Handbook of Language Development in a Social Context
Publisher: Cambridge University Press.
Editors: James Law, Sheena Reilly and Cristina McKean

CHAPTER 3

Genetic studies of language disorders

Hayley Mountford[1], Ruth Braden[2], Angela Morgan[2], Dianne Newbury[1]

Oxford Brookes University[1], Murdoch Children's Research Institute[2]

Corresponding email address: hmountford@brookes.ac.uk

Introduction

The ability to learn complex language is arguably the most distinguishing evolutionary feature of the human species. Children commence learning language prior to birth, and while most children continue on a typical trajectory of language development, a subset of children fail to acquire language abilities at the same level as their peers. These developmental language disorders (DLD) have long term effects on social emotional skills, literacy, employment and wellbeing. Despite years of research, we continue to lack critical understanding of the biological underpinnings of language. Language disorders appear to be highly heritable, with a complex interaction of genetic and environmental factors thought to be the cause. Unlike childhood apraxia of speech (CAS), which commonly co-occurs with DLD and has been associated with multiple single gene causes, no monogenic cause has been identified to date. Linkage and association analyses have implicated a handful of genes in DLD populations, although small sample sizes have thus far restricted the number of candidate genes identified. Current heterogeneity within and between available cohorts makes comparison and replication of results difficult to achieve.

This chapter provides an overview of the genetic bases of developmental language disorders, with an emphasis on the importance of defining the phenotype of DLD to inform gene discovery. As such, the first part of this chapter is dedicated to reviewing the specific features (the *phenotype*) of DLD in the genetic literature. The influence of historic variation in

diagnostic inclusion criteria on researchers' abilities to compare and replicate genotype-phenotype studies will also be discussed.

The second part of this chapter is concerned with genes implicated in DLD. An exhaustive list of possible gene pathways and genetic syndromes associated with DLD is beyond the scope of this chapter. Rather, this chapter will (i) provide an overview of more recently identified gene pathways in populations with DLD; (ii) discuss multiple approaches to genetic analysis based on the hypothesized architecture of DLD and (iii) discuss the influence of epigenetics in this field of research. This chapter emphasizes the important role of estimating the genetic architecture of DLD in order to decipher underlying genetic associations.

Phenotype of Developmental Language Disorder

Language is a multidomain skill, encompassing both receptive and expressive abilities across the domains of morphology, phonology, syntax, semantics and pragmatics. Atypical language is associated with a range of aetiologies such as (i) acquired brain injury (e.g., stroke; Liégeois et al., 2019), traumatic brain injury (Haarbauer-Krupa et al., 2019; Liégeois et al., 2013; Catroppa and Anderson, 2004; Ewing-Cobbs and Barnes, 2002) and brain tumour (Docking et al., 2016); (ii) neurodevelopmental conditions, such as Cerebral Palsy (Mei et al., 2016), Autism Spectrum Disorder (Brignell et al., 2018) intellectual disability and Attention Deficit Hyperactivity Disorder (Sciberras et al., 2014); known genetic syndromes (St John et al., 2019; Morgan et al., 2015; Cleland et al 2010; White et al., 2010) and (iii) hearing impairment (Wake et al., 2005), to name but a few. By contrast, the term 'idiopathic' language disorder is used to describe children with language impairments not associated with any known cause such as those listed earlier. The potential genetic basis of idiopathic language disorder is the focus of this chapter. Historical terminology for this condition has included Specific Language Impairment (SLI), while recent nomenclature defines these conditions as Developmental Language Disorders (DLDs; Bishop et al., 2017; Bishop et al., 2016) (See Chapter 1 for further discussion of terminological history). DLDs can be persistent in nature, leading to long-term functional limitations on literacy, social skills, employment and emotional wellbeing, and overall lower quality of life (Eadie et al 2018; Beitchman et al., 2014; Bretherton et al., 2014; Clegg et al., 2005; McKean et al., 2017; St Clair et al., 2011). Prevalence estimates of DLD range between 7 to 19% of children affected in the pre-school and early school years (McKean et al., 2017; Reilly et al., 2010; Tomblin et

al., 1997). Both genetic and environmental risk and protective factors are thought to influence a child's acquisition and trajectory of language development (Morgan et al., 2015; Reilly et al., 2010; Reilly et al., 2007), leading to vast heterogeneity amongst DLD populations. This chapter will focus on the current understanding of the genetics of language disorder.

Symptomatology and Diagnosis of Developmental Language Disorder

As described earlier, Specific Language Impairment (SLI) is the historical term previously used to describe deficits in language ability relative to non-verbal intelligence, in children with idiopathic language disorder (Bishop et al., 2017; Bishop et al., 2016). Yet there has never been a clear consensus regarding the specific level of language deficit or non-verbal ability required for an SLI diagnosis. Most commonly in the SLI research literature, language testing scores greater than 1.25 standard deviations below the mean and Performance IQ scores of 85 or higher were suggested to be the required discrepancy for SLI diagnosis (Reilly et al., 2014). This differed from specified cut-off criteria in other diagnostic manuals, such as that of the ICD-10 (WHO, 2010), where diagnosis was determined based on language abilities greater than two standard deviations below the mean and at least one standard deviation lower than non-verbal IQ. In 2017, a Delphi study led by the CATALISE Consortium recommended *Developmental Language Disorder* (DLD) as the preferred diagnostic terminology for cases of language impairment where a child demonstrated a deficit in language ability, not associated with another biomedical aetiology (Bishop et al., 2017; Bishop et al., 2016).

Young children who are diagnosed with DLD typically present with a developmental history of delayed communication milestones such as the age of first words, production of two-word combinations and sentence generation. As these children reach primary school age, DLD generally manifests as low scores on standardised measures of receptive and expressive language, including vocabulary, grammar, reading and pragmatics (Conti-Ramsden et al., 2001).There is high comorbidity of DLD with other developmental conditions, including speech sound disorders (estimated comorbidity between 11-77%, discussed in Shriberg et al. 1999; 40.8% in Eadie et al. 2018) and literacy disorders such as dyslexia (estimated comorbidity between 17-29%; Catts et al., 2005). This comorbidity could be suggestive of shared genetic aetiologies between these conditions. In this next section, we briefly discuss phenotyping approaches used to characterise cohorts of children with DLD/SLI in studies

examining the genetic contribution of the condition. A more detailed explanation of the specific gene pathways associated with DLD will then be discussed in Part 2 of this chapter.

Phenotyping in genetic studies of Developmental Language Disorder

The historical changes in nomenclature and a lack of consensus over diagnostic criteria for DLD have led to varied phenotyping approaches in studies examining the genetic architecture of this condition. A summary of the variation in phenotypic inclusion criteria for key genetic studies of SLI/DLD is presented in Table 3.1.

This variation makes comparison and replication of results across genetic studies challenging. Further, it hinders the amalgamation of cohorts to increase sample sizes to lead to adequately powered meta-analyses. Clearly defined diagnostic criteria for language disorders is a critical aspect of methodological design for investigative genetic studies of DLD. The following section details genetic analysis approaches used in the field to date, and gene pathways revealed to be associated with DLD using these methods.

Authors	**Study type**	**Nomenclature**	**Inclusion criteria: Language**	**Inclusion criteria: Cognition**	**Exclusion criteria**
Bishop et al. (1995)	Twin study	SLI	Based on DSR-III-R criteria: SS ≤80 on language measure. Must have significant impairment on ≥1 of 4 language measures.	Discrepancy of ≥20 points between non-verbal IQ and language measure.	Mental retardation; ASD; SNHL; structural abnormality of articulators; serious visual impairment; medical syndrome; EAL status.
Bartlett et al. (2002)	Linkage study	SLI	Spoken Language Quotient SS ≤85	Performance IQ ≥80 + Performance IQ ≥ Spoken Language Quotient	Hearing impairment; Motor impairments or oral structural deviations affecting speech or non-speech movement of the articulators; Diagnosis of ASD, schizophrenia, psychoses, or neurological disorder.
Falcaro et al. (2008)	Linkage	SLI	Part of parent longitudinal study (Conti-Ramsden & Botting 1999; Conti-Ramsden et al. 1997) + Attending language units in United Kingdom + Language SS≤1SD at 1 time point during longitudinal parent study	Performance IQ ≥80	Sensorineural hearing loss; EAL status; Medical condition likely to affect language; ASD diagnosis.
Newbury et al. (2009)	Linkage study	SLI	* CELF-R expressive or receptive SS ≥1.5SD below normative mean	*PIQ≥80	*MZ twinning, chronic illness requiring multiple hospital visits or admissions, deafness, an ICD-10/DSM-IV diagnosis of childhood autism, EAL, care provision by local authorities, and known neurological disorders

Villanueva et al. (2011)	GWAS	SLI	Phonology and expressive and receptive morphosyntax SS>2 SD below population mean on	PIQ>80th percentile	Hearing impairment; oral motor or structural; Diagnosis of ASD, emotional difficulties, or neurological disorder.
Luciano et al. (2013)	GWAS (Population)	Quantitative language across population	Population study, low language determined based on non-word repetition tasks.	-	-
Eicher et al. (2013)	GWAS	Language Impairment (+/- RD)	z-score ≤-1 on ≥2 of 3 language tasks (phoneme deletion, verbal comprehension, non-word repetition)	IQ ≥76	-
Gialluisi et al. (2014)	GWAS	Language Impairment (+/- RD)	3 cohorts with varied inclusion criteria: *Cohort 1:* SLIC * CELF-R expressive or receptive SS ≥1.5SD below normative mean. *Cohort 2*: UK Reading Disability (UK-RD): diagnosis RD. *Cohort 3*: Colorado Learning Disabilities Research Centre (CLDRC): 2 datasets, one recruited on basis of diagnosis of RD, one on diagnosis of ADHD. Language SS≥3SD sample mean.	*Cohort 1*: * PIQ≥80 *Cohort 2: Reading IQ discrepancy and/or IQ>90* *Cohort 3*: FSIQ ≥70	*Cohort 1*: *MZ twinning, chronic illness requiring multiple hospital visits or admissions, deafness, an ICD-10/DSM-IV diagnosis of childhood autism, EAL, care provision by local authorities, and known neurological disorders *Cohort 2:* - *Cohort 3:* If ≥3 SS were ≥3SD from mean
Harlaar et al. (2014)	GWAS (Population)	Quantitative language across population	Population study, low language determined using receptive language measures included in the cognitive test battery.	-	-
St Pourcain et al. (2014)	GWAS	Quantitative language across population	Population study, low language determined using MCDI.	-	-
Nudel et al. (2014)	GWAS	SLI	* CELF-R expressive or receptive SS ≥1.5SD below normative mean	* PIQ≥80	*MZ twinning, chronic illness requiring multiple hospital visits or admissions, deafness, an ICD-

					10/DSM-IV diagnosis of childhood autism, EAL, care provision by local authorities, and known neurological disorders
Evans et al. (2015)	Linkage study	Poor language	Recruited from a longitudinal language study. Overall language composite score calculated based on 3 composite language scores representing overall language, vocabulary and sentence use.	Performance IQ >70	-
Kornilov et al. (2016)	GWAS Isolated population ~400	DLD	Met impairment criterion (z-score<-1) on ≥2 quantitative phenotypes obtained via analysis of semi-structured speech samples.	-	Children attending specialist education settings.
Devanna et al. (2017)	Sequencing study	SLI	* CELF-R expressive or receptive SS ≥1.5SD below normative mean	* PIQ≥80	*MZ twinning, chronic illness requiring multiple hospital visits or admissions, deafness, an ICD-10/DSM-IV diagnosis of childhood autism, EAL, care provision by local authorities, and known neurological disorders
Chen et al. (2017)	Sequencing study (SLIC Cohort)	Severe SLI	* CELF-R expressive or receptive SS ≥1.5SD below normative mean	*PIQ≥80	*MZ twinning, chronic illness requiring multiple hospital visits or admissions, deafness, an ICD-10/DSM-IV diagnosis of childhood autism, EAL, care provision by local authorities, and known neurological disorders

Table 3.1: Summary of phenotypic inclusion criteria for key studies of DLD/SLI. * = Study used SLIC cohort criteria; ADHD = Attention Deficit Hyperactivity Disorder; ASD = Autism Spectrum Disorder; EAL = English as an Additional Language; IQ = Intelligence Quotient; MCDI = MacArthur-Bates Communicative Development Inventory; RD = Reading disorder; SD = Standard Deviation; SNHL = Sensorineural Hearing Loss; SS = Standard Score.

The genetics of Developmental Language Disorder

Heritability of DLD

Contribution of twin studies for estimating heritability

Across most conditions, the earliest models to interrogate strength of inheritance were twin models. If a trait is strongly genetic in origin, the concordance between monozygotic twins will be 100%, and dizygotic around 50%. If a trait is entirely due to shared environment, then the concordance between monozygotic twins will be very similar to that of the dizygotic twins as we assume that each twin pair is given a similar environment in which to grow up and develop.

