skills*, 79(3 Pt 2):1611–6, 1994.
- L. El Zein, A. Ait-Lounis, L. Morle, J. Thomas, B. Chhin, N. Spassky, W. Reith, and B. Durand. Rfx3 governs growth and beating efficiency of motile cilia in mouse and controls the expression of genes involved in human ciliopathies. *J Cell Sci*, 122(Pt 17):3180–9, 2009.
- C. D. Elliot, D. J. Murray, and L. S. Pearson. *British Abilities Scales*. NFER-Nelson, Windsor, UK, 1983. British Abilities Scales.
- J. T. Eppig, J. A. Blake, C. J. Bult, J. E. Richardson, J. A. Kadin, and M. Ringwald. Mouse genome informatics (mgi) resources for pathology and toxicology. *Toxicol Pathol*, 35(3):456–7, 2007.
- A. Erhardt, L. Czibere, D. Roeske, S. Lucae, P. G. Unschuld, S. Ripke, M. Specht, M. A. Kohli, S. Kloiber, M. Ising, A. Heck, H. Pfister, P. Zimmermann, R. Lieb, B. Putz, M. Uhr, P. Weber, J. M. Deussing, M. Gonik, M. Bunck, M. S. Kebler, E. Frank, C. Hohoff, K. Domschke, P. Krakowitzky, W. Maier, B. Bandelow, C. Jacob, J. Deckert, S. Schreiber, J. Strohmaier, M. Nothen, S. Cichon, M. Rietschel, T. Bettecken, M. E. Keck, R. Landgraf, B. Muller-Myhsok, F. Holsboer, and E. B. Binder. Tmem132d, a new candidate for anxiety phenotypes: evidence from human and mouse studies. *Mol Psychiatry*, 16(6):647–63, 2011.
- A. Erhardt, N. Akula, J. Schumacher, D. Czamara, N. Karbalai, B. Muller-Myhsok, O. Mors, A. Borglum, A. S. Kristensen, D. P. Woldbye, P. Koefoed, E. Eriksson, E. Maron, A. Metspalu, J. Nurnberger, R. A. Philibert, J. Kennedy, K. Domschke, A. Reif, J. Deckert, T. Otowa, Y. Kawamura, H. Kaiya, Y. Okazaki, H. Tani, K. Tokunaga, T. Sasaki, J. P. Ioannidis, F. J. McMahon, and E. B. Binder. Replication and meta-analysis of tmem132d gene variants in panic disorder. *Transl Psychiatry*, 2:e156, 2012.
- N. Eriksson, J. M. Macpherson, J. Y. Tung, L. S. Hon, B. Naughton, S. Saxonov, L. Avey, A. Wojcicki, I. Pe'er, and J. Mountain. Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genet*, 6(6):e1000993, 2010.
- B. Ewing, L. Hillier, M. C. Wendl, and P. Green. Base-calling of automated sequencer traces using phred. i. accuracy assessment. *Genome Res*, 8(3):175–85, 1998.
- J. Fagard. The nature and nurture of human infant hand preference. *Ann N Y Acad Sci*, 1288(1): 114–23, 2013.
- H. Fairweather. Sex-differences in cognition. *Cognition*, 4(3):231–280, 1976.
- C. Faurie, W. Schiefenhovel, S. Le Bomin, S. Billiard, and M. Raymond. Variation in the frequency of left-handedness in traditional societies. *Current Anthropology*, 46(1):142–147, 2005.

- A. J. Fawcett and R. I. Nicolson. Performance of dyslexic children on cerebellar and cognitive tests. *J Mot Behav*, 31(1):68–78, 1999.
- A. J. Fawcett, R. I. Nicolson, and P. Dean. Impaired performance of children with dyslexia on a range of cerebellar tasks. *Annals of Dyslexia*, 46:259–283, 1996.
- S. Feng, S. Wang, C. C. Chen, and L. Lan. Gwapower: a statistical power calculation software for genome-wide association studies with quantitative traits. *BMC Genet*, 12:12, 2011.
- S. Fiddy, D. Cattermole, D. Xie, X. Y. Duan, and R. Mott. An integrated system for genetic analysis. *Bmc Bioinformatics*, 7:210, 2006.
- S. E. Fisher and G. F. Marcus. The eloquent ape: genes, brains and the evolution of language. *Nature Reviews Genetics*, 7(1):9–20, 2006.
- S. E. Fisher, C. Francks, A. J. Marlow, I. L. MacPhie, D. F. Newbury, L. R. Cardon, Y. Ishikawa-Brush, A. J. Richardson, J. B. Talcott, J. Gayan, R. K. Olson, B. F. Pennington, S. D. Smith, J. C. DeFries, J. F. Stein, and A. P. Monaco. Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia. *Nat Genet*, 30(1):86–91, 2002.
- W. T. Fitch and S. N. Braccini. Primate laterality and the biology and evolution of human handedness: a review and synthesis. *Ann N Y Acad Sci*, 1288:70–85, 2013.
- M. Fliegauf, T. Benzing, and H. Omran. When cilia go bad: cilia defects and ciliopathies. *Nature reviews. Molecular cell biology*, 8(11):880–93, 2007.
- B. F. Fong, A. J. Buis, G. J. Savelsbergh, and J. I. de Vries. Influence of breech presentation on the development of fetal arm posture. *Early Human Development*, 81(6):519–27, 2005a.
- B. F. Fong, G. J. Savelsbergh, H. P. van Geijn, and J. I. de Vries. Does intra-uterine environment influence fetal head-position preference? a comparison between breech and cephalic presentation. *Early Human Development*, 81(6):507–17, 2005b.
- C. L. Fox and D. W. Frayer. Non-dietary marks in the anterior dentition of the krapina neanderthals. *International Journal of Osteoarchaeology*, 7(2):133–149, 1997.
- C. Francks, S. E. Fisher, I. L. MacPhie, A. J. Richardson, A. J. Marlow, J. F. Stein, and A. P. Monaco. A genomewide linkage screen for relative hand skill in sibling pairs. *American Journal of Human Genetics*, 70(3):800–805, 2002.
- C. Francks, L. E. DeLisi, S. E. Fisher, S. H. Laval, J. E. Rue, J. F. Stein, and A. P. Monaco. Confirmatory evidence for linkage of relative hand skill to 2p12-q11. *Am J Hum Genet*, 72(2):499–502, 2003a.
- C. Francks, S. E. Fisher, A. J. Marlow, I. L. MacPhie, K. E. Taylor, A. J. Richardson, J. F. Stein, and A. P. Monaco. Familial and genetic effects on motor coordination, laterality, and reading-related cognition. *Am J Psychiatry*, 160(11):1970–7, 2003b.
- C. Francks, S. Paracchini, S. D. Smith, A. J. Richardson, T. S. Scerri, L. R. Cardon, A. J. Marlow, I. L. MacPhie, J. Walter, B. F. Pennington, S. E. Fisher, R. K. Olson, J. C. DeFries, J. F. Stein, and A. P. Monaco. A 77-kilobase region of chromosome 6p22.2 is associated with dyslexia in families from the united kingdom and from the united states. *Am J Hum Genet*, 75(6):1046–58, 2004.
- C. Francks, S. Maegawa, J. Lauren, B. S. Abrahams, A. Velayos-Baeza, S. E. Medland, S. Colella, M. Groszer, E. Z. McAuley, T. M. Caffrey, T. Timmusk, P. Pruunsild, I. Koppel, P. A. Lind, N. Matsumoto-Itaba, J. Nicod, L. Xiong, R. Joobor, W. Enard, B. Krinsky, E. Nanba, A. J. Richardson, B. P. Riley, N. G. Martin, S. M. Strittmatter, H. J. Moller, D. Rujescu, D. St Clair, P. Muglia, J. L. Roos, S. E. Fisher, R. Wade-Martins, G. A. Rouleau, J. F. Stein, M. Karayiorgou, D. H. Geschwind, J. Ragoussis, K. S. Kendler, M. S. Airaksinen, M. Oshimura, L. E. DeLisi, and A. P. Monaco. Lrrtm1 on chromosome 2p12 is a maternally suppressed gene that is associated paternally with handedness and schizophrenia. *Molecular Psychiatry*, 12(12):1129–1139, 2007.

- D. W. Frayer, I. Fiore, C. Lalueza-Fox, J. Radovic, and L. Bondioli. Right handed neandertals: Vindija and beyond. *Journal of Anthropological Sciences*, 88:113–127, 2010.
- D. W. Frayer, M. Lozano, J. M. B. de Castro, E. Carbonell, J. L. Arsuaga, J. Radovic, I. Fiore, and L. Bondioli. More than 500,000 years of right-handedness in europe. *Laterality*, 17(1):51–69, 2012.
- A. M. Galaburda. Ordinary and extraordinary brain development: Anatomical variation in developmental dyslexia. *Ann Dyslexia*, 39(1):65–80, 1989. Ordinary and extraordinary brain development: Anatomical variation in developmental dyslexia.
- A. M. Galaburda, G. F. Sherman, G. D. Rosen, F. Aboitiz, and N. Geschwind. Developmental dyslexia: four consecutive patients with cortical anomalies. *Ann Neurol*, 18(2):222–33, 1985.
- A. M. Galaburda, J. LoTurco, F. Ramus, R. H. Fitch, and G. D. Rosen. From genes to behavior in developmental dyslexia. *Nat Neurosci*, 9(10):1213–7, 2006.
- K. Gao, A. Masuda, T. Matsuura, and K. Ohno. Human branch point consensus sequence is yunay. *Nucleic Acids Res*, 36(7):2257–67, 2008.
- A. Gazit, A. Yaniv, A. Bafico, T. Pramila, M. Igarashi, J. Kitajewski, and S. A. Aaronson. Human frizzled 1 interacts with transforming wnts to transduce a tcf dependent transcriptional response. *Oncogene*, 18(44):5959–66, 1999.
- N. Geschwind. The organization of language and the brain. *Science*, 170(3961):940–4, 1970.
- N. Geschwind and P. Behan. Left-handedness: association with immune disease, migraine, and developmental learning disorder. *Proc Natl Acad Sci U S A*, 79(16):5097–100, 1982.
- N. Getchell, P. Pabreja, K. Neeld, and V. Carrio. Comparing children with and without dyslexia on the movement assessment battery for children and the test of gross motor development. *Perceptual and Motor Skills*, 105(1):207–214, 2007.
- G. Gibson. Rare and common variants: twenty arguments. *Nat Rev Genet*, 13(2):135–45, 2011.
- A. Giljov, K. Karenina, and Y. Malashichev. Limb preferences in a marsupial, macropus rufogriseus: evidence for postural effect. *Animal Behaviour*, 83(2):525–534, 2012.
- S. Girirajan, J. A. Rosenfeld, G. M. Cooper, F. Antonacci, P. Siswara, A. Itsara, L. Vives, T. Walsh, S. E. McCarthy, C. Baker, H. C. Mefford, J. M. Kidd, S. R. Browning, B. L. Browning, D. E. Dickel, D. L. Levy, B. C. Ballif, K. Platky, D. M. Farber, G. C. Gowans, J. J. Wetherbee, A. Asamoah, D. D. Weaver, P. R. Mark, J. Dickerson, B. P. Garg, S. A. Ellingwood, R. Smith, V. C. Banks, W. Smith, M. T. McDonald, J. J. Hoo, B. N. French, C. Hudson, J. P. Johnson, J. R. Ozmores, J. B. Moeschler, U. Surti, L. F. Escobar, D. El-Khechen, J. L. Gorski, J. Kussmann, B. Salbert, Y. Lacassie, A. Biser, D. M. McDonald-McGinn, E. H. Zackai, M. A. Deardorff, T. H. Shaikh, E. Haan, K. L. Friend, M. Fichera, C. Romano, J. Gecz, L. E. DeLisi, J. Sebat, M. C. King, L. G. Shaffer, and E. E. Eichler. A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat Genet*, 42(3):203–9, 2010.
- S. Girirajan, Z. Brkanac, B. P. Coe, C. Baker, L. Vives, T. H. Vu, N. Shafer, R. Bernier, G. B. Ferrero, M. Silengo, S. T. Warren, C. S. Moreno, M. Fichera, C. Romano, W. H. Raskind, and E. E. Eichler. Relative burden of large cnvs on a range of neurodevelopmental phenotypes. *PLoS Genet*, 7(11):e1002334, 2011.
- J. Golding, M. Pembrey, and R. Jones. Alspac—the avon longitudinal study of parents and children. i. study methodology. *Paediatr Perinat Epidemiol*, 15(1):74–87, 2001.
- F. Gomi and Y. Uchida. Map1b 1-126 interacts with tubulin isoforms and induces neurite outgrowth and neuronal death of cultured cortical neurons. *Brain Res*, 1433:1–8, 2012.

- D. Gordon, C. Abajian, and P. Green. Consed: a graphical tool for sequence finishing. *Genome Res*, 8(3):195–202, 1998.
- K. L. Gunderson, S. Kruglyak, M. S. Graige, F. Garcia, B. G. Kermani, C. Zhao, D. Che, T. Dickinson, E. Wickham, J. Bierle, D. Doucet, M. Milewski, R. Yang, C. Siegmund, J. Haas, L. Zhou, A. Oliphant, J. B. Fan, S. Barnard, and M. S. Chee. Decoding randomly ordered dna arrays. *Genome Res*, 14(5):870–7, 2004.
- M. Guven, D. D. Elalmis, S. Binokay, and U. Tan. Population-level right-paw preference in rats assessed by a new computerized food-reaching test. *Int J Neurosci*, 113(12):1675–89, 2003.
- D. F. Halpern and S. Coren. Do right-handers live longer? *Nature*, 333(6170):213, 1988.
- D. F. Halpern and S. Coren. Handedness and life span. *N Engl J Med*, 324(14):998, 1991.
- E. Hampson and J. S. Sankar. Hand preference in humans is associated with testosterone levels and androgen receptor gene polymorphism. *Neuropsychologia*, 50(8):2018–25, 2012.
- K. Hannula-Jouppi, N. Kaminen-Ahola, M. Taipale, R. Eklund, J. Nopola-Hemmi, H. Kaariainen, and J. Kere. The axon guidance receptor gene *robo1* is a candidate gene for developmental dyslexia. *PLoS Genet*, 1(4):e50, 2005.
- L. J. Harris. Do left-handers die sooner than right-handers? commentary on coren and halpern's (1991) "left-handedness: a marker for decreased survival fitness". *Psychol Bull*, 114(2):203–34; discussion 235–47, 1993.
- C. J. Haycraft, B. Banizs, Y. Aydin-Son, Q. Zhang, E. J. Michaud, and B. K. Yoder. Gli2 and gli3 localize to cilia and require the intraflagellar transport protein polaris for processing and function. *PLoS Genet*, 1(4):e53, 2005.
- S.E. Henderson and D.A. Sugden. *Movement Assessment Battery for Children manual*. The Psychological Corporation, Sidcup, 1992. Movement Assessment Battery for Children manual.
- B. S. Hensler, C. Schatschneider, J. Taylor, and R. K. Wagner. Behavioral genetic approach to the study of dyslexia. *Journal of Developmental and Behavioral Pediatrics*, 31(7):525–532, 2010.
- P. G. Hepper, S. Shahidullah, and R. White. Handedness in the human fetus. *Neuropsychologia*, 29(11):1107–11, 1991.
- P. G. Hepper, G. R. McCartney, and E. A. Shannon. Lateralised behaviour in first trimester human foetuses. *Neuropsychologia*, 36(6):531–4, 1998.
- P. G. Hepper, D. L. Wells, and C. Lynch. Prenatal thumb sucking is related to postnatal handedness. *Neuropsychologia*, 43(3):313–5, 2005.
- R. Hering. Migraine, tension headache and left-handedness. *Cephalalgia*, 15(2):145–6, 1995.
- R. A. Hicks, C. Johnson, T. Cuevas, D. Deharo, and J. Bautista. Do right-handers live longer? an updated assessment of baseball player data. *Percept Mot Skills*, 78(3 Pt 2):1243–7, 1994.
- L. A. Hindorff, J. MacArthur, J. Morales, H. A. Junkins, P.N. Hall, A.K. Klemm, and T. A. Manolio. A catalog of published genome-wide association studies. 2013/04/18.
- A. S. Hinrichs, D. Karolchik, R. Baertsch, G. P. Barber, G. Bejerano, H. Clawson, M. Diekhans, T. S. Furey, R. A. Harte, F. Hsu, J. Hillman-Jackson, R. M. Kuhn, J. S. Pedersen, A. Pohl, B. J. Raney, K. R. Rosenbloom, A. Siepel, K. E. Smith, C. W. Sugnet, A. Sultan-Qurraie, D. J. Thomas, H. Trumbower, R. J. Weber, M. Weirauch, A. S. Zweig, D. Haussler, and W. J. Kent. The ucsc genome browser database: update 2006. *Nucleic Acids Res*, 34(Database issue):D590–8, 2006.
- N. Hirokawa, Y. Tanaka, Y. Okada, and S. Takeda. Nodal flow and the generation of left-right asymmetry. *Cell*, 125(1):33–45, 2006.

- N. Hirokawa, Y. Tanaka, and Y. Okada. Cilia, kif3 molecular motor and nodal flow. *Current opinion in cell biology*, 24(1):31–9, 2012.
- J. N. Hirschhorn and M. J. Daly. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*, 6(2):95–108, 2005.
- R. A. Hoh, T. R. Stowe, E. Turk, and T. Stearns. Transcriptional program of ciliated epithelial cells reveals new cilium and centrosome components and links to human disease. *PLoS One*, 7(12):e52166, 2012.
- B. Hopkins, W. Lems, B. Janssen, and G. Butterworth. Postural and motor asymmetries in newborns. *Human Neurobiology*, 6(3):153–156, 1987.
- B. Hopkins, Y. L. Lems, J. B. Hoeksma, O. Kardaun, and G. Butterworth. Development of head position preference during early infancy - a longitudinal-study in the daily life situation. *Dev Psychobiol*, 23(1):39–53, 1990.
- W. D. Hopkins. Comparative and familial analysis of handedness in great apes. *Psychological Bulletin*, 132(4):538–559, 2006.
- B. Howie, J. Marchini, and M. Stephens. Genotype imputation with thousands of genomes. *G3 (Bethesda)*, 1(6):457–70, 2011.
- B. N. Howie, P. Donnelly, and J. Marchini. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*, 5(6):e1000529, 2009.
- J. A. Hubacek, B. J. Piper, H. Pikhart, A. Peasey, R. Kubinova, and M. Bobak. Lack of an association between left-handedness and apoe polymorphism in a large sample of adults: Results of the czech hapiee study. *Laterality*, 2012.
- A. Ihara, M. Hirata, N. Fujimaki, T. Goto, Y. Umekawa, N. Fujita, Y. Terazono, A. Matani, Q. Wei, T. Yoshimine, S. Yorifuji, and T. Murata. Neuroimaging study on brain asymmetries in situs inversus totalis. *J Neurol Sci*, 288(1-2):72–8, 2010.
- S. Illingworth and D. V. Bishop. Atypical cerebral lateralisation in adults with compensated developmental dyslexia demonstrated using functional transcranial doppler ultrasound. *Brain Lang*, 111(1):61–5, 2009.
- S. Iversen, K. Berg, B. Ellertsen, and F. E. Tonnessen. Motor coordination difficulties in a municipality group and in a clinical sample of poor readers. *Dyslexia*, 11(3):217–31, 2005.
- A. E. Ivliev, P. A. t Hoen, W. M. van Roon-Mom, D. J. Peters, and M. G. Sergeeva. Exploring the transcriptome of ciliated cells using in silico dissection of human tissues. *PLoS One*, 7(4):e35618, 2012.
- S. Jastak and G. Wilkinson. *Wide-range achievement test (revised)*. Jastak Associates, Wilmington, DE, 1984. Wide-range achievement test (revised).
- J. M. Johnson, J. Castle, P. Garrett-Engele, Z. Kan, P. M. Loerch, C. D. Armour, R. Santos, E. E. Schadt, R. Stoughton, and D. D. Shoemaker. Genome-wide survey of human alternative pre-mrna splicing with exon junction microarrays. *Science*, 302(5653):2141–4, 2003.
- D. E. Kaplan, J. Gayan, J. Ahn, T. W. Won, D. Pauls, R. K. Olson, J. C. DeFries, F. Wood, B. F. Pennington, G. P. Page, S. D. Smith, and J. R. Gruen. Evidence for linkage and association with reading disability on 6p21.3-22. *Am J Hum Genet*, 70(5):1287–98, 2002.
- D. N. Kennedy, K. M. O’Craven, B. S. Ticho, A. M. Goldstein, N. Makris, and J. W. Henson. Structural and functional brain asymmetries in human situs inversus totalis. *Neurology*, 53(6):1260–5, 1999.

- M. C. Kiefer, J. E. Tucker, R. Joh, K. E. Landsberg, D. Saltman, and P. J. Barr. Identification of a second human subtilisin-like protease gene in the fes/fps region of chromosome 15. *DNA Cell Biol*, 10(10):757–69, 1991.
- A. J. S. Klar. A single locus, rght, specifies preference for hand utilization in humans. *Cold Spring Harbor Symposia on Quantitative Biology*, 61:59–65, 1996.
- S. Knecht, B. Drager, M. Deppe, L. Bobe, H. Lohmann, A. Floel, E. B. Ringelstein, and H. Henningsen. Handedness and hemispheric language dominance in healthy humans. *Brain*, 123 Pt 12:2512–8, 2000.
- L. Kruglyak and D. A. Nickerson. Variation is the spice of life. *Nat Genet*, 27(3):234–6, 2001.
- N. Kurotaki, J. J. Shen, M. Touyama, T. Kondoh, R. Visser, T. Ozaki, J. Nishimoto, T. Shiihara, K. Uetake, Y. Makita, N. Harada, S. Raskin, C. W. Brown, P. Hoglund, N. Okamoto, and J. R. Lupski. Phenotypic consequences of genetic variation at hemizygous alleles: Sotos syndrome is a contiguous gene syndrome incorporating coagulation factor twelve (fxii) deficiency. *Genet Med*, 7(7):479–83, 2005.
- T. LaFramboise. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Res*, 37(13):4181–93, 2009.
- K. N. Laland. Exploring gene-culture interactions: insights from handedness, sexual selection and niche-construction case studies. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 363(1509):3577–3589, 2008.
- K. N. Laland, J. Kumm, J. D. Van Horn, and M. W. Feldman. A gene-culture model of human handedness. *Behav Genet*, 25(5):433–45, 1995.
- E. S. Lander. The new genomics: global views of biology. *Science*, 274(5287):536–9, 1996.
- K. Landerl and K. Moll. Comorbidity of learning disorders: prevalence and familial transmission. *J Child Psychol Psychiatry*, 51(3):287–94, 2010.
- S. H. Laval, J. C. Dann, R. J. Butler, J. Loftus, J. Rue, S. J. Leask, N. Bass, M. Comazzi, A. Vita, S. Nanko, S. Shaw, P. Peterson, G. Shields, A. B. Smith, J. Stewart, L. E. DeLisi, and T. J. Crow. Evidence for linkage to psychosis and cerebral asymmetry (relative hand skill) on the x chromosome. *Am J Med Genet*, 81(5):420–7, 1998.
- P. Letzkus, W. A. Ribi, J. T. Wood, H. Zhu, S. W. Zhang, and M. V. Srinivasan. Lateralization of olfaction in the honeybee *apis mellifera*. *Curr Biol*, 16(14):1471–6, 2006.
- P. Letzkus, N. Boeddeker, J. T. Wood, S. W. Zhang, and M. V. Srinivasan. Lateralization of visual learning in the honeybee. *Biol Lett*, 4(1):16–8, 2008.
- K. R. Lin, V. Prabhu, H. Shah, A. Kamath, and B. Joseph. Handedness in diplegic cerebral palsy. *Dev Neurorehabil*, 15(5):386–9, 2012.
- S. L. Lipscombe and T. Prior. Is there any relationship between handedness and unilateral headache in migraine? *Cephalalgia*, 22(2):146–8, 2002.
- K. Lovell. Cerebral-dominance and its relation to psychological function - zangwill,ol. *British Journal of Educational Psychology*, 30(3):296–296, 1960.
- M. Lozano, M. Mosquera, J. M. B. de Castro, J. L. Arsuaga, and E. Carbonell. Right handedness of homo heidelbergensis from sima de los huesos (atapuerca, spain) 500,000 years ago. *Evolution and Human Behavior*, 30(5):369–376, 2009.
- M. Luciano, P. A. Lind, D. L. Duffy, A. Castles, M. J. Wright, G. W. Montgomery, N. G. Martin,

- and T. C. Bates. A haplotype spanning k1aa0319 and ttrap is associated with normal variation in reading and spelling ability. *Biol Psychiatry*, 62(7):811–7, 2007.
- J. R. Lupski and P. Stankiewicz. Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS Genet*, 1(6):e49, 2005.
- J. R. Lupski, C. A. Wise, A. Kuwano, L. Pentao, J. T. Parke, D. G. Glaze, D. H. Ledbetter, F. Greenberg, and P. I. Patel. Gene dosage is a mechanism for charcot-marie-tooth disease type-1a. *Nature Genetics*, 1(1):29–33, 1992.
- R. Magi and A. P. Morris. Gwama: software for genome-wide association meta-analysis. *Bmc Bioinformatics*, 11:288, 2010.
- J. M. Maisog, E. R. Einbinder, D. L. Flowers, P. E. Turkeltaub, and G. F. Eden. A meta-analysis of functional neuroimaging studies of dyslexia. *Ann N Y Acad Sci*, 1145:237–59, 2008.
- T. A. Manolio, F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, J. H. Cho, A. E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C. N. Rotimi, M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G. Gibson, J. L. Haines, T. F. Mackay, S. A. McCarroll, and P. M. Visscher. Finding the missing heritability of complex diseases. *Nature*, 461(7265):747–53, 2009.
- L. F. Marchant and W. C. McGrew. Human handedness: an ethological perspective. *Human Evolution*, 13(3-4):221–228, 1998. Human handedness: an ethological perspective.
- L. F. Marchant and W. C. McGrew. Handedness is more than laterality: lessons from chimpanzees. *Ann N Y Acad Sci*, 1288:1–8, 2013.
- J. Marchini and B. Howie. Genotype imputation for genome-wide association studies. *Nat Rev Genet*, 11(7):499–511, 2010.
- J. Marchini, B. Howie, S. Myers, G. McVean, and P. Donnelly. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*, 39(7):906–13, 2007.
- C. Marino, A. Citterio, R. Giorda, A. Facchetti, G. Menozzi, L. Vanzin, M. L. Lorusso, M. Nobile, and M. Molteni. Association of short-term memory with a variant within *dyx1c1* in developmental dyslexia. *Genes Brain Behav*, 6(7):640–6, 2007.
- A. J. Marlow, S. E. Fisher, A. J. Richardson, C. Francks, J. B. Talcott, A. P. Monaco, J. F. Stein, and L. R. Cardon. Investigation of quantitative measures related to reading disability in a large sample of sib-pairs from the uk. *Behav Genet*, 31(2):219–30, 2001.
- S. Massinen, M. E. Hokkanen, H. Matsson, K. Tammimies, I. Tapia-Paez, V. Dahlstrom-Heuser, J. Kuja-Panula, J. Burghoorn, K. E. Jeppsson, P. Swoboda, M. Peyrard-Janvid, R. Toftgard, E. Castren, and J. Kere. Increased expression of the dyslexia candidate gene *dcac2* affects length and signaling of primary cilia in neurons. *PLoS One*, 6(6):e20580, 2011.
- M. I. McCarthy, G. R. Abecasis, L. R. Cardon, D. B. Goldstein, J. Little, J. P. Ioannidis, and J. N. Hirschhorn. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet*, 9(5):356–69, 2008.
- S. E. McCarthy, V. Makarov, G. Kirov, A. M. Addington, J. McClellan, S. Yoon, D. O. Perkins, D. E. Dickel, M. Kusenda, O. Krastoshevsky, V. Krause, R. A. Kumar, D. Grozeva, D. Malhotra, T. Walsh, E. H. Zackai, P. Kaplan, J. Ganesh, I. D. Krantz, N. B. Spinner, P. Rocanova, A. Bhandari, K. Pavon, B. Lakshmi, A. Leotta, J. Kendall, Y. H. Lee, V. Vacic, S. Gary, L. M. Iakoucheva, T. J. Crow, S. L. Christian, J. A. Lieberman, T. S. Stroup, T. Lehtimaki, K. Puura, C. Haldeman-Englert, J. Pearl, M. Goodell, V. L. Willour, P. Derosse, J. Steele, L. Kassem, J. Wolff, N. Chitkara, F. J. McMahon, A. K. Malhotra, J. B. Potash, T. G. Schulze, M. M. Nothen, S. Cichon, M. Rietschel, E. Leibenluft, V. Kustanovich, C. M. Lajonchere, J. S. Sutcliffe, D. Skuse, M. Gill, L. Gallagher,