Bishop et al (1995) conducted a seminal early twin study in the field. She examined the heritability of SLI in 90 pairs of twins where at least one twin met the criteria for SLI. They found that the concordance (where both twins met the criteria for SLI) in monozygotic twins (identical twins who share 100% of their DNA) was almost 100%, whereas dizygotic twins (who share on average 50% of their DNA as with regular siblings) concordance was around 50%. These early findings demonstrated a strong inherited contribution to SLI. A number of subsequent studies have supported this early finding (Lewis and Thompson, 1992, Hayiou-Thomas et al., 2005) and today DLD is widely accepted to consistently show heritability of around ~75%. This heritability rate is considered extremely high for a complex and multifactorial disorder.

Further evidence of the inherited basis of language disorders was shown by Stromswold (1998) who found the risk of language disorder increases if a first degree relative also has a language disorder. Language disorders appear more heritable than general language ability within the normal range across the general population, perhaps suggesting that these driving genetic factors have more impact upon risk of disorder than on typical language ability (Dale et al., 1998, Spinath et al., 2004). These studies illustrate the importance of correct diagnosis of language disorder as they indicate that the effects of genetic factors may differ between subsets of individuals.

Complexity of language disorders

Heritability studies have clearly shown the important contribution of genetics to language disorders, and to a lesser extent to language ability in general. However, these studies rely upon a falsely binary idea of genetics, our inherited material, and environment, which

encapsulates everything else. This is an oversimplification of an enormously complex and dynamic system that changes over time in response to subtle signals that we are only beginning to understand. Furthermore, these approaches are limited in that they do not allow us to ascertain which variants, genes, molecular pathways or cell types are driving these complex systems.

Despite there being a strong genetic contribution, the majority of language disorders cannot be classified as 'genetic' or 'environmental'. Conceptually, language disorders are separated into two groups reflecting the expected underlying genetic contributions. In monogenic disorders, a single change in the DNA sequence is necessary and sufficient to cause disorder. All individuals carrying the variants will develop the disorder. It should be noted however, that even in these rare cases, the genetic background and environment can still play a role and are able to modify the clinical presentation of that variant. Most cases of DLD are considered as complex disorders. This means that they involve many inherited genetic variants that together confer a susceptibility (or risk), while being influenced by environmental factors. Within such complex disorders, variations in the genetic sequence are viewed as 'risk factors'. The more genetic risk variants an individual carries, the more likely they are to develop DLD. Each of these risk variants usually has only a very small effect and increase in risk. This genetic 'risk' can be further modified by the environment, which may amplify or diminish the genetic risk. For example, environmental factors such as socioeconomic status (Locke et al., 2002, Law et al., 2011) may interact with genetic 'risk' to put a child at a higher risk of DLD. Complex disorders are sometimes described as a 'perfect storm' of circumstances in which each risk matrix is personalised to the individual (Virgin and Todd, 2011).

The advantage of conceptualising language disorders in this monogenic versus complex way is that is that the assumed genetic architecture can guide gene identification approaches. The identification of genes underlying monogenic disorders has traditionally been performed within large families and, more recently through the sequencing of trios (an affected individual and both parents - explained in more detail in the following section). If a disorder is assumed to be genetically complex, then association methods may be more appropriate.

Mendelian Genetics Primer

To fully understand how these genetic models work, we must first briefly cover the genetics of inheritance. The study of families has been the traditional starting point to look for variants

that cause disease, leading to the identification of genes and molecular pathways involved in language. It is based on the premise that related individuals share their DNA with their family members. A child shares half of their DNA with their mother, half with their father, and, on average half with each sibling. In families where there is an inherited disease or disorder, this information can be used to find the variant that is causing the disease – termed a variant, mutation, or pathogenic allele. Genetic studies typically focus on large families where a severe disorder is inherited in a characteristic way, known as 'Mendelian disorders' because they show recessive or dominant inheritance patterns first recorded by Gregor Mendel in the 1800's.

We each carry two copies of our DNA. One copy is inherited from our Father and the other from our Mother. Across each copy of DNA there are sites in the sequence that typically vary between individuals. These sites can be easily characterised using DNA sequencing methods which can characterise every single base pair of every single chromosome (whole genome sequencing) or every single base pair of every single gene (i.e. not the 'spacer' DNA found between genes - exome sequencing). Large-scale sequencing projects show that on average, we each carry around 1,000,000 sequence variants that differ from the common version found within our ethnic grouping. It is expected that about 2,000 of these will have an effect of the protein they encode (The 1000 Genomes Project Consortium, 2015). Most of these variants are considered to be harmless and make us who we are. In a monogenic disorder, where a given variant is considered both necessary and sufficient to cause disorder, therefore the primary challenge is to figure out which of these potential 2,000 variants that may have a function are directly involved in our disease of interest.

Inherited Family Models

There are two common forms of inheritance that are most relevant to monogenic language disorders; recessive and dominant. Recessive inheritance requires both copies of a gene to be non-functional for disorder to occur; as long as some protein is produced (even half), the cell can function. In this instance, both parents usually carry a single deleterious variant. They do not present with a disorder as both copies of the gene need to be disrupted to have an effect. But if a child inherits both dysfunctional gene copies disorder will occur even though the parents were unaffected. In contrast, dominant disorders occur in the presence of a single dysfunctional gene copy. In this case one functional copy of the gene is not enough for the cell to function properly. In dominant disorders, affected children will usually inherit the

disorder from an affected parent who also carries the dysfunctional gene. However, an interesting feature of dominant disorders is that inheriting a dominant variant may not always result in the disease phenotype. This is termed 'incomplete penetrance'. A variant is said to be 'fully penetrant' when inheriting the variant always results in the disease. Fully penetrant dominant conditions are rare and in practice most dominant disorders show incomplete penetrance, often making them difficult to identify.

De Novo Inheritance Model

Although most pathogenic variants are inherited within the models described above, an alternative does exist; *de novo* variants occur spontaneously in an embryo rather than being inherited. During foetal development, all individuals acquire new variants that become fixed within that individual. It is estimated that approximately 100 so-called *de novo* variants occur in every individual. These variants are not inherited from parents but can alter the risk of disorder if they have a significant effect upon protein function. The *de novo* sequencing paradigm focuses upon families in which a child is affected by a severe neurodevelopmental disorder for which there is no family history - the trio sequencing approach of the affected child and both unaffected parents. The paradigm asserts that such patterns of disorder are likely to arise from *de novo* changes that affect gene function. This is a powerful approach because the number of *de novo* variants that fit this expectation is usually low allowing the effective narrowing of the dataset.

Rare variants in Language Disorders

The best characterised and most studied mode of inheritance is the Mendelian inheritance of rare variants, showing either recessive (two copies of the variant) or dominant inheritance (one copy of the variant, where one working copy is not enough). The types of causative variants in both recessive and dominant inheritance patterns tend to be very rare in the general population; usually found in less than 1% of the population, but often more like 0.1%, or never previously reported. In addition, these rare variants have a detrimental impact upon the cell i.e. a large effect size resulting in a clearly defined phenotype (the top left of the infographic in Figure 1). The best example of a rare Mendelian variant in language disorders is the *FOXP2* gene where a variant led to childhood apraxia of speech (CAS) phenotype in the KE family (Lai et al., 2001). CAS is a "deficit in speech-motor programming and/or planning that affects a child's ability to perform the spatiotemporal parameters of movement sequences, resulting in errors in speech sound production and prosody" (American Speech-

Language-Hearing Association, 2007, p 1.). This dominantly inherited variant was shared by affected family members but never by unaffected members, and rather unusually was found to be fully penetrant where all carriers were affected by CAS. Variants in *FOXP2* are a well characterised cause of the CAS phenotype with many independent cases being described in the literature (MacDermot et al., 2005, Moralli et al., 2015, Reuter et al., 2017, Tomblin et al., 2009, Turner et al., 2013, or for review see Morgan et al., 2017). Individuals with reduced levels of the FOXP2 protein always have expressive language difficulties with delayed and unintelligible speech but typical non-verbal intelligence, fitting the diagnosis of CAS (Morgan et al 2017). Despite the apraxic presentation, affected individuals have intact gross- and fine-motor skills. Although highly penetrant, *FOXP2* mutations are extremely rare; it is estimated that they account for about 2% of CAS cases (MacDermot et al., 2005) and are not thought to contribute to other forms of language disorder (Eising et al., 2018).

The example of *FOXP2* is an extremely unusual case, where a fully penetrant and dominant variant results in a clearly defined phenotype, and most unusually in a large multi-generational family. No other studies of rare variants in speech and language disorders have been able to replicate this exceptional finding. With the advent of genomic sequencing technologies, studies have begun to apply both exome sequencing and *de novo* sequencing approaches to speech and language disorders. After the identification of *FOXP2*, a proliferation of studies naturally focused upon known partners or targets of the *FOXP2* gene such as *CNTNAP2* and *FOXP1* (Hamdan et al., 2010, Horn et al., 2010, O'Roak et al., 2011, Bacon and Rappold, 2012, Srivastava et al., 2014, Lozano et al., 2015, Sollis et al., 2015). Since these genes are known to function in similar cellular roles to FOXP2, they represent attractive candidates for speech and language-related traits. While such studies uncovered relevant cases, these were affected by widespread neurodevelopmental difficulties, often accompanied by global developmental delay rather than the specific speech and language disorders attributed to *FOXP2* disruption (Sollis et al., 2015).

Current approaches to gene identification utilise high-throughput sequencing. The strength of these studies lies in their ability to evaluate variation at a genome-wide level, providing complete sequencing data across all known genes. Although it is necessary to filter the large number of variants catalogued, the targeting of already known candidate genes restricts the potential to identify new candidate genes. As population sequence data become more readily available, filtering often relies upon the identification of rare variants with predicted deleterious effects. One recent example of this was a study by Chen et al. (2017) who

performed whole exome sequencing on 43 individuals with severe forms of SLI. They identified three individuals with variants in genes (*ERC1*, *GRIN2A* and *SRPX2*) that fully explained their SLI status. They also identified a number of variants in new candidate genes, showing an overrepresentation of variants in genes involved in molecular pathways already linked to neurodevelopmental disorders. Interestingly, many of the identified variants showed incomplete segregation, meaning that unaffected family members may carry them. This suggests that these rare variants maybe risk factors rather than Mendelian variants *per se*.

Eising et al. (2018) applied the *de novo* paradigm described above in 19 CAS cases allowing the identification of rare de novo variants in the *CHD3*, *SETD1A* and *WDR5* genes. The relevance of *CHD3* has since been confirmed in a follow-up study that collated 34 cases with *CHD3* variants (Snijders Blok et al., 2018). Loss-of-function variants in this gene lead to speech and language deficits accompanied by macrocephaly often in the presence of severe neurodevelopmental difficulties (Snijders Blok et al., 2018). Interestingly, CHD3 has previously been shown to interact with the FOXP2 protein indicating that shared pathways may exist between candidate genes (Estruch et al., 2016). It therefore follows that the investigation of shared functional pathways represents a viable approach to gene identification. Accordingly, Eising et al. (2018) characterised functional pathways in brain development allowing the identification of a further five candidates variants in their CAS cases, namely *KAT6A*, *SETBP1*, *ZFHX4*, *TNRC6B* and *MKL2*. Some of these genes again overlap with previous studies and candidates. *SETBP1* was identified as a candidate gene in a screen of Russian families affected by SLI and a variant was also identified in an unrelated CAS case (Kornilov et al., 2016). Disruption of *SETBP1* has previously been described in individuals with motor delay and intellectual disability. Such overlaps illustrate the importance of a growing body of evidence; the more comparable datasets available, the easier it is to spot patterns between cases. This is particularly important when individual variants are rare.

There are a large number of Mendelian rare variants that have been shown to cause language disorders, and a complete review is outside the scope of this chapter. These are reviewed in detail by Fisher and Scharff (2009), Graham and Fisher (2013), Barnett and van Bon (2015), and most recently Mountford and Newbury (2018).

The Role of Rare Variants in Complex Disease

Traditionally, geneticists have thought about genetically influenced disorders as either Mendelian or complex in nature, and considered them to be binary. This arbitrary distinction reflects the tools used to study them; family-based small scale studies for Mendelian disorders to identify high effect size and extremely rare variants, and larger cohorts of cases and controls to identify shared common variants or region which confer a moderate effect size. It is currently extremely technically challenging to identify common or rare variants which only confer a small (or even moderate) effect size, as none of our current genetic approaches have the resolution to detect these smaller contributory effects. In practice, very few disorders actually fit this binary division.