- N. R. Mendell, N. Craddock, M. J. Owen, M. C. O'Donovan, T. H. Shaikh, E. Susser, L. E. Delisi, P. F. Sullivan, C. K. Deutsch, J. Rapoport, D. L. Levy, M. C. King, and J. Sebat. Microduplications of 16p11.2 are associated with schizophrenia. *Nat Genet*, 41(11):1223–7, 2009.
- W. C. McGrew, W. Schiefenhover, and L. F. Marchant. Introduction to the evolution of human handedness. *Ann N Y Acad Sci*, 1288:v–vi, 2013.
- P. McGuffin, P. Asherson, M. Owen, and A. Farmer. The strength of the genetic effect. is there room for an environmental influence in the aetiology of schizophrenia? *Br J Psychiatry*, 164(5):593–9, 1994.
- C. McManus. *Right hand, left hand: The origins of asymmetry in brains, bodies, atoms and cultures*. Harvard University Press, 2004. 0674016130.
- I. C. McManus, A. Davison, and J. A. Armour. Multilocus genetic models of handedness closely resemble single-locus models in explaining family data and are compatible with genome-wide association studies. *Ann N Y Acad Sci*, 1288:48–58, 2013.
- S. E. Medland, D. L. Duffy, A. B. Spurdle, M. J. Wright, G. M. Geffen, G. W. Montgomery, and N. G. Martin. Opposite effects of androgen receptor cag repeat length on increased risk of left-handedness in males and females. *Behav Genet*, 35(6):735–44, 2005.
- S. E. Medland, D. L. Duffy, M. J. Wright, G. M. Geffen, D. A. Hay, F. Levy, C. E. M. Van-Beijsterveldt, G. Willemsen, G. C. Townsend, V. White, A. W. Hewitt, D. A. Mackey, J. M. Bailey, W. S. Slutske, D. R. Nyholt, S. A. Treloar, N. G. Martin, and D. I. Boomsma. Genetic influences on handedness: Data from 25,732 australian and dutch twin families. *Neuropsychologia*, 47(2):330–337, 2009.
- A. Meixner, S. Haverkamp, H. Wassele, S. Fuhrer, J. Thalhammer, N. Kropf, R. E. Bittner, H. Lassmann, G. Wiche, and F. Propst. Map1b is required for axon guidance and is involved in the development of the central and peripheral nervous system. *J Cell Biol*, 151(6):1169–78, 2000.
- H. Meng, K. Hager, M. Held, G. P. Page, R. K. Olson, B. F. Pennington, J. C. DeFries, S. D. Smith, and J. R. Gruen. Tdt-association analysis of ekn1 and dyslexia in a colorado twin cohort. *Hum Genet*, 118(1):87–90, 2005.
- H. B. Messinger, M. I. Messinger, and J. R. Graham. Migraine and left-handedness: is there a connection? *Cephalalgia*, 8(4):237–44, 1988.
- A. Miklosi and R. J. Andrew. Right eye use associated with decision to bite in zebrafish. *Behav Brain Res*, 105(2):199–205, 1999.
- B. Mohapatra, B. Casey, H. Li, T. Ho-Dawson, L. Smith, S. D. Fernbach, L. Molinari, S. R. Niesh, J. L. Jefferies, W. J. Craigen, J. A. Towbin, J. W. Belmont, and S. M. Ware. Identification and functional characterization of nodal rare variants in heterotaxy and isolated cardiovascular malformations. *Hum Mol Genet*, 18(5):861–71, 2009.
- K. L. O. Mueller, S. D. Marion, L. K. Paul, and W. S. Brown. Bimanual motor coordination in agenesis of the corpus callosum. *Behav Neurosci*, 123(5):1000–1011, 2009.
- M. A. Nakaya, K. Biris, T. Tsukiyama, S. Jaime, J. A. Rawls, and T. P. Yamaguchi. Wnt3a links left-right determination with segmentation and anteroposterior axis elongation. *Development*, 132(24):5425–36, 2005.
- D. F. Newbury, L. Winchester, L. Addis, S. Paracchini, L. L. Buckingham, A. Clark, W. Cohen, H. Cowie, K. Dworzynski, A. Everitt, I. M. Goodyer, E. Hennessy, A. D. Kindley, L. L. Miller, J. Nasir, A. O'Hare, D. Shaw, Z. Simkin, E. Simonoff, V. Slonims, J. Watson, J. Ragoussis, S. E. Fisher, J. R. Seckl, P. J. Helms, P. F. Bolton, A. Pickles, G. Conti-Ramsden, G. Baird, D. V. Bishop, and A. P. Monaco. Cmp1 and atp2c2 modulate phonological short-term memory in language impairment. *Am J Hum Genet*, 85(2):264–72, 2009.

- A. Ng, M. Wong, B. Viviano, J. M. Erlich, G. Alba, C. Pflederer, P. Y. Jay, and S. Saunders. Loss of glypican-3 function causes growth factor-dependent defects in cardiac and coronary vascular development. *Dev Biol*, 335(1):208–15, 2009.
- R. I. Nicolson and A. J. Fawcett. Automaticity - a new framework for dyslexia research. *Cognition*, 35(2):159–182, 1990.
- R. I. Nicolson and A. J. Fawcett. Developmental dyslexia, learning and the cerebellum. *J Neural Transm Suppl*, (69):19–36, 2005.
- S. Nonaka, H. Shiratori, Y. Saijoh, and H. Hamada. Determination of left-right patterning of the mouse embryo by artificial nodal flow. *Nature*, 418(6893):96–9, 2002.
- R. C. Oldfield. The assessment and analysis of handedness: The edinburgh inventory. *Neuropsychologia*, 9(1):97–113, 1971.
- R. O’Rahilly and F. Muller. Developmental stages in human embryos: Revised and new measurements. *Cells Tissues Organs*, 192(2):73–84, 2010.
- M. Papadatou-Pastou, M. Martin, M. R. Munafo, and G. V. Jones. Sex differences in left-handedness: a meta-analysis of 144 studies. *Psychol Bull*, 134(5):677–99, 2008.
- S. Paracchini, T. Scerri, and A. P. Monaco. The genetic lexicon of dyslexia. *Annu Rev Genomics Hum Genet*, 8:57–79, 2007.
- S. Paracchini, C. D. Steer, L. L. Buckingham, A. P. Morris, S. Ring, T. Scerri, J. Stein, M. E. Pembrey, J. Ragoussis, J. Golding, and A. P. Monaco. Association of the *kiaa0319* dyslexia susceptibility gene with reading skills in the general population. *Am J Psychiatry*, 165(12):1576–84, 2008.
- S. Paracchini, Q. W. Ang, F. J. Stanley, A. P. Monaco, C. E. Pennell, and A. J. Whitehouse. Analysis of dyslexia candidate genes in the raine cohort representing the general australian population. *Genes Brain Behav*, 10(2):158–65, 2011.
- Silvia Paracchini. Molecular genetics of dyslexia. *eLS*, 2009. 047001590X.
- T. Pastinen, M. Raitio, K. Lindroos, P. Tainola, L. Peltonen, and A. C. Syvanen. A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Res*, 10(7):1031–42, 2000.
- S. Pence. Paw preference in rats. *J Basic Clin Physiol Pharmacol*, 13(1):41–9, 2002.
- P. Pennekamp, C. Karcher, A. Fischer, A. Schweickert, B. Skryabin, J. Horst, M. Blum, and B. Dworniczak. The ion channel polycystin-2 is required for left-right axis determination in mice. *Curr Biol*, 12(11):938–43, 2002.
- S. A. Petrill, K. Deater-Deckard, L. A. Thompson, C. Schatschneider, L. S. Dethorne, and D. J. Vandenberg. Longitudinal genetic analysis of early reading: The western reserve reading project. *Reading and Writing*, 20(1-2):127–146, 2007.
- D. Pinto, A. T. Pagnamenta, L. Klei, R. Anney, D. Merico, R. Regan, J. Conroy, T. R. Magalhaes, C. Correia, B. S. Abrahams, J. Almeida, E. Bacchelli, G. D. Bader, A. J. Bailey, G. Baird, A. Battaglia, T. Berney, N. Bolshakova, S. Bolte, P. F. Bolton, T. Bourgeron, S. Brennan, J. Brian, S. E. Bryson, A. R. Carson, G. Casallo, J. Casey, B. H. Chung, L. Cochrane, C. Corsello, E. L. Crawford, A. Crossett, C. Cytrynbaum, G. Dawson, M. de Jonge, R. Delorme, I. Drmic, E. Duketis, F. Duque, A. Estes, P. Farrar, B. A. Fernandez, S. E. Folstein, E. Fombonne, C. M. Freitag, J. Gilbert, C. Gillberg, J. T. Glessner, J. Goldberg, A. Green, J. Green, S. J. Guter, H. Hakonarson, E. A. Heron, M. Hill, R. Holt, J. L. Howe, G. Hughes, V. Hus, R. Igliozzi, C. Kim, S. M. Klauck, A. Kolevzon, O. Korvatska, V. Kustanovich, C. M. Lajonchere, J. A. Lamb, M. Laskawiec, M. Leboyer, A. Le Couteur, B. L. Leventhal, A. C. Lionel, X. Q. Liu, C. Lord, L. Lotspeich, S. C. Lund, E. Maestrini, W. Mahoney, C. Mantoulan, C. R. Marshall, H. McConachie, C. J. McDougle,

- J. McGrath, W. M. McMahon, A. Merikangas, O. Migita, N. J. Minshew, G. K. Mirza, J. Munson, S. F. Nelson, C. Noakes, A. Noor, G. Nygren, G. Oliveira, K. Papanikolaou, J. R. Parr, B. Parrini, T. Paton, A. Pickles, and M. Pilorge. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*, 466(7304):368–72, 2010.
- D. Pinto, K. Darvishi, X. Shi, D. Rajan, D. Rigler, T. Fitzgerald, A. C. Lionel, B. Thiruvahindrapuram, J. R. Macdonald, R. Mills, A. Prasad, K. Noonan, S. Gribble, E. Prigmore, P. K. Donahoe, R. S. Smith, J. H. Park, M. E. Hurles, N. P. Carter, C. Lee, S. W. Scherer, and L. Feuk. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nat Biotechnol*, 29(6):512–20, 2011.
- B. J. Piper, A. L. Yasen, A. E. Taylor, J. R. Ruiz, J. W. Gaynor, C. A. Dayger, M. Gonzalez-Gross, O. D. Kwon, L. G. Nilsson, I. N. Day, J. Raber, and J. K. Miller. Non-replication of an association of apolipoprotein e2 with sinistrality. *Laterality*, 18(2):251–61, 2013.
- K. Poikolainen. Risk factors for alcohol dependence: a case-control study. *Alcohol Alcohol*, 35(2):190–6, 2000.
- J. Preslar, H. I. Kushner, L. Marino, and B. Pearce. Autism, lateralisation, and handedness: A review of the literature and meta-analysis. *Laterality*, 2013.
- C. J. Price. The anatomy of language: a review of 100 fmri studies published in 2009. *Year in Cognitive Neuroscience 2010*, 1191:62–88, 2010.
- R. J. Pruim, R. P. Welch, S. Sanna, T. M. Teslovich, P. S. Chines, T. P. Gliedt, M. Boehnke, G. R. Abecasis, and C. J. Willer. Locuszoom: Regional visualization of genome-wide association scan results. *Bioinformatics*, 2010.
- S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, and P. C. Sham. Plink: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, 81(3):559–75, 2007.
- A. Quaranta, M. Siniscalchi, A. Frate, and G. Vallortigara. Paw preference in dogs: relations between lateralised behaviour and immunity. *Behav Brain Res*, 153(2):521–5, 2004.
- J. Raber, Y. Huang, and J. W. Ashford. Apoe genotype accounts for the vast majority of ad risk and ad pathology. *Neurobiol Aging*, 25(5):641–50, 2004.
- F. Ramus and G. Szenkovits. What phonological deficit? *Quarterly Journal of Experimental Psychology*, 61(1):129–141, 2008.
- F. Ramus, E. Pidgeon, and U. Frith. The relationship between motor control and phonology in dyslexic children. *Journal of Child Psychology and Psychiatry*, 44(5):712–722, 2003.
- M. Raymond and D. Pontier. Is there geographical variation in human handedness? *Laterality*, 9(1):35–51, 2004.
- R. Redon, S. Ishikawa, K. R. Fitch, L. Feuk, G. H. Perry, T. D. Andrews, H. Fiegler, M. H. Shapero, A. R. Carson, W. Chen, E. K. Cho, S. Dallaire, J. L. Freeman, J. R. Gonzalez, M. Gratacos, J. Huang, D. Kalaitzopoulos, D. Komura, J. R. MacDonald, C. R. Marshall, R. Mei, L. Montgomery, K. Nishimura, K. Okamura, F. Shen, M. J. Somerville, J. Tchinda, A. Valsesia, C. Woodward, F. Yang, J. Zhang, T. Zerjal, L. Armengol, D. F. Conrad, X. Estivill, C. Tyler-Smith, N. P. Carter, H. Aburatani, C. Lee, K. W. Jones, S. W. Scherer, and M. E. Hurles. Global variation in copy number in the human genome. *Nature*, 444(7118):444–54, 2006.
- E. Reissmann, H. Jornvall, A. Blokzijl, O. Andersson, C. Chang, G. Minichiotti, M. G. Persico, C. F. Ibanez, and A. H. Brivanlou. The orphan receptor alk7 and the activin receptor alk4 mediate signaling by nodal proteins during vertebrate development. *Genes & development*, 15(15):2010–22, 2001.

- F. Richlan, M. Kronbichler, and H. Wimmer. Structural abnormalities in the dyslexic brain: A meta-analysis of voxel-based morphometry studies. *Hum Brain Mapp*, 2012.
- E. Roessler, M. V. Ouspenskaia, J. D. Karkera, J. I. Velez, A. Kantipong, F. Lacbawan, P. Bowers, J. W. Belmont, J. A. Towbin, E. Goldmuntz, B. Feldman, and M. Muenke. Reduced nodal signaling strength via mutation of several pathway members including foxh1 is linked to human heart defects and holoprosencephaly. *Am J Hum Genet*, 83(1):18–29, 2008.
- L. J. Rogers and L. Workman. Footedness in birds. *Animal Behaviour*, 45(2):409–411, 1993.
- L. Ronnqvist and B. Hopkins. Head position preference in the human newborn: A new look. *Child Dev*, 69(1):13–23, 1998.
- R. Rosenthal. The file drawer problem and tolerance for null results. *Psychol Bull*, 86(3):638–641, 1979.
- B. Royer-Zemmour, M. Ponsolle-Lenfant, H. Gara, P. Roll, C. Leveque, A. Massacrier, G. Ferracci, J. Cillario, A. Robaglia-Schlupp, R. Vincentelli, P. Cau, and P. Szepetowski. Epileptic and developmental disorders of the speech cortex: ligand/receptor interaction of wild-type and mutant srpx2 with the plasminogen activator receptor upar. *Hum Mol Genet*, 17(23):3617–30, 2008.
- S. Sacco, M. L. Moutard, and J. Fagard. Agenesis of the corpus callosum and the establishment of handedness. *Dev Psychobiol*, 48(6):472–81, 2006.
- S. Saha, A. G. Barnett, C. Foldi, T. H. Burne, D. W. Eyles, S. L. Buka, and J. J. McGrath. Advanced paternal age is associated with impaired neurocognitive outcomes during infancy and childhood. *PLoS Med*, 6(3):e40, 2009.
- F. Sanger, S. Nicklen, and A. R. Coulson. Dna sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, 74(12):5463–7, 1977.
- P. Satz and J. M. Fletcher. Left-handedness and dyslexia - an old myth revisited. *Journal of Pediatric Psychology*, 12(2):291–298, 1987.
- A. Scally, J. Y. Dutheil, L. W. Hillier, G. E. Jordan, I. Goodhead, J. Herrero, A. Hobolth, T. Lapalainen, T. Mailund, T. Marques-Bonet, S. McCarthy, S. H. Montgomery, P. C. Schwalie, Y. A. Tang, M. C. Ward, Y. Xue, B. Yngvadottir, C. Alkan, L. N. Andersen, Q. Ayub, E. V. Ball, K. Beal, B. J. Bradley, Y. Chen, C. M. Clee, S. Fitzgerald, T. A. Graves, Y. Gu, P. Heath, A. Heger, E. Karakoc, A. Kolb-Kokocinski, G. K. Laird, G. Lunter, S. Meader, M. Mort, J. C. Mullikin, K. Munch, T. D. O'Connor, A. D. Phillips, J. Prado-Martinez, A. S. Rogers, S. Sajjadian, D. Schmidt, K. Shaw, J. T. Simpson, P. D. Stenson, D. J. Turner, L. Vigilant, A. J. Vilella, W. Whitener, B. Zhu, D. N. Cooper, P. de Jong, E. T. Dermitzakis, E. E. Eichler, P. Flicek, N. Goldman, N. I. Mundy, Z. Ning, D. T. Odom, C. P. Ponting, M. A. Quail, O. A. Ryder, S. M. Searle, W. C. Warren, R. K. Wilson, M. H. Schierup, J. Rogers, C. Tyler-Smith, and R. Durbin. Insights into hominid evolution from the gorilla genome sequence. *Nature*, 483(7388):169–75, 2012.
- T. S. Scerri and G. Schulte-Korne. Genetics of developmental dyslexia. *Eur Child Adolesc Psychiatry*, 19(3):179–97, 2010.
- T. S. Scerri, S. E. Fisher, C. Francks, I. L. MacPhie, S. Paracchini, A. J. Richardson, J. F. Stein, and A. P. Monaco. Putative functional alleles of dyx1c1 are not associated with dyslexia susceptibility in a large sample of sibling pairs from the uk. *J Med Genet*, 41(11):853–7, 2004.
- T. S. Scerri, S. Paracchini, A. Morris, I. L. MacPhie, J. Talcott, J. Stein, S. D. Smith, B. F. Pennington, R. K. Olson, J. C. DeFries, A. P. Monaco, and A. J. Richardson. Identification of candidate genes for dyslexia susceptibility on chromosome 18. *PLoS One*, 5(10):e13712, 2010.
- T. S. Scerri, A. P. Morris, L. L. Buckingham, D. F. Newbury, L. L. Miller, A. P. Monaco, D. V. Bishop,

- and S. Paracchini. Dcdc2, kiaa0319 and cmip are associated with reading-related traits. *Biol Psychiatry*, 70(3):237–45, 2011.
- W. Schiefenhovel. Biased semantics for right and left in 50 indo-european and non-indo-european languages. *Ann N Y Acad Sci*, 1288:135–52, 2013.
- J. Schumacher, H. Anthoni, F. Dahdouh, I. R. Konig, A. M. Hillmer, N. Kluck, M. Manthey, E. Plume, A. Warnke, H. Remschmidt, J. Hulsmann, S. Cichon, C. M. Lindgren, P. Propping, M. Zucchelli, A. Ziegler, M. Peyrard-Janvid, G. Schulte-Korne, M. M. Nothen, and J. Kere. Strong genetic evidence of dcdc2 as a susceptibility gene for dyslexia. *Am J Hum Genet*, 78(1):52–62, 2006.
- J. Sebat, B. Lakshmi, D. Malhotra, J. Troge, C. Lese-Martin, T. Walsh, B. Yamrom, S. Yoon, A. Krasnitz, J. Kendall, A. Leotta, D. Pai, R. Zhang, Y. H. Lee, J. Hicks, S. J. Spence, A. T. Lee, K. Puura, T. Lehtimaki, D. Ledbetter, P. K. Gregersen, J. Bregman, J. S. Sutcliffe, V. Jobanputra, W. Chung, D. Warburton, M. C. King, D. Skuse, D. H. Geschwind, T. C. Gilliam, K. Ye, and M. Wigler. Strong association of de novo copy number mutations with autism. *Science*, 316(5823):445–9, 2007.
- S. J. Segalowitz, B. E. Berge, S. Lawson, and D. Brown. If you can replicate the handedness-immune disorder effect, the more power to you. *Brain Cogn*, 26(2):217–27, 1994.
- A. V. Segre, L. Groop, V. K. Mootha, M. J. Daly, and D. Altshuler. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet*, 6(8), 2010.
- C. C. Sexton, H. L. Gelhorn, J. A. Bell, and P. M. Classi. The co-occurrence of reading disorder and adhd: epidemiology, treatment, psychosocial impact, and economic burden. *J Learn Disabil*, 45(6):538–64, 2012.
- M. B. Shannon, B. L. Patton, S. J. Harvey, and J. H. Miner. A hypomorphic mutation in the mouse laminin alpha5 gene causes polycystic kidney disease. *J Am Soc Nephrol*, 17(7):1913–22, 2006.
- J. Shapleske, S. L. Rossell, P. W. Woodruff, and A. S. David. The planum temporale: a systematic, quantitative review of its structural, functional and clinical significance. *Brain Res Brain Res Rev*, 29(1):26–49, 1999.
- M. Shinawi and S. W. Cheung. The array cgh and its clinical applications. *Drug Discov Today*, 13(17-18):760–70, 2008.
- A. E. Sieg, E. Zandona, V. M. Izzo, F. V. Paladino, and J. R. Spotila. Population level "flipperedness" in the eastern pacific leatherback turtle. *Behav Brain Res*, 206(1):135–8, 2010.
- J. W. Smoller, N. Craddock, K. Kendler, P. H. Lee, B. M. Neale, J. I. Nurnberger, S. Ripke, S. Santangelo, and P. F. Sullivan. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet*, 381(9875):1371–9, 2013.
- M. Snowling. Developmental dyslexia - a cognitive-developmental perspective. *Bulletin of the British Psychological Society*, 40:A23–A23, 1987.
- M. J. Snowling. The development of grapheme-phoneme correspondence in normal and dyslexic readers. *J Exp Child Psychol*, 29(2):294–305, 1980.
- M. J. Snowling. From language to reading and dyslexia. *Dyslexia*, 7(1):37–46, 2001.
- P. J. Snyder and L. J. Harris. Lexicon size and its relation to foot preference in the african grey parrot (psittacus erithacus). *Neuropsychologia*, 35(6):919–26, 1997.
- I. Sommer, N. Ramsey, R. Kahn, A. Aleman, and A. Bouma. Handedness, language lateralisation and anatomical asymmetry in schizophrenia: meta-analysis. *Br J Psychiatry*, 178:344–51, 2001.

- V. A. Sovrano. A note on asymmetric use of the forelimbs during feeding in the european green toad (*bufo viridis*). *Laterality*, 12(5):458–63, 2007.
- M. Spielmann and E. Klopocki. Cnvs of noncoding cis-regulatory elements in human disease. *Curr Opin Genet Dev*, 23(3):249–56, 2013.
- M. Spielmann, F. Brancati, P. M. Krawitz, P. N. Robinson, D. M. Ibrahim, M. Franke, J. Hecht, S. Lohan, K. Dathe, A. M. Nardone, P. Ferrari, A. Landi, L. Wittler, B. Timmermann, D. Chan, U. Mennen, E. Klopocki, and S. Mundlos. Homeotic arm-to-leg transformation associated with genomic rearrangements at the *pitx1* locus. *Am J Hum Genet*, 91(4):629–35, 2012.
- G. Stancher, E. Clara, L. Regolin, and G. Vallortigara. Lateralized righting behavior in the tortoise (*testudo hermanni*). *Behav Brain Res*, 173(2):315–9, 2006.
- P. Stankiewicz and J. R. Lupski. Structural variation in the human genome and its role in disease. *Annu Rev Med*, 61:437–55, 2010.
- F. J. Steemers, W. Chang, G. Lee, D. L. Barker, R. Shen, and K. L. Gunderson. Whole-genome genotyping with the single-base extension assay. *Nat Methods*, 3(1):31–3, 2006.
- J. T. Stock, M. K. Shirley, L. A. Sarringhaus, T. G. Davies, and C. N. Shaw. Skeletal evidence for variable patterns of handedness in chimpanzees, human hunter-gatherers, and recent british populations. *Ann N Y Acad Sci*, 1288:86–99, 2013.
- C. J. Stoodley and J. F. Stein. A processing speed deficit in dyslexic adults? evidence from a peg-moving task. *Neurosci Lett*, 399(3):264–7, 2006.
- C. J. Stoodley and J. F. Stein. The cerebellum and dyslexia. *Cortex*, 47(1):101–16, 2011.
- C. J. Stoodley, A. J. Fawcett, R. I. Nicolson, and J. F. Stein. Impaired balancing ability in dyslexic children. *Experimental Brain Research*, 167(3):370–80, 2005.
- B. E. Stranger, E. A. Stahl, and T. Raj. Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics*, 187(2):367–83, 2011.
- F. Strockens, O. Gunturkun, and S. Ocklenburg. Limb preferences in non-human vertebrates. *Laterality*, 2012.
- T. Sun and C. A. Walsh. Molecular approaches to brain asymmetry and handedness. *Nat Rev Neurosci*, 7(8):655–62, 2006.
- C. J. Tabin and K. J. Vogan. A two-cilia model for vertebrate left-right axis specification. *Genes & development*, 17(1):1–6, 2003.
- M. Taipale, N. Kaminen, J. Nopola-Hemmi, T. Haltia, B. Myllyluoma, H. Lyytinen, K. Muller, M. Kaaranen, P. J. Lindsberg, K. Hannula-Jouppi, and J. Kere. A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. *Proc Natl Acad Sci U S A*, 100(20):11553–8, 2003.
- D. Takao, T. Nemoto, T. Abe, H. Kiyonari, H. Kajiura-Kobayashi, H. Shiratori, and S. Nonaka. Asymmetric distribution of dynamic calcium signals in the node of mouse embryo during left-right axis formation. *Dev Biol*, 376(1):23–30, 2013.
- A. C. Tang and T. Verstynen. Early life environment modulates ‘handedness’ in rats. *Behav Brain Res*, 131(1-2):1–7, 2002.
- A. Tarkar, N. T. Loges, C. E. Slagle, R. Francis, G. W. Dougherty, J. V. Tamayo, B. Shook, M. Cantino, D. Schwartz, C. Jahnke, H. Olbrich, C. Werner, J. Raidt, P. Pennekamp, M. Abouhamed, R. Hjeij, G. Kohler, M. Griesse, Y. Li, K. Lemke, N. Klena, X. Liu, G. Gabriel, K. Tobita, M. Jaspers, L. C. Morgan, A. J. Shapiro, S. J. Letteboer, D. A. Mans, J. L. Carson, M. W. Leigh, W. E. Wolf, S. Chen,

- J. S. Lucas, A. Onoufriadis, V. Plagnol, M. Schmidts, K. Boldt, R. Roepman, M. A. Zariwala, C. W. Lo, H. M. Mitchison, M. R. Knowles, R. D. Burdine, J. J. Loturco, and H. Omran. Dyx1c1 is required for axonemal dynein assembly and ciliary motility. *Nat Genet*, 45(9):995–1003, 2013.
- The International HapMap Consortium. The international hapmap project. *Nature*, 426(6968):789–96, 2003.
- The International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*, 455(7210):237–41, 2008.
- The Wellcome Trust Case-Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447(7145):661–78, 2007.
- L. Tommasi and G. Vallortigara. Footedness in binocular and monocular chicks. *Laterality*, 4(1):89–95, 1999.
- J. Torgersen. Situs inversus, asymmetry, and twinning. *Am J Hum Genet*, 2(4):361–70, 1950.
- S. R. Tymanskyj, T. M. Scales, and P. R. Gordon-Weeks. Map1b enhances microtubule assembly rates and axon extension rates in developing neurons. *Mol Cell Neurosci*, 49(2):110–9, 2012.
- R. Uher. The role of genetic variation in the causation of mental illness: an evolution-informed framework. *Mol Psychiatry*, 14(12):1072–82, 2009.
- T. Van Agtmael, S. M. Forrest, and R. Williamson. Parametric and non-parametric linkage analysis of several candidate regions for genes for human handedness. *European Journal of Human Genetics*, 10(10):623–630, 2002.
- N. Van der Aa, G. Vandeweyer, E. Reyniers, S. Kenis, L. Dom, G. Mortier, L. Rooms, and R. F. Kooy. Haploinsufficiency of cmip in a girl with autism spectrum disorder and developmental delay due to a de novo deletion on chromosome 16q23.2. *Autism Res*, 5(4):277–81, 2012.
- F. Vargha-Khadem, A. M. O’Gorman, and G. V. Watters. Aphasia and handedness in relation to hemispheric side, age at injury and severity of cerebral lesion during childhood. *Brain*, 108 (Pt 3):677–96, 1985.
- A. Velayos-Baeza, C. Toma, S. Paracchini, and A. P. Monaco. The dyslexia-associated gene kiaa0319 encodes highly n- and o-glycosylated plasma membrane and secreted isoforms. *Hum Mol Genet*, 17(6):859–71, 2008.
- S. C. Vernes, E. Spiteri, J. Nicod, M. Groszer, J. M. Taylor, K. E. Davies, D. H. Geschwind, and S. E. Fisher. 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(6):1232–50, 2007.
- I. A. P. Ververs, J. I. P. Devries, H. P. Vangeijn, and B. Hopkins. Prenatal head position from 12–38 weeks .1. developmental aspects. *Early Human Development*, 39(2):83–91, 1994.
- R. K. Wagner and J. K. Torgesen. The nature of phonological processing and its causal role in the acquisition of reading-skills. *Psychol Bull*, 101(2):192–212, 1987.
- T. Walsh, J. M. McClellan, S. E. McCarthy, A. M. Addington, S. B. Pierce, G. M. Cooper, A. S. Nord, M. Kusenda, D. Malhotra, A. Bhandari, S. M. Stray, C. F. Rippey, P. Roccanova, V. Makarov, B. Lakshmi, R. L. Findling, L. Sikich, T. Stromberg, B. Merriman, N. Gogtay, P. Butler, K. Eckstrand, L. Noory, P. Gochman, R. Long, Z. Chen, S. Davis, C. Baker, E. E. Eichler, P. S. Meltzer, S. F. Nelson, A. B. Singleton, M. K. Lee, J. L. Rapoport, M. C. King, and J. Sebat. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science*, 320 (5875):539–43, 2008.

- K. Wang, M. Li, D. Hadley, R. Liu, J. Glessner, S. F. Grant, H. Hakonarson, and M. Bucan. Penncnv: an integrated hidden markov model designed for high-resolution copy number variation detection in whole-genome snp genotyping data. *Genome Res*, 17(11):1665–74, 2007.
- D. M. Warren, M. Stern, R. Duggirala, T. D. Dyer, and L. Almasy. Heritability and linkage analysis of hand, foot, and eye preference in mexican americans. *Laterality*, 11(6):508–24, 2006.
- K. Watanabe, S. Hamada, C. Bianco, M. Mancino, T. Nagaoka, M. Gonzales, V. Bailly, L. Strizzi, and D. S. Salomon. Requirement of glycosylphosphatidylinositol anchor of cripto-1 for trans activity as a nodal co-receptor. *The Journal of biological chemistry*, 282(49):35772–86, 2007.
- T. Watson. An account of some cases of transposition observed in the human body. *Lond Med Gazette*, 18:393–403, 1836. London Medical Gazette.
- C. Webber. Functional enrichment analysis with structural variants: Pitfalls and strategies. *Cytogenet Genome Res*, 2011.
- C. Webber, J. Y. Hehir-Kwa, D. Q. Nguyen, B. B. de Vries, J. A. Veltman, and C. P. Ponting. Forging links between human mental retardation-associated cnvs and mouse gene knockout models. *PLoS Genet*, 5(6):e1000531, 2009.
- D. L. Wells. Lateralised behaviour in the domestic dog, *canis familiaris*. *Behav Processes*, 61(1-2): 27–35, 2003.
- B. L. White, P. Castle, and R. Held. Observations on the development of visually-directed reaching. *Child Dev*, 35:349–64, 1964.
- K. G. Wigg, J. M. Couto, Y. Feng, B. Anderson, T. D. Cate-Carter, F. Macciardi, R. Tannock, M. W. Lovett, T. W. Humphries, and C. L. Barr. Support for ekn1 as the susceptibility locus for dyslexia on 15q21. *Mol Psychiatry*, 9(12):1111–21, 2004.
- N. M. Williams, I. Zaharieva, A. Martin, K. Langley, K. Mantripragada, R. Fossdal, H. Stefansson, K. Stefansson, P. Magnusson, O. O. Gudmundsson, O. Gustafsson, P. Holmans, M. J. Owen, M. O'Donovan, and A. Thapar. Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. *Lancet*, 376(9750):1401–8, 2010.
- B. Winney, A. Boumertit, T. Day, D. Davison, C. Echeta, I. Evseeva, K. Hutnik, S. Leslie, K. Nicodemus, E. C. Royrvik, S. Tonks, X. Yang, J. Cheshire, P. Longley, P. Mateos, A. Groom, C. Relton, D. T. Bishop, K. Black, E. Northwood, L. Parkinson, T. M. Frayling, A. Steele, J. R. Sampson, T. King, R. Dixon, D. Middleton, B. Jennings, R. Bowden, P. Donnelly, and W. Bodmer. People of the british isles: preliminary analysis of genotypes and surnames in a uk-control population. *European Journal of Human Genetics*, 20(2):203–10, 2012.
- J. L. Wrana, L. Attisano, R. Wieser, F. Ventura, and J. Massague. Mechanism of activation of the tgfbeta receptor. *Nature*, 370(6488):341–7, 1994.
- H. Yamashita, P. ten Dijke, P. Franzen, K. Miyazono, and C. H. Heldin. Formation of hetero-oligomeric complexes of type i and type ii receptors for transforming growth factor-beta. *The Journal of biological chemistry*, 269(31):20172–8, 1994.
- J. Yang, B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. E. Goddard, and P. M. Visscher. Common snps explain a large proportion of the heritability for human height. *Nat Genet*, 42(7):565–9, 2010.
- C. Yeo and M. Whitman. Nodal signals to smads through cripto-dependent and cripto-independent mechanisms. *Mol Cell*, 7(5):949–57, 2001.
- S. Yoshida, H. Shiratori, I. Y. Kuo, A. Kawasumi, K. Shinohara, S. Nonaka, Y. Asai, G. Sasaki, J. A. Belo, H. Sasaki, J. Nakai, B. Dworniczak, B. E. Ehrlich, P. Pennekamp, and H. Hamada. Cilia at

- the node of mouse embryos sense fluid flow for left-right determination via *pkd2*. *Science*, 338(6104):226–31, 2012.
- J. Zhou, F. Yang, N. A. Leu, and P. J. Wang. *Mns1* is essential for spermiogenesis and motile ciliary functions in mice. *PLoS Genet*, 8(3):e1002516, 2012.
- J. C. Ziegler, C. Perry, A. Ma-Wyatt, D. Ladner, and G. Schulte-Korne. Developmental dyslexia in different languages: language-specific or universal? *J Exp Child Psychol*, 86(3):169–93, 2003.
- S. Zollner and J. K. Pritchard. Overcoming the winner’s curse: estimating penetrance parameters from case-control data. *Am J Hum Genet*, 80(4):605–15, 2007.

Chapter 11

Appendix

11.1 Fine mapping of *PCSK6* locus

Table 11.1: SNPs identified in the 3' UTR or near the exons in the 3' end of *PCSK6*. SNPs that failed genotyping are denoted by a dash (-). For the single nucleotide deletion polymorphism rs34318939, the deletion is represented by a dot (·).

<i>Individual</i>	<i>PegQ</i>	<i>rs11855415</i>	<i>rs2073595</i>	<i>rs45488201</i>	<i>rs45592533</i>	<i>rs36022203</i>	<i>rs34318939</i>	<i>rs1030</i>
1	-0.16	T/T	G/A	T/T	T/T	G/G	A/A	C/C
2	-0.16	T/T	G/A	T/T	T/T	G/G	A/A	-
3	0.067	T/T	G/A	T/T	T/T	G/T	A/A	C/C
4	0.07	T/T	G/G	T/T	T/T	G/G	A/A	C/C
5	0.194	T/T	G/A	T/T	T/T	G/T	A/A	C/C
6	0.198	T/T	G/A	-	-	G/T	A/·	C/C
7	-0.129	T/A	G/G	-	-	G/T	A/·	C/T
8	-0.08	T/A	G/A	G/T	G/T	G/T	A/·	C/T
9	0.066	T/A	A/A	-	-	G/T	A/·	C/T
10	0.07	T/A	G/A	G/T	G/T	G/T	A/·	C/T
11	0.2	T/A	G/G	T/T	T/T	G/G	-	C/T
12	0.2	T/A	G/G	T/T	T/T	G/G	A/A	C/T
13	0.08	A/A	G/G	T/T	T/T	G/G	A/·	C/T
14	0.08	A/A	G/G	T/T	T/T	G/G	·/·	T/T
15	0.25	A/A	G/G	-	-	G/G	·/·	-
16	0.38	A/A	G/G	-	-	G/G	·/·	T/T

Table 11.2: **SNPs identified in the secondary promoter of *PCSK6***. SNPs that failed genotyping are denoted by a dash (-).

<i>Individual</i>	<i>PegQ</i>	<i>rs11855415</i>	<i>rs12900794</i>	<i>rs12916087</i>	<i>VNTR (number of repeats)</i>
1	-0.24	T/T	C/C	AG	9/9
2	-0.08	T/T	T/T	AA	9/9
3	-0.01	T/T	C/T	AA	9/6
4	0.03	T/T	C/C	AG	9/9
5	0.06	T/T	C/C	AG	9/6
6	0.08	T/T	C/C	AA	9/9
7	0.11	T/T	C/C	AG	9/9
8	0.12	T/T	C/C	AG	9/6
9	0.16	T/T	C/C	GG	9/9
10	0.32	T/T	T/T	AA	9/9
11	-0.13	T/A	C/T	AG	9/9
12	0.00	T/A	T/T	AA	9/9
13	0.05	T/A	T/T	AA	9/9
14	0.08	T/A	C/T	AG	9/9
15	0.11	T/A	C/T	AG	9/9
16	0.14	T/A	-	-	9/9
17	0.17	T/A	C/T	AG	9/9
18	0.29	T/A	C/T	AA	9/6
19	0.07	A/A	T/T	AA	9/9
20	0.09	A/A	T/T	AA	9/9
21	0.14	A/A	T/T	AA	9/9
22	0.17	A/A	T/T	AA	9/9
23	0.25	A/A	T/T	AA	9/9
24	0.28	A/A	T/T	AA	9/9

Table 11.3: **Genotyping of cell lines for the most highly associated SNP, rs11855415, and the VNTR identified in the secondary promoter of PCSK6.** SNPs that failed genotyping are denoted by a dash (-).

<i>Cell Line</i>	<i>rs11855415</i>	<i>VNTR</i>
293T	TT	9/9
CHP212	TT	9/9
LAN5	TT	9/6
LMR32	AT	9/9
KELLY	TT	9/6
M17	TT	9/9
MRC5V2	AT	9/9
NT2	AT	9/6
SHY-5Y	TT	-
SKNF1	AT	-
SKNMC	TT	6/6
SKNAS	TT	9/9
SKNDZ	TT	9/9

11.2 Samples that failed QC

Controls	Genotyping Failure	Discordant Sex	Duplicated / Related	Divergent Ancestry	Reason for Exclusion
DYS_128_3					Call rate 95.2%
DYS_1_4					Call rate 95.8%
DYS_162_3					Call rate 95.8%
DYS_163_4					Call rate 97.6%
DYS_165_4					Call rate 96.2%
DYS_166_3					Call rate 96.0%
DYS_180_4					Call rate 96.5%
DYS_206_4					Call rate 93.4%
DYS_219_4					Call rate 97.0%
DYS_2_4					Call rate 97.0%
DYS_31_3					Call rate 97.0%
DYS_32_4					Call rate 96.1%
DYS_341_4					Call rate 95.7%
DYS_3_4					Call rate 97.8%
DYS_357_4					Call rate 97.1%
DYS_35_7					Call rate 97.0%
DYS_363_4					Call rate 97.9%
DYS_36_3					Call rate 97.9%
DYS_4_3					Call rate 95.8%
DYS_7_4					Call rate 96.4%
DYS_366_4					Female when expected male
DYS_369_4					Female when expected male
DYS_66_4					Female when expected male
DYS_69_4					Female when expected male
DYS_109_3					Non-European ancestry
DYS_213_3					Non-European ancestry
DYS_338_4					Non-European ancestry
DYS_85_3					Non-European ancestry

Table 11.4: List of 550K cases (cohort 1) that failed quality control.

Controls	Genotyping Failure	Discordant Sex	Duplicated / Related	Divergent Ancestry	Reason for Exclusion
WTCCC66113					Call rate 57.1%
WTCCC66171					Call rate 93.3%
WTCCC66173					Call rate 96.4%
WTCCC66175					Call rate 95.7%
WTCCC66178					Call rate 95.3%
WTCCC66186					Call rate 96.1%
WTCCC66196					Call rate 94.9%
WTCCC66200					Call rate 95.1%
WTCCC66208					Call rate 94%
WTCCC66209					Call rate 95.4%
WTCCC66215					Call rate 94.5%
WTCCC66220					Call rate 95.5%
WTCCC66225					Call rate 95.6%
WTCCC66227					Call rate 97.9%
WTCCC66262					Call rate 52.3%
WTCCC66266					Call rate 53%
WTCCC66281					Call rate 51.8%
WTCCC66282					Call rate 50.3%
WTCCC66285					Call rate 55.7%
WTCCC66370					Call rate 95.6%
WTCCC66408					Call rate 97%
WTCCC66411					Call rate 61.5%
WTCCC66416					Call rate 94.4%
WTCCC66439					Call rate 92.6%
WTCCC66448					Call rate 94.9%
WTCCC66478					Call rate 96%
WTCCC66481					Call rate 96.9%
WTCCC66494					Call rate 95.5%
WTCCC66506					Call rate 96.7%
WTCCC66513					Call rate 97.5%
WTCCC66516					Call rate 78.3%
WTCCC66568					Call rate 97.6%
WTCCC66604					Call rate 97.1%
WTCCC66608					Call rate 97.7%
WTCCC66616					Call rate 94.6%
WTCCC66648					Call rate 94.8%
WTCCC66653					Call rate 97.8%
WTCCC66658					Call rate 97.5%
WTCCC66660					Call rate 95.3%
WTCCC66663					Call rate 97.1%
WTCCC66674					Call rate 97.2%
WTCCC66676					Call rate 95.4%
WTCCC66677					Call rate 97.5%
WTCCC66687					Call rate 97.6%
WTCCC66693					Call rate 97.6%
WTCCC66697					Call rate 97.2%
WTCCC66703					Call rate 96.9%
WTCCC66705					Call rate 95.1%
WTCCC66719					Call rate 97.2%
WTCCC66719					Call rate 97.2%
WTCCC66772					Call rate 97.7%
WTCCC66788					Call rate 97.9%
WTCCC66841					Call rate 98%
WTCCC66849					Call rate 97.9%
WTCCC66882					Call rate 74.9%
WTCCC66915					Call rate 77.4%
WTCCC67029					Call rate 96.3%
WTCCC67025					Call rate 94.3%
WTCCC67029					Call rate 95.3%
WTCCC67041					Call rate 95.4%
WTCCC67046					Call rate 95.2%
WTCCC67048					Call rate 78.7%
WTCCC67053					Call rate 96.2%
WTCCC67058					Call rate 95.8%
WTCCC67059					Call rate 94.9%
WTCCC67062					Call rate 95.7%
WTCCC67074					Call rate 75%
WTCCC67079					Call rate 97%
WTCCC67084					Call rate 97%
WTCCC67085					Call rate 96.4%
WTCCC67165					Call rate 95.2%
WTCCC88226					Call rate 98%
WTCCC88266					Call rate 76.7%
WTCCC88287					Call rate 97.3%
WTCCC88526					Call rate 97.7%
WTCCC88550					Call rate 97.6%
WTCCC88565					Call rate 97.8%
WTCCC88590					Call rate 94.6%
WTCCC88594					Call rate 97.1%
WTCCC88625					Call rate 97.8%
WTCCC88640					Call rate 96.1%
WTCCC88642					Call rate 93.7%
WTCCC88725					Call rate 97.8%
WTCCC88727					Call rate 94.9%
WTCCC88743					Call rate 96.7%
WTCCC88803					Call rate 97.6%
WTCCC88816					Call rate 96.9%
WTCCC88853					Call rate 97.8%
WTCCC88879					Call rate 97.8%
WTCCC88893					Call rate 96.6%
WTCCC89008					Call rate 92.3%
WTCCC89095					Call rate 96.6%
WTCCC89105					Call rate 95.6%
WTCCC89110					Call rate 97.2%
WTCCC89179					Call rate 95.5%
WTCCC89181					Call rate 93.1%
WTCCC89185					Call rate 93.4%
WTCCC89202					Call rate 92%
WTCCC89206					Call rate 94%
WTCCC89308					Call rate 97.7%
WTCCC89620					Ambiguous sex
WTCCC67127					Ambiguous sex
22Q13203436					1st degree relative of 22Q13203460
WTCCC88789					WTCCC88825
WTCCC88814					WTCCC88931
WTCCC88874					WTCCC88894
WTCCC88924					WTCCC88851
WTCCC86105					Non-European ancestry
WTCCC66130					Non-European ancestry
WTCCC66147					Non-European ancestry
WTCCC66421					Non-European ancestry
WTCCC66438					Non-European ancestry
WTCCC66453					Non-European ancestry
WTCCC66588					Non-European ancestry
WTCCC66618					Non-European ancestry
WTCCC66684					Non-European ancestry
WTCCC66865					Non-European ancestry
WTCCC66883					Non-European ancestry
WTCCC66928					Non-European ancestry
WTCCC67155					Non-European ancestry
WTCCC88223					Non-European ancestry
WTCCC88265					Non-European ancestry
WTCCC88323					Non-European ancestry
WTCCC88393					Non-European ancestry
WTCCC89142					Non-European ancestry
WTCCC89186					Non-European ancestry
WTCCC89210					Non-European ancestry

Table 11.5: List of WTCCC 550K controls that failed quality control.

Cases	Genotyping Failure	Discordant Sex	Duplicated / Related	Divergent Ancestry	Reason for Exclusion	
DYS_369_3					Call rate 92.4%	
DYS_398_3					Call rate 55.9%	
DYS_BER_055					Call rate 95.0%	
DYS_BER_135					Call rate 83.0%	
DYS_BER_138					Call rate 88.9%	
DYS_BER_146					Call rate 97.9%	
DYS_65_3					Female when expected male	
DYS_BER_017_rep2					Male when expected female	
DYS_BER_041_4					Female when expected male	
DYS_BER_194					Male when expected female	
DYS_206_4					Call rate 58.1%	
DYS_BER_009					Call rate 70.6%	
DYS_BER_128					Call rate 68.1%	
DYS_BER_275					Call rate 89.0%	
DYS_109_4					Non-European Ancestry	
DYS_213_4					Non-European Ancestry	
DYS_301_4					Non-European Ancestry	
DYS_302_3					Non-European Ancestry	
DYS_338_3					Non-European Ancestry	
DYS_356_3					Non-European Ancestry	
DYS_356_4					Non-European Ancestry	
DYS_363_3					Non-European Ancestry	
DYS_363_5					Non-European Ancestry	
DYS_85_5					Non-European Ancestry	
DYS_AST_010					Non-European Ancestry	
DYS_AST_015					Non-European Ancestry	
DYS_AST_037					Non-European Ancestry	
DYS_AST_063					Non-European Ancestry	
DYS_BER_120_3					Non-European Ancestry	
DYS_BER_120_34					Non-European Ancestry	
DYS_BER_120_4					Non-European Ancestry	
DYS_BER_122					Non-European Ancestry	
DYS_BER_151					Non-European Ancestry	
DYS_BER_187					Non-European Ancestry	
DYS_BER_228					Non-European Ancestry	
DYS_BER_258					Non-European Ancestry	
DYS_BER_259					Non-European Ancestry	
DYS_BER_267					Non-European Ancestry	
DYS_BER_274					Non-European Ancestry	
DYS_107_4					sib of DYS_207_3	
DYS_120_3					Identical twin with DYS_120_4	
DYS_120_34					Identical twin with DYS_120_4	
DYS_207_3					sib of DYS_107_4	
DYS_55_3					Half Sib of 55_4 and 55_5	
DYS_BER_001					rep2 had a better call rate	
DYS_BER_001_rep1					rep2 had a better call rate	
DYS_BER_027					rep2 had a better call rate	
DYS_BER_027_rep1					rep2 had a better call rate	
DYS_BER_048					rep2 had a better call rate	
DYS_BER_048_rep1					rep2 had a better call rate	
DYS_BER_092					rep2 had a better call rate	
DYS_BER_092_rep1					rep2 had a better call rate	
DYS_BER_114					DYS_BER_127	
DYS_BER_114_rep1					DYS_BER_127	
DYS_BER_131					rep2 had a better call rate	
DYS_BER_131_rep1					rep2 had a better call rate	
DYS_BER_132					rep2 had a better call rate	
DYS_BER_132_rep1					rep2 had a better call rate	
DYS_BER_161					rep2 had a better call rate	
DYS_BER_161_rep1					rep2 had a better call rate	
DYS_BER_162					rep2 had a better call rate	
DYS_BER_162_rep1					rep2 had a better call rate	
DYS_BER_246					I mixed up with DYS_BER_247	
DYS_BER_247					I mixed up with DYS_BER_246	
DYS_179_4					sib of DYS_181_4	
DYS_BER_017					rep2 had a better call rate	
DYS_BER_017_rep1					rep2 had a better call rate	
DYS_BER_246_rep1					I mixed up with DYS_BER_247	
DYS_BER_246_rep2					I mixed up with DYS_BER_247	
DYS_BER_247_rep1					I mixed up with DYS_BER_246	
DYS_BER_247_rep2					I mixed up with DYS_BER_246	
DYS_321_4					Call Rate 80.3%	

Table 11.6: List of OmniExpress cases (cohort 2) and siblings that failed quality control.

Controls	Genotyping Failure	Discordant Sex	Duplicated / Related	Divergent Ancestry	Reason for Exclusion
FOD013					Call rate 96.9%
FOD021					Call rate 53.9%
FOD023					Call rate 91.3%
FOD059					Call rate 59.6%
NIR050					Call rate 74.5%
NOR088					Call rate 96.5%
RAN009					Call rate 80.9%
WAL012					Call rate 91.2%
CHE505					Call rate 90.8%
FOD048					Ambiguous when expected female
LIN625					Male when expected Female
LIN701					Female when expected male
LIN706					Male when expected Female
NOT519					Male when expected Female
NOT520					Female when expected male
ORK751					Ambiguous when expected female
RAN007					Male when expected Female
SOM003					Female when expected male
ARG376					Male when expected Female
FOD011					Call rate 76.8%
GLO055					Call rate 80.8%
LIN818					Call rate 68.6%
NOR090					Call rate 66.5%
IOM003					Non-European Ancestry
LAN551					ARG376 1st degree relative
COR606					COR293 1st degree relative
NEA575					RAN408 1st degree relative
RAN081					RAN080 1st degree relative
NPE041					NPE053 1st degree relative
NWA152					NWA148 1st degree relative
ORK717					COR638
ORK758					ORK757 1st degree relative
ORK776					ORK775 1st degree relative
ORK779					ORK780 1st degree relative

Table 11.7: List of POBI OmniExpress controls that failed quality control.

11.3 Most highly associated markers in each GWAS

Table 11.8: **A list of all SNPs with $P \leq 1 \times 10^{-4}$ from the 550K dyslexia case-control GWAS.** All the genes within 100kb to the left or right of the SNP are listed.