The landscape on rare variants is slowly transforming, and they are increasingly being seen as relevant to complex disorders and contributing to an individual's overall risk of developing a disease phenotype (Blair et al., 2013, Franic et al., 2015). This way of thinking is highlighted by a recent paper on the contribution of rare variants to ASD (Constantino, 2018), exploring the way in which combinations of rare variants contribute to a shift in risk matrix.

Recently, a study looked at the contribution of rare variants to two traits body mass index (BMI) and height traditionally thought of as complex traits (Wainschtein et al., 2019). Heritability for BMI was estimated at 0.4 and therefore 0.6 due to environment, and 0.79 for height with 0.21 due to environment. Staggeringly, they found that rare variants with allele frequencies of between 0.001% and 1% in the population account for 0.54 of heritability for height and 0.51 for BMI, respectively. This astonishing study shows that rare variants account for a large proportion of the 'missing heritability' that we encounter in complex disease, and that they are as relevant to complex disease as to Mendelian diseases. A recent, large GWAS study found that the type of variant is highly related to the subtype of ASD: *de novo* variants were more common in individuals with ASD and intellectual disability compared to those with ASD only, and that that common variants appeared to contribute more to high functioning subtypes of ASD (Grove et al., 2019). Their findings suggest that rare variants likely play a bigger role in severe disorder, whereas common variants are present in less severe phenotypes.

Similarly, genetic studies show the overlap between other disorders and often implicate a gene in more than one subtype of neurodevelopmental disorder suggesting a shared genetic aetiology. The principal example of this is the gene *CNTNAP2*, which was first identified as a

gene of interest because it is regulated by FOXP2 (Vernes et al., 2008). The *CNTNAP2* gene has been associated with performance in language-related tasks in families with language disorders (Devanna et al., 2017), and its role in increased risk of neurodevelopmental disorder replicated in both population (Vernes et al., 2008, Whitehouse et al., 2011), dyslexia (Peter et al., 2011, Newbury et al., 2011), and ASD (Alarcon et al., 2008, Arking et al., 2008, Centanni et al., 2015). This substantial phenotypic overlap shows that a single gene can play a role in many disorders, and therefore we may need to consider neurodevelopmental disorders as a whole rather than sub-categorising into specific phenotypes. It also suggests that FOXP2-related pathways are likely to be involved in a more global range of neurodevelopmental processes, and not just specific to language development.

Copy Number Variants (CNVs)

Copy number variants (CNVs) are submicroscopic deletions or duplications of genetic material that range from a few hundred bases of DNA to entire chromosomes. As well as carrying a number of variants in our DNA, we also each carry between six and ten CNVs. Some of these are inherited from our parents, but some of them are not inherited and occur sporadically in an individual's DNA (termed *de novo*). CNVs have been definitively linked to developmental disorders (Beckmann et al., 2007) particularly large and rare CNVs have been found to be enriched in individuals with autism (Sanders et al., 2015), intellectual disability (Coe et al., 2014) and attention deficit hyperactivity disorder (Lionel et al., 2011).

Overall burden of CNVs has been shown to play a role in DLD. Simpson et al. (2015) detected increased CNV burden in cases and their unaffected relatives, again in the SLIC Cohort. More recently, (Kalnak et al., 2018) found that DLD individuals harboured more and larger rare CNVs that the typically developing controls. No repetitive common CNVs account for the burden, and CNVs tend to be individually rare. These findings indicate that an increased burden of CNVs may play a key role in mediating an increased risk of language disorder.

The potential effect of an individual CNV can be difficult to determine and it depends upon which genes are affected, and how much it impacts upon the function of those genes. Many CNVs are tolerated and have little to no discernible effect on cellular function, whereas some are extremely damaging and result in clear genetic conditions. These micro-deletion/-duplication syndromes often affect global neurodevelopment and have been associated with severe speech and language disorders. In particular, deletions of chromosome 16p11 have

been associated with a penetrant form of CAS (Raca et al., 2013, Newbury et al., 2013, Fedorenko et al., 2016) although these apraxic features are usually accompanied by other neurobehavioral deficits (Mei et al., 2018).

Disruptions of the *FOXP2* gene by deletion or duplication invariably lead to CAS (Zeesman et al., 2006, Feuk et al., 2006, Lennon et al., 2007). Other rare CNVs reported in single isolated CAS cases have led to the identification of new candidate genes *BCL11A* (Peter et al., 2014), *ERC1* (Thevenon et al., 2013) and *SEMA6D* (Ercan-Sencicek et al., 2012), and validated through later sequencing studies (Chen et al., 2017, Soblet et al., 2018), although their specific role in language disorders is yet to be elucidated. Similarly, Morgan et al., (2018) showed that patients with Koolen de Vries Syndrome (Koolen et al., 2004) caused by either a 17q21.31 microdeletion or variants in the *KANSL1* gene lead to striking CAS and dysarthria. The frequency of CNVs involved in language disorders may range from extremely rare through to fairly common. It is therefore difficult to place them on the infographic (Figure 3.1) as the effect size and frequency of each CNV needs to be considered individually.

Common Genetic Model

We have come to understand that the inheritance model for DLDs likely involves a number of variants which together interact with environment to contribute to genetic susceptibility or risk. Termed the 'complex genetic model' it means that each variant contributes to an individual's overall level of risk of developing a language disorder.

Genetics studies of common variants in complex diseases broadly follow two approaches: linkage studies and genome-wide association studies (GWAS). Linkage studies are based on identifying chromosome regions where individuals who are more similar at the genetic level are more similar at the phenotypic level. This type of genetic study was popular in the early 2000's and has been highly fruitful in the identification of genes involved in language disorders which we will explore in the next section. GWAS analyses identify genetic variants that are more common in affected cases than in unaffected controls. Both of these approaches assume that the disease is caused by a small number of shared genes, and that the study participants are relatively similar to each other i.e. the same ethnicity or from the same geographical region.

Linkage analyses

Linkage studies are used to identify regions of the genome which are shared between affected individuals and/or family members. Within the complex genetic model of common variants, linkage studies are well suited to the identification of common variants present in more than 10% of the population and with a moderate to large effect size. Their particular strength is the identification of shared variants that may be common within a particular population and contributing to a common disorder. Major examples of successful linkage studies in SLI/DLD come from the Specific Language Impairment Consortium (SLIC) cohort (S. L. I. Consortium, 2002). By looking for shared regions of DNA across the whole genome, the authors were able to identify a strong association to between the 16q24 (called SLI1) chromosomal region and nonword repetition, and also the 19q13 region (SLI2) and the CELF-R measure of expressive language (shown in Figure 2A). Further refining of these regions identified clear signals in two genes; C-mad inducing protein (*CMIP*) and calcium-transporting ATPase type 2C member 2 (*ATP2C2*) (Newbury et al., 2009). Both of these contain common alleles which confer a moderate effect size of increased risk (Figure 1). Additional studies have further supported the role of *CMIP* (Van der Aa et al., 2012) and *ATP2C2* (Smith, 2011) in language disorders through the identification of cases carrying CNVs.

Interestingly, Newbury et al. (2009) found that *CMIP* was associated with language, reading and spelling in the SLIC cohort and the general population perhaps contributing to phonological language skills across the entire range of language ability. In contrast, *ATP2C2* was associated with language measures in the SLIC cohort, but only showed association within the language-impaired group in the general population.

This suggests that *ATP2C2* may be associated with language ability specifically in language impaired individuals. This idea of a gene having different roles is termed ‘pleiotropy’, meaning a gene may have several phenotypes associated with it which can express under certain circumstances. See Scerri et al. (2011) and Newbury et al. (2011) for further reading on pleiotropy.

A small number of other linkage studies have been able to successfully identify candidate regions. Bartlett et al. (2002) performed an association on five large Canadian families with a history of familial SLI. Both language and reading phenotypes were tested for association. Chromosomal region 13q21 showed the strongest association with a reading-impaired

phenotype, and two additional regions 2p22 and 17q23 showed more moderate associations to the overall phenotype associated language delayed phenotype (Figure 2B). Most recently, Evans et al. (2015) examined 147 sibling pairs where at least one sibling had a diagnosis of SLI. The strongest association was identified for phonological memory at chromosome regions 10q23.33 and 13q33.3 (Figure 2C).

Linkage studies are notoriously difficult to replicate in other populations. Particularly with disorders as heterogeneous as DLD, it is not unexpected that a single or small number of chromosomal regions consistently show association with language disorder phenotypes (Reader et al., 2014). The power to statistically detect a genetic signal in this way is extremely limited if that signal is not shared consistently between family units.

Genome-Wide Association Studies

A more common approach to identifying genomic regions associated with DLD is to perform a genome-wide association study (GWAS). GWAS studies use up-to-date genetic technologies to look for shared variants in affected participants compared to unaffected controls. The density of markers in this method gives a finer genomic resolution, often simultaneously looking at 300,000 variants instead of the thousands used in linkage studies. Similarly, the number of study participants tends to be an order of magnitude larger than seen in linkage studies. A recent study GWAS on schizophrenia detected over 100 novel regions implicated in the disease (Ripke et al., 2014). To achieve this impressive level of resolution, they recruited 37,000 schizophrenia cases and 113,000 unaffected controls. None of the GWAS performed of DLDs or related phenotypes have been on this scale, however, the community recognises that a concerted and collaborative effort needs to be made to achieve such a thing. A recent GWAS of ASD tested 18,000 Danish individuals in the iPSYCH consortia with a diagnosis of ASD, and almost 28,000 unaffected controls (Grove et al., 2019). They were able to identify a number of novel loci which overlapped those identified by previous GWAS studies into schizophrenia, depression and educational attainment strongly suggestive of neuropsychiatric and neurodevelopmental disorders aetiological commonality.

A number of GWAS have been performed on SLI/DLD, and overlapping phenotypes, and the identified genetic regions of are summarised in Table 3.2 (evaluated and reviewed by Carrion-Castillo et al. (2016)). As with the linkage studies, these show little consistency in the genomic regions found to be associated between the seven studies.

Study	Sample no.	Cohort Type	Chr. assoc.
Luciano et al. (2013)	~6,500	Population	21
Eicher et al. (2013)	~170	Selected reading and language impaired	3, 4, 13
Nudel et al. (2014)	~250	Selected (parent of origin)	5, 14*
St Pourcain et al. (2014)	~10,000	Population	3*
Gialluisi et al. (2014)	~1,800	Selected reading and language impaired	7, 21
Harlaar et al. (2014)	~2,000	Population	2, 10
Kornilov et al. (2016)	~400	Isolated population	9, 21

Table 3.2: Summary of findings from Genome-wide Association studies of spoken language disorders.

This characteristic lack of replicability between the chromosomal regions between studies reflects a number of key factors. Firstly, there are substantial differences in the methods used between studies and therefore differences in phenotyping. The lack of gold-standard diagnostic criteria for SLI/DLD mean that the phenotypes between cohorts vary dramatically, and it is therefore uninformative to compare between them. Secondly, it reflects the aetiology of DLDs being underpinned by the small effect size of many contributing variants. It suggests that there is no one, or small number of genes driving DLDs, but that there may be many variants each conferring a small increase in risk. In general, common variants carry only a small effect size, and the risk variants detected by the linkage and GWAS studies are no exception - in Figure 1, these variants lie in the bottom right corner falling in the 'Common Variants with Small Effects Identified by GWAS' region of the infographic. Genes identified from GWAS regions inform investigations into monogenic language disorders, and, vice versa, genes implicated in monogenic language disorders can inform GWAS studies.

As the list of candidate genes grows, so does our understanding of the underlying biological mechanisms and our ability to predict risk. By considering all variants implicated within a GWAS, even those that do not reach significance, we are able to derive "polygenic profiles" which allow us to consider differences and similarities between disorders. These scores are typically derived from extremely large GWAS studies, from which it is possible to identify a

set of common variants that maximally capture genetic effects upon the outcome of interest. Although we are not quite there with respect to DLD yet, early studies indicate that this is a promising area of research. Shared genetic effects have been shown to exist between cognitive ability and educational outcomes and between language development and psychosocial outcomes (Newbury et al., 2019). As polygenic profiles are developed, they become more sensitive and specific allowing more accurate inferences. For example, initial profiling of educational attainment explained approximately 2% of variance (Rietveld et al., 2013) while current scores explain as much as 13% (Lee et al., 2018). Polygenic methods are currently being developed with regard to both language disorder (Newbury et al., 2019) and dyslexia (Gialluisi et al., 2019).

Missing heritability and outstanding issues

The common variant model of inheriting a number of risk variants each conferring a small or moderate effect size only partly explains the genetics of language disorders. This knowledge gap is often referred to as the 'missing heritability', reflecting the lack of understanding as to how genetic differences drive the majority of language disorder cases. In this chapter, we will cover some of the other models that underpin language disorders covering Mendelian disorders through to less well studied phenomena such as epigenetic effects.

Gene-Gene Interactions

Gene-gene interactions, also known as epistasis, are when two alleles interact with each other to result in a phenotype not seen when only one of the alleles is present. These are sometimes referred to modifiers or genetic background when in the context of dominant variants, or a 'second hit' of another variant in a different gene with is also having an effect. Conceptually, both of these interactions fit the Mendelian disease models but it could be equally applied to complex disorders to increase risk and heterogeneity of symptoms. Gene-gene effects have not been studied in language disorders as they require an *a priori* knowledge of which genes to look at, but they are known to play a role in the variable severity of microdeletion disorders (Veltman and Brunner, 2010) and sex chromosome trisomy cases (Rocca et al., 2016, Le Gall et al., 2017). Thus, it seems extremely likely that gene-gene interactions will contribute to language disorders and this relatively new field of research shows great promise to understand the interactions between key language genes.

Gene-Environment Interactions

Gene-environment interactions refer to a genetic effect which is modulated in response to an environmental effect. Again, there are no examples of this in language disorders, but it is highly likely to play an important role. The commonly referenced example of gene-environment interaction is in the serotonin transporter *SLC6A4*, also known as 5-HTTLPR which was linked to depression. The widely accepted theory was that there are two versions of the SLC6A4 protein; a short and a long form. The short form has been associated with increased risk of poor developmental outcomes which only express when profound environmental stress is experienced, whereas the short version is associated with positive developmental outcomes in a positive environment (van Ijzendoorn et al., 2012). Recently, this body of work was called into question by a large study of >600,000 participants that failed to replicate links between depression and *SLC6A4* (Border et al., 2019). This example provides evidence of the complexities of substantiating genetic effects and interactions, which may vary between tissues types and over developmental course.

Epigenetics

Epigenetics is a relatively new field of research which looks at changes in the regulation of a gene in response to environment. There are a number of mechanisms that exert epigenetic control over the regulation of a gene, some may be passed onto offspring and some are not. Two of the best characterised epigenetic mechanisms are DNA methylation and histone modification. Both of these regulatory mechanisms have been shown to play a role in neurological degenerative disorders; Alzheimers disease (Nicolia et al., 2017) and Parkinson's disease (Park et al., 2016), consecutively. Epigenetic regulatory mechanisms have been proposed by a number of groups as a likely to play a role in DLDs (Smith, 2011, Rice, 2012, Kraft and DeThorne, 2014). This represents a very plausible layer of genetic control that is likely to contribute to the complexity and heterogeneity of language disorders. To date, no studies have successfully shown an association between epigenetic regulation and DLDs. However, as we generate a better understanding of individual differences in epigenetic markers (Gunasekara et al., 2019), we are starting to be able to identify systematic changes that may be related to disorder risk. Early evidence suggests that prenatal epigenetic changes can persist throughout life (Heijmans et al., 2008) and that these changes may be important in brain development (Kupers et al., 2019). These findings indicate that this area has strong potential for future research.

Summary

This chapter has provided a phenotypic overview of DLD and outlined current understanding of the genetic underpinnings of this disorder. While advances in the fields of genetic technologies and bioinformatics have allowed for more in-depth research and the discovery of associated gene pathways, the picture has become equally more complex. A handful of genes have been implicated in DLD populations, yet rather than the monogenic causes implicated in previous studies of disorder, Genome-Wide Association Studies have shown that DLDs are more likely the result of combinations of many variants, each conferring a small increase in risk. To date, small sample sizes have restricted the number of candidate genes identified in these studies, thus recruitment of larger-scale cohorts will be key to uncovering further variants. Past studies have been based on varied inclusion criteria, reflecting historical lack of a consensus definition or classification system for DLD. The importance of consistent, fine-grain phenotyping in genetic studies of these populations going forward is thus important in order that studies may be replicated, and cohorts can be considered collectively. Sophisticated deep phenotyping of the language, speech and cognitive abilities will be critical in understanding the genotype-phenotype interactions of candidate genes. Epigenetic influences also likely contribute to the phenotype observed and will be an important area of future research. Greater understanding of the genetic underpinnings of DLD is of critical importance in the age of personalised medicine. This work has direct implications in the clinical setting for informing timely diagnoses and genetic counselling, while allowing for the development of targeted therapies and improved long-term outcomes for the DLD population.

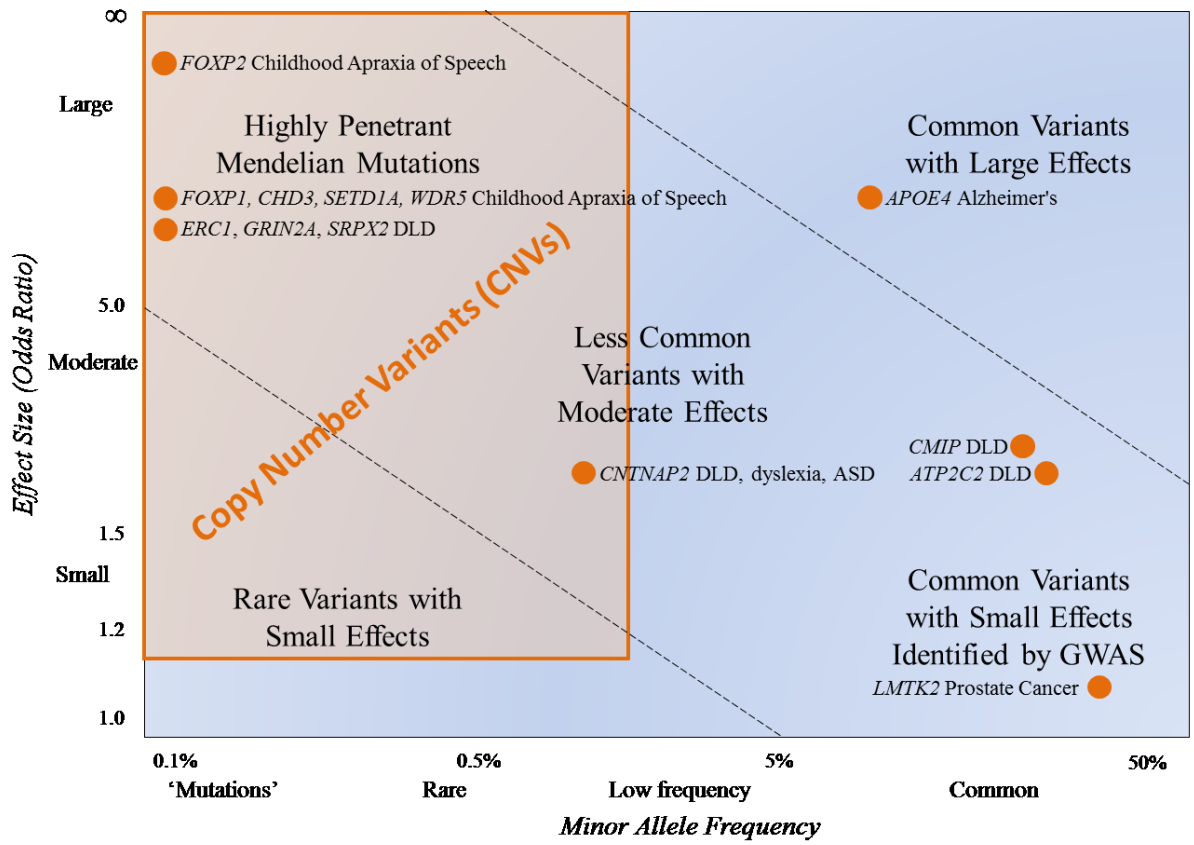
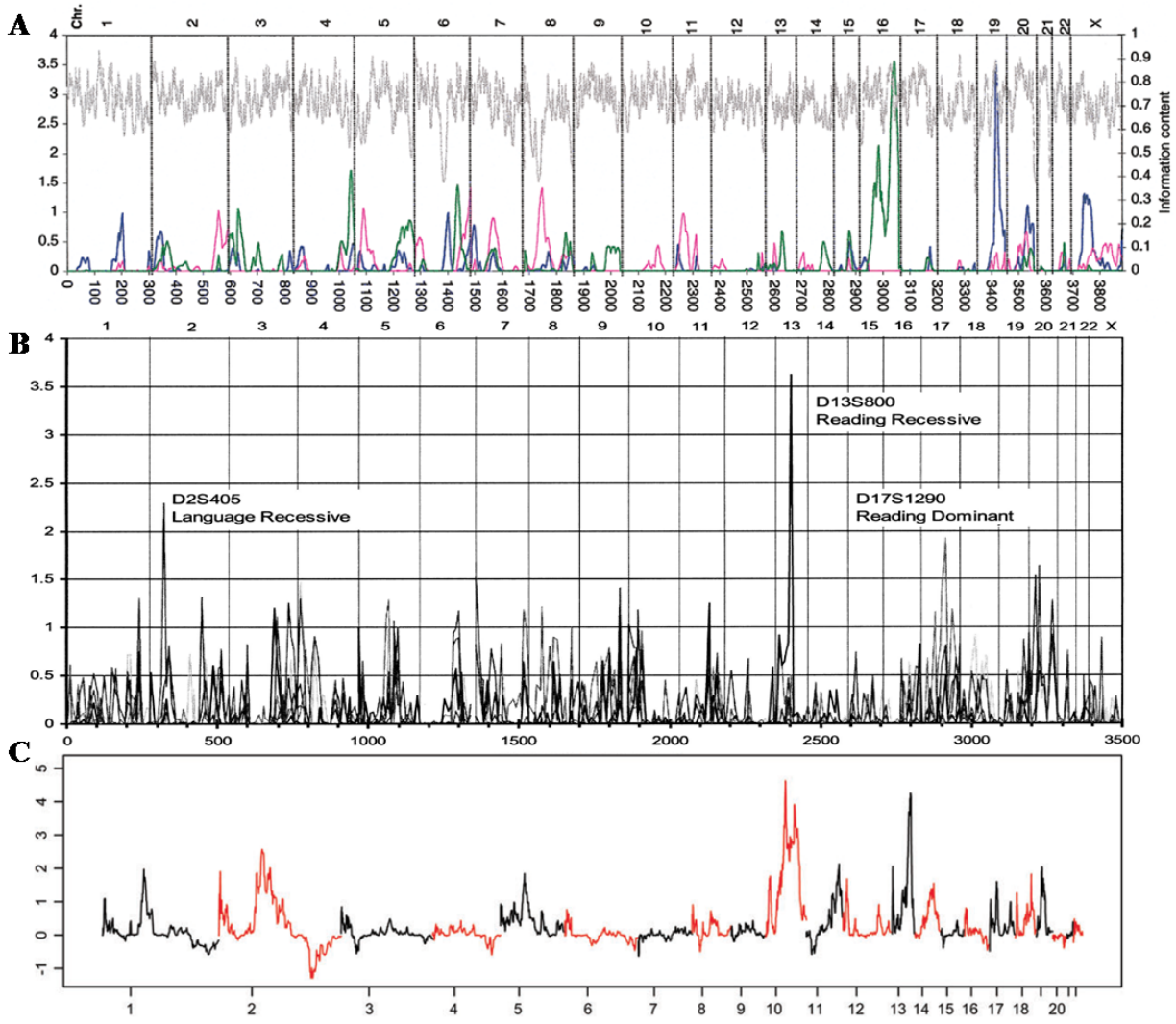


Figure 3.1 *shows an infographic showing the categories of genetic models. Along the* ***x*** *axis is allele frequency which shows how frequently a variant is present in the population. Each person carries two copies of every letter of our DNA, and each of these is termed an 'allele'. Where a person carries the most common form of the allele, this is referred to as a wild-type allele but can be thought of as the regular 'standard' version. We are interested in the variant non-typical form of the allele which is termed the 'minor allele'. It is the frequency of these 'minor alleles' in the population that is of particular interest in genetics. In a complex genetic model, we would expect that a variant might be carried by many people in the population. Usually a common allele is considered to be carried by between 5% and 50% of the population, or in other words a minor allele frequency of 5-50%. The* ***y*** *axis shows the effect size of the variant. This reflects the severity of the impact of carrying that particular variant (or allele), so a variant with a small effect size may be easily tolerated whereas a variant with a large effect size will be very likely to result in a phenotype. A number of familiar diseases fall into this complex genetic model, where the variant is very common in the population. One particularly strong example of a common variant with a large effect size is the apolipoprotein E variant ε4 (APOE-ε4) in Alzheimer's disease (Lambert et al., 2009). This is a very common variant of APOE which is carried by about 10-15% of people.*

Carrying this allele of APOE doubles the chance of developing Alzheimer's disease, which in complex genetics terms is considered to be a large effect size. Conversely, some common variants have been found to have a small effect size, where carrying the variant increases risk by a small amount. One example of this is the Lemur Tyrosine Kinase 2 gene, LMTK2, which is present in around 43% of the population, and slightly increases the chance of developing prostate cancer (Eeles et al., 2008). Each of these individual genetic variants increases risk of developing a disease to various extents, but lifestyle and environmental factors still contribute to the overall risk of developing the disease.