<i>Chr</i>	<i>rsnum</i>	<i>Position</i>	<i>OR</i>	<i>OR 95L</i>	<i>OR 95U</i>	<i>P</i>	<i>Gene</i>	<i>Left Genes</i>	<i>Right Genes</i>
3	rs838600	144698162	2.5	1.7	3.7	2.32×10^{-6}	SLC9A9	-	-
7	rs13232561	27611685	0.27	0.16	0.47	3.31×10^{-6}	HIBADH	-	-
21	rs240486	16694093	3	1.9	4.7	4.21×10^{-6}	LINC00478	-	-
2	rs17742513	211830373	2.5	1.7	3.6	5.44×10^{-6}	-	-	-
12	rs7975671	81962244	2.4	1.7	3.6	6.28×10^{-6}	TMTC2	-	-
8	rs2513980	98831608	2.5	1.7	3.7	6.9×10^{-6}	-	MTDH	LAPTM4B
10	rs11012119	20860835	0.15	0.068	0.35	7.67×10^{-6}	-	-	MIR4675
10	rs953093	20867111	0.15	0.068	0.35	7.75×10^{-6}	-	-	MIR4675
2	rs1836729	211883394	2.3	1.6	3.4	1.14×10^{-5}	-	-	ERBB4
16	rs4441292	78648869	2.9	1.8	4.6	1.34×10^{-5}	-	-	-
12	rs7304484	81971798	2.3	1.6	3.4	1.47×10^{-5}	TMTC2	-	-
10	rs7081156	20882037	0.1	0.036	0.29	1.58×10^{-5}	-	MIR4675	-
6	rs12203200	147893667	2.9	1.8	4.6	1.6×10^{-5}	SAMD5	-	-
10	rs11598854	20937909	0.11	0.04	0.3	1.77×10^{-5}	-	MIR4675	-
2	rs3816784	216993652	14	4.1	46	1.85×10^{-5}	SMARCA1	MARCH4	RPL37A
10	rs16920624	20850531	0.1	0.037	0.29	1.89×10^{-5}	-	-	MIR4675
5	rs39930	11393507	2.3	1.6	3.4	1.9×10^{-5}	CTNND2	-	-
7	rs2306848	129749650	0.13	0.049	0.33	2×10^{-5}	CPA4	CPA2	CPA1,CPA5,CEP41
2	rs16840590	133215749	2.5	1.6	3.9	2.18×10^{-5}	NCKAP5	GPR39,LYPD1	-
13	rs1001034	75105566	0.29	0.16	0.52	2.24×10^{-5}	LMO7	UCHL3,COMMD6	-
10	rs7910614	20848829	0.16	0.071	0.38	2.28×10^{-5}	-	-	MIR4675
19	rs2457785	21818119	4.7	2.3	9.7	2.31×10^{-5}	ZNF43	ZNF100,ZNF43,LOC641367	-
8	rs1398295	139351291	2.3	1.6	3.4	2.36×10^{-5}	FAM135B	-	-
5	rs3805715	41194246	0.25	0.13	0.47	2.39×10^{-5}	C6	HEATR7B2	-
5	rs32121	11403359	2.3	1.6	3.5	2.73×10^{-5}	CTNND2	-	-
5	rs32125	11397903	2.3	1.5	3.3	2.91×10^{-5}	CTNND2	-	-
10	rs2140380	68348217	3.7	2	6.8	3.09×10^{-5}	CTNNA3	-	LRRTM3
8	rs388769	17626684	0.39	0.25	0.61	3.65×10^{-5}	MTUS1	MTUS1,PDGFRL	-
1	rs12069869	167293554	0.34	0.21	0.57	3.84×10^{-5}	-	-	ATP1B1,NME7
14	rs8019477	78711561	2.3	1.5	3.4	3.95×10^{-5}	NRXN3	-	-
6	rs6908087	7562937	0.39	0.25	0.61	3.96×10^{-5}	-	DSP,SNRNP48	-
1	rs11118262	217545735	2.1	1.5	3	4.01×10^{-5}	-	LYPLAL1	-
3	rs9821051	144727698	2.2	1.5	3.1	4.16×10^{-5}	SLC9A9	-	-

3	rs9872790	144721564	2.2	1.5	3.2	4.27×10^{-5}	SLC9A9	-	-
2	rs4287713	211854014	2.3	1.5	3.3	5×10^{-5}	-	-	ERBB4
11	rs11034689	4904013	2.4	1.6	3.6	5.06×10^{-5}	-	OR51S1,OR51G2,OR51A7,OR51G1,OR51T1	OR51A4,OR51L1,OR51A2,MMP26
8	rs7460061	66540229	0.31	0.17	0.54	5.07×10^{-5}	-	-	LOC286186
2	rs2102101	227118593	0.35	0.21	0.58	5.12×10^{-5}	-	-	-
8	rs2468008	98812839	2.3	1.5	3.4	5.21×10^{-5}	-	MTDH	LAPTM4B
21	rs7509909	21388284	0.4	0.26	0.63	5.45×10^{-5}	NCAM2	-	-
12	rs12815690	81982661	0.43	0.29	0.65	5.71×10^{-5}	TMTC2	-	-
5	rs32138	11388373	2.3	1.5	3.4	5.88×10^{-5}	CTNND2	-	-
5	rs32128	11396384	2.2	1.5	3.2	6.07×10^{-5}	CTNND2	-	-
8	rs2013938	6463312	2.1	1.5	3.1	6.32×10^{-5}	MCPH1	ANGPT2	AGPAT5
2	rs1500993	227157219	0.35	0.21	0.58	6.37×10^{-5}	-	-	MIR5702
7	rs739898	27879585	0.35	0.21	0.59	6.8×10^{-5}	JAZF1	TAX1BP1	-
5	rs325845	41120483	0.27	0.15	0.52	7.12×10^{-5}	-	HEATR7B2	C6
5	rs161649	55225787	0.46	0.32	0.68	7.21×10^{-5}	IL31RA	DDX4	IL6ST
1	rs1344190	90136529	2.6	1.6	4.1	7.53×10^{-5}	LRRC8D	-	ZNF326,GEMIN8P4
1	rs1547502	217630980	0.45	0.3	0.67	7.62×10^{-5}	-	-	-
9	rs10981955	115631293	5.1	2.3	11	7.66×10^{-5}	-	-	ZNF618
8	rs10875421	139360276	2.2	1.5	3.3	7.98×10^{-5}	FAM135B	-	-
14	rs17109461	79105347	0.077	0.022	0.28	8.05×10^{-5}	NRXN3	-	-
2	rs2595214	46071174	0.39	0.24	0.62	8.24×10^{-5}	PRKCE	-	-
16	rs4533299	78626960	2.5	1.6	4	8.31×10^{-5}	-	-	-
14	rs7151991	31705323	2.4	1.5	3.6	8.53×10^{-5}	-	ARHGAP5,ARHGAP5-AS1	-
8	rs2568409	130486806	2.6	1.6	4.1	8.65×10^{-5}	-	-	-
5	rs17624740	90105434	3.5	1.9	6.5	8.74×10^{-5}	GPR98	-	-
2	rs10497933	211849925	0.36	0.22	0.6	8.75×10^{-5}	-	-	ERBB4
8	rs2912065	6478073	2.1	1.5	3.1	9.1×10^{-5}	MCPH1	ANGPT2	AGPAT5
21	rs9979231	21399144	0.38	0.23	0.61	9.4×10^{-5}	NCAM2	-	-
3	rs512541	23018041	0.48	0.33	0.69	9.46×10^{-5}	-	-	-
8	rs4735795	424668	0.44	0.29	0.66	9.53×10^{-5}	-	FBXO25	C8orf42

Table 11.9: A list of all SNPs with $P \leq 1 \times 10^{-4}$ from the Illumina OmniExpress dyslexia case-control GWAS. All the genes within 100kb to the left or right of the SNP are listed.

<i>Chr</i>	<i>rsnum</i>	<i>Position</i>	<i>OR</i>	<i>OR 95L</i>	<i>OR 95U</i>	<i>P value</i>	<i>Gene</i>	<i>Left Genes</i>	<i>Right Genes</i>
18	rs8083477	8517847	2	1.5	2.6	6.18×10^{-7}	-	-	RAB12
18	rs7241049	8516997	1.9	1.5	2.5	2.05×10^{-6}	-	-	RAB12
22	rs2413204	34227271	2	1.5	2.7	2.53×10^{-6}	LARGE	LOC100506195	-
22	rs1002048	34253393	2	1.5	2.7	2.88×10^{-6}	LARGE	-	-
18	rs1426310	28898294	1.8	1.4	2.2	5.95×10^{-6}	DSG1	-	DSG4
8	rs8218	99114970	1.9	1.4	2.4	6.89×10^{-6}	HRSP12	RPL30,C8orf47,SNORA72,MATN2	NIPAL2,POP1
18	rs1426309	28897353	1.7	1.4	2.2	7.55×10^{-6}	-	-	DSG4,DSG1
18	rs1426311	28898567	1.7	1.4	2.2	7.56×10^{-6}	DSG1	-	DSG4
18	rs12967407	28913599	1.7	1.4	2.2	7.66×10^{-6}	DSG1	-	DSG4
18	rs2199301	28910615	1.7	1.4	2.2	7.66×10^{-6}	DSG1	-	DSG4
6	rs2006798	1316451	0.54	0.41	0.71	8.34×10^{-6}	-	FOXQ1	FOXQ2
1	rs2066108	216877492	2.1	1.5	2.9	9.2×10^{-6}	ESRRG	-	-
13	rs4119478	73254212	1.8	1.4	2.3	1.09×10^{-5}	-	-	BORA,DIS3,MZT1
4	rs16852239	40477729	0.53	0.39	0.7	1.19×10^{-5}	RBM47	-	MIR4802
7	rs41334	28703795	1.8	1.4	2.4	1.32×10^{-5}	CREB5	-	CREB5
7	rs850569	145384643	1.6	1.3	2.1	1.73×10^{-5}	-	-	-
13	rs9564996	74039264	0.49	0.36	0.68	2.01×10^{-5}	-	-	-
10	rs11191659	105101701	2	1.4	2.7	2.08×10^{-5}	PCGF6	INA,LOC729020	TAF5,PDCD11,MIR1307,USMG5
8	rs2469367	3448547	0.62	0.5	0.77	2.17×10^{-5}	CSMD1	-	-
5	rs17211996	144100737	2.6	1.7	4.1	2.2×10^{-5}	-	-	-
18	rs9952617	28881801	0.6	0.47	0.76	2.29×10^{-5}	-	-	DSG4,DSG1
13	rs9543067	73256676	1.8	1.4	2.3	2.33×10^{-5}	-	-	PIBF1,BORA,DIS3,MZT1
18	rs1966760	8525350	1.7	1.3	2.2	2.48×10^{-5}	-	-	RAB12
18	rs2042720	28838984	1.7	1.3	2.1	2.59×10^{-5}	-	DSC1	DSG1
21	rs11910489	18046904	6.7	2.8	16	2.76×10^{-5}	-	MIR125B2,LINC00478	-
18	rs7240897	28816434	1.7	1.3	2.1	2.86×10^{-5}	-	DSC1	DSG1
4	rs278959	40488449	0.57	0.43	0.74	3.04×10^{-5}	RBM47	-	MIR4802
14	rs873915	51472624	1.7	1.3	2.2	3.55×10^{-5}	TRIM9	PYGL	-
12	rs1532395	117186316	0.53	0.39	0.72	3.73×10^{-5}	RNFT2	C12orf49	-
12	rs1163929	17602872	1.8	1.3	2.3	3.76×10^{-5}	-	-	-
18	rs11663134	28869498	1.7	1.3	2.1	3.81×10^{-5}	-	-	DSG4,DSG1
13	rs9543080	73268372	1.7	1.3	2.1	3.84×10^{-5}	-	-	PIBF1,BORA,DIS3,MZT1
18	rs7242742	28820025	1.7	1.3	2.1	4.22×10^{-5}	-	DSC1	DSG1

12	rs4907432	132011715	0.46	0.31	0.67	4.36×10^{-5}	-	-	-
10	rs16915736	33884138	1.6	1.3	2	4.5×10^{-5}	-	-	-
18	rs9950233	8540983	1.6	1.3	2.1	4.58×10^{-5}	-	-	RAB12
18	rs7229974	28890717	0.61	0.48	0.77	4.76×10^{-5}	-	-	DSG4,DSG1
14	rs12431994	52312518	1.9	1.4	2.5	4.77×10^{-5}	-	-	GNG2
21	rs220159	43547873	0.61	0.48	0.77	4.81×10^{-5}	UMODL1	C21orf128	ABCG1
18	rs1866104	28636005	0.62	0.49	0.78	4.84×10^{-5}	-	DSC3	DSC1,DSC2
11	rs2034691	68810837	0.36	0.22	0.59	5.06×10^{-5}	-	MRGPRD,MRGPRF	TPCN2
14	rs1877044	45495763	1.6	1.3	2.1	5.09×10^{-5}	FAM179B	KLHL28	FKBP3,PRPF39,SNORD127
13	rs2479563	27600884	0.61	0.48	0.78	5.19×10^{-5}	-	-	USP12
6	rs11242674	1307846	0.61	0.48	0.78	5.24×10^{-5}	-	-	FOXQ1,FOXF2
1	rs10465691	211920624	1.8	1.4	2.4	5.3×10^{-5}	LPGAT1	NEK2	-
1	rs11119806	211921229	1.8	1.4	2.4	5.3×10^{-5}	LPGAT1	NEK2	-
13	rs9573559	76052329	1.7	1.3	2.2	5.43×10^{-5}	TBC1D4	-	UCHL3,COMMD6
14	rs930988	25219124	0.61	0.48	0.78	5.55×10^{-5}	-	-	STXBP6
18	rs12964839	28809552	1.6	1.3	2.1	5.55×10^{-5}	-	DSC1	DSG1
18	rs9946611	28814294	1.6	1.3	2.1	5.55×10^{-5}	-	DSC1	DSG1
13	rs10507815	74039881	0.45	0.3	0.66	5.72×10^{-5}	-	-	-
20	rs6064566	56146796	0.6	0.47	0.77	5.73×10^{-5}	-	PCK1,CTCFL	PMEPA1,ZBP1
8	rs10088925	99128077	1.7	1.3	2.2	5.76×10^{-5}	HRSP12	RPL30,C8orf47,SNORA72,MATN2	NIPAL2,POP1
1	rs10798774	180516881	1.8	1.3	2.3	5.82×10^{-5}	-	ACBD6	XPR1
13	rs9518128	101321244	1.6	1.3	2.1	5.95×10^{-5}	TMTC4	A2LD1	NALCN-AS1
11	rs573264	114092582	0.63	0.5	0.79	6.28×10^{-5}	ZBTB16	-	NNMT
12	rs12423929	104146343	2.4	1.5	3.6	6.38×10^{-5}	STAB2	-	NT5DC3,GNN
19	rs10422807	1262978	0.51	0.37	0.71	6.55×10^{-5}	-	C19orf26,ATP5D,MIDN,STK11,SBNO2	EFNA2,CIRBP-AS1,MUM1,CIRBP,C19orf24
6	rs12213197	90673538	1.6	1.3	2	6.55×10^{-5}	BACH2	CASP8AP2,GJA10	-
18	rs1126214	28639548	0.62	0.49	0.79	6.66×10^{-5}	-	DSC3	DSC1,DSC2
1	rs7538424	216829908	1.8	1.4	2.5	6.93×10^{-5}	ESRRG	-	-
1	rs7550534	216839426	1.8	1.4	2.5	6.93×10^{-5}	ESRRG	-	-
1	rs946468	216836576	1.8	1.4	2.5	6.93×10^{-5}	ESRRG	-	-
7	rs10240570	71954779	2.7	1.6	4.3	6.97×10^{-5}	-	CALN1	TYW1B
18	rs966376	28832838	1.6	1.3	2.1	6.99×10^{-5}	-	DSC1	DSG1
21	rs225410	43700907	1.6	1.3	2	7.01×10^{-5}	ABCG1	-	TFF1,TMPRSS3,TFF2,TFF3
6	rs9479087	151943014	0.55	0.41	0.74	7.53×10^{-5}	-	CCDC170	ESR1
10	rs1922139	53363713	1.6	1.3	2.1	7.64×10^{-5}	PRKG1	-	CSTF2T
18	rs10468901	28921297	0.61	0.48	0.78	7.74×10^{-5}	DSG1	-	DSG4
9	rs7047997	5703394	0.54	0.39	0.73	7.74×10^{-5}	KIAA1432	-	ERMP1

2	rs340740	16469823	1.6	1.3	2	8.13×10^{-5}	-	-	-
6	rs9689020	1335969	0.62	0.48	0.78	8.49×10^{-5}	-	FOXQ1	FOXF2
18	rs9954315	28873403	1.6	1.3	2.1	8.62×10^{-5}	-	-	DSG4,DSG1
5	rs1863982	67768380	0.56	0.42	0.75	8.65×10^{-5}	-	-	-
8	rs12541730	41007462	1.7	1.3	2.2	8.74×10^{-5}	-	-	-
16	rs16951383	79816194	0.55	0.4	0.74	8.86×10^{-5}	-	-	-
2	rs6546674	71334998	0.57	0.43	0.76	8.96×10^{-5}	-	NAGK,OR7E91P	MCEE,MPHOSPH10,PAIP2B
7	rs1543989	156327884	0.61	0.47	0.78	9.4×10^{-5}	-	LOC285889	LINC00244
9	rs2760415	5696021	0.53	0.38	0.73	9.69×10^{-5}	KIAA1432	-	ERMP1
18	rs7230461	28823765	1.6	1.3	2	9.76×10^{-5}	-	DSC1	DSG1
21	rs492338	43701977	1.6	1.2	2	9.84×10^{-5}	ABCG1	-	TFF1,TMPRSS3,TFF2,TFF3
1	rs7543366	180506038	1.6	1.3	2.1	9.86×10^{-5}	-	MIR3121,ACBD6	XPR1
4	rs12509327	11156010	0.5	0.35	0.71	9.86×10^{-5}	-	-	-
4	rs7659670	11175752	0.5	0.35	0.71	9.86×10^{-5}	-	-	-
2	rs990276	78542946	0.6	0.46	0.78	9.95×10^{-5}	-	-	-
2	rs13012094	16631289	1.6	1.3	2.1	9.97×10^{-5}	-	-	FAM49A

Table 11.10: A list of the 25 most highly associated SNPs from the Illumina 550K dyslexia case-control GWAS after imputation. All the genes within 100kb to the left or right of the SNP are listed. An _i next to the chromosome number indicates that the SNP is imputed.

<i>Chr</i>	<i>Variant</i>	<i>Position</i>	<i>case MAF</i>	<i>control MAF</i>	<i>OR</i>	<i>OR 95L</i>	<i>OR 95U</i>	<i>P value</i>	<i>Info</i>	<i>Gene</i>	<i>Left Genes</i>	<i>Right Genes</i>
10_i	rs138057753	20846666	0.018	0.053	0.34	0.16	0.71	3.57×10^{-7}	0.82	-	MIR4675	-
3_i	rs11922395	150159417	0.031	0.013	2.5	1.3	5	6.68×10^{-7}	0.84	TSC22D2	-	-
11_i	MERGED_DEL_2_63353	4968197	0.45	0.4	1.2	0.99	1.5	7.56×10^{-7}	0.93	OR51A4	OR51S1,OR51T1,OR51G2,OR51A7,OR51G1	OR51L1,OR52J3,OR51A2,MMP26
21_i	rs240487	17772777	0.25	0.18	1.5	1.2	2	1.05×10^{-6}	0.92	LINC00478	-	-
2_i	rs12623753	212129606	0.49	0.35	1.8	1.5	2.3	1.21×10^{-6}	0.99	-	-	-
2_i	rs137900157	212132713	0.49	0.35	1.8	1.5	2.3	1.21×10^{-6}	0.99	-	-	-
2_i	rs75339535	212132803	0.49	0.35	1.8	1.5	2.3	1.21×10^{-6}	0.99	-	-	-
10_i	rs55841737	20887242	0.046	0.1	0.43	0.26	0.69	1.25×10^{-6}	1	-	MIR4675	-
10_i	chr10:20835199:D	20835199	0.025	0.065	0.37	0.19	0.7	1.36×10^{-6}	0.98	-	-	MIR4675
7_i	chr7:27653025:D	27653025	0.14	0.18	0.72	0.53	0.98	1.44×10^{-6}	0.95	HIBADH	-	-
8_i	rs2513979	98757391	0.41	0.37	1.2	0.95	1.5	1.46×10^{-6}	0.96	-	MTDH	LAPTM4B
14_i	rs28723500	83435306	0.1	0.17	0.56	0.4	0.79	1.57×10^{-6}	0.98	-	-	-
3	rs838600	143215472	0.57	0.46	1.6	1.3	2	1.58×10^{-6}	1	SLC9A9	-	-
1_i	rs116114323	212594063	0.031	0.014	2.3	1.2	4.6	1.61×10^{-6}	0.84	-	TMEM206,SNORA16B,PPP2R5A	NENF
9_i	rs11790623	22184585	0.0066	0.035	0.18	0.054	0.62	1.73×10^{-6}	0.87	-	CDKN2B-AS1	-
2_i	rs62178255	212133729	0.51	0.36	1.8	1.5	2.3	1.88×10^{-6}	0.99	-	-	-
10_i	rs186924582	20849607	0.022	0.058	0.37	0.19	0.73	1.89×10^{-6}	0.97	-	MIR4675	-
5_i	rs41077	11344287	0.49	0.42	0.76	0.61	0.94	1.91×10^{-6}	0.96	CTNND2	-	-
4_i	rs148695918	9012083	0.0072	0.013	0.57	0.17	1.9	1.94×10^{-6}	0.88	-	LOC650293,LOC100288392	-
8_i	rs7463317	66402871	0.11	0.18	1.7	1.2	2.3	2.08×10^{-6}	0.96	-	-	LOC286186
10_i	rs10159544	20820909	0.042	0.097	0.4	0.24	0.67	2.09×10^{-6}	1	-	-	MIR4675
10_i	rs12247928	20820721	0.042	0.097	0.4	0.24	0.67	2.09×10^{-6}	1	-	-	MIR4675
10_i	rs7092233	20831779	0.042	0.097	0.41	0.24	0.67	2.09×10^{-6}	1	-	-	MIR4675
10	rs11012119	20820829	0.042	0.098	0.4	0.24	0.66	2.09×10^{-6}	1	-	-	MIR4675
10_i	rs12241173	20820467	0.042	0.097	0.4	0.24	0.67	2.1×10^{-6}	1	-	-	MIR4675

Table 11.11: **A list of the 25 most highly associated SNPs from the Illumina OmniExpress dyslexia case-control GWAS after imputation.** All the genes within 100kb to the left or right of the SNP are listed. An _i next to the chromosome number indicates that the SNP is imputed.

<i>Chr</i>	<i>Variant</i>	<i>Position</i>	<i>case MAF</i>	<i>control MAF</i>	<i>OR</i>	<i>OR 95L</i>	<i>OR 95U</i>	<i>P value</i>	<i>Info</i>	<i>Gene</i>	<i>Left Genes</i>	<i>Right Genes</i>
18	rs8083477	8517847	0.32	0.2	1.8	1.4	2.4	8.76×10^{-8}	1	-	-	RAB12
18_i	rs8083486	8517859	0.32	0.2	1.8	1.4	2.4	8.78×10^{-8}	1	-	-	RAB12
18_i	rs6506591	8517515	0.32	0.2	1.8	1.4	2.4	1.07×10^{-7}	1	-	-	RAB12
18_i	rs16953675	8521071	0.32	0.2	1.9	1.4	2.4	1.26×10^{-7}	1	-	-	RAB12
18_i	rs11663377	8535676	0.31	0.2	1.9	1.4	2.4	1.39×10^{-7}	0.99	-	-	RAB12
18_i	rs11081392	8523490	0.31	0.2	1.8	1.4	2.3	1.94×10^{-7}	1	-	-	RAB12
18_i	rs7227856	8517147	0.31	0.2	1.8	1.4	2.3	2.23×10^{-7}	1	-	-	RAB12
18_i	rs1372779	8525618	0.31	0.2	1.8	1.4	2.3	2.39×10^{-7}	1	-	-	RAB12
18_i	rs72927113	8516944	0.31	0.2	1.8	1.4	2.3	2.67×10^{-7}	1	-	-	RAB12
18_i	rs1470247	8525169	0.31	0.2	1.8	1.4	2.3	2.69×10^{-7}	1	-	-	RAB12
18_i	rs7227112	8517049	0.31	0.2	1.8	1.4	2.3	2.69×10^{-7}	1	-	-	RAB12
18	rs7241049	8516997	0.31	0.2	1.8	1.4	2.3	2.69×10^{-7}	1	-	-	RAB12
18_i	rs17581114	8518846	0.31	0.2	1.8	1.4	2.3	2.7×10^{-7}	1	-	-	RAB12
18_i	rs59286512	8521075	0.31	0.2	1.8	1.4	2.3	2.7×10^{-7}	1	-	-	RAB12
18_i	rs62085132	8518985	0.31	0.2	1.8	1.4	2.3	2.7×10^{-7}	1	-	-	RAB12
18_i	rs7227270	8517877	0.31	0.2	1.8	1.4	2.3	2.7×10^{-7}	1	-	-	RAB12
18_i	rs11081393	8524099	0.31	0.2	1.8	1.4	2.3	2.71×10^{-7}	1	-	-	RAB12
18_i	rs62085137	8537880	0.32	0.21	1.8	1.4	2.3	5.79×10^{-7}	0.98	-	-	RAB12
1_i	rs3843275	180530607	0.33	0.22	0.56	0.43	0.71	5.84×10^{-7}	0.92	-	ACBD6	XPR1
18_i	rs17496786	8537404	0.31	0.2	1.8	1.4	2.3	6.97×10^{-7}	0.99	-	-	RAB12
18_i	rs11665236	8537779	0.32	0.21	1.8	1.4	2.3	7.78×10^{-7}	0.98	-	-	RAB12
1_i	rs4652514	180529542	0.33	0.22	0.56	0.44	0.72	8.09×10^{-7}	0.93	-	ACBD6	XPR1
18_i	rs11665212	8537623	0.31	0.2	1.8	1.4	2.3	8.47×10^{-7}	0.99	-	-	RAB12
6_i	chr6:1315019:I	1315019	0.19	0.3	1.9	1.4	2.4	8.48×10^{-7}	0.99	-	FOXQ1	FOXF2
6	rs2006798	1316451	0.19	0.3	1.9	1.4	2.4	8.63×10^{-7}	1	-	FOXQ1	FOXF2

Table 11.12: **A list of SNPs with $P \leq 5 \times 10^{-6}$ from the combined dyslexia case-control GWA data-sets.** All the genes within 100kb to the left or right of the SNP are listed. The number of samples tested includes the combined number of cases and controls for both studies. ++ indicates that the effect allele is associated with an increased risk for dyslexia in both studies, and -- indicates the effect allele is protective against dyslexia.

<i>Chr</i>	<i>Position</i>	<i>Variant</i>	<i>Effect Allele Frequency</i>	<i>OR</i>	<i>OR 95L</i>	<i>OR 95U</i>	<i>SE</i>	<i>P value</i>	<i>Number of Samples</i>	<i>Effects</i>	<i>Gene</i>	<i>Left Genes</i>	<i>Right Genes</i>
6	1555810	rs9502966	0.12	2.3	1.6	3.1	0.31	5.39×10^{-7}	1903	++	-	-	GMDS,FOXC1
1	216877492	rs2066108	0.82	0.52	0.4	0.68	0.062	1.38×10^{-6}	1905	--	ESRRG	-	-
1	216876964	rs78851238	0.18	1.9	1.5	2.5	0.23	1.45×10^{-6}	1904	++	ESRRG	-	-
1	216877312	rs2813666	0.82	0.52	0.4	0.68	0.062	1.45×10^{-6}	1904	--	ESRRG	-	-
1	216876843	rs17043436	0.18	1.9	1.5	2.5	0.23	1.46×10^{-6}	1905	++	ESRRG	-	-
1	216878133	chr1:216878133:D	0.18	1.9	1.5	2.5	0.23	1.47×10^{-6}	1903	++	ESRRG	-	-
1	216878297	rs2789722	0.82	0.52	0.4	0.68	0.062	1.47×10^{-6}	1905	--	ESRRG	-	-
1	216883063	rs61817046	0.17	2	1.5	2.6	0.24	1.48×10^{-6}	1903	++	ESRRG	-	-
1	216876475	rs2820882	0.82	0.52	0.4	0.68	0.062	1.53×10^{-6}	1904	--	ESRRG	-	-
1	216879911	rs2813711	0.83	0.51	0.39	0.67	0.062	1.63×10^{-6}	1905	--	ESRRG	-	-
1	216879359	rs2999503	0.82	0.52	0.4	0.68	0.062	1.77×10^{-6}	1904	--	ESRRG	-	-
1	216880104	rs2789721	0.82	0.52	0.4	0.68	0.063	1.88×10^{-6}	1903	--	ESRRG	-	-
1	216880168	rs2820881	0.82	0.52	0.4	0.68	0.063	1.88×10^{-6}	1903	--	ESRRG	-	-
1	216878845	rs17043466	0.18	1.9	1.5	2.5	0.23	1.92×10^{-6}	1903	++	ESRRG	-	-
1	216887443	rs61817047	0.18	1.9	1.5	2.6	0.24	1.95×10^{-6}	1903	++	ESRRG	-	-
1	216888789	rs74642143	0.18	1.9	1.5	2.6	0.24	1.98×10^{-6}	1903	++	ESRRG	-	-
17	54771370	rs9891351	0.022	0.19	0.097	0.38	0.049	2.37×10^{-6}	1903	--	-	NOG	C17orf67
2	76880944	rs7576696	0.74	0.55	0.43	0.71	0.062	2.63×10^{-6}	1903	--	-	-	LRRTM4
4	187000815	rs35990575	0.25	1.8	1.4	2.2	0.19	3.26×10^{-6}	1903	++	TLR3	-	FAM149A
9	138577916	rs515074	0.28	1.8	1.4	2.2	0.19	3.27×10^{-6}	1903	++	-	LCN9,LOC100130954,GLT6D1	SOHLH1,KCNT1
1	216903512	rs61819764	0.17	1.9	1.5	2.5	0.23	3.83×10^{-6}	1904	++	ESRRG	ESRRG	-
1	69057627	rs2206780	0.29	1.7	1.3	2.1	0.17	4.29×10^{-6}	1904	++	-	DEPDC1	-
1	69058576	rs9436401	0.29	1.7	1.3	2.1	0.17	4.99×10^{-6}	1904	++	-	DEPDC1	-

Table 11.13: **A list of SNPs with $P \leq 1 \times 10^{-5}$ from the CCI GWA meta-analysis.** All the genes within 100kb to the left or right of the SNP are listed. The number of samples tested includes the combined number of cases and controls for both studies. ++ indicates that the effect allele is associated with an increased CCI score in both studies, and -- indicates the effect allele is associated with an decreased CCI score.