Figure 3.2 *shows Manhattan plots showing regions of genetic association. Figure 3.2A shows regions 16q24 (SLI1) associated with nonword repetition and 19q13 region (SLI2) associated with the CELF-R measure of expressive language (adapted from S. L. I. Consortium (2002)). Figure 3.2B region 13q21 associated with a reading-impaired phenotype, and 2p22 and 17q23 associated with language delay (adapted from Bartlett et al. (2002)). Figure 3.2C shows regions 10q23.33 and 13q33.3 associated with phonological memory (adapted from Evans et al. (2015)).*



REFERENCES

Alarcon, M., Abrahams, B. S., Stone, J. L., Duvall, J. A., Perederiy, J. V., Bomar, J. M., Sebat, J., Wigler, M., Martin, C. L., Ledbetter, D. H., Nelson, S. F., Cantor, R. M. and Geschwind, D. H. (2008), 'Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene', *American Journal of Human Genetics*, vol. 82, no. 1, pp. 150-159 [Online]. DOI: 10.1016/j.ajhg.2007.09.005 (Accessed 3 June 2019).

American Speech-Language-Hearing Association (ASHA). (2007). *Ad Hoc Committee on Childhood Apraxia of Speech*. [Online] Rockville (MD). Available at http://www.asha.org/policy/PS2007-00277/. (Accessed 28 May 2019).

Arking, D. E., Cutler, D. J., Brune, C. W., Teslovich, T. M., West, K., Ikeda, M., Rea, A., Guy, M., Lin, S., Cook, E. H. and Chakravarti, A. (2008), 'A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism', *American Journal of Human Genetics*, vol. 82, no. 1, pp. 160-164 [Online]. DOI: 10.1016/j.ajhg.2007.09.015 (Accessed 3 June 2019).

Bacon, C. and Rappold, G. A. (2012), 'The distinct and overlapping phenotypic spectra of FOXP1 and FOXP2 in cognitive disorders', *Human Genetics*, vol. 131**,** no. 11, pp. 1687-1698 [Online]. DOI: 10.1007/s00439-012-1193-z (Accessed 3 June 2019).

Barnett, C. P. and Van Bon, B. W. (2015), 'Monogenic and chromosomal causes of isolated speech and language impairment', *Journal of Medical Genetics*, vol. 52**,** no. 11, pp. 719-729 [Online]. DOI: 10.1136/jmedgenet-2015-103161 (Accessed 3 June 2019).

Bartlett, C. W., Flax, J. F., Logue, M. W., Vieland, V. J., Bassett, A. S., Tallal, P. and Brzustowicz, L. M. (2002), 'A major susceptibility locus for specific language impairment is located on 13q21', *American Journal of Human Genetics*, vol. 71**,** no. 1 pp. 45-55 [Online]. DOI: 10.1086/341095 (Accessed 3 June 2019).

Beckmann, J. S., Estivill, X. and Antonarakis, S. E. (2007), 'Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability', *Nature Reviews Genetics*, vol. 8**,** no. 8, pp. 639-646 [Online]. DOI: 10.1038/nrg2149 (Accessed 3 June 2019).

Beitchman, J. H., Brownlie, E. B. and Bao, L. (2014) 'Age 31 mental health outcomes of childhood language and speech disorders', *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 53, no. 10, pp. 1102-1110 [Online]. DOI: 10.1016/j.jaac.2014.07.006 (Accessed 28 May 2019).

Bishop, D. V. M., North, T. and Donlan, C. (1995), 'Genetic basis of specific language impairment: evidence from a twin study', *Developmental Medicine & Child Neurology*, vol. 37, no. 1, pp. 56-71 [Online]. DOI: 10.1111/j.1469-8749.1995.tb11932.x (Accessed 3 June 2019).

Bishop, D. V. M., Snowling, M. J., Thompson, P. A. and Greenhalgh, T. (2016), 'CATALISE: A multinational and multidisciplinary Delphi consensus study;

Identifying language impairments in children', *PLoS One*, vol. 11, no. 7 [Online]. DOI: 10.1371/journal.pone.0158753 (Accessed 28 May 2019).

Bishop, D. V. M., Snowling, M. J., Thompson, P. A. and Greenhalgh, T. (2017), 'Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology', *Journal of Child Psychology and Psychiatry*, vol. 58, no. 10, pp. 1068-1080 [Online]. DOI: 10.1111/jcpp.12721 (Accessed 28 May 2019).

Blair, D. R., *et. al.* (2013), 'A nondegenerate code of deleterious variants in mendelian loci contributes to complex disease risk', *Cell*, vol. 155, no. 1, pp. 70-80 [Online]. DOI: 10.1016/j.cell.2013.08.030 (Accessed 3 June 2019).

Border, R., Johnson, E. C., Evans, L. M., Smolen, A., Berley, N., Sullivan, P. F. and Keller, M. C. (2019), 'No support for historical candidate gene or candidate gene-by-interaction hypotheses for major depression across multiple large samples', *American Journal of Psychiatry*, vol. 176, no. 5, pp. 376-387 [Online]. DOI: 10.1176/appi.ajp.2018.18070881 (Accessed 3 June 2019).

Bretherton, L., Prior, M., Bavin, E., Cini, E., Eadie, P. and Reilly, S. (2014), 'Developing relationships between language and behaviour in preschool children from the Early Language in Victoria Study: Implications for intervention', *Emotional and Behavioural Difficulties*, vol. 19, no. 1, pp. 7-27 [Online]. DOI: 10.1080/13632752.2013.854956 (Accessed 28 May 2019).

Brignell, A., Williams, K., Jachno, K., Prior, M., Reilly, S. and Morgan, A. T. (2018), 'Patterns and predictors of language development from 4 to 7 years in verbal children with and without Autism Spectrum Disorder', *Journal of Autism and Developmental Disorders*, vol. 10, pp 3282 [Online]. DOI: 10.1007/s10803-018-3565-2 (Accessed 28 May 2019).

Carrion-Castillo, A., Van Bergen, E., Vino, A., Van Zuijen, T., de Jong, P. F., Francks, C. and Fisher, S. E. (2016), 'Evaluation of results from genome-wide studies of language and reading in a novel independent dataset', *Genes, Brain and Behavior*, vol. 15, no. 6, pp. 531-541 [Online]. DOI: 10.1111/gbb.12299 (Accessed 3 June 2019).

Catroppa, C. and Anderson, V. (2004), 'Recovery and predictors of language skills two years following pediatric traumatic brain injury', *Brain and Language,* vol. 88, no. 1, pp. 68-78 [Online]. DOI: 10.1016/S0093-934X(03)00159-7 (Accessed 28 May 2019).

Catts, H. W., Adlof, S. M., Hogan, T. P. and Weismer, S. E. (2005), 'Are Specific Language Impairment and Dyslexia distinct disorders?', *Journal of Speech, Language, and Hearing Research,* vol. 48, no. 6**,** pp. 1378-1396 [Online]. DOI: 10.1044/1092-4388(2005/096) (Accessed 28 May 2019).

Centanni, T. M., Green, J. R., Iuzzini-Seigel, J., Bartlett, C. W. and Hogan, T. P. (2015), 'Evidence for the multiple hits genetic theory for inherited language impairment: a case study', *Frontiers in Genetics*, vol. 6**,** pp. 272 [Online]. DOI: 10.3389/fgene.2015.00272 (Accessed 3 June 2019).

Chen, X. S., Reader, R. H., Hoischen, A., Veltman, J. A., Simpson, N. H., Francks, C., Newbury, D. F. and Fisher, S. E. (2017), 'Next-generation DNA sequencing identifies novel gene variants and pathways involved in specific language impairment', *Scientific Reports*, vol. 7**,** pp. 46105 [Online]. DOI: 10.1038/srep46105 (Accessed 3 June 2019).

Clegg, J., Hollis, C., Mawhood, L. and Rutter, M. (2005), 'Developmental language disorders – a follow-up in later adult life. Cognitive, language and psychosocial outcomes', *Journal of Child Psychology and Psychiatry,* vol. 46, no. 2**,** pp. 128-149 [Online]. DOI: 10.1111/j.1469-7610.2004.00342.x (Accessed 28 May 2019).

Cleland, J., Wood, S., Hardcastle, W., Wishart, J. and Timmins, C. (2010), 'Relationship between speech, oromotor, language and cognitive abilities in children with Down's syndrome', *International Journal of Language & Communication Disorders,* vol. 45, no. 1**,** pp. 83-95 [Online]. DOI: 10.3109/13682820902745453 (Accessed 28 May 2019).

Coe, B. P., *et. al.* (2014), 'Refining analyses of copy number variation identifies specific genes associated with developmental delay', *Nature Genetics*, vol. 46**,** no. 10, pp. 1063 [Online]. DOI: 10.1038/ng.3092 (Accessed 3 June 2019).

Constantino, J. N. (2018), 'Deconstructing autism: from unitary syndrome to contributory developmental endophenotypes', *International Review of Psychiatry*, vol. 30**,** no. 1, pp. 18-24 [Online]. DOI: 10.1080/09540261.2018.1433133 (Accessed 3 June 2019).

Conti-Ramsden, G., Botting, N., Simkin, Z. and Knox, E. (2001) 'Follow-up of children attending infant language units: Outcomes at 11 years of age', *International Journal of Language & Communication Disorders,* vol. 36, no.2**,** pp. 207-219 [Online]. DOI: doi.org/10.1080/13682820121213136828209027454 53 (Accessed 28 May 2019).

Dale, P. S., Simonoff, E., Bishop, D. V., Eley, T. C., Oliver, B., Price, T. S., Purcell, S., Stevenson, J. and Plomin, R. (1998), 'Genetic influence on language delay in two-year-old children', *Nature Neuroscience*, vol. 1**,** no. 4., pp. 324-328 [Online]. DOI: 10.1038/1142 (Accessed 3 June 2019).

Devanna, P., Chen, X. S., Ho, J., Gajewski, D., Smith, S. D., Gialluisi, A., Francks, C., Fisher, S. E., Newbury, D. F. and Vernes, S. C. (2017), 'Next-gen sequencing identifies non-coding variation disrupting miRNA-binding sites in neurological disorders'. *Molecular Psychiatry*, vol. 23, no. 5, pp. 1375-1384 [Online]. DOI: 10.1038/mp.2017.30 (Accessed 3 June 2019).

Docking, K., Munro, N., Marshall, T. and Togher, L. (2016) 'Narrative skills of children treated for brain tumours: The impact of tumour and treatment related variables on microstructure and macrostructure', *Brain Injury,* vol. 30, no. 8**,** pp. 1005-1018 [Online]. DOI: 10.3109/02699052.2016.1147602 (Accessed 28 May 2019).

Eadie, P., Conway, L., Hallenstein, B., Mensah, F. K., McKean, C. and Reilly, S. (2018) 'Quality of Life in Children with Developmental Language Disorder', *International Journal of Language & Communication Disorders,* vol. 53, no. 4**,** pp. 799-810 [Online]. DOI: 10.1111/1460-6984.12385 (Accessed 28 May 2019).

Eadie, P., Morgan, A. T., Ukoumunne, O. C., Ttofari Eecen, K., Wake, M. and Reilly, S. (2015) 'Speech sound disorder at 4 years: Prevalence, comorbidities, and predictors in a community cohort of children', *Developmental Medicine & Child Neurology,* vol. 57, no. 6**,** pp. 578-584 [Online]. DOI: 10.1111/dmcn.12635 (Accessed 28 May 2019).

Eeles, R. A., *et. al.* (2008), 'Multiple newly identified loci associated with prostate cancer susceptibility', *Nature Genetics*, vol. 40**,** no. 3, pp. 316-321 [Online]. DOI: 10.1038/ng.90 (Accessed 3 June 2019).

Eicher, J. D., Powers, *et. al.* (2013). 'Genome-wide association study of shared components of reading disability and language impairment,' *Genes, Brain and Behavior*, vol. 12**,** no. 8, pp. 792-801 [Online]. DOI: 10.1111/gbb.12085 (Accessed 3 June 2019).

Eising, E., *et. al.* (2018), 'A set of regulatory genes co-expressed in embryonic human brain is implicated in disrupted speech development', *Molecular Psychiatry***,** pp. 1 [Online]. DOI: 10.1038/s41380-018-0020-x (Accessed 3 June 2019).

Ercan-Sencicek, A. G., Davis Wright, N. R., Sanders, S. J., Oakman, N., Valdes, L., Bakkaloglu, B., Doyle, N., Yrigollen, C. M., Morgan, T. M. and Grigorenko, E. L. (2012), 'A balanced t(10;15) translocation in a male patient with developmental language disorder', *European Journal of Medical Genetics*, vol. 55**,** no. 2, pp. 128-131 [Online]. DOI: 10.1016/j.ejmg.2011.12.005 (Accessed 3 June 2019).

Estruch, S. B., Graham, S. A., Deriziotis, P. and Fisher, S. E. (2016), 'The language-related transcription factor FOXP2 is post-translationally modified with small ubiquitin-like modifiers', *Scientific Reports*, vol. 6**,** pp. 20911 [Online]. DOI: 10.1038/srep20911 (Accessed 3 June 2019).

Evans, P. D., Mueller, K. L., Gamazon, E. R., Cox, N. J. and Tomblin, J. B. (2015), 'A genome-wide sib-pair scan for quantitative language traits reveals linkage to chromosomes 10 and 13', *Genes, Brain and Behavior*, vol. 14, no. 5, pp. 387-397 [Online]. DOI: 10.1111/gbb.12223 (Accessed 3 June 2019).

Ewing-Cobbs, L. and Barnes, M. (2002) 'Linguistic outcomes following traumatic brain injury in children', *Seminars in Pediatric Neurology*, vol. 9, no. 3, pp. 209-217 [Online]. DOI: 10.1053/spen.2002.35502 (Accessed 28 May 2019).

Falcaro, M., Pickles, A., Newbury, D. F., Addis, L., Banfield, E., Fisher, S. E., Monaco, A. P., Simkin, Z., Conti-Ramsden, G. and SLI Consortium. (2008) 'Genetic and phenotypic effects of phonological short-term memory and grammatical morphology in Specific Language Impairment', *Genes, Brain and Behavior,* vol. 7, no. 4**,** pp. 393-402 [Online]. DOI: 10.1111/j.1601-183X.2007.00364.x (Accessed 28 May 2019).

Fedorenko, E., Morgan, A., Murray, E., Cardinaux, A., Mei, C., Tager-Flusberg, H., Fisher, S. E. and Kanwisher, N. (2016), 'A highly penetrant form of childhood apraxia of speech due to deletion of 16p11.2', *European Journal of Human Genetics*, vol. 24, no. 2, pp. 302-306 [Online]. DOI: 10.1038/ejhg.2015.149 (Accessed 3 June 2019).

Feuk, L., *et. al.* (2006), 'Absence of a paternally inherited FOXP2 gene in developmental verbal dyspraxia', *American Journal of Human Genetics*, vol. 79**,** no. 5, pp. 965-972 [Online]. DOI: 10.1086/508902 (Accessed 3 June 2019).

Fisher, S. E. and Scharff, C. (2009), 'FOXP2 as a molecular window into speech and language', *Trends in Genetics*, vol. 25**,** no. 4, pp. 166-177 [Online]. DOI: 10.1016/j.tig.2009.03.002 (Accessed 3 June 2019).

Franic, S., *et. al.* (2015), 'Intelligence: shared genetic basis between mendelian disorders and a polygenic trait', *European Journal of Human Genetics*, vol. 23**,** no. 10, pp. 1378-1383 [Online]. DOI: 10.1038/ejhg.2015.3 (Accessed 3 June 2019).

Gialluisi, A., *et. al.* (2019), 'Genome-wide association scan identifies new variants associated with a cognitive predictor of dyslexia', *Translational Psychiatry*, vol. 9, no. 77, pp. 1-77 [Online]. DOI: 10.1038/s41398-019-0402-0 (Accessed 12 September 2019).

Gialluisi, A., *et. al.* (2014), 'Genome-wide screening for DNA variants associated with reading and language traits', *Genes, Brain and Behavior*, vol. 13**,** no. 7, pp. 686-701 [Online]. DOI: 10.1111/gbb.12158 (Accessed 3 June 2019).

Graham, S. A. and Fisher, S. E. (2013), 'Decoding the genetics of speech and language', *Current Opinions in Neurobiology*, vol. 23**,** no. 1, pp. 43-51 [Online]. DOI: 10.1016/j.conb.2012.11.006 (Accessed 3 June 2019).

Grove, J., *et. al.* (2019), 'Identification of common genetic risk variants for autism spectrum disorder', *Nature Genetics*, vol. 51**,** no. 3, pp. 431-444 [Online]. DOI: 10.1038/s41588-019-0344-8 (Accessed 3 June 2019).

Gunasekara, C. J., *et. al.* (2019), 'A genomic atlas of systemic interindividual epigenetic variation in humans', *Genome Biology*, vol. 20, no. 1, pp. 105-116 [Online]. DOI: 10.1186/s13059-019-1708-1 (Accessed 12 September 2019).

Haarbauer-Krupa, J., King, T. Z., Wise, J., Gillam, S., Trapani, J., Weissman, B. and DePompei, R. (2019) 'Early elementary school outcome in children with a history of traumatic brain injury before age 6 years', *The Journal of Head Trauma Rehabilitation,* vol. 34, no. 2**,** pp. 111-121 [Online]. DOI: 10.1097/HTR.0000000000000414 (Accessed 28 May 2019).

Hamdan, F. F., *et. al.* (2010), 'De novo mutations in foxp1 in cases with intellectual disability, autism, and language impairment', *American Journal of Human Genetics*, vol. 87**,** no. 8, pp. 671-678 [Online]. DOI: 10.1016/j.ajhg.2010.09.017 (Accessed 3 June 2019).

Harlaar, N., Meaburn, E. L., Hayiou-Thomas, M. E., Wellcome Trust Case Control Consortium, Davis, O. S., Docherty, S., Hanscombe, K. B., Haworth, C. M., Price, T. S., Trzaskowski, M., Dale, P. S. and Plomin, R. (2014), 'Genome-wide association study of receptive language ability of 12-year-olds', *Journal of Speech Language and Hearing Research*, vol. 57**,** no. 1, pp. 96-105 [Online]. DOI: 10.1044/1092-4388(2013/12-0303) (Accessed 3 June 2019).

Hayiou-Thomas, M. E., Oliver, B. and Plomin, R. (2005), 'Genetic influences on specific versus nonspecific language impairment in 4-year-old twins', *Journal of Learning Disabilities*, vol. 38**,** no. 3, pp. 222-232 [Online]. DOI: 10.1177/00222194050380030401 (Accessed 3 June 2019).

Heijmans, B.T., Tobi, E. W., Stein, A. D., Putter, H., Blauw, G. J., Susser, E. S., Slagboom, P. E. and Lumey, L. H. (2008), 'Persistent epigenetic differences associated with prenatal exposure to famine in humans', *Proceeding of the National Academy of Science USA*, vol. 105, no. 44, pp. 17046-17049 [Online]. DOI: 10.1073/pnas.0806560105 (Accessed 12 September 2019).

Horn, D, *et. al.* (2010), 'Identification of FOXP1 deletions in three unrelated patients with mental retardation and significant speech and language deficits', *Human Mutation*, vol. 31**,** no. 11, pp. e1851-60 [Online]. DOI: 10.1002/humu.21362 (Accessed 3 June 2019).

Kalnak, N., Stamouli, S., Peyrard-Janvid, M., Rabkina, I., Becker, M., Klingberg, T., Kere, J., Forssberg, H. and Tammimies, K. (2018), 'Enrichment of rare copy number variation in children with developmental language disorder', *Clinical Genetics*, vol. 94**,** no. 3-4, pp. 313-320 [Online]. DOI: 10.1111/cge.13389 (Accessed 3 June 2019).

Koolen, D. A, Vissers, L. E., Nillesen, W., Smeets, D., van Ravenswaaij, C. M., Sistermans, E. A., Veltman, J. A. and de Vries, B. B. (2004), ' A novel microdeletion, del(2)(q22.3q23.3) in a mentally retarded patient, detected by array-based comparative genomic hybridization', *Clinical Genetics*, vol. 65, no. 5, pp. 429-429 [Online]. DOI: 10.1111/j.0009-9163.2004.00245.x (Accessed 3 June 2019).

Kornilov, S. A., Rakhlin, N., Koposov, R., Lee, M, Yrigollen, C., Caglayan, A. O., Magnuson, J. S., Mane, S., Chang, J. T. and Grigorenko, E. L. (2016), 'Genome-wide association and exome sequencing study of language disorder in an isolated population', *Pediatrics*, vol. 137, no. 4 [Online]. DOI: 10.1542/peds.2015-2469 (Accessed 3 June 2019).

Kraft, S. J. and DeThorne, L. S. (2014), 'The brave new world of epigenetics: embracing complexity in the study of speech and language disorders', *Current Developmental Disorders Reports*, vol. 1**,** no. 3, pp. 207-214 [Online]. DOI: 10.1007/s40474-014-0024-4 (Accessed 3 June 2019).

Krapohl, E., Patel, H., Newhouse, S., Curtis, C. J., von Stumm, S., Dale, P. S., Zabeneh, D., Breen, G., O'Reilly, P. F. and Plomin, R. (2018), 'Multi-polygenic score approach to trait prediction', *Molecular Psychiatry*, vol. 23, no. 5, pp. 1368-1374 [Online]. DOI: 10.1038/mp.2017.163

Küpers, L. K., *et. al.* (2019), 'Meta-analysis of epigenome-wide association studies in neonates reveals widespread differential DNA methylation associated with birthweight', *Nature Communications*, vol. 10, no. 1, pp. 1893-1904 [Online]. DOI: 10.1038/s41467-019-09671-3 (Accessed 12 September 2019).

Lai, C. S., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F. and Monaco, A. P. (2001), 'A forkhead-domain gene is mutated in a severe speech and language disorder', *Nature*, vol. 413**,** no. 6855, pp. 519-523 [Online]. DOI: 10.1038/35097076 (Accessed 3 June 2019).

Lambert, J. C., *et. al.* (2009), 'Genome-wide association study identifies variants at CLU and CR1 associated with alzheimer's disease', *Nature Genetics*, vol. 41**,** no. 10, pp. 1094-1099 [Online]. DOI: 10.1038/ng.439 (Accessed 3 June 2019).

Law, J., Mcbean, L. and Rush, R. (2011), 'Communication skills in a population of primary school-aged children raised in an area of pronounced social disadvantage', *International Journal of Language & Communication Disorders*, vol. 46**,** no. 6, pp. 657-664 [Online]. DOI: 10.1111/j.1460-6984.2011.00036.x (Accessed 3 June 2019).

Lee, J. J., *et. al.* (2018), 'Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals', *Nature Genetics*, vol. 50, no. 8, pp. 1112–1121 [Online]. DOI: 10.1038/s41588-018-0147-3 (Accessed 12 September 2019).

Le Gall, J., Nizon, M., Pichon, O., Andrieux, J., Audebert-Bellanger, S., Baron, S., Beneteau, C., Bilan, F., Boute, O. and Busa, T. (2017), 'Sex chromosome aneuploidies and copy-number variants: a further explanation for neurodevelopmental prognosis variability?', *European Journal of Human Genetics*, vol. 25**,** no. 8, pp. 930-934 [Online]. DOI: 10.1038/ejhg.2017.93 (Accessed 3 June 2019).

Lennon, P. A., Cooper, M. L., Peiffer, D. A., Gunderson, K. L., Patel, A., Peters, S., Cheung, S. W. and Bacino, C. A. (2007), 'Deletion of 7q31.1 supports involvement of foxp2 in language impairment: clinical report and review', *American Journal of Medical Genetics Part A*, vol. 143a, no. 8, pp. 791-798 [Online]. DOI: 10.1002/ajmg.a.31632 (Accessed 3 June 2019).

Lewis, B. A. and Thompson, L. A. (1992), 'A study of developmental speech and language disorders in twins', *Journal of Speech Hearing Research*, vol. 35**,** no. 5, pp. 1086-1094 [Online]. DOI: 10.1044/jshr.3505.1086 (Accessed 3 June 2019).

Lionel, A. C., *et. al.* (2011), 'Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD', *Science Translational Medicine*, vol. 3**,** no. 95, pp. 95ra75 [Online]. DOI: 10.1126/scitranslmed.3002464 (Accessed 3 June 2019).

Locke, A., Ginsborg, J. and Peers, I. (2002), 'Development and disadvantage: implications for the early years and beyond', *International Journal of Language & Communication Disorders*, vol. 37**,** no. 1, pp. 3-15 [Online]. DOI: 10.1080/13682820110089911 (Accessed 3 June 2019).

Lozano, R., Vino, A., Lozano, C., Fisher, S. E. and Deriziotis, P. (2015), 'A de novo FOXP1 variant in a patient with autism, intellectual disability and severe speech and language impairment', *European Journal of Human Genetics*, vol. 23**,** no. 12, pp. 1702-1707 [Online]. DOI: 10.1038/ejhg.2015.66 (Accessed 3 June 2019).