Chr	Position	Variant	Effect Allele Frequency	β	β 95L	β 95U	SE	P value	Number of Samples	Effects	Gene	Left Genes	Right Genes
12	129940481	rs10773664	0.45	0.26	0.16	0.36	0.052	4.84×10^{-7}	573	++	TMEM132D	-	-
2	205733879	rs78578260	0.12	0.4	0.24	0.56	0.081	7.7×10^{-7}	572	++	PARD3B	-	-
1	95065538	rs1747659	0.35	0.26	0.15	0.36	0.054	1.31×10^{-6}	571	++	-	ABCD3,F3	-
1	95057013	rs1408341	0.4	0.25	0.15	0.36	0.053	1.34×10^{-6}	572	++	-	ABCD3,F3	-
1	95065869	rs841709	0.65	-0.26	-0.36	-0.15	0.054	1.34×10^{-6}	572	--	-	ABCD3,F3	-
2	169070298	rs6751410	0.59	0.25	0.14	0.35	0.053	3.55×10^{-6}	571	++	STK39	-	-
7	134214293	rs706180	0.33	-0.27	-0.38	-0.15	0.058	4.34×10^{-6}	571	--	AKR1B10	AKR1B1	AKR1B15
9	85518592	rs12552755	0.064	0.51	0.29	0.72	0.11	4.47×10^{-6}	572	++	-	-	RASEF
1	95062959	rs1747653	0.34	0.25	0.14	0.35	0.054	5.51×10^{-6}	571	++	-	ABCD3,F3	-
7	134198377	rs67604615	0.28	-0.27	-0.39	-0.15	0.059	6.11×10^{-6}	571	--	-	AKR1B1	AKR1B10,AKR1B15
11	120086918	rs499434	0.86	0.34	0.19	0.49	0.075	6.23×10^{-6}	571	++	OAF	TRIM29	LOC649133,POU2F3
8	67554357	rs11787419	0.056	0.51	0.29	0.73	0.11	6.92×10^{-6}	573	++	VCPIP1	MYBL1	C8orf44,SGK3,C8orf44-SGK3
2	169070337	rs55705213	0.4	-0.24	-0.34	-0.13	0.053	7.63×10^{-6}	571	--	STK39	-	-
3	34987073	rs11924798	0.68	-0.24	-0.35	-0.14	0.055	8.55×10^{-6}	572	--	-	-	-
14	86410463	rs2153078	0.5	-0.25	-0.36	-0.14	0.056	8.94×10^{-6}	571	--	-	-	-
12	13046125	rs10772610	0.69	0.25	0.14	0.36	0.056	9.08×10^{-6}	573	++	GPRC5A	RPL13AP20,DDX47	GPRC5D,HEBP1,MIR614,LOC100506314
2	169071034	rs11897414	0.34	-0.24	-0.34	-0.13	0.054	9.12×10^{-6}	572	--	STK39	-	-
1	95059479	rs1778211	0.27	0.26	0.14	0.37	0.058	9.33×10^{-6}	572	++	-	ABCD3,F3	-

Table 11.14: A list of SNPs with $P \leq 1 \times 10^{-5}$ from the single-word reading GWA meta-analysis. All the genes within 100kb to the left or right of the SNP are listed. The number of samples tested includes the combined number of cases and controls for both studies. ++ indicates that the effect allele is associated with an increased reading score in both studies, and -- indicates the effect allele is associated with an decreased reading score.

Chr	Position	Variant	Effect Allele Frequency	β	β 95L	β 95U	SE	P value	Number of Samples	Effects	Gene	Left Genes	Right Genes
3	122779147	chr3:122779147:D	0.12	-0.32	-0.46	-0.18	0.07	4.82×10^{-6}	577	--	-	SEMA5B	PDIA5
3	122779141	chr3:122779141:D	0.12	-0.32	-0.46	-0.18	0.07	4.91×10^{-6}	577	--	-	SEMA5B	PDIA5
1	237833352	rs57749080	0.36	0.21	0.12	0.3	0.046	5.64×10^{-6}	577	++	RYR2	-	-
1	237834701	rs2012813	0.39	0.2	0.11	0.28	0.044	6.28×10^{-6}	578	++	RYR2	-	-
11	120086918	rs499434	0.86	0.27	0.15	0.39	0.061	6.9×10^{-6}	578	++	OAF	TRIM29	POU2F3,LOC649133
22	29710360	rs13056977	0.17	-0.25	-0.37	-0.14	0.057	7.88×10^{-6}	579	--	RASL10A	4 genes	MIR3653,SNORD125,AP1B1
22	29753322	rs5752902	0.18	-0.25	-0.36	-0.14	0.056	9.17×10^{-6}	577	--	AP1B1	7 genes	RFPL1,RFPL1-AS1
17	7804405	rs62062524	0.16	-0.27	-0.4	-0.15	0.062	9.79×10^{-6}	577	--	CHD3	5 genes	5 genes

Table 11.15: A list of SNPs with $P \leq 1 \times 10^{-5}$ from the single-word spelling GWA meta-analysis. For an explanation see Table 11.14.

Chr	Position	Variant	Effect Allele Frequency	β	β 95L	β 95U	SE	P value	Number of Samples	Effects	Gene	Left Genes	Right Genes
2	169095090	rs10930316	0.29	-0.21	-0.29	-0.13	0.042	7.47×10^{-7}	569	--	STK39	-	-
2	169096848	rs1517324	0.29	-0.21	-0.29	-0.12	0.042	1.07×10^{-6}	569	--	STK39	-	-
2	169094179	rs72884536	0.28	-0.21	-0.29	-0.12	0.043	1.15×10^{-6}	569	--	STK39	-	-
2	169091327	chr2:169091327:D	0.29	-0.2	-0.28	-0.12	0.042	1.79×10^{-6}	569	--	STK39	-	-
2	169095296	rs1400635	0.29	-0.2	-0.28	-0.12	0.042	1.82×10^{-6}	568	--	STK39	-	-
13	100302314	rs4772226	0.44	0.19	0.11	0.27	0.041	3.56×10^{-6}	568	++	CLYBL	TM9SF2,MIR4306	-
13	55428734	rs77690573	0.09	-0.32	-0.45	-0.18	0.07	5.92×10^{-6}	568	--	-	-	-
13	100287551	rs4418926	0.41	0.18	0.1	0.26	0.04	6.26×10^{-6}	570	++	CLYBL	TM9SF2	MIR4306
23	86987258	rs6521964	0.11	0.22	0.12	0.31	0.048	6.88×10^{-6}	563	++	-	KLHL4	-
13	100269204	rs9634517	0.38	0.18	0.1	0.26	0.041	7.02×10^{-6}	569	++	CLYBL	TM9SF2	MIR4306
3	142807143	chr3:142807143:I	0.17	0.25	0.14	0.36	0.056	8.46×10^{-6}	568	++	-	U2SURP,LOC100289361	CHST2
2	222326174	chr2:222326174:D	0.7	0.2	0.11	0.29	0.046	8.93×10^{-6}	568	++	EPHA4	-	-

11.4 Published papers

To date I have three published first-name author papers, which are appended to the end of the thesis.

Scerri, T.S.*, **Brandler, W.M.***, Paracchini, S.*, et al. (2011). *PCSK6* is Associated with Handedness in Individuals with Dyslexia. **Human Molecular Genetics** 20(3): 608-614 (*co-first author).

Brandler, W.M., Morris, A.P., Evans, D.M., Scerri, T.S., Kemp, J.P., et al. (2013). Common Variants in Left/Right Asymmetry Genes and Pathways Are Associated with Relative Hand Skill. **PLoS Genetics** 9(9): e1003751.

doi:10.1371/journal.pgen.1003751

Brandler, W.M., Paracchini, S. (2013). The genetic relationship between handedness and neurodevelopmental disorders. **Trends in Molecular Medicine**.

doi:10.1016/j.molmed.2013.10.008

PCSK6 is associated with handedness in individuals with dyslexia

Thomas S. Scerri^{1,†}, William M. Brandler^{1,†}, Silvia Paracchini^{1,†}, Andrew P. Morris¹, Susan M. Ring², Alex J. Richardson⁵, Joel B. Talcott³, John Stein⁴ and Anthony P. Monaco^{1,*}

¹Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK, ²Department of Social Medicine, University of Bristol, Bristol BS8 2BN, UK, ³School of Life and Health Sciences, Aston University, Birmingham, B4 7ET, UK, ⁴Department of Physiology, University of Oxford, Parks Road, Oxford, OX1 3PT, UK and ⁵Department of Social Policy and Social Work, University of Oxford, OX1 2ER, UK

Received August 23, 2010; Revised and Accepted November 1, 2010

Approximately 90% of humans are right-handed. Handedness is a heritable trait, yet the genetic basis is not well understood. Here we report a genome-wide association study for a quantitative measure of relative hand skill in individuals with dyslexia [reading disability (RD)]. The most highly associated marker, rs11855415 ($P = 4.7 \times 10^{-7}$), is located within *PCSK6*. Two independent cohorts with RD show the same trend, with the minor allele conferring greater relative right-hand skill. Meta-analysis of all three RD samples is genome-wide significant ($n = 744$, $P = 2.0 \times 10^{-8}$). Conversely, in the general population ($n = 2666$), we observe a trend towards reduced laterality of hand skill for the minor allele ($P = 0.0020$). These results provide molecular evidence that cerebral asymmetry and dyslexia are linked. Furthermore, *PCSK6* is a protease that cleaves the left–right axis determining protein NODAL. Functional studies of *PCSK6* promise insights into mechanisms underlying cerebral lateralization and dyslexia.

INTRODUCTION

The population bias in right-handedness is a characteristic feature of humans (1). There is some variance across cultures (2,3); however, there are no known populations in the world in which left-handers are the majority.

Handedness is correlated with cerebral asymmetries and right-handedness implies a dominance of the left hemisphere for motor function. Since Paul Broca reported in 1861 the case of a patient who had aphasia caused by a lesion in the left hemisphere (4), there has been significant interest in the idea that language laterality and handedness are linked. It has been proposed that handedness emerged as a consequence of the evolution of language (5). Moreover, theories dating back almost a century posit a connection between handedness and neurodevelopmental disorders such as specific language impairment (SLI) and dyslexia (6), two disorders affecting language and reading skills, respectively, with a prevalence between 5 and 10% (7).

However, no convincing association has been found between either hand preference or hand skill and neurodevelopmental disorders (8,9). Functional brain imaging studies have shown a weak correlation between handedness and cerebral dominance for language; with 96% of strong right-handers, compared with 73% of strong left-handers, showing left-hemisphere dominance for language (10).

There is suggestive, but mixed, evidence indicating that there may be atypical cerebral asymmetry in patients with dyslexia or SLI (reviewed in 11,12). With respect to reading disability (RD), Galaburda *et al.* (13,14) reported eight consecutive post-mortem specimens with RD that had reduced planum temporale asymmetry. Subsequent structural MRI studies have been inconsistent in replicating this finding (reviewed in 11). However, a meta-analysis of 17 functional neuroimaging studies comparing RD individuals to controls during reading tasks suggests that underactivation of the left hemisphere is found in the inferior parietal, superior temporal, middle and inferior temporal and

*To whom correspondence should be addressed at: Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK. Tel: +44 1865287503; Fax: +44 1865287650; Email: anthony.monaco@well.ox.ac.uk

[†]These authors contributed equally.

fusiform regions (15). Conversely, there were no abnormalities of activation observed in the right hemisphere or cerebellum (15). One study included in this meta-analysis demonstrated that adults with RD have reduced activation of the left middle and inferior temporal regions during reading tasks (16). The same authors subsequently showed using voxel-based morphometry that these same regions show cortical structural disorganization (17), providing a link between functional and structural differences of the left hemisphere in individuals with RD.

Understanding the genetic basis of handedness can help define the relationships between handedness, language, cerebral asymmetry and neurodevelopmental disorders. Handedness is a heritable trait, with additive genetic effects accounting for about a quarter of the variance (9,18). Molecular studies are consistent with a polygenic model of handedness. Linkage analyses have identified several loci including 2p12-q11 (19,20), 10q26 (21) and 12q21-23 (22). To date, only two specific genes have been suggested as candidates for handedness. The imprinted gene *LRRTM1* in 2p12-q11 has been associated with handedness and schizophrenia (23), and a candidate gene approach identified the X-linked androgen receptor (24). However, a recent genome-wide association study (GWAS) (25) found no single nucleotide polymorphisms (SNPs) were associated with handedness at P -values below 5×10^{-8} , the standard genome-wide threshold for significance.

Here we report a GWAS for a quantitative measure of relative hand skill (peg-board task). The most highly associated SNP is in an intron of *PCSK6*, a gene that encodes a protein involved in left-right axis determination (26). We then replicate this association in two other independent samples with RD reaching overall genome-wide significance.

RESULTS

We genotyped 197 unrelated individuals with RD (stage 1) on the Illumina 550k SNP array and used imputation to test over 2 million SNPs for association with a quantitative measure of relative hand skill (peg-board task) (27). The samples were selected as part of an ongoing study of dyslexia from a large sample of families, each with at least two siblings who show symptoms of dyslexia (28,29). The peg-board task measures the time taken by the subjects to move a row of 10 pegs from one location to another with the left hand (L) and right hand (R) separately (30). From these data, we derived the measure PegQ [$2(L - R)/(L + R)$], which adjusts for overall differences in hand skill between subjects (19). This task produces an approximately normally distributed variable with a positive mean (Supplementary Material, Fig. S1); a positive PegQ indicates superior relative right-hand skill, and a negative PegQ indicates superior relative left-hand skill.

After the initial filtering and quality control stages, we tested 192 individuals for association with PegQ. There was no evidence of population structure and the genomic control statistic (λ) was 1.0031. No SNPs gave P -values below 5×10^{-8} in stage 1 (Supplementary Material, Table S1, Figs S2 and S3). The strongest association signal in the genome for directly genotyped SNPs was with rs9806256 ($P = 1.1 \times 10^{-6}$), while for imputed SNPs peak association was observed

with rs11855415 ($P = 4.7 \times 10^{-7}$). Both SNPs are located within a cluster of five highly correlated SNPs, spanning 12 kb from introns 14–18 of *PCSK6* (proprotein convertase subtilisin/kexin type 6, also known as *PACE4*, Fig. 1) on chromosome 15q26. Individuals with the minor (derived) allele have significantly greater relative right-hand skill compared with those carrying the major (ancestral) allele. The mean effect size of each copy of the minor allele is 0.60 standard deviations (SD) to the positive (right-handed) end of the PegQ distribution (see ' β ' in Table 1).

PCSK6 is known to play a key role in regulating left-right axis specification (26), making it a highly attractive candidate gene for involvement in handedness. We therefore followed up the initial association by genotyping rs11855415 and rs9806256 in an independent sample of individuals with RD who had also performed the same peg-board task (stage 2, $n = 376$). The results showed the same trend for increased right-hand skill in carriers of the minor allele; the association was nominally significant with rs11855415 ($P = 0.033$, $\beta = 0.19$, Table 1), but did not meet significance for rs9806256. Next, we genotyped these same two SNPs in individuals from the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort who have performed a similar peg-board task, and have been characterized by similar reading measures to the two previous stages. We selected individuals (stage 3, $n = 197$) to closely match the ascertainment criteria for RD used in stages 1 and 2. The analyses showed a significant increase in relative right-hand skill associated with the minor allele of both rs11855415 ($P = 0.0025$, $\beta = 0.36$, Table 1) and rs9806256 ($P = 0.00067$, $\beta = 0.40$, Table 1). Meta-analysis of all three stages yielded a P -value of 2.0×10^{-8} (Table 1, Supplementary Material, Fig. S4) for rs11855415, exceeding accepted genome-wide thresholds for significance. The data indicated that in the total sample of RD individuals, the average effect size of the minor allele is 0.35 SD towards the positive end of the PegQ distribution.

Then we investigated whether the effect of rs11855415 that we had observed in people with RD could be detected in the general population. We analyzed all the available ALSPAC children with a PegQ score ($n = 2666$), excluding those with neurodevelopmental disorders and performance IQ < 85 (see Supplementary Methods). The PegQ distribution for this sample compared with the ALSPAC children with RD in stage 3 shows neither a significant difference in the mean ($P = 0.54$), consistent with a previous study (9), nor a difference in the likelihood of being left- or right-handed ($P = 0.69$). In this general population sample, we did not detect significant association between rs11855415 and relative hand skill ($P = 0.25$, $\beta = -0.038$). However, the distribution of PegQ for each genotype in this general population sample shows that the PegQ scores for carriers of the minor allele appear clustered tightly around the mean (Supplementary Material, Fig. S5). This suggests that carriers of the minor allele in the general population cohort show reduced variability in relative hand skill, centered on the mean, in contrast to the RD sample which shows an increase in relative right-hand skill. To test this, we ran the association analysis for rs11855415 using the absolute value of the standardized PegQ score as a quantitative trait (Supplementary Material, Fig. S6). This compared scores closer to the mean against scores at both tails of the PegQ distribution.

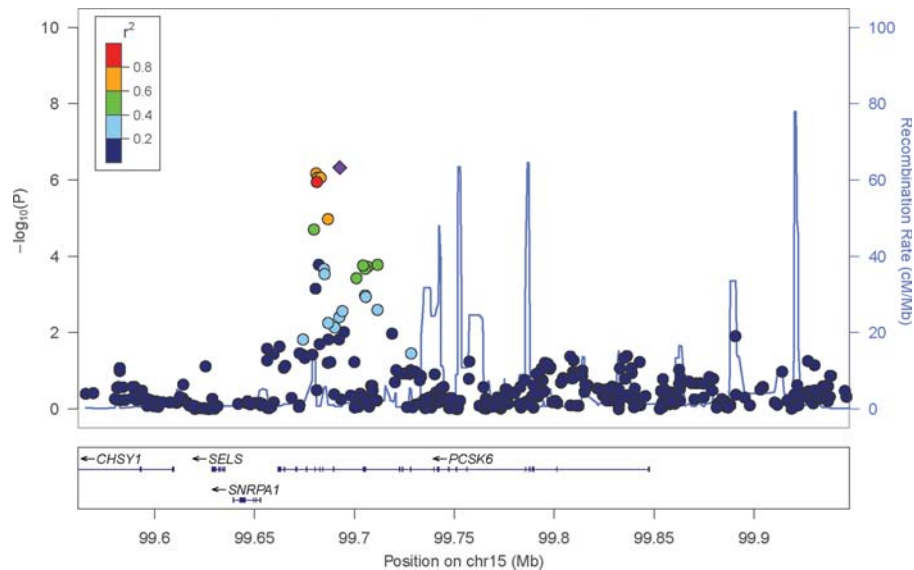


Figure 1. Visualization of P -values at the *PCSK6* locus. Negative $\log_{10}$ of the P -values for all genotyped and imputed SNPs around *PCSK6* plotted in LocusZoom (53). Linkage disequilibrium (r^2) with the most highly associated SNP, rs11855415 (purple diamond), is calculated based on the HapMap phase II CEU population and is shown by the colour of the SNPs. The recombination rate is shown by the blue line and the locations of genes in this locus are shown in the panel below the plot.

This analysis showed a nominally significant trend towards the mean of the PegQ distribution for carriers of the minor allele ($P = 0.0020$). Future studies will establish whether a similar pattern of findings is seen in other population cohorts.

DISCUSSION

We have performed the first GWAS for a quantitative measure of human handedness, as assessed by the peg-board task (30). The most highly associated SNP, rs11855415, is within an intron of *PCSK6*. The minor allele of this SNP confers increased relative right-hand skill. This association replicated in two independent groups with RD and is the first SNP for handedness to be identified at a genome-wide significant level. This effect was specific for people with RD, while the minor allele of rs11855415 showed a significant trend towards reduced laterality of hand skill in the general population.

PCSK6 encodes a proprotein convertase that processes proteins such as NODAL from their latent precursors into biologically active products (31). NODAL is a TGF β -related protein that specifies both the anteroposterior and left–right axes (32). In a proportion of mouse embryos null for *Pcsk6*, the normally asymmetrically expressed *Nodal*, *Lefty* and *Pitx2* mRNAs are bilaterally expressed (26). Some embryos subsequently display laterality defects such as situs ambiguous (26). *PCSK6* is widely expressed in humans, with particularly high expression in the liver, corpus callosum (the band of axon fibers that connect the left and right hemispheres) and spinal cord (33). Mutations in other *NODAL* pathway members have been linked to lateralization defects such as holoprosencephaly, which is an incomplete separation of the forebrain into discrete hemispheres (34,35).

The observation that the increase in relative right-hand skill associated with rs11855415 in *PCSK6* is specific to RD suggests

there may be epistatic interaction between *PCSK6* and RD susceptibility genes. There have been a number of candidate genes proposed for RD, including *DCDC2* (36,37), *DYX1C1* (38), *KIAA0319* (39,40) and *ROBO1* (41). These genes are all involved in neuronal migration or axon guidance (42). It may be that the functional change in *PCSK6* subtly alters the initial left–right patterning of the early embryo, and this has a downstream effect during neuronal migration on the development of cerebral asymmetry. Future studies will seek to explain the difference between RD and general population subjects.

The genome-wide significance of SNP associations in *PCSK6*, coupled with the role of this gene in a known biological pathway for left–right asymmetry, make it a strong candidate for involvement in cerebral asymmetry and thus handedness. Further work will seek to identify the functional genetic change in *PCSK6*, along with its mechanism of action. Our results are consistent with the emerging view that handedness is determined by multiple interacting genetic and environmental factors. Interestingly, *PCSK6* has been found to be a direct target of *FOXP2* (43), while *PCSK6* may interact with *SRPX2* (44). Mutation in either *FOXP2* or *SRPX2* may lead to severe forms of language or speech impairment. It would therefore be interesting to investigate the effect of *PCSK6* on handedness in individuals with other neurodevelopmental disorders such as SLI. Understanding the genes that modify the effects of *PCSK6* in RD individuals will offer us a significant insight into the etiology of both cerebral asymmetry and dyslexia.

MATERIALS AND METHODS

Study participants

Individuals were initially selected from our collection of families with RD (45) for severity of phenotype to undergo

Table 1. Association analysis results for the two markers in *PCSK6* in individuals with RD

Marker	Stage 1			Stage 2			Stage 3			Meta-analysis		
	<i>n</i>	<i>P</i> -value	β^a	SE ^b	<i>n</i>	<i>P</i> -value	β^a	SE	<i>n</i>	<i>P</i> -value	β^a	β 95L ^c
rs11855415	191	4.7×10^{-7}	0.60	0.12	368	0.033	0.19	0.09	185	0.0025	0.36	0.23
rs9806256	192	1.1×10^{-6}	0.56	0.12	364	0.18	0.12	0.09	182	0.00067	0.40	0.19
												0.43

^aThe mean effect size of each copy of the minor allele measured in standard deviations.^bStandard error.^c95% confidence intervals for β (L, lower, U, upper).

a GWAS for RD. Genotype data were generated for 197 individuals (stage 1). Families were recruited from the Dyslexia Clinic of the Royal Berkshire Hospital and range in age from 6 to 25 years. A second sample of 376 unrelated individuals with RD (stage 2) was recruited from either the Dyslexia Research Centre clinics in Oxford and Reading or the Aston Dyslexia and Development Clinic in Birmingham. The majority of these individuals are between 8 and 18 years old. The third sample of unrelated individuals came from the Avon Longitudinal Study of Parents and Children (ALSPAC). For ALSPAC, all pregnant women in the Bristol, UK, area with a then expected delivery date between 1 April 1991 and 31 December 1992 were approached for participation in the study that has today resulted in a general population cohort of about 14 000 children (46). For the present study, we assigned individuals from the ALSPAC cohort into a sub-group with RD ($n = 197$, stage 3), and a sub-group representing the general population without a neurodevelopmental disorder ($n = 2667$). Complete details of the ascertainment criteria for the RD and general population sub-groups are given as Supplementary Material.

Relative hand skill phenotype

The Pegboard test (30) was administered to the individuals who were analyzed in stages 1 and 2. The test involved the measurement of the time taken by the subjects to move, with each hand, a row of 10 pegs on a board from one location to another. This test was repeated five times for each hand in all the subjects. From these data, we derived the measure of relative hand skill (PegQ) for each subject, which was calculated as the difference between the average times for the left hand (L) and the right hand (R), $L - R$, divided by the average time for both hands combined, $(L + R)/2$, to correct for the overall difference in hand skill between the subjects (19). It has previously been shown that although overall hand skill improves with age (i.e. the peg test is performed in a shorter time with both hands), relative hand proficiency measured in this way does not change significantly between the ages of 3 and 15 (47). Therefore, although there is a large range of ages in our study (6–25 years), we do not expect this to significantly affect our results. The distribution of PegQ in stage 1 had a mean of 0.0796 (SD = 0.099), and in stage 2 it was 0.061 (SD = 0.091). The mean PegQ score is similar to Francks' (19) sample, which had a mean of 0.072. The PegQ distribution is unimodal, continuous and approximately normal, making it suitable as a quantitative phenotype for a GWAS (Supplementary Material, Fig. S1).

A peg test similar to that used in stages 1 and 2 was performed in the ALSPAC sample as part of a battery of manual dexterity tests known as Movement ABC (48). In this case, there were 12 pegs on the table that the child picked up one-at-a-time and placed in a peg board. After initial practice, children performed the test once (due to time constraints) with each hand. The mean PegQ was slightly different in stage 3 at 0.113 (SD = 0.199), while the distribution remains unimodal, continuous and approximately normal. One outlier was removed from further analysis (PegQ = 1.50).

Genotyping and quality control

The 197 individuals (stage 1) were genotyped on the Illumina® 550k Platform according to the manufacturer's instructions. One individual was excluded based on having a call rate <95%. We filtered out all SNPs with either minor allele frequency (MAF) < 0.05 (due to the small sample size), or successful genotype call rate < 0.95, or violation of Hardy–Weinberg equilibrium ($P < 5.7 \times 10^{-7}$; in autosomes). In total, 477 347 SNPs passed quality control. Principal component analysis (PCA) to test for individuals with divergent ancestry compared with CEU HapMap3 samples was performed as detailed previously (49), and as a result four further individuals were excluded (Supplementary Material, Fig. S7). In order to investigate population structure within our sample, we repeated our PCA without the HapMap3 samples. We tested for association between PegQ and the first two principal components (PC, adjusting for age and sex) in a linear regression model. There was no evidence of association with either component ($P = 0.244$ for PC1 and $P = 0.184$ for PC2). We thus did not adjust for these components in our subsequent GWAS association analysis.