Luciano, M., Evans, D. M., Hansell, N. K., Medland, S. E., Montgomery, G. W., Martin, N. G., Wright, M. J. and Bates, T. C. (2013), 'A genome-wide association study for reading and language abilities in two population cohorts', *Genes, Brain and Behavior*, vol. 12**,** no. 6, pp. 645-652 [Online]. DOI: 10.1111/gbb.12053 (Accessed 3 June 2019).

MacDermot, K. D., Bonora, E., Sykes, N., Coupe, A. M., Lai, C. S., Vernes, S. C., Vargha-Khadem, F., McKenzie, F., Smith, R. L., Monaco, A. P. and Fisher, S. E. (2005), 'Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits', *American Journal of Human Genetics*, vol. 76**,** no. 6, pp. 1074-1080 [Online]. DOI: 10.1086/430841 (Accessed 3 June 2019).

McKean, C., Reilly, S., Bavin, E. L., Bretherton, L., Cini, E., Conway, L., Cook, F., Eadie, P., Prior, M., Wake, M. and Mensah, F. (2017) 'Language Outcomes at 7 Years: Early Predictors and Co-Occurring Difficulties', *Pediatrics,* vol. 139, no.3**,** pp. e20161684 [Online]. DOI: 10.1542/peds.2016-1684 (Accessed 28 May 2019).

Mei, C., Fedorenko, E., Amor, D. J., Boys, A., Hoeflin, C., Carew, P., Burgess, T., Fisher, S. E. and Morgan, A. T. (2018), 'Deep phenotyping of speech and language skills in individuals with 16p11.2 deletion', *European Journal of Human Genetics*, vol. 26**,** no. 5, pp. 676-686 [Online]. DOI: 10.1038/s41431-018-0102-x (Accessed 3 June 2019).

Mei, C., Reilly, S., Reddihough, D., Mensah, F., Pennington, L. and Morgan, A. T. (2016) 'Language outcomes of children with cerebral palsy aged 5 years and 6 years: a population-based study. Vol 58, no 6, pp 605-611 [Online]. DOI: 10.1111/dmcn.12957 (Accessed 28 May 2019).

Moralli, D., Nudel, R., Chan, M. T. M., Green, C. M., Volpi, E. V., Benítez-Burraco, A., Newbury, D. F. and García-Bellido, P. (2015), 'Language impairment in a case of a complex chromosomal rearrangement with a breakpoint downstream of FOXP2', *Molecular Cytogenetics*, vol. 8**,** no. 1, pp. 36 [Online]. DOI: 10.1186/s13039-015-0148-1 (Accessed 3 June 2019).

Morgan, A. T., Fisher, S. E., Scheffer, I. E., and Hildebrand, M. (2017) *FOXP2-related speech and language disorders* [Online], University of Washington, Seattle. Available at GeneReviews® [Internet], https://www.ncbi.nlm.nih.gov/books/NBK368474/ (Accessed 28 May 2019).

Morgan, A. T., Mei, C., Da Costa, A., Fifer, J., Lederer, D., Benoit, V., McMillin, M. J., Buckingham, K. J., Bamshad, M. J., Pope, K. and White, S. M. (2015) 'Speech and language in a genotyped cohort of individuals with Kabuki syndrome', *American Journal of Medical Genetics, Part A,* vol. 167, no. 7**,** pp. 1483-92 [Online]. DOI: (Accessed 28 May 2019).

Morgan, A. T., van Haaften, L., van Hulst, K., Edley, C., Mei, C., Tan, T. Y., Amor, D., Fisher, S. E. and Koolen, D. A. (2018), 'Early speech development in Koolen de Vries syndrome limited by oral praxis and hypotonia', *European Journal of Human Genetics*, vol. 26, pp. 75-84 [Online]. DOI: 10.1038/s41431-017-0035-9 (Accessed 3 June 2019).

Mountford, H. S. and Newbury, D. F. (2018), 'The genomic landscape of language: insights into evolution', *Journal of Language Evolution*, vol. 3**,** no. 1, pp. 49-58 [Online]. DOI: 10.1093/jole/lzx019 (Accessed 3 June 2019).

Newbury, D. F., Gibson, J. L., Conti-Ramsden, G., Pickles, A., Durkin, K. and Toseeb, U. (2019), 'Using Polygenic Profiles to Predict Variation in Language and Psychosocial Outcomes in Early and Middle Childhood', *Journal of Speech, Language, and Hearing Research*, epub ahead of print [Online]. DOI: 10.1044/2019_JSLHR-L-19-0001 (Accessed 12 September 2019).

Newbury, D. F., Mari, F., Sadighi Akha, E., MacDermot, K. D., Canitano, R., Monaco, A. P., Taylor, J. C., Renieri, A., Fisher, S. E. and Knight, S. J. (2013), 'Dual copy number variants involving 16p11 and 6q22 in a case of childhood apraxia of speech and pervasive developmental disorder', *European Journal of Human Genetics*, vol. 21**,** no. 4, pp. 361-365 [Online]. DOI: 10.1038/ejhg.2012.166 (Accessed 3 June 2019).

Newbury, D. F., Paracchini, S., Scerri, T. S., Winchester, L., Addis, L., Richardson, A. J., Walter, J., Stein, J. F., Talcott, J. B. and Monaco, A. P. (2011), 'Investigation of dyslexia and SLI risk variants in reading- and language-impaired subjects', *Behavior Genetics*, vol. 41**,** no. 1, pp. 90-104 [Online]. DOI: 10.1007/s10519-010-9424-3 (Accessed 3 June 2019).

Newbury, D. F., *et. al.* (2009), 'CMIP and ATP2C2 modulate phonological short-term memory in language impairment', *American Journal of Human Genetics*, vol. 85**,** no. 2, pp. 264-272 [Online]. DOI: 10.1016/j.ajhg.2009.07.004 (Accessed 3 June 2019).

Nicolia, V., Cavallaro, R. A., Lopez-Gonzalez, I., Maccarrone, M., Scarpa, S., Ferrer, I. and Fuso, A. (2017), 'DNA methylation profiles of selected pro-inflammatory cytokines in alzheimer disease', *Journal of Neuropathology & Experimental Neurology*, vol. 76**,** no. 1, pp. 27-31 [Online]. DOI: 10.1093/jnen/nlw099 (Accessed 3 June 2019).

Nudel, R., *et. al.* (2014), 'Genome-wide association analyses of child genotype effects and parent-of-origin effects in specific language impairment', *Genes, Brain and Behavior*, vol. 13**,** no. 4, pp. 418-429 [Online]. DOI: 10.1111/gbb.12127 (Accessed 3 June 2019).

O'Roak, B. J., *et. al.* (2011), 'Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations', *Nature Genetics*, vol. 43**,** no. 6, pp. 585-589 [Online]. DOI: 10.1038/ng.835 (Accessed 3 June 2019).

Park, G., Tan, J., Garcia, G., Kang, Y., Salvesen, G. and Zhang, Z. (2016), 'Regulation of histone acetylation by autophagy in parkinson disease', *Journal of Biological*

Chemistry, vol. 291**,** no. 7, pp. 3531-3540 [Online]. DOI: 10.1074/jbc.M115.675488 (Accessed 3 June 2019).

Peter, B., Matsushita, M., Oda, K. and Raskind, W. (2014), 'De novo microdeletion of BCL11A is associated with severe speech sound disorder', *American Journal of Medical Genetics Part A*, vol. 164a**,** no. 8, pp. 2091-2096 [Online]. DOI: 10.1002/ajmg.a.36599 (Accessed 3 June 2019).

Peter, B., Raskind, W. H., Matsushita, M., Lisowski, M., Vu, T., Berninger, V. W., Wijsman, E. M. and Brkanac, Z. (2011), 'Replication of CNTNAP2 association with nonword repetition and support for FOXP2 association with timed reading and motor activities in a dyslexia family sample', *Journal of Neurodevelopmental Disorders*, vol. 3**,** no. 1, pp. 39-49 [Online]. DOI: 10.1007/s11689-010-9065-0 (Accessed 3 June 2019).

Raca, G., Baas, B. S., Kirmani, S., Laffin, J. J., Jackson, C. A., Strand, E. A., Jakielski, K. J. and Shriberg, L. D. (2013), 'Childhood apraxia of speech (CAS) in two patients with 16p11.2 microdeletion syndrome', *European Journal of Human Genetics*, vol. 21**,** no. 4, pp. 455-459 [Online]. DOI: 10.1038/ejhg.2012.165 (Accessed 3 June 2019).

Reader, R. H., Covill, L. E., Nudel, R. and Newbury, D. F. (2014), 'Genome-wide studies of specific language impairment', *Current Behavioral Neuroscience Reports*, vol. 1**,** no. 4, pp. 242-250 [Online]. DOI: 10.1007/s40473-014-0024-z (Accessed 3 June 2019).

Reilly, S., Tomblin, B., Law, J., McKean, C., Mensah, F. K., Morgan, A. T., Goldfeld, S., Nicholson, J. M. and Wake, M. (2014) 'Specific language impairment: A convenient label for whom?', *International Journal of Language & Communication Disorders,* vol. 49, no. 4**,** pp. 416-451 [Online]. DOI: 10.1111/1460-6984.12102 (Accessed 28 May 2019).

Reilly, S., Wake, M., Bavin, E. L., Prior, M., Williams, J., Bretherton, L., Eadie, P., Barrett, Y. and Ukoumunne, O. C. (2007) 'Predicting language at 2 years of age: A prospective community study', *Pediatrics,* vol. 120, no. 6**,** pp. 1441-1449 [Online]. DOI: 10.1542/peds.2007-0045 (Accessed 28 May 2019).

Reilly, S., Wake, M., Ukoumunne, O. C., Bavin, E., Prior, M., Cini, E., Conway, L., Eadie, P. and Bretherton, L. (2010) 'Predicting language outcomes at 4 years of age: Findings from the Early Language in Victoria Study', *Pediatrics,* vol. 126, no. 6**,** pp. 1530-1537 [Online]. DOI: 10.1542/peds.2010-0254 (Accessed 28 May 2019).

Reuter, M. S., *et. al.* (2017), 'FOXP2 variants in 14 individuals with developmental speech and language disorders broaden the mutational and clinical spectrum', *Journal of Medical Genetics*, vol. 54**,** no. 1, pp. 64-72 [Online]. DOI: 10.1136/jmedgenet-2016-104094 (Accessed 3 June 2019).

Rice, M. L. (2012), 'Toward epigenetic and gene regulation models of specific language impairment: looking for links among growth, genes, and impairments', *Journal of Neurodevelopmental Disorders*, vol. 4**,** no. 1, pp. 27 [Online]. DOI: 10.1186/1866-1955-4-27 (Accessed 3 June 2019).

Rietveld, C. A., *et. al.* (2013), 'GWAS of 126,559 individuals identifies genetic variants associated with educational attainment', *Science*, vol. 340, no. 6139, pp. 1467-1471 [Online]. DOI: 10.1126/science.1235488 (Accessed 12 September 2019).

Rocca, M. S., Pecile, V., Cleva, L., Speltra, E., Selice, R., Di Mambro, A., Foresta, C. and Ferlin, A. (2016), 'The klinefelter syndrome is associated with high recurrence of copy number variations on the X chromosome with a potential role in the clinical phenotype', *Andrology*, vol. 4**,** no. 2, pp. 328-334 [Online]. DOI: 10.1111/andr.12146 (Accessed 3 June 2019).

Sanders, S. J., *et. al.* (2015), 'Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci', *Neuron*, vol. 87**,** no. 6, pp. 1215-1233 [Online]. DOI: 10.1016/j.neuron.2015.09.016 (Accessed 3 June 2019).

Scerri, T. S., Morris, A. P., Buckingham, L. L., Newbury, D. F., Miller, L. L., Monaco, A. P., Bishop, D. V. and Paracchini, S. (2011), 'DCDC2, KIAA0319 and CMIP are associated with reading-related traits', *Biological Psychiatry*, vol. 70**,** no. 3, pp. 237-245 [Online]. DOI: 10.1016/j.biopsych.2011.02.005 (Accessed 3 June 2019).

Sciberras, E., Mueller, K. L., Efron, D., Bisset, M., Anderson, V., Schilpzand, E. J., Jongeling, B. and Nicholson, J. M. (2014) 'Language problems in children with ADHD: A community-based study,' *Pediatrics,* vol.5**,** no. pp 793 [Online]. DOI: 10.1542/peds.2013-3355 (Accessed 28 May 2019).

Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014), 'Biological insights from 108 schizophrenia-associated genetic loci', *Nature*, vol. 511**,** no. 7510, pp. 421-427 [Online]. DOI: 10.1038/nature13595 (Accessed 3 June 2019).