The 376 unrelated cases (stage 2) were genotyped for the two SNPs, rs11855415 and rs9806256, as part of a larger study using iPLEX assays from Sequenom® (San Diego, CA, USA), according to the manufacturer's instructions. We also re-genotyped the original 197 individuals (stage 1) with these assays, and found 99% concordance for the directly genotyped rs9806256 and 98% concordance for the imputed SNP rs11855415. The Hardy–Weinberg equilibrium P -values for the SNPs were rs11855415 $P = 0.243$ and rs9806256 $P = 0.820$. The integrated genotyping system (50) was used to store genotypes.

The same two SNPs were genotyped in the ALSPAC children cohort by KBiosciences, Hertfordshire, UK (www.kbioscience.co.uk date last accessed 12/11/10), using a fluorescence-based competitive allele-specific PCR (KASPar) assay. The Hardy–Weinberg P -values for the two SNPs were rs11855415 $P = 0.018$ and rs9806256 $P = 0.248$.

Imputation

We used IMPUTE [v2.0.5 for autosomes, (27) and v1.0.0 for the X chromosome, (51)] to estimate the genotypes of SNPs not directly genotyped on the Illumina 550k platform. This involved combining information from (i) the directly genotyped SNPs, (ii) a densely genotyped reference panel (60 CEPH founders from HapMap Phase II, build 35) and (iii) a fine-scale recombination map, to infer the missing genotypes. Imputed SNPs were filtered out if they had an MAF < 0.05 or proper info < 0.5, allowing for us to test a further 1 709 209 SNPs for association.

Statistical analysis

Directly genotyped and imputed SNPs were tested for association with the quantitative measure PegQ under an additive model using SNPTEST (v1.1.4, www.stats.ox.ac.uk/~marchini/software/gwas/snpctest.html date last accessed 12/11/10) simultaneously specifying sex and age as covariates. For the X chromosome, analysis in stage 1 was conducted

by running the association test separately for the two sexes in SNPTEST and combined the results using GWAMA [v1.4, (52)], which performs fixed effects meta-analysis using inverse-variance weighting of the parameter estimates. The results for the three stages were then meta-analyzed using GWAMA (52). The values of PegQ were standardized in all stages separately to give a mean of 0, and a SD of 1. For the two SNPs genotyped in ALSPAC in the general population, the absolute value of the standardized PegQ score was also run as a quantitative phenotype.

To rule out any underlying confounding effects, we tested rs11855415 and rs9806256 for association with measures of hand motor skill, IQ and reading ability (CCI, CCN, OLSON, READ, SPELL, SPOON, defined in Supplementary Material), in each of the three stages as well as in the general population sample whenever data were available, and found no significant association (Supplementary Material, Table S2).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We are extremely grateful to all the families who took part in this study. We thank Janet Walter for collecting many of the samples and carrying out many of the psychometric tests in stages 1 and 2. For the ALSPAC study, we wish to thank the midwives for their help in recruiting the families, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. We thank Laura Miller and Bryan Howie for technical assistance, and Dorothy Bishop, Simon Fisher and Jonathan Flint for their constructive comments regarding the manuscript.

Conflict of Interest statement. None declared.

FUNDING

This research was specifically funded by the Wellcome Trust (076566/Z/05/Z); (075491/Z/04), the Medical Research Council (G0800523/86473) and the EU (Neurodys, 018696). Core support for ALSPAC is provided by the UK Medical Research Council (Grant ref: 74882), the Wellcome Trust (Grant ref: 076467) and the University of Bristol. Funding to pay the Open Access publication charges for this article was provided by the Wellcome Trust.

REFERENCES

1. Gilbert, A.N. and Wysocki, C.J. (1992) Hand preference and age in the United States. *Neuropsychologia*, **30**, 601–608.
2. Marchant, L.F. and McGrew, W.C. (1998) Human handedness: an ethological perspective. *Hum. Evol.*, **13**, 221–228.
3. Raymond, M. and Pontier, D. (2004) Is there geographical variation in human handedness? *Laterality*, **9**, 35–51.

4. Broca, P. (1861) Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole). *Bull. Soc. Anatom.*, **6**, 330–357.
5. Corballis, M.C. (1991) *The Lopsided Ape: Evolution of the Generative Mind*. Oxford University Press, New York.
6. Orton, S.T. (1966) *Word-blindness in School Children and Other Papers on Strophosymbolia (Specific Language Disability-Dyslexia) 1925–1946*. Orton Society, Pomfret, Conn.
7. Pennington, B.F. and Bishop, D.V. (2009) Relations among speech, language, and reading disorders. *Annu. Rev. Psychol.*, **60**, 283–306.
8. Bishop, D.V.M. (1990) *Handedness and Developmental Disorder*. Mac Keith Press, Lavenham.
9. Francks, C., Fisher, S.E., Marlow, A.J., MacPhie, I.L., Taylor, K.E., Richardson, A.J., Stein, J.F. and Monaco, A.P. (2003) Familial and genetic effects on motor coordination, laterality, and reading-related cognition. *Am. J. Psychiatry*, **160**, 1970–1977.
10. Knecht, S., Dräger, B., Deppe, M., Bobe, L., Lohmann, H., Floel, A., Ringelstein, E.B. and Henningsen, H. (2000) Handedness and hemispheric language dominance in healthy humans. *Brain*, **123**, 2512–2518.
11. Leonard, C.M. and Eckert, M.A. (2008) Asymmetry and dyslexia. *Dev. Neuropsychol.*, **33**, 663–681.
12. Friederici, A.D. (2006) The neural basis of language development and its impairment. *Neuron*, **52**, 941–952.
13. Galaburda, A.M., Sherman, G.F., Rosen, G.D., Aboitiz, F. and Geschwind, N. (1985) Developmental dyslexia: four consecutive patients with cortical anomalies. *Ann. Neurol.*, **18**, 222–233.
14. Galaburda, A.M. (1989) Ordinary and extraordinary brain development: anatomical variation in developmental dyslexia. *Ann. Dyslexia*, **39**, 65–80.
15. Richlan, F., Kronbichler, M. and Wimmer, H. (2009) Functional abnormalities in the dyslexic brain: a quantitative meta-analysis of neuroimaging studies. *Hum. Brain Mapp.*, **30**, 3299–3308.
16. Paulesu, E., Demonet, J.F., Fazio, F., McCrory, E., Chanoine, V., Brunswick, N., Cappa, S.F., Cossu, G., Habib, M., Frith, C.D. *et al.* (2001) Dyslexia: cultural diversity and biological unity. *Science*, **291**, 2165–2167.
17. Silani, G., Frith, U., Demonet, J.F., Fazio, F., Perani, D., Price, C., Frith, C.D. and Paulesu, E. (2005) Brain abnormalities underlying altered activation in dyslexia: a voxel based morphometry study. *Brain*, **128**, 2453–2461.
18. Medland, S.E., Duffy, D.L., Wright, M.J., Geffen, G.M., Hay, D.A., Levy, F., Van-Beijsterveldt, C.E.M., Willemsen, G., Townsend, G.C., White, V. *et al.* (2009) Genetic influences on handedness: data from 25,732 Australian and Dutch twin families. *Neuropsychologia*, **47**, 330–337.
19. Francks, C., Fisher, S.E., MacPhie, I.L., Richardson, A.J., Marlow, A.J., Stein, J.F. and Monaco, A.P. (2002) A genomewide linkage screen for relative hand skill in sibling pairs. *Am. J. Hum. Genet.*, **70**, 800–805.
20. Francks, C., DeLisi, L.E., Fisher, S.E., Laval, S.H., Rue, J.E., Stein, J.F. and Monaco, A.P. (2003) Confirmatory evidence for linkage of relative hand skill to 2p12-q11. *Am. J. Hum. Genet.*, **72**, 499–502.
21. Van Agetmael, T., Forrest, S.M. and Williamson, R. (2002) Parametric and non-parametric linkage analysis of several candidate regions for genes for human handedness. *Eur. J. Hum. Genet.*, **10**, 623–630.
22. Warren, D.M., Stern, M., Duggirala, R., Dyer, T.D. and Almasy, L. (2006) Heritability and linkage analysis of hand, foot, and eye preference in Mexican Americans. *Laterality*, **11**, 508–524.
23. Francks, C., Maegawa, S., Lauren, J., Abrahams, B.S., Velayos-Baeza, A., Medland, S.E., Colella, S., Groszer, M., McAuley, E.Z., Caffrey, T.M. *et al.* (2007) LRRTM1 on chromosome 2p12 is a maternally suppressed gene that is associated paternally with handedness and schizophrenia. *Mol. Psychiatry*, **12**, 1129–1139.
24. Medland, S.E., Duffy, D.L., Spurdle, A.B., Wright, M.J., Geffen, G.M., Montgomery, G.W. and Martin, N.G. (2005) Opposite effects of androgen receptor CAG repeat length on increased risk of left-handedness in males and females. *Behav. Genet.*, **35**, 735–744.
25. Eriksson, N., Macpherson, J.M., Tung, J.Y., Hon, L.S., Naughton, B., Saxonov, S., Avey, L., Wojcicki, A., Pe'er, I. and Mountain, J. (2010) Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genet.*, **6**, e1000993.
26. Constan, D.B. and Robertson, E.J. (2000) SPC4/PACE4 regulates a TGFbeta signaling network during axis formation. *Genes Dev.*, **14**, 1146–1155.
27. Howie, B.N., Donnelly, P. and Marchini, J. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.*, **5**, e1000529.
28. Fisher, S.E., Marlow, A.J., Lamb, J., Maestrini, E., Williams, D.F., Richardson, A.J., Weeks, D.E., Stein, J.F. and Monaco, A.P. (1999) A quantitative-trait locus on chromosome 6p influences different aspects of developmental dyslexia. *Am. J. Hum. Genet.*, **64**, 146–156.
29. Marlow, A.J., Fisher, S.E., Richardson, A.J., Francks, C., Talcott, J.B., Monaco, A.P., Stein, J.F. and Cardon, L.R. (2001) Investigation of quantitative measures related to reading disability in a large sample of sib-pairs from the UK. *Behav. Genet.*, **31**, 219–230.
30. Annett, M. (1985) *Left, Right, Hand and Brain: The Right Shift Theory*. Psychology Press, London.
31. Kiefer, M.C., Tucker, J.E., Joh, R., Landsberg, K.E., Saltman, D. and Barr, P.J. (1991) Identification of a second human subtilisin-like protease gene in the fes/fps region of chromosome 15. *DNA Cell Biol.*, **10**, 757–769.
32. Beck, S., Le Good, J.A., Guzman, M., Ben Haim, N., Roy, K., Beermann, F. and Constam, D.B. (2002) Extraembryonic proteases regulate Nodal signalling during gastrulation. *Nat. Cell Biol.*, **4**, 981–985.
33. Johnson, J.M., Castle, J., Garrett-Engle, P., Kan, Z., Loerch, P.M., Armour, C.D., Santos, R., Schadt, E.E., Stoughton, R. and Shoemaker, D.D. (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science*, **302**, 2141–2144.
34. Roessler, E., Ouspenskaia, M.V., Karkera, J.D., Velez, J.I., Kantipong, A., Lacbawan, F., Bowers, P., Belmont, J.W., Towbin, J.A., Goldmuntz, E. *et al.* (2008) Reduced NODAL signaling strength via mutation of several pathway members including FOXH1 is linked to human heart defects and holoprosencephaly. *Am. J. Hum. Genet.*, **83**, 18–29.
35. Mohapatra, B., Casey, B., Li, H., Ho-Dawson, T., Smith, L., Fernbach, S.D., Molinari, L., Niesh, S.R., Jefferies, J.L., Craigen, W.J. *et al.* (2009) Identification and functional characterization of NODAL rare variants in heterotaxy and isolated cardiovascular malformations. *Hum. Mol. Genet.*, **18**, 861–871.
36. Meng, H., Smith, S.D., Hager, K., Held, M., Liu, J., Olson, R.K., Pennington, B.F., DeFries, J.C., Gelemtier, J., O'Reilly-Pol, T. *et al.* (2005) DCDC2 is associated with reading disability and modulates neuronal development in the brain. *Proc. Natl Acad. Sci. USA*, **102**, 17053–17058.
37. Schumacher, J., Anthoni, H., Dahdouh, F., Konig, I.R., Hillmer, A.M., Kluck, N., Manthey, M., Plume, E., Warnke, A., Remschmidt, H. *et al.* (2006) Strong genetic evidence of DCDC2 as a susceptibility gene for dyslexia. *Am. J. Hum. Genet.*, **78**, 52–62.
38. Taipale, M., Kaminen, N., Nopola-Hemmi, J., Haltia, T., Myllyluoma, B., Lyytinen, H., Muller, K., Kaaranen, M., Lindsberg, P.J., Hannula-Jouppi, K. *et al.* (2003) A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. *Proc. Natl Acad. Sci. USA*, **100**, 11553–11558.
39. Francks, C., Paracchini, S., Smith, S.D., Richardson, A.J., Scerri, T.S., Cardon, L.R., Marlow, A.J., MacPhie, I.L., Walter, J., Pennington, B.F. *et al.* (2004) A 77-kilobase region of chromosome 6p22.2 is associated with dyslexia in families from the United Kingdom and from the United States. *Am. J. Hum. Genet.*, **75**, 1046–1058.
40. Cope, N., Harold, D., Hill, G., Moskvina, V., Stevenson, J., Holmans, P., Owen, M.J., O'Donovan, M.C. and Williams, J. (2005) Strong evidence that KIAA0319 on chromosome 6p is a susceptibility gene for developmental dyslexia. *Am. J. Hum. Genet.*, **76**, 581–591.
41. Hannula-Jouppi, K., Kaminen-Ahola, N., Taipale, M., Eklund, R., Nopola-Hemmi, J., Kaariainen, H. and Kere, J. (2005) The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. *PLoS Genet.*, **1**, e50.
42. Galaburda, A.M., LoTurco, J., Ramus, F., Fitch, R.H. and Rosen, G.D. (2006) From genes to behavior in developmental dyslexia. *Nat. Neurosci.*, **9**, 1213–1217.
43. 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.
44. Royer-Zemmour, B., Ponsolle-Lenfant, M., Gara, H., Roll, P., Leveque, C., Massacrier, A., Ferracci, G., Cillario, J., Robaglia-Schlupp, A., Vincentelli, R. *et al.* (2008) Epileptic and developmental disorders of the speech cortex: ligand/receptor interaction of wild-type and mutant SRPX2

- with the plasminogen activator receptor uPAR. *Hum. Mol. Genet.*, **17**, 3617–3630.
45. Fisher, S.E., Francks, C., Marlow, A.J., MacPhie, I.L., Newbury, D.F., Cardon, L.R., Ishikawa-Brush, Y., Richardson, A.J., Talcott, J.B., Gayan, J. *et al.* (2002) Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia. *Nat. Genet.*, **30**, 86–91.
46. Golding, J., Pembrey, M. and Jones, R. (2001) ALSPAC—the Avon Longitudinal Study of Parents and Children. I. Study methodology. *Paediatr. Perinat. Epidemiol.*, **15**, 74–87.
47. Annett, M. (1970) The growth of manual preference and speed. *Br. J. Psychol.*, **61**, 545–558.
48. Henderson, S.E. and Sugden, D.A. (1992) *Movement Assessment Battery for Children Manual*. The Psychological Corporation, Sidcup.
49. Anderson, C.A., Pettersson, F.H., Clarke, G.M., Cardon, L.R., Morris, A.P. and Zondervan, K.T. (2010) Data quality control in genetic case-control association studies. *Nat. Protoc.*, **5**, 1564–1573.
50. Fiddy, S., Cattermole, D., Xie, D., Duan, X.Y. and Mott, R. (2006) An integrated system for genetic analysis. *BMC Bioinformatics*, **7**, 210.
51. Marchini, J., Howie, B., Myers, S., McVean, G. and Donnelly, P. (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.*, **39**, 906–913.
52. Magi, R. and Morris, A.P. (2010) GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics*, **11**, 288.
53. 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.

Common Variants in Left/Right Asymmetry Genes and Pathways Are Associated with Relative Hand Skill

William M. Brandler^{1,2}, Andrew P. Morris¹, David M. Evans^{3,4}, Thomas S. Scerri⁵, John P. Kemp^{3,4}, Nicholas J. Timpson^{3,4}, Beate St Pourcain⁴, George Davey Smith^{3,4}, Susan M. Ring³, John Stein⁶, Anthony P. Monaco¹, Joel B. Talcott⁷, Simon E. Fisher^{8,9}, Caleb Webber^{2*}, Silvia Paracchini^{10*}

1 Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom, **2** MRC Functional Genomics Unit, Department of Physiology, Anatomy & Genetics, University of Oxford, Oxford, United Kingdom, **3** MRC Centre for Causal Analyses in Translational Epidemiology, University of Bristol, Bristol, United Kingdom, **4** School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom, **5** The Walter and Eliza Hall Research Institute of Medical Research, Melbourne, Australia, **6** Department of Physiology, University of Oxford, Oxford, United Kingdom, **7** School of Life and Health Sciences, Aston University, Birmingham, United Kingdom, **8** Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, **9** Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands, **10** School of Medicine, University of St Andrews, St Andrews, United Kingdom

Abstract

Humans display structural and functional asymmetries in brain organization, strikingly with respect to language and handedness. The molecular basis of these asymmetries is unknown. We report a genome-wide association study meta-analysis for a quantitative measure of relative hand skill in individuals with dyslexia [reading disability (RD)] ($n = 728$). The most strongly associated variant, rs7182874 ($P = 8.68 \times 10^{-9}$), is located in *PCSK6*, further supporting an association we previously reported. We also confirmed the specificity of this association in individuals with RD; the same locus was not associated with relative hand skill in a general population cohort ($n = 2,666$). As *PCSK6* is known to regulate NODAL in the development of left/right (LR) asymmetry in mice, we developed a novel approach to GWAS pathway analysis, using gene-set enrichment to test for an over-representation of highly associated variants within the orthologs of genes whose disruption in mice yields LR asymmetry phenotypes. Four out of 15 LR asymmetry phenotypes showed an over-representation ($FDR \leq 5\%$). We replicated three of these phenotypes; situs inversus, heterotaxia, and double outlet right ventricle, in the general population cohort ($FDR \leq 5\%$). Our findings lead us to propose that handedness is a polygenic trait controlled in part by the molecular mechanisms that establish LR body asymmetry early in development.

Citation: Brandler WM, Morris AP, Evans DM, Scerri TS, Kemp JP, et al. (2013) Common Variants in Left/Right Asymmetry Genes and Pathways Are Associated with Relative Hand Skill. *PLoS Genet* 9(9): e1003751. doi:10.1371/journal.pgen.1003751

Editor: Daniel H. Geschwind, University of California Los Angeles, United States of America

Received: March 7, 2013; **Accepted:** July 10, 2013; **Published:** September 12, 2013

Copyright: © 2013 Brandler et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the University of St Andrews, the UK Medical Research Council (grant number G0800523/86473 to SP), the Max Planck Society, and the the EU (Neurodys, 018696). Genotyping at the Wellcome Trust Centre for Human Genetics was supported by the Wellcome Trust (090532/Z/09/Z) and a Medical Research Council Hub Grant (G0900747 91070). Core support for ALSPAC was provided by the UK Medical Research Council and the Wellcome Trust (092731) and the University of Bristol. SP is a Royal Society University Research Fellow. CW is also funded by the UK Medical Research Funding and the EU (GENCODYS, 241995). APMor was supported by the Wellcome Trust (grant numbers WT075491, WT090532, and WT098017). WMB is the recipient of a Nuffield Department of Medicine Prize Studentship. JPK is funded by a Wellcome Trust PhD studentship (WT083431MA). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Caleb.Webber@dpag.ox.ac.uk (CW); sp58@st-andrews.ac.uk (SP)

Introduction

Humans display such striking cerebral asymmetries that we have been labeled, “The Lopsided Ape” [1]. The prevalence of right-handedness for tool use in excess of 85% across all human societies implies left-hemisphere dominance for controlling hand preference [2,3]. The same degree of collective population bias is not observed in the other great apes [4,5], suggesting right-handedness evolved along the human lineage [6]. Studies of fossil hominins suggest that the evolution of right-handedness dates back over half a million years [7], and is shared with *Homo neanderthalensis* [8,9] and *Homo heidelbergensis* [10].

Handedness correlates with asymmetries of both the motor cortex and language-related areas of the brain, such as the planum temporale [11–13]. Furthermore, there is a weak correlation between language lateralization and handedness: in a study of 326 individuals, 96% of strong right-handers, as compared with 73%

of strong left-handers, show left-hemisphere dominance for language [14]. Because language, like handedness, evolved primarily along the human lineage, it has been proposed that the two evolved together [15]. It has also been hypothesized that atypical cerebral asymmetries are associated with neurodevelopmental disorders. Although many associations appear to be spurious [16], dyslexia has been associated with structural and functional asymmetry differences, particularly in language related regions like the superior temporal gyrus (which includes the planum temporale) [17,18]. Schizophrenia has been associated with both reduced planum temporale asymmetry [19–21] and an increased prevalence of left-handedness [22].

Both cerebral asymmetry and handedness are heritable traits, with additive genetic effects estimated to account for about a half and a quarter of the variance, respectively [23–25]. Heritability estimates are consistent with theories that propose a single causal gene underlies handedness [26–28], however population genomics

Author Summary

Humans have developed a population level bias towards right-handedness for tool-use. Understanding the genetic basis of handedness can help explain why this bias exists and may offer clues into the evolution of handedness and brain asymmetry. We have tested for correlation between relative hand skill and hundreds of thousands of genetic variants in a cohort of individuals with reading disability. The strongest associated variant is in the gene *PCSK6*, an enzyme that cleaves NODAL into an active form. NODAL plays a key role during the establishment of left/right (LR) asymmetry in diverse species, from snails to mammals. *Pcsk6* knock-out mice display LR asymmetry defects like heterotaxia (abnormal organ positioning). We uncovered further variants associated with relative hand skill in the human versions of genes that also cause the LR asymmetry phenotypes heterotaxia, and situs inversus (reversal of organ asymmetry) when knocked out in mice. These results replicate in an independent general population cohort without reading disability. We propose that handedness is under the control of many variants, some of which are in genes that also contribute to the determination of body LR asymmetry.

theory predicts that a significant proportion of the genetic variance in common traits is the results of multiple common variants of small effect [29–31]. Indeed, the first direct molecular studies of handedness support the view that it has a complex multifactorial basis. To date, positional cloning and candidate gene studies have implicated *LRRMT1* and *AR*, each of which would explain only a small proportion of the genetic variance in handedness. The imprinted *LRRMT1* at 2p12 has been associated with both relative hand skill and schizophrenia [32]. A candidate gene approach found that CAG repeat length in the X-linked androgen receptor (*AR*) has opposite effects on the probability of left-handedness in males and females [33]. The same polymorphism in *AR* was found to be associated with ‘mixed-handedness’ in an independent cohort [34]. A genome-wide association study (GWAS) of 4,268 individuals for hand preference [35] found no single nucleotide polymorphisms (SNPs) associated at $P \leq 5 \times 10^{-8}$, the standard genome-wide threshold for significance. Based on GWAS and linkage studies, both published and unpublished, it has been estimated that at least 40 loci underlie the genetic basis of handedness [36].

We recently identified the first association with handedness through a GWAS for relative hand skill (peg-board task, see **Materials and Methods**) [37]. The associated variant, rs11855415, is located in an intron of the proprotein convertase subtilisin/kexin type 6 (*PCSK6*, also known as *PACE4*). *PCSK6* is a protease that cleaves NODAL proprotein into an active form. NODAL determines the development of left/right (LR) asymmetry in bilaterians from snails to vertebrates [38,39]. Mice in which *Pcsk6* is disrupted display asymmetry defects like heterotaxia (abnormal distribution of visceral organs) [40].

In our previous study, we analyzed a cohort of individuals originally selected for reading disability (RD; dyslexia) that were characterized with handedness measures. The most highly associated SNP, rs11855415, was then genotyped in two independent cohorts with RD, reaching genome-wide significance in the combined meta-analysis [37]. However, the association was specific to individuals with RD; rs11855415 was not associated with hand skill in a general population cohort unaffected with RD. A study of an independent general population cohort has

identified a variable number tandem repeat (VNTR), close to rs11855415, associated with degree of handedness, providing further evidence that *PCSK6* is important in the development of handedness [41].

Here we have genotyped the replication cohorts from our original study genome-wide, and perform a meta-analysis after imputation with reference panels from the 1000 Genomes Project [42]. The *PCSK6* locus is the only genome-wide significant signal observed. Because *PCSK6* plays a key role in regulating the activity of NODAL during the development of LR asymmetry in vertebrates [40], we hypothesize that other genes in this pathway are relevant for the development of handedness. Performing gene-set enrichment analysis, we identify an over-representation of SNPs strongly associated with relative hand skill in the human orthologs of genes that cause asymmetry defects when disrupted in mice. We replicate these results in a large independent general population cohort without reading disability. Our results suggest that handedness is a complex trait controlled in part by genes responsible for the establishment of LR body asymmetry early in development.

Results

The study participants are derived from four cohorts (**Table 1**), three with RD (cohorts 1–3, $n = 185$, 374, and 169 respectively), and a general population cohort (cohort 4, $n = 2,666$). The samples from cohorts 1 and 2 were selected as part of an ongoing study of dyslexia [43,44]. The samples from cohorts 3 and 4 are part of the Avon Longitudinal Study of Parents and Children (ALSPAC). We divided them into individuals with RD (cohort 3), and those unaffected with RD (cohort 4), based on their IQ and reading scores (see **Materials and Methods**). All have had their relative hand skill assessed.

Relative Hand Skill Phenotype

The relative hand skill peg-board task measures the time taken by the subjects to move a row of ten pegs from one location to another with the left hand (L) and right hand (R) separately [26]. Each hand was tested five times, from the average time for each hand we derived the measure PegQ [$2(L-R)/(L+R)$]. This task produces an approximately normally distributed variable with a positive mean; a positive PegQ indicates superior relative right-hand skill, and a negative PegQ indicates superior relative left-hand skill. In cohort 1, the PegQ distribution has a mean of 0.080 ($SD = 0.099$) [37], and in cohort 2 it is 0.061 ($SD = 0.091$). A similar peg-board test involving moving 12 pegs was performed just once (after practice) in ALSPAC (cohorts 3 and 4, see **Materials and Methods**). Since the test was slightly different, the mean PegQ was slightly different to cohorts 1 and 2 (**Table 2**), while the distribution remains unimodal, continuous and approximately normal.

We investigated the correlation between binary hand preference and relative hand skill using logistic regression in all the individuals in our cohorts for which we have data. The PegQ phenotype is highly correlated with the preferred hand of the participants. In cohorts 1 & 2, for every 1 standard deviation shift to the right of the Peg distribution, individuals are 13.4 times more likely to be right handed (95% Confidence intervals: 6.11–36.49). In ALSPAC (cohorts 3 & 4), they are 3.23 times more likely to be right handed (95% Confidence intervals: 2.83–3.70). The weaker correlation in ALSPAC is likely due to the fact that the peg-board task was performed only once, as opposed to five times in cohorts 1 and 2 (see **Materials and Methods**). As the peg-board tests fine motor control, PegQ also correlates with dexterity, measured as the time

Table 1. GWAS study design.

Cohort	n	Genotyping Array	Imputation Reference Panel	Imputation Software	Reading Disabled
1	185	Illumina 550K	1000 Genomes March 2012	Impute	Yes
2	374	Illumina OmniExpress	1000 Genomes March 2012	Impute	Yes
3	169	Illumina 550K	HapMap Phase II	MACH	Yes
4	2,666	Illumina 550K	HapMap Phase II	MACH	No

Three cohorts with reading disability and a further general population cohort were genotyped and tested after imputation for association with relative hand skill.
doi:10.1371/journal.pgen.1003751.t001

taken to perform the task with either the right hand (cohorts 1 & 2, Pearson's $r = -0.42$, cohorts 3 and 4 $r = -0.50$) or the left hand (cohorts 1 & 2, Pearson's $r = 0.27$, cohorts 3 and 4 $r = 0.59$).

Individuals with RD (cohort 3) are not significantly more likely to be left-handed ($f = 11.4\%$) than those in the general population (cohort 4, $f = 10.9\%$, two-tailed Fisher's exact test $P = 0.82$). Furthermore, they are not more likely to be shifted toward the left-handed distribution of the peg board task, however their average time for both hands is slower, indicating lower overall motor skill ($P = 0.001$, one tailed Student's t-test, **Table 2**).

GWAS for Relative Hand Skill

Cohorts 1 and 2 have been genotyped on the Illumina 550K and OmniExpress SNP arrays, respectively, and we imputed genotypes using the 1000 Genomes reference panel (Phase 1, March 2012 release). The samples from cohorts 3 and 4 come from the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort. The participants have performed a similar peg-board task (see **Materials and Methods**), been characterized by similar reading measures to cohorts 1 and 2, been genotyped on the Illumina 550K SNP array, and had genotypes imputed using the HapMap II CEU reference panel. We selected individuals (cohort 3) to closely match the ascertainment criteria for RD used in the other two cohorts. Furthermore we selected a general population cohort, excluding individuals showing phenotypes compatible with a neurodevelopmental disorder diagnosis (cohort 4, consistent with the methodology in our previous report [37]).

We tested 2,095,209 SNPs for association with relative hand skill in a meta-analysis of the three RD cohorts. The marker rs7182874 within *PCSK6* is the most highly associated SNP ($P = 8.68 \times 10^{-9}$, **Figures 1, S1, Tables 3, S1**), confirming our previous findings [37]). Each copy of the minor (derived) allele shifts an individual 0.37 standard deviations toward the right-handed end of the distribution of relative hand skill. No SNP is genome-wide significant in the general population cohort; the most highly associated SNP, rs7883190 ($P = 2.08 \times 10^{-6}$) is located approximately 6 Kb upstream of the gene *GPC3* on the X chromosome, a gene that causes heart and lung asymmetry defects when disrupted in mice (cohort 4, **Table S2**) [45].

Gene-Set Enrichment Analysis (GSEA)

Disruption of *PCSK6* in mice causes asymmetry defects [40,46]. We therefore employed gene-set enrichment analysis (GSEA) to test for an over-representation of highly associated SNPs within the human orthologs of mouse genes whose disruption causes LR asymmetry defects, as annotated by the Mouse Genome Informatics (MGI) resource [47]. GSEA was implemented in MAGENTA, which combines variant association P values into individual gene scores, correcting for confounders such as gene size, variant number, and linkage disequilibrium [48]. It then ranks all gene scores and tests for an overrepresentation of a given set of genes above the 75th percentile of all gene P values. We chose this cutoff because it yields the best power for detecting enrichment in traits that are highly polygenic with many associations of weak effect [48]. Genes were defined as from 20 Kb upstream of the most 5' transcription start site to 10 Kb downstream of the most 3' untranslated region, to include adjacent cis-regulatory elements. We tested for an over-representation of genes annotated to each of 15 morphological asymmetry phenotypes (with between 10 and 100 annotated genes per phenotype) in the meta-analysis of cohorts 1–3. Four phenotype gene sets out of the 15 tested were significant at an $FDR \leq 5\%$ in the combined RD meta-analysis (**Table 4**); abnormal right lung morphology (MP:0009570), double outlet right ventricle (MP:0000284), heterotaxia (MP:0004133), and situs inversus (MP:0002766). MAGENTA also provides the overrepresentation of genes above the 95th percentile cutoff by default. Despite the fact that the expected number of genes above this threshold was very low (typically around one or two genes per phenotype), reducing our power to detect any enrichment, we observed five phenotypes at $FDR \leq 5\%$ (**Table S3**). The genes that comprise the LR asymmetry phenotype sets overlap considerably; thus FDR correction for multiple testing is conservative. A list of the ten most highly associated genes from the four phenotypes above the 75th percentile cutoff is given in **Table 5**. After removing *PCSK6* from the analysis, the phenotypes abnormal right lung morphology, situs inversus, and double outlet right ventricle, still show an over-representation of associated SNPs ($FDR \leq 5\%$, **Table S4**). Therefore the association in *PCSK6* is not solely driving the results.

Table 2. Individuals with reading disability (cohort 3, $n = 169$) are no more skilled with their left- or right- hand (PegQ) compared to the general population (cohort 4, $n = 2,666$), however they perform the relative hand skill task slower (motor skill).

Phenotype	General Population Ave	General Population SD	Reading Disabled Ave	Reading Disabled SD	P
PegQ	0.11	0.19	0.10	0.18	0.32
Motor Skill	23.57	3.18	24.39	3.65	0.0011

We used a one-tailed Student's t-test, not assuming equal variance.
doi:10.1371/journal.pgen.1003751.t002

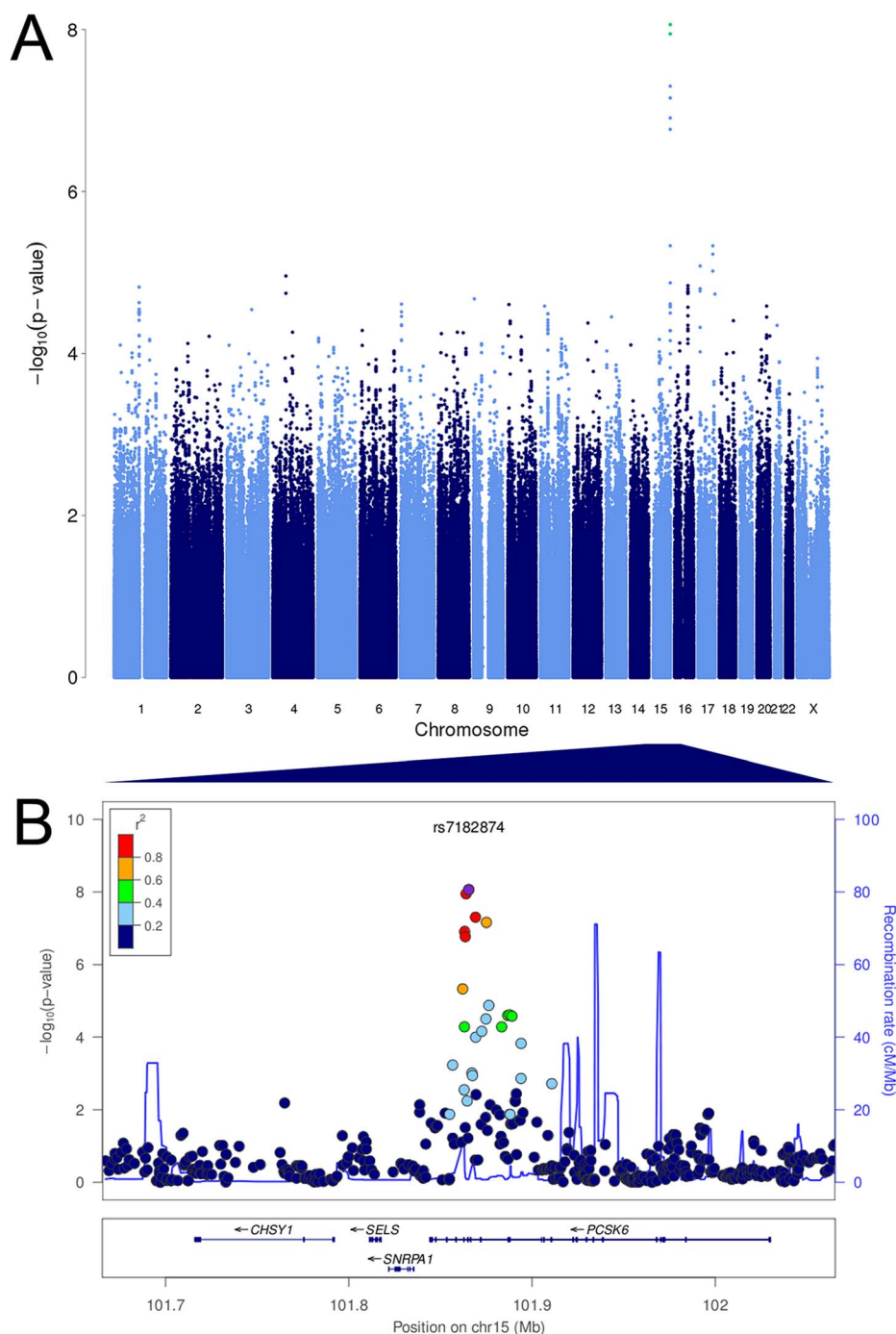


Figure 1. *PCSK6* is associated with relative hand skill in individuals with reading disability. Manhattan plot and visualization of P-values at the *PCSK6* locus. (A) Manhattan plot of the P values for the meta-analysis of the RD Peg relative hand skill GWAS data-sets (cohorts 1–3). The observed P values are plotted against chromosome and position along the genome. The green dots represent markers that are genome-wide significant at $P \leq 5 \times 10^{-8}$. (B) Negative $\log_{10}$ of the P-values for all genotyped and imputed SNPs around *PCSK6* plotted in LocusZoom. Linkage disequilibrium (r^2) with the most highly associated SNP, rs7182874 (purple diamond), is shown by the color of the SNPs. The recombination rate is shown by the blue line and the locations of genes in this locus are shown in the panel below the plot.
doi:10.1371/journal.pgen.1003751.g001

We took forward the four phenotypes that show an over-representation of genes with low P -value SNPs in the RD cohorts for testing in the general population (cohort 4). Three out of the four phenotypes show enrichment at ($FDR \leq 5\%$), double outlet right ventricle ($P = 0.019$, adjusted for multiple testing), heterotaxia (adjusted $P = 0.026$), and situs inversus (adjusted $P = 0.021$) (**Table**

S5). We have therefore shown that our findings from the RD cohort hold true for the general population.

There are 116 unique genes whose disruption in their mouse ortholog is annotated with these three phenotypes. 47 genes in the RD cohorts, and 42 in the general population cohort, show P values above the 75th percentile of all gene P values. Of these, 16

Table 3. The intronic marker rs7182874 (ancestral allele = C; derived, minor allele = T) within the gene *PCSK6* on chromosome 15 is associated with relative hand skill in individuals with reading disability.

Cohort	n	P	β	S.E.
1	185	2.30E-06	0.55	0.12
2	374	0.0099	0.26	0.10
3	169	0.0049	0.34	0.12
Meta-Analysis	728	8.68E-09	0.37	0.064
General Population (4)	2,666	0.77	-0.01	0.032

β = effect size of each copy of the minor allele in standard deviations, S.E. = Standard Error.

doi:10.1371/journal.pgen.1003751.t003

overlap between the RD meta-analysis and the general population cohort (one-tailed Fisher's exact test, $P=0.72$). Therefore, although the same phenotypes show enrichments, the associated SNPs and genes are different, i.e. there is both allelic and locus heterogeneity between the cohorts.

Four out of the five most highly associated genes from the mouse phenotypes tested (**Table 5**) are involved in ciliogenesis (*RFX3*, *MNS1*, *GLI3*, and *PKD2*). Because ciliopathies cause cerebral phenotypes like agenesis of the corpus callosum and hydrocephaly, we also tested these two phenotypes in both the meta-analysis of the individuals with RD (cohorts 1–3), and separately in the general population (cohort 4). The absent corpus callosum (MP:0002196) phenotype showed an over-representation in the RD cohorts (adjusted $P=0.016$, **Table S6**), but not in the general population (cohort 4). Of the associated genes from the absent corpus callosum phenotype, only *GLI3* overlaps with the

associated genes in **Table 5**. Hydrocephaly (MP:0001891) did not show an over-representation in either cohort.

Discussion

We have performed a genome-wide association study (GWAS) meta-analysis for a quantitative measure of handedness and found association with variants that implicate genes involved in the determination of left/right body asymmetry. To achieve this, we developed a novel approach to GWAS pathway analysis, coupling gene-set enrichment analysis (GSEA) with mouse phenotype data. In the RD meta-analysis, we showed there was an over-representation of variants with low P -values within human orthologs of genes that cause LR asymmetry phenotypes when knocked-out in mice. We replicated the same over-representation in the general population. We also confirmed that the gene *PCSK6* was associated with relative hand skill at genome-wide significance (RD meta-analysis, cohorts 1–3, rs7182874 $P=8.68 \times 10^{-9}$).

The genes identified by GSEA in the meta-analysis of cohorts with RD are involved in the determination of LR asymmetry in mice (**Table 5**). *Mns1* and the transcription factor *Rfx3* both play key roles in ciliogenesis [46,49]. The transcription factor *Gli3* is expressed at primary cilia, localizing to both the distal tip of the cilia as well as the nucleus, and is involved in Sonic hedgehog signal transduction [50]. The unidirectional fluid flow created by cilia rotation that breaks asymmetry is detected via the Ca^{2+} channel PKD2 [51]. This flow concentrates the morphogen NODAL on the left, which is cleaved into an active form by the protease PCSK6 [52]. The NODAL signal is then transduced through the transmembrane receptor ACVR2B [53]. Retinoic acid acts through receptors like RARA to affect expression of *Nodal*, *Pitx2*, and *Lefty* in mice [54]. FZD1 also acts upstream of NODAL, it is a receptor for WNT3A [55], which activates the Delta/Notch pathway to regulate expression of *Nodal* in mice [56].

Table 4. In the relative hand skill GWAS meta-analysis of individuals with reading disability, there is an over-representation of highly associated SNPs within human orthologs of genes involved in causing LR asymmetry phenotypes when knocked out in mice.

Mouse Phenotype ID	Phenotype	Genes	Median Size/KB	Mean Size	Expected	Observed	Adjusted P
MP:0010429	abnormal heart left ventricle outflow tract morphology	47	42	83	12	13	0.4
MP:0003922	abnormal heart right atrium morphology	20	50	87	5	4	0.74
MP:0009569	abnormal left lung morphology	22	51	89	6	8	0.2
MP:0001706	abnormal left-right axis patterning	62	48	73	16	20	0.17
MP:0009570	abnormal right lung morphology	32	68	120	8	16	0.0087
MP:0000644	Dextrocardia	26	64	113	7	11	0.07
MP:0000284	double outlet right ventricle	87	47	92	22	35	0.016
MP:0002625	heart left ventricle hypertrophy	48	29	107	12	12	0.6
MP:0000276	heart right ventricle hypertrophy	25	18	96	6	3	0.93
MP:0004133	Heterotaxia	42	46	80	11	18	0.025
MP:0000542	left-sided isomerism	17	64	92	4	8	0.068
MP:0004158	right aortic arch	36	27	80	9	14	0.07
MP:0000531	right pulmonary isomerism	15	39	77	4	5	0.35
MP:0000508	right-sided isomerism	16	46	88	4	5	0.37
MP:0002766	situs inversus	24	63	103	6	11	0.046

15 phenotypes that contain between 10 and 100 genes, and cause asymmetry defects when knocked out in mice, were tested using MAGENTA (see **Materials and Methods**). The observed number of genes were compared to the expected using a cut-off that equals the 75th percentile of all gene P values. P values were then adjusted for multiple testing using an FDR correction.

doi:10.1371/journal.pgen.1003751.t004

Table 5. LR asymmetry genes are associated with relative hand skill (meta-analysis of cohorts 1–3).

Gene	Chromosome	Gene Size/kb	Gene P	Best SNP	SNP P	MAF	β
<i>PCSK6</i>	15	186	3.94E-08	rs7182874	8.68E-09	0.29	0.37
<i>PKD2</i>	4	70	0.00034	rs2728105	0.00015	0.46	0.22
<i>MNS1</i>	15	36	0.00087	rs6493858	0.00032	0.44	−0.21
<i>GLI3</i>	7	276	0.0056	rs2282920	0.00016	0.41	0.23
<i>RFX3</i>	9	301	0.0079	rs2991306	0.00032	0.11	−0.33
<i>ACVR2B</i>	3	39	0.017	rs7645704	0.0038	0.07	0.34
<i>MGRN1</i>	16	66	0.019	rs2279408	0.0036	0.20	0.20
<i>LAMA5</i>	20	58	0.019	rs477859	0.0032	0.22	0.20
<i>RARA</i>	17	48	0.025	rs2715553	0.0085	0.47	−0.15
<i>FZD1</i>	7	4	0.027	rs6977978	0.0055	0.14	−0.20

We've listed the ten lowest gene *P* values that are also within one of the four enriched phenotypes from **Table 4** in the RD meta-analysis. MAF = minor allele frequency, β = effect size of each copy of the minor allele, in standard deviations.
doi:10.1371/journal.pgen.1003751.t005

Similarly, in a significant proportion of mouse embryos where *Mgm1* is disrupted, *Pitx2*, *Lefty1*, and *Lefty2* are abnormally expressed [57]. *LAMA5* is a component of the extracellular matrix, its precise role in the development of LR asymmetry is unknown. However, mice in which *Lama5* has been disrupted develop Polycystic Kidney Disease (PKD) [58], a phenotype comorbid with LR asymmetry defects caused by ciliopathies [59].

We have shown that our findings from the RD cohorts hold true for the general population, three of the four phenotype gene sets tested are significant at $FDR \leq 5\%$ (heterotaxia, situs inversus, and double outlet right ventricle). However, the associated variants and genes are different. For example, the most highly associated marker in the general population GWAS (rs7883190, $P = 2.08 \times 10^{-6}$, **Table S2**) is located 6 Kb upstream of *GPC3* on the X chromosome, a gene that causes the double outlet right ventricle phenotype when disrupted in mice [45].

If genes that are critical to the breaking of LR body symmetry influence handedness and brain asymmetry, individuals with Mendelian disorders that cause asymmetry defects like situs inversus (SI) should have a greater prevalence of left-handedness. However, studies dating as far back as 1836 have shown that individuals with SI present with a prevalence of left-handedness equal to the rest of the population [60–62]. A more recent study of individuals with primary ciliary dyskinesia (a ciliopathy that causes SI) has confirmed this finding [63]. However, two brain imaging studies of individuals with SI have shown a significant reversal of the typical pattern of right-frontal and left-occipital petalia asymmetry [64,65], of which one study also showed a significant reversal of language dominance [65]. However, each study only assessed three individuals with SI, and given its rarity (~ 1 in 10,000 [62]), large-scale studies have yet to be performed. Furthermore, a study of gene expression in fetal brains identified 27 differentially expressed genes between the two hemispheres, none of which are known to be involved in body asymmetry development [66]. It is possible that compensatory mechanisms allow for the normal development of handedness in individuals with situs inversus and the development of handedness and brain asymmetry is more complex than just involving early LR asymmetry determining genes.

To help explain the relationship between LR asymmetry determination, brain asymmetry, and handedness, we hypothesize that the same molecular mechanism for visceral LR asymmetry is involved during cerebral development. In support of this, *RFX3*

regulates the expression of *GLI3* in the telencephalon in mice, which in turn regulates the distribution of guidepost neurons necessary for corpus callosum formation [67]. *Rfx3* deficient mice show corpus callosum agenesis and disorganization of guidepost neurons necessary for axon pathfinding across the brain's midline [67]. Humans with ciliopathies commonly present with two specific structural changes within the brain; hydrocephaly and an absent corpus callosum [59]. Because four of the five most strongly associated genes shown in **Table 5** are involved in ciliogenesis, we therefore performed GSEA on these two phenotypes as well. Enrichment was detected in absent corpus callosum, but not in hydrocephaly in the RD cohort, although this did not replicate in the general population cohort. The only gene that overlaps with the LR asymmetry genes in **Table 5**, annotated to the absent corpus callosum phenotype, is *GLI3*. However, *RFX3* has not yet been annotated to this phenotype in the mouse phenotype ontology [67], while *PCSK6* is also known to have high expression in the corpus callosum [68]. Both postmortem measurements and MRI studies have independently indicated the tendency for the corpus callosum to be larger in left-handers [69,70], suggesting that the corpus callosum plays a role in the expression of handedness. Our study provides genetic evidence in support of this.

The relationship between these genes and both corpus callosum development and brain asymmetry are not well understood in humans. However, infants with agenesis of the corpus callosum appear to show a similar bias in right-handedness and performance on unimanual hand skill tasks to controls, although they perform slower at bimanual tasks [71,72]. Brain asymmetries have been better studied in zebrafish and have revealed that these processes are under the control of Nodal-related genes, suggesting a conserved pathway controlling both visceral and cerebral asymmetries [73]. A particular line of zebrafish, *frequent situs inversus (fsi)*, displays reversal of both body and neuronal asymmetries, as well as of some characteristic behavioral phenotypes, like eye preference and direction of prey approach [74]. Our study implicates Nodal-related genes in the development of handedness and therefore brain asymmetry in humans.

The main limitations of this study are dependent on the currently available datasets, characterized by a degree of genotype and phenotype heterogeneity, and a small sample size. Genotypes from ALSPAC were imputed with HapMap phase II instead of the 1000 genomes reference panel. This means we were restricted to

analyzing only ~2 million SNPs instead of over 6 million common SNPs in the 1000 genomes project. The imputation accuracy is also lower with HapMap II, however we restricted our study to common variants ($MAF \geq 5\%$), for which imputation accuracy is high (see box 4 of [75]). Not only was the relative hand skill phenotype slightly different in ALSPAC (see **Materials and Methods**), it was only measured once instead of five times. Furthermore the phenotype is only a proxy for one specific aspect of handedness, namely manipulating objects quickly with each hand separately. It is strongly correlated with both hand preference and dexterity (see **Results**), and it is not clear whether the associations observed are due to handedness or dexterity or a combination of the two. Although we interrogated a quantitative phenotype, which is more powerful than a binary case-control GWAS such as that performed by Eriksson et al. [35], the numbers of participants were relatively small (discovery cohorts: $n = 744$ and replication: $n = 2,666$). Analysis of phenotypes like PegQ in much larger cohorts will be needed to investigate comprehensively the biological complexity of laterality.

Based on our results we propose that the same biological mechanism underlying the development of LR asymmetry in the body plays a role in determining handedness. However, although we replicated the same LR asymmetry phenotypes in the general population as the RD cohorts, different SNPs located in different genes within these same pathways were responsible for driving the association. Furthermore, the genome-wide significant SNP in *PCSK6* does not even show the same trend in the general population (**Table 3**). This suggests there is both allelic and locus heterogeneity between the RD and general population cohorts. We suggest this heterogeneity may derive from epistasis between genes involved in dyslexia and those involved in handedness. In support of this, a cilia-related coexpression module derived from publically available microarray datasets found that the dyslexia associated genes, *DCDC2*, *DYX1C1*, and *KIAA0319* are coexpressed in cilia, suggesting they interact with genes involved in the determination of LR asymmetry [76]. Furthermore, overexpression of *Dcdc2* increases the length of cilia in rats [77]. Also, the Zebrafish ortholog of the dyslexia candidate gene *DYX1C1* has been shown to play a key role in cilia growth and function, and the inhibition of *dyx1c1* produces asymmetry phenotypes like situs inversus [78]. We have confirmed, as previously reported [16,23], that no consistent association has been found between either hand preference or relative hand skill and dyslexia. However, early work by Galaburda et al. reported eight consecutive post-mortem specimens with RD that had reduced planum temporale asymmetry [79,80]. This suggests subtle structural asymmetries might be involved in dyslexia. It has also been shown that reading performance is correlated with motor abilities on the peg-board test [23,81]. We have replicated this finding by comparing cohorts 3 and 4 in the current study – demonstrating that individuals with RD are slower at the peg board test than those in the general population ($P = 0.001$, **Table 2**). This is consistent with a broad literature showing motor skill difficulties in children with dyslexia [82–86].

The precise relationship between handedness, cerebral asymmetry, and neurodevelopmental disorders like dyslexia remains to be determined. However, our results have implicated some of the same genes involved in body asymmetry in handedness, and therefore the development of cerebral asymmetry in both individuals with RD and the general population. Larger scale GWAS or sequencing studies of handedness, coupled with a deeper understanding of the genetic basis of language-related disorders like dyslexia will help determine whether handedness and language do indeed develop hand-in-hand.

Materials and Methods

Ethics Statement

Ethical approval for cohorts 1 and 2 was acquired from the Oxfordshire Psychiatric Research Ethics Committee (OPREC O01.02). For the ALSPAC cohorts, ethical approval was obtained from the ALSPAC Law and Ethics committee and relevant local ethics committees, and written informed consent provided by all parents.

Study Participants

The initial cohort was selected for severity of phenotype to undergo a GWAS for RD as described before [37]. Briefly, families were recruited from the Dyslexia clinic of the Royal Berkshire Hospital and one child per family with dyslexia was selected ($n = 185$) ranging in age from 6 to 25 years. Unrelated individuals with RD (cohort 2, $n = 374$) were recruited from either the Dyslexia Research Centre clinics in Oxford and Reading or the Aston Dyslexia and Development Clinic in Birmingham [43]. The majority of these individuals are between 8 and 18 years old. The third cohort of unrelated individuals came from the Avon Longitudinal Study of Parents and Children (ALSPAC). ALSPAC is a general population cohort that started collecting data on around 14,000 pregnant women in the Bristol area with a then expected delivery date between 1 April 1991 and 31 December 1992 [87]. This cohort is described in detail on the website (<http://www.alspac.bris.ac.uk>) and elsewhere [88]. The children are being followed and measured for many different phenotypes including laterality measures. For the present study, we assigned individuals from the ALSPAC cohort into a sub-group with RD ($n = 169$, cohort 3), and a sub-group representing the general population who do not show phenotypes compatible with a neurodevelopmental disorder ($n = 2666$) consistent with our previous study [37].

Relative Hand Skill Phenotype

The peg-board test [26] was administered to the individuals who were analyzed in cohorts 1 and 2. The relative hand skill peg-board task measures the time taken by the subjects to move a row of ten pegs from one location to another with the left hand (L) and right hand (R) separately [26]. Each hand was tested five times, from the average time for each hand we derived the measure PegQ [$2(L-R)/(L+R)$]. In cohort 1, the distribution has a mean of 0.080 ($SD = 0.099$) [37], and in cohort 2 it is 0.061 ($SD = 0.091$). A similar peg-board test was performed in ALSPAC (cohorts 3 and 4) as part of a battery of manual dexterity tests known as Movement ABC [89]. In this case there were 12 pegs on the table that the child picked up one-at-a-time and placed in a peg-board. After initial practice, children performed the test once (due to time constraints) with each hand. Since the test was slightly different, the mean PegQ was slightly different to cohorts 1 and 2 (**Table 2**), while the distribution remains unimodal, continuous and approximately normal. One outlier was removed from further analysis (PegQ = 1.50).

Genotyping and Quality Control

Genotypes for cohort 1 were generated on the Illumina 550K SNP array [37]. Genotypes for cohort 2 were generated for the present study on the Illumina OmniExpress SNP array (730,525 markers), according to the manufacturer's instructions. We followed the quality control protocol set out by Anderson et al., 2010 [90] excluding individuals with: discordant sex information, a low genotyping success rate ($\leq 98\%$, **Figure S2**), duplicated or related to other individuals (identity by state IBS metric ≥ 0.1875),

divergent ancestry from CEU hapmap3 reference samples based on principal component analysis (PCA, **Figure S3**).

Furthermore, we excluded SNPs that: have an excessive missing genotype rate (call rate ≤ 0.98 , **Figure S4**), show a significant deviation from Hardy-Weinberg Equilibrium ($P \leq 10^{-6}$), and have a minor allele frequency (MAF) ≤ 0.05 .

After QC, 458,236 (cohort 1) and 589,768 (cohort 2) single nucleotide polymorphisms (SNPs) were taken forward for imputation.

Genotyping for the ALSPAC (cohorts 3 and 4) was carried out using the Illumina HumanHap 550K array. Individuals were excluded on the basis of the following: sex mismatches, minimal or excessive heterozygosity, disproportionate levels of individual missingness, cryptic relatedness measured as proportion of identity by descent, and insufficient sample replication. The remaining individuals were assessed for evidence of population stratification by multidimensional scaling analysis and compared with Hapmap II (release 22) European descent (CEU), Han Chinese, Japanese and Yoruba reference populations; all individuals with non-European ancestry were removed in the ALSPAC dataset. SNPs with a minor allele frequency of $<1\%$, a call rate of $<95\%$ or evidence for violations of Hardy-Weinberg equilibrium were removed.

Imputation

In cohorts 1 and 2, haplotypes were inferred prior to imputation using the Segmented HAPlotype Estimation and Imputation Tool (shapeit v2 – <http://www.shapeit.fr/> [91]). We used IMPUTE [42] (v2.2.2 – http://mathgen.stats.ox.ac.uk/impute/impute_v2.html) to estimate the genotypes of SNPs not directly genotyped. Imputed SNPs were filtered out if they had an MAF ≤ 0.05 or imputation info ≤ 0.8 . We then used the 1000 Genomes Phase I integrated variant set (March 2012, build 37) as the reference panel for imputation, testing 5,829,682 SNPs for association.

In ALSPAC autosomal genotypic data were imputed in ALSPAC using Markov Chain Haplotyping software (MACH v1.0.16) and phased haplotype data from CEU individuals (Hapmap release 22, Phase II NCBI B36, dbSNP 126) based on a cleaned dataset of 8,365 individuals and 464,311 autosomal SNPs. After imputation, all poorly imputed SNPs (RSQR < 0.30) and those with low minor allele frequency (MAF ≤ 0.01) were excluded from analyses.

We were not able to obtain 1000 Genomes imputation data for the ALSPAC cohort, however we did perform imputation using the HapMap II reference panel for cohorts 1 and 2, and obtained very similar results, suggesting our overall conclusions are not an artifact of using different imputation panels (**Tables S7, S8**).

Statistical Analysis

Association with PegQ was tested under an additive model using SNPTEST (v2.4.1, <http://www.stats.ox.ac.uk/~marchini/software/gwas/snpctest.html>) simultaneously specifying sex and age as covariates. The genomic inflation factor (λ) was calculated for directly genotyped and imputed SNPs combined, and showed no evidence of population structure, λ was 1.01 for cohort 1, 1.00 for cohort 2, 1.01 for cohort 3, and 1.00 for cohort 4. The values of PegQ were standardized in all cohorts separately to give a mean of 0, and a SD of 1. The results for the three cohorts were then meta-analyzed under a fixed-effects model with inverse variance weighting using GWAMA [92]. In total, 2,095,209 SNPs that were either directly genotyped or imputed and overlapped in all three studies and passed quality control were tested for association.

We performed pathway analysis using Meta-Analysis Gene-set Enrichment of variaNT Associations (MAGENTA) [48].

MAGENTA first maps SNPs to genes (defined as from 20 Kb upstream of the 5' UTR to 10 Kb downstream), and then selects the most strongly associated SNP, after correcting for confounding factors like gene size, the number of SNPs, linkage disequilibrium, number of recombination hotspots, in order to calculate a P value for each gene. It then ranks all gene P values and tests if there is an overrepresentation of a given set of genes above a user-specified cutoff of all genes. We chose the 75th percentile as a cutoff because it yields the best power for detecting enrichment in traits that are highly polygenic with many associations with weak effect [48]. We tested all pathways in the mouse genome informatics database (MGI, $n = 15$) that cause asymmetry defects when knocked out in mice and contain more than 10 and less than 100 genes, against a background of all genes in the MGI database (6,573 genes as of August 2012, <http://www.informatics.jax.org/> [47]). Mouse phenotypic measurements were chosen because they closely resemble observations from human clinical examination of patients with mutations in orthologous genes [93].

Finally, in order to compare relative hand skill and hand motor skill in cohorts 3 and 4, we used a one-tailed Student's T-Test, not assuming equal variance.

Supporting Information

Figure S1 QQ plot of the P values for the relative hand skill GWAS meta-analysis in individuals with reading disability (RD). The observed P values are plotted against the expected P values; a deviation from the null (diagonal line) indicates an enrichment of genuine associations. (PDF)

Figure S2 Genotype failure rate plotted against heterozygosity. Shading indicates sample density and dashed lines denote quality control thresholds. (PDF)

Figure S3 Principal component analysis. Ancestry Clustering based on principle component analysis of the original GWAS data. HapMap3 reference samples; CEU (red), JPT & CHB (purple) and YRI (green), are plotted alongside all individuals in the GWAS (black crosses). Individuals with a second principal component score less than 0.072 (black dashed line) were removed from the study. (PDF)

Figure S4 SNP call rate. Histogram of the fraction of missing data for each SNP across all individuals that passed the 'per-individual' quality control (see Materials and Methods). The dashed vertical line represents the threshold (2%) at which SNPs were removed from subsequent analysis because of an excess failure rate. (PDF)

Table S1 A list of SNPs with $P \leq 1 \times 10^{-4}$ from the relative hand skill GWAS meta-analysis in individuals with RD ($n = 728$). All the genes within 100 kb to the left or right of the SNP are listed, if no gene is present, it is denoted by a dash. EAF = Effect Allele Frequency, β = effect size (standard deviation shift per copy of the effect allele, 95 percent confidence intervals included), SE = standard error. (XLSX)

Table S2 A list of SNPs with $P \leq 1 \times 10^{-4}$ in the relative hand skill GWAS in the general population ($n = 2,666$). All the genes within 100 kb to the left or right of the SNP are listed, if no gene is present, it is denoted by a dash. EAF = Effect Allele Frequency,

β = effect size (standard deviation shift per copy of the effect allele, 95 percent confidence intervals included), SE = standard error. (XLSX)

Table S3 Fifteen phenotypes that contain between 10 and 100 genes, and cause asymmetry defects when knocked out in mice, were tested using MAGENTA for an enrichment of SNPs with low P values. The observed number of genes were compared to the expected using a cut-off that equals the 95th percentile of all gene P values (compare to the 75th percentile cutoff in Table 4). P values were then adjusted for multiple testing using an FDR correction. (XLSX)

Table S4 After removing the known association in *PCSK6* from the analysis, in the relative hand skill GWAS meta-analysis of individuals with reading disability, there remains an over-representation of highly associated SNPs within human orthologs of genes involved in causing LR asymmetry phenotypes when knocked out in mice. Fifteen phenotypes that contain between 10 and 100 genes, and cause asymmetry defects when knocked out in mice, were tested for an enrichment of SNPs with low P values. We used an enrichment cut-off that equals the 75th percentile of all gene P values. P values were then adjusted for multiple testing using an FDR correction. (XLSX)

Table S5 In the relative hand skill general population GWAS, there is an over-representation of highly associated SNPs within human orthologs of genes involved in causing LR asymmetry phenotypes when knocked out in mice. We tested four phenotypes that contain between 10 and 100 genes, causing asymmetry defects when knocked out in mice, and showed enrichment of SNPs with low P values in the RD cohorts meta-analysis. We used an enrichment cut-off that equals the 75th percentile of all gene P values. P values were then adjusted for multiple testing using an FDR correction. (XLSX)

Table S6 In the relative hand skill GWAS meta-analysis of individuals with reading disability, there is an over-representation of highly associated SNPs within human orthologs of genes involved in causing an absent corpus callosum when knocked out in mice. We tested two cerebral phenotypes that are often observed in individuals with ciliopathies for an enrichment of SNPs with low P values. We used an enrichment cut-off that equals

the 75th percentile of all gene P values. P values were then adjusted for multiple testing using an FDR correction. (XLSX)

Table S7 GWAS results based on imputations from HapMap data. A list of SNPs with $P \leq 1 \times 10^{-4}$ from the relative hand skill meta-analysis in individuals with RD ($n = 728$). All the genes within 100 kb to the left or right of the SNP are listed, if no gene is present, it is denoted by a dash. EAF = Effect Allele Frequency, β = effect size (standard deviation shift per copy of the effect allele, 95 percent confidence intervals included), SE = standard error. Results are very similar to those obtained using 1000 Genomes data for imputation (see Table S1). (XLSX)

Table S8 MAGENTA pathway analysis based on imputations from HapMap data. Fifteen phenotypes that contain between 10 and 100 genes, and cause asymmetry defects when knocked out in mice, were tested using MAGENTA for an enrichment of SNPs with low P values. P values were then adjusted for multiple testing using an FDR correction. Results are very similar to those obtained using 1000 Genomes data for imputations (see Table 4). (XLSX)

Acknowledgments

We are grateful to all the families and sample donors who took part in this study. For the ALSPAC study, we wish to thank the midwives for their help in recruiting the families, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. We thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics for the generation of genotyping data. We thank the Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and also 23andMe for generating the ALSPAC genotyping data. We'd also like to thank Alessandro Gialluisi and Clyde Francks of the Max Planck Institute for their helpful contributions with regards to the data quality control and analysis. This publication is the work of the authors and SP will serve as guarantor for the contents of this paper.

Author Contributions

Analyzed the data: WMB DME. Wrote the paper: WMB CW SEF SP. Designed research: SP WMB APMor CW SEF TSS JBT JS APMon. Were involved in the collection and processing of data: WMB TSS SP SEF JPK NJT BSP SMR DME GDS. Performed research: WMB. Supervised the study: SP CW SEF APMor APMon.

References

- Corballis MC (1993) The lopsided ape: Evolution of the generative mind: Oxford University Press, USA.
- Marchant LF, McGrew WC (1998) Human handedness: an ethological perspective. *Human Evolution* 13: 221–228.
- Raymond M, Pontier D (2004) Is there geographical variation in human handedness? *Laterality* 9: 35–51.
- McGrew WC, Marchant LF (1997) On the other hand: Current issues in and meta-analysis of the behavioral laterality of hand function in nonhuman primates. *Yearbook of Physical Anthropology*, Vol 40 - 1997 40: 201–232.
- Cashmore L, Uomini N, Chapelain A (2008) The evolution of handedness in humans and great apes: a review and current issues. *Journal of Anthropological Sciences* 86: 7–35.
- Uomini NT (2009) The prehistory of handedness: archaeological data and comparative ethology. *J Hum Evol* 57: 411–419.
- Frayner DW, Lozano M, de Castro JMB, Carbonell E, Arsuaga JL, et al. (2012) More than 500,000 years of right-handedness in Europe. *Laterality* 17: 51–69.
- Frayner DW, Fiore I, Lalueza-Fox C, Radovic J, Bondioli L (2010) Right handed Neandertals: Vindija and beyond. *Journal of Anthropological Sciences* 88: 113–127.
- Fox CL, Frayer DW (1997) Non-dietary marks in the anterior dentition of the Krapina neandertals. *International Journal of Osteoarchaeology* 7: 133–149.
- Lozano M, Mosquera M, de Castro JMB, Arsuaga JL, Carbonell E (2009) Right handedness of Homo heidelbergensis from Sima de los Huesos (Atapuerca, Spain) 500,000 years ago. *Evolution and Human Behavior* 30: 369–376.
- Amunts K, Schlaug G, Schleicher A, Steinmetz H, Dabringhaus A, et al. (1996) Asymmetry in the human motor cortex and handedness. *Neuroimage* 4: 216–222.
- Dassonville P, Zhu XH, Uurbil K, Kim SG, Ashe J (1997) Functional activation in motor cortex reflects the direction and the degree of handedness. *Proc Natl Acad Sci U S A* 94: 14015–14018.
- Herve PY, Crivello F, Percey G, Mazoyer B, Tzourio-Mazoyer N (2006) Handedness and cerebral anatomical asymmetries in young adult males. *Neuroimage* 29: 1066–1079.
- Knecht S, Dräger B, Deppe M, Bobe L, Lohmann H, et al. (2000) Handedness and hemispheric language dominance in healthy humans. *Brain* 123: 2512–2518.
- Corballis MC (2003) From mouth to hand: gesture, speech, and the evolution of right-handedness. *Behav Brain Sci* 26: 199–208; discussion 208–160.
- Bishop DVM (1990) Handedness and developmental disorder. Lavenham: Mac Keith Press. xi, 213 p. p.
- Maisog JM, Einbinder ER, Flowers DL, Turkeltaub PE, Eden GF (2008) A meta-analysis of functional neuroimaging studies of dyslexia. *Ann N Y Acad Sci* 1145: 237–259.

18. Richlan F, Kronbichler M, Wimmer H (2012) Structural abnormalities in the dyslexic brain: A meta-analysis of voxel-based morphometry studies. *Hum Brain Mapp*: doi: 10.1002/hbm.22127.
19. Shapleske J, Rossell SL, Woodruff PW, David AS (1999) The planum temporale: a systematic, quantitative review of its structural, functional and clinical significance. *Brain Res Brain Res Rev* 29: 26–49.
20. Sommer I, Ramsey N, Kahn R, Aleman A, Bouma A (2001) Handedness, language lateralisation and anatomical asymmetry in schizophrenia: meta-analysis. *Br J Psychiatry* 178: 344–351.
21. Clark GM, Crow TJ, Barrick TR, Collinson SL, James AC, et al. (2010) Asymmetry loss is local rather than global in adolescent onset schizophrenia. *Schizophr Res* 120: 84–86.
22. Dragovic M, Hammond G (2005) Handedness in schizophrenia: a quantitative review of evidence. *Acta Psychiatr Scand* 111: 410–419.
23. Francks C, Fisher SE, Marlow AJ, MacPhie IL, Taylor KE, et al. (2003) Familial and genetic effects on motor coordination, laterality, and reading-related cognition. *Am J Psychiatry* 160: 1970–1977.
24. Medland SE, Duffy DL, Wright MJ, Geffen GM, Hay DA, et al. (2009) Genetic influences on handedness: Data from 25,732 Australian and Dutch twin families. *Neuropsychologia* 47: 330–337.
25. Geschwind DH, Miller BL, DeCarli C, Carmelli D (2002) Heritability of lobar brain volumes in twins supports genetic models of cerebral laterality and handedness. *Proc Natl Acad Sci U S A* 99: 3176–3181.
26. Annett M (1985) Left, Right, Hand and Brain: The Right Shift Theory. London: Psychology Press. 488 p. p.
27. Klar AJS (1996) A single locus, RGHT, specifies preference for hand utilization in humans. *Cold Spring Harbor Symposia on Quantitative Biology* 61: 59–65.
28. McManus C (2004) Right hand, left hand: The origins of asymmetry in brains, bodies, atoms and cultures: Harvard University Press.
29. Chakravarti A (1999) Population genetics—making sense out of sequence. *Nat Genet* 21: 56–60.
30. Reich DE, Lander ES (2001) On the allelic spectrum of human disease. *Trends Genet* 17: 502–510.
31. Lander ES (1996) The new genomics: global views of biology. *Science* 274: 536–539.
32. Francks C, Maegawa S, Lauren J, Abrahams BS, Velayos-Baeza A, et al. (2007) LRRTM1 on chromosome 2p12 is a maternally suppressed gene that is associated paternally with handedness and schizophrenia. *Molecular Psychiatry* 12: 1129–1139.
33. Medland SE, Duffy DL, Spurdle AB, Wright MJ, Geffen GM, et al. (2005) Opposite effects of androgen receptor CAG repeat length on increased risk of left-handedness in males and females. *Behav Genet* 35: 735–744.
34. Hampson E, Sankar JS (2012) Hand preference in humans is associated with testosterone levels and androgen receptor gene polymorphism. *Neuropsychologia* 50: 2018–2025.
35. Eriksson N, Macpherson JM, Tung JY, Hon LS, Naughton B, et al. (2010) Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genet* 6: e1000993.
36. McManus IC, Davison A, Armour JAL (2013) Multilocus genetic models of handedness closely resemble single-locus models in explaining family data and are compatible with genome-wide association studies. *Ann NY Acad Sci* 1288:48–58.
37. Scerri TS, Brandler WM, Paracchini S, Morris AP, Ring SM, et al. (2011) PCSK6 is associated with handedness in individuals with dyslexia. *Hum Mol Genet* 20: 608–614.
38. Grande C, Patel NH (2009) Nodal signalling is involved in left-right asymmetry in snails. *Nature* 457: 1007–1011.
39. Mercola M, Levin M (2001) Left-right asymmetry determination in vertebrates. *Annu Rev Cell Dev Biol* 17: 779–805.
40. Constam DB, Robertson EJ (2000) SPC4/PACE4 regulates a TGFbeta signaling network during axis formation. *Genes Dev* 14: 1146–1155.
41. Arning L, Ocklenburg S, Schulz S, Ness V, Gerding WM, et al. (2013) VNTR Polymorphism Is Associated with Degree of Handedness but Not Direction of Handedness. *PLoS One* 8: e67251.
42. Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 5: e1000529.
43. Scerri TS, Paracchini S, Morris A, MacPhie IL, Talcott J, et al. (2010) Identification of candidate genes for dyslexia susceptibility on chromosome 18. *PLoS One* 5: e13712.
44. Francks C, Paracchini S, Smith SD, Richardson AJ, Scerri TS, et al. (2004) A 77-kilobase region of chromosome 6p22.2 is associated with dyslexia in families from the United Kingdom and from the United States. *Am J Hum Genet* 75: 1046–1058.
45. Ng A, Wong M, Viviano B, Erlich JM, Alba G, et al. (2009) Loss of glypican-3 function causes growth factor-dependent defects in cardiac and coronary vascular development. *Dev Biol* 335: 208–215.
46. Zhou J, Yang F, Leu NA, Wang PJ (2012) MNS1 is essential for spermiogenesis and motile ciliary functions in mice. *PLoS Genet* 8: e1002516.
47. Blake JA, Bult CJ, Eppig JT, Kadin JA, Richardson JE (2009) The Mouse Genome Database genotypes:phenotypes. *Nucleic Acids Res* 37: D712–719.
48. Segre AV, Groop L, Mootha VK, Daly MJ, Altshuler D (2010) Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet* 6: e1001058.
49. El Zein L, Ait-Lounis A, Morle L, Thomas J, Chhin B, et al. (2009) RFX3 governs growth and beating efficiency of motile cilia in mouse and controls the expression of genes involved in human ciliopathies. *J Cell Sci* 122: 3180–3189.
50. Haycraft CJ, Banizs B, Aydin-Son Y, Zhang Q, Michaud EJ, et al. (2005) Gli2 and Gli3 localize to cilia and require the intraflagellar transport protein polaris for processing and function. *PLoS Genet* 1: e53.
51. Yoshida S, Shiratori H, Kuo IY, Kawasumi A, Shinohara K, et al. (2012) Cilia at the node of mouse embryos sense fluid flow for left-right determination via Pkd2. *Science* 338: 226–231.
52. Constam DB, Robertson EJ (1999) Regulation of bone morphogenetic protein activity by pro domains and proprotein convertases. *J Cell Biol* 144: 139–149.
53. Yeo C, Whitman M (2001) Nodal signals to Smads through Cripto-dependent and Cripto-independent mechanisms. *Mol Cell* 7: 949–957.
54. Chazaud C, Chambon P, Dolle P (1999) Retinoic acid is required in the mouse embryo for left-right asymmetry determination and heart morphogenesis. *Development* 126: 2589–2596.
55. Gazit A, Yaniv A, Bafico A, Pramila T, Igarashi M, et al. (1999) Human frizzled 1 interacts with transforming Wnts to transduce a TCF dependent transcriptional response. *Oncogene* 18: 5959–5966.
56. Nakaya MA, Biris K, Tsukiyama T, Jaime S, Rawls JA, et al. (2005) Wnt3a links left-right determination with segmentation and anteroposterior axis elongation. *Development* 132: 5425–5436.
57. Cota CD, Bagheri P, Pelc P, Smith CO, Bodner CR, et al. (2006) Mice with mutations in Mahogunin ring finger-1 (Mgri1) exhibit abnormal patterning of the left-right axis. *Dev Dyn* 235: 3438–3447.
58. Shannon MB, Patton BL, Harvey SJ, Miner JH (2006) A hypomorphic mutation in the mouse laminin alpha5 gene causes polycystic kidney disease. *J Am Soc Nephrol* 17: 1913–1922.
59. Badano JL, Mitsuma N, Beales PL, Katsanis N (2006) The ciliopathies: an emerging class of human genetic disorders. *Annu Rev Genomics Hum Genet* 7: 125–148.
60. Watson T (1836) An account of some cases of transposition observed in the human body. *Lond Med Gazette* 18: 393–403.
61. Cockayne EA (1938) The genetics of transposition of the viscera. *Q J Med* 31: 479–493.
62. Torgersen J (1950) Situs inversus, asymmetry, and twinning. *Am J Hum Genet* 2: 361–370.
63. McManus IC, Martin N, Stubbings GF, Chung EM, Mitchison HM (2004) Handedness and situs inversus in primary ciliary dyskinesia. *Proc Biol Sci* 271: 2579–2582.
64. Kennedy DN, O'Craven KM, Ticho BS, Goldstein AM, Makris N, et al. (1999) Structural and functional brain asymmetries in human situs inversus totalis. *Neurology* 53: 1260–1265.
65. Ihara A, Hirata M, Fujimaki N, Goto T, Umekawa Y, et al. (2010) Neuroimaging study on brain asymmetries in situs inversus totalis. *J Neurol Sci* 288: 72–78.
66. Sun T, Patoine C, Abu-Khalil A, Visvader J, Sum E, et al. (2005) Early asymmetry of gene transcription in embryonic human left and right cerebral cortex. *Science* 308: 1794–1798.
67. Benadiba C, Magnani D, Niquille M, Morle L, Valloton D, et al. (2012) The ciliogenic transcription factor RFX3 regulates early midline distribution of guidepost neurons required for corpus callosum development. *PLoS Genet* 8: e1002606.
68. Johnson JM, Castle J, Garrett-Engel P, Kan Z, Loerch PM, et al. (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* 302: 2141–2144.
69. Witelson SF (1985) The brain connection: the corpus callosum is larger in left-handers. *Science* 229: 665–668.
70. Habib M, Gayraud D, Oliva A, Regis J, Salamon G, et al. (1991) Effects of handedness and sex on the morphology of the corpus callosum: a study with brain magnetic resonance imaging. *Brain Cogn* 16: 41–61.
71. Mueller KLO, Marion SD, Paul LK, Brown WS (2009) Bimanual Motor Coordination in Agenesis of the Corpus Callosum. *Behav Neurosci* 123: 1000–1011.
72. Sacco S, Moutard ML, Fagard J (2006) Agenesis of the corpus callosum and the establishment of handedness. *Dev Psychobiol* 48: 472–481.
73. Concha ML, Burdine RD, Russell C, Schier AF, Wilson SW (2000) A nodal signaling pathway regulates the laterality of neuroanatomical asymmetries in the zebrafish forebrain. *Neuron* 28: 399–409.
74. Barth KA, Miklosi A, Watkins J, Bianco IH, Wilson SW, et al. (2005) fsi zebrafish show concordant reversal of laterality of viscera, neuroanatomy, and a subset of behavioral responses. *Current biology : CB* 15: 844–850.
75. Marchini J, Howie B (2010) Genotype imputation for genome-wide association studies. *Nat Rev Genet* 11: 499–511.
76. Ivliev AE, t Hoen PA, van Roon-Mom WM, Peters DJ, Sergeeva MG (2012) Exploring the transcriptome of ciliated cells using in silico dissection of human tissues. *PLoS One* 7: e35618.
77. Massinen S, Hokkanen ME, Matsson H, Tammimies K, Tapia-Paez I, et al. (2011) Increased expression of the dyslexia candidate gene DCDC2 affects length and signaling of primary cilia in neurons. *PLoS One* 6: e20580.
78. Chandrasekar G, Vesterlund L, Hultenby K, Tapia-Paez I, Kere J (2013) The Zebrafish Orthologue of the Dyslexia Candidate Gene *DIX1C1* Is Essential for Cilia Growth and Function. *PLoS One* 8: e63123.

79. Galaburda AM, Sherman GF, Rosen GD, Aboitiz F, Geschwind N (1985) Developmental dyslexia: four consecutive patients with cortical anomalies. *Ann Neurol* 18: 222–233.
80. Galaburda AM (1989) Ordinary and extraordinary brain development: Anatomical variation in developmental dyslexia. *Ann Dyslexia* 39: 65–80.
81. Stoodley CJ, Stein JF (2006) A processing speed deficit in dyslexic adults? Evidence from a peg-moving task. *Neurosci Lett* 399: 264–267.
82. Ramus F, Pidgeon E, Frith U (2003) The relationship between motor control and phonology in dyslexic children. *Journal of Child Psychology and Psychiatry* 44: 712–722.
83. Fawcett AJ, Nicolson RI, Dean P (1996) Impaired performance of children with dyslexia on a range of cerebellar tasks. *Annals of Dyslexia* 46: 259–283.
84. Iversen S, Berg K, Ellertsen B, Tonnessen FE (2005) Motor coordination difficulties in a municipality group and in a clinical sample of poor readers. *Dyslexia* 11: 217–231.
85. Nicolson RI, Fawcett AJ (1990) Automaticity - a New Framework for Dyslexia Research. *Cognition* 35: 159–182.
86. Getchell N, Pabreja P, Neeld K, Carrio V (2007) Comparing children with and without dyslexia on the movement assessment battery for children and the test of gross motor development. *Perceptual and Motor Skills* 105: 207–214.
87. Golding J, Pembrey M, Jones R (2001) ALSPAC—the Avon Longitudinal Study of Parents and Children. I. Study methodology. *Paediatr Perinat Epidemiol* 15: 74–87.
88. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, et al. (2012) Cohort Profile: The 'Children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol* 42:111–27.
89. Henderson SE, Sugden DA (1992) Movement Assessment Battery for Children manual. Sidcup: The Psychological Corporation.
90. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, et al. (2010) Data quality control in genetic case-control association studies. *Nat Protoc* 5: 1564–1573.
91. Delaneau O, Marchini J, Zagury JF (2012) A linear complexity phasing method for thousands of genomes. *Nat Methods* 9: 179–181.
92. Magi R, Morris AP (2010) GWAMA: software for genome-wide association meta-analysis. *Bmc Bioinformatics* 11: 288.
93. Webber C, Hehir-Kwa JY, Nguyen DQ, de Vries BB, Veltman JA, et al. (2009) Forging links between human mental retardation-associated CNVs and mouse gene knockout models. *PLoS Genet* 5: e1000531.

For copyright reasons the article formerly located here has been removed.
Brandler, W. M. & Paracchini, S. (2014). 'The genetic relationship between handedness and neurodevelopmental disorders', Trends in Molecular Medicine, 20 (2), 83-90. Available at: <http://dx.doi.org/10.1016/j.molmed.2013.10.008>