Shriberg, L. D., Tomblin, J. B. and McSweeny, J. L. (1999) 'Prevalence of dpeech delay in 6-year-old children and comorbidity with language impairment', *Journal of Speech, Language, and Hearing Research,* vol. 42, no. 6**,** pp. 1461-1481 [Online]. DOI: 10.1044/jslhr.4206.1461 (Accessed 28 May 2019).

Simpson, N. H., *et. al.* (2015), 'Genome-wide analysis identifies a role for common copy number variants in specific language impairment', *European Journal of Human Genetics*, vol. 23**,** no. 10, pp. 1370-1377 [Online]. DOI: 10.1038/ejhg.2014.296 (Accessed 3 June 2019).

S. L. I. Consortium (2002), 'A genomewide scan identifies two novel loci involved in specific language impairment', *American Journal of Human Genetics*, vol. 70**,** no. 2, pp. 384-398 [Online]. DOI: 10.1086/338649 (Accessed 3 June 2019).

Smith, S. D. (2011), 'Approach to epigenetic analysis in language disorders', *Journal of Neurodevelopmental Disorders*, vol. 3**,** no. 4, pp. 356-64 [Online]. DOI: 10.1007/s11689-011-9099-y (Accessed 3 June 2019).

Snijders Blok, L., *et. al.* (2018), 'CHD3 helicase domain mutations cause a neurodevelopmental syndrome with macrocephaly and impaired speech and

language', *Nature Communications*, vol. 9**,** no. 1, pp. 4619 [Online]. DOI: 10.1038/s41467-018-06014-6 (Accessed 3 June 2019).

Soblet, J., Dimov, I., Graf von Kalckreuth, C., Cano-Chervel, J., Baijot, S., Pelc, K., Sottiaux, M., Vilain, C., Smits, G. and Deconinck, N. (2018), 'BCL11a frameshift mutation associated with dyspraxia and hypotonia affecting the fine, gross, oral, and speech motor systems', *American Journal of Medical Genetics Part A*, vol. 176**,** no. 1, pp. 201-208 [Online]. DOI: 10.1002/ajmg.a.38479 (Accessed 3 June 2019).

Sollis, E., Graham, S. A., Vino, A., Froehlich, H., Vreeburg, M., Dimitropoulou, D., Gilissen, C., Pfundt, R., Rappold, G. A., Brunner, H. G., Deriziotis, P. and Fisher, S. E. (2015), 'Identification and functional characterization of de novo FOXP1 variants provides novel insights into the etiology of neurodevelopmental disorder', *Human Molecular Genetics*, vol. 25**,** no. 3, pp. 546-557 [Online]. DOI: 10.1093/hmg/ddv495 (Accessed 3 June 2019).

Spinath, F. M., Price, T. S., Dale, P. S. and Plomin, R. (2004), 'The genetic and environmental origins of language disability and ability', *Child Development*, vol. 75**,** no. 2, pp. 445-454 [Online]. DOI: 10.1111/j.1467-8624.2004.00685.x (Accessed 3 June 2019).

Srivastava, S., Cohen, J. S., Vernon, H., Baranano, K., McClellan, R., Jamal, L., Naidu, S. and Fatemi, A. (2014), 'Clinical whole exome sequencing in child neurology practice', *Annals of Neurology*, vol. 76**,** no. 4, pp. 473-483 [Online]. DOI: 10.1002/ana.24251(Accessed 3 June 2019).

St Clair, M. C., Pickles, A., Durkin, K. and Conti-Ramsden, G. (2011) 'A longitudinal study of behavioral, emotional and social difficulties in individuals with a history of Specific Language Impairment (SLI)', *Journal of communication disorders,* vol. 44, no. 2**,** pp. 186-199 [Online]. DOI: 10.1016/j.jcomdis.2010.09.004 (Accessed 28 May 2019).

St John, M., Ponchard, C., van Reyk, O., Mei, C., Pigdon, L., Amor, D. J. and Morgan, A. T. (2019) 'Speech and language in children with Klinefelter syndrome', *Journal of Communication Disorders,* vol. 78, pp. 84-96 [Online]. DOI: 10.1016/j.jcomdis.2019.02.003 (Accessed 28 May 2019).

St Pourcain, B., *et. al.* (2014), 'Common variation near ROBO2 is associated with expressive vocabulary in infancy', *Nature Communications*, vol. 5**,** no. 5, pp. 4831[Online]. DOI: 10.1038/ncomms5831 (Accessed 3 June 2019).

Stromswold, K. (1998), 'Genetics of spoken language disorders', *Human Biology*, vol. 70**,** no. 2, pp. 297-324 [Online]. Available at www.jstor.org/stable/41465640 (Accessed 3 June 2019).

The 1000 genomes project consortium (2015), 'A global reference for human genetic variation', *Nature*, vol. 526**,** no. 7571, pp. 68-74 [Online]. DOI: 10.1038/nature15393 (Accessed 3 June 2019).

Thevenon, J., *et. al.* (2013), '12p13.33 microdeletion including ELKS/ERC1, a new locus associated with childhood apraxia of speech', *European Journal of Human Genetics*, vol. 21**,** no. 1, pp 82-88 [Online]. DOI: 10.1038/ejhg.2012.116 (Accessed 3 June 2019).

Tomblin, J. B., O'Brien, M., Shriberg, L. D., Williams, C., Murray, J., Patil, S., Bjork, J., Anderson, S. and Ballard, K. (2009), 'Language features in a mother and daughter of a chromosome 7;13 translocation involving FOXP2', *Journal of Speech Language and Hearing Research*, vol. 52**,** no. 5, pp. 1157-1174 [Online]. DOI: 10.1044/1092-4388(2009/07-0162) (Accessed 3 June 2019).

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E. and O'Brien, M. (1997) Prevalence of Specific Language Impairment in kindergarten children', *Journal of Speech, Language, and Hearing Research,* vol. 40, no. 6**,** pp. 1245-1260 [Online]. DOI: 10.1044/jslhr.4006.1245 (Accessed 28 May 2019).

Turner, S. J., Hildebrand, M. S., Block, S., Damiano, J., Fahey, M., Reilly, S., Bahlo, M., Scheffer, I. E. and Morgan, A. T. (2013), 'Small intragenic deletion in foxp2 associated with childhood apraxia of speech and dysarthria', *American Journal of Medical Genetics Part A*, vol. 161a**,** no. 9, pp. 2321-2326 [Online]. DOI: 10.1002/ajmg.a.36055 (Accessed 3 June 2019).

Van der Aa, N., Vandeweyer, G., Reyniers, E., Kenis, S., Dom, L., Mortier, G., Rooms, L. and Kooy, R. F. (2012), 'Haploinsufficiency of CMIP in a girl with autism spectrum disorder and developmental delay due to a de novo deletion on chromosome 16q23.2', *Autism Research*, vol. 5**,** no. 4, pp. 277-281 [Online]. DOI: 10.1002/aur.1240 (Accessed 3 June 2019).

Van Ijzendoorn, M. H., Belsky, J. and Bakermans-Kranenburg, M. J. (2012), 'Serotonin transporter genotype 5HTTLPR as a marker of differential susceptibility? A meta-analysis of child and adolescent gene-by-environment studies', *Translational Psychiatry*, vol. 2**,** pp. e147 [Online]. DOI: 10.1038/tp.2012.73 (Accessed 3 June 2019).

Veltman, J. A. and Brunner, H. G. (2010), 'Understanding variable expressivity in microdeletion syndromes', *Nature Genetics*, vol. 42**,** no. 3, pp. 192-193 [Online]. DOI: 10.1038/ng0310-192 (Accessed 3 June 2019).

Vernes, S. C., Newbury, D. F., Abrahams, B. S., Winchester, L., Nicod, J., Groszer, M., Alarcon, M., Oliver, P. L., Davies, K. E., Geschwind, D. H., Monaco, A. P. and Fisher, S. E. (2008), 'A functional genetic link between distinct developmental language disorders', *New England Journal of Medicine*, vol. 359**,** no. 22, pp. 2337-2345 [Online]. DOI: 10.1056/NEJMoa0802828 (Accessed 3 June 2019).

Villanueva, P., Newbury, D. F., Jara1, L., De Barbieri, Z., Mirza, G., Palomino, H. M., Fernandez, M. A., Cazier, J., Monaco, A.P. and Palomino H. (2011) 'Genome-wide analysis of genetic susceptibility to language impairment in an isolated Chilean population', *European Journal of Human Genetics,* European Journal of Human

Genetics, vol.19, pp. 687–695 [Online] DOI: 10.1038/ejhg.2010.251 (Accessed 7 June 2019)

Virgin, H. W. and Todd, J. A. (2011), 'Metagenomics and personalized medicine', *Cell*, vol. 147**,** no. 1, pp. 44-56 [Online]. DOI: 10.1016/j.cell.2011.09.009 (Accessed 3 June 2019).

Wainschtein, P., *et. al.* (2019), 'Recovery of trait heritability from whole genome sequence data', *bioRxiv***,** 588020 [Online]. DOI: 10.1101/588020 (Accessed 3 June 2019).

Wake, M., Poulakis, Z., Hughes, E. K., Carey-Sargeant, C. and Rickards, F. W. (2005) 'Hearing impairment: A population study of age at diagnosis, severity, and language outcomes at 7-8 years', *Archives of Disease in Childhood,* vol. 90, no. 3**,** pp. 238-44 [Online]. DOI: 10.1136/adc.2003.039354 (Accessed 28 May 2019).

White, S. M., Morgan, A. T., Da Costa, A., Lacombe, D., Knight, S. J. L., Houlston, R., Whiteford, M. L., Newbury-Ecob, R. A. and Hurst, J. A. (2010) 'The phenotype of Floating-Harbor syndrome in 10 patients', *American Journal of Medical Genetics, Part A,* vol. 152A, no. 4**,** pp. 821-829 [Online]. DOI: 10.1002/ajmg.a.33294. (Accessed 28 May 2019).

Whitehouse, A. J., Bishop, D. V., Ang, Q. W., Pennell, C. E. and Fisher, S. E. (2011), 'CNTNAP2 variants affect early language development in the general population', *Genes, Brain and Behavior*, vol. 10**,** no. 4, pp. 451-456 [Online]. DOI: 10.1111/j.1601-183X.2011.00684.x (Accessed 3 June 2019).

The World Health Organisation (2010). *ICD-10: International Statistical Classification of Diseases and Related Health Problems (10th Rev. Edn).* Geneva, World Health Organization. (Available at: https://icd.who.int/browse10/2010/en#/F80-F89) (Accessed 28 May 2019).

Zeesman, S., Nowaczyk, M. J., Teshima, I., Roberts, W., Cardy, J. O., Brian, J., Senman, L., Feuk, L., Osborne, L. R. and Scherer, S. W. (2006), 'Speech and language impairment and oromotor dyspraxia due to deletion of 7q31 that involves FOXP2', *American Journal of Medical Genetics Part A*, vol. 140**,** no. 5, pp. 509-514 [Online]. DOI: 10.1002/ajmg.a.31110 (Accessed 3 June 2019).

List of items for glossary:

Allele: a variation of a genetic sequence that is a single variant or region of DNA relevant to a disorder or characteristic
Childhood Apraxia of Speech: a neurological speech sound disorder resulting from deficits in speech-motor programming and/or planning
Complex Genetic Disorder: a disorder which results from a combination of genetic and environmental factors
Copy Number Change (CNV): submicroscopic losses and gains of genetic material
De novo genetic model: a genetic model in which the causative variant is not inherited from either parent, but has occurred spontaneously
Dizygotic Twins: non-identical twins who share approximately half of their DNA
Dominant Inheritance: when inheriting one copy of a causative variant is enough to result in a genetic disorder
Epigenetics: a field of genetic study which examines factors outside of the DNA sequence that can affect the way in which genes work. For example, the regulatory control of genes in which they are turned on and off
Gene-gene interaction: also known as epistasis – where variants in two genes interact with each other, resulting in a phenotype that would not occur in the presence of only one of the alleles
Gene-environment interaction: where a genetic variant or allele is modulated by environmental effects
Genome-Wide Association Study: a method for studying regions of the genome commonly shared by affected individuals in complex disease
Fully penetrant: see dominant inheritance – when a disorder always results from inheriting causative variant(s)
Heritability: the extent to which a trait is thought to be genetically inherited
Incomplete penetrance: see dominant inheritance – when a disorder does not always result from inheriting a causative variant(s)
Linkage Study: a method for studying regions of the genome that are shared between affected relatives to aid identification of genes involved in a characteristic
Mendelian Inheritance: inheriting high impact variant(s) from parents that result in a phenotype
Missing heritability: the proportion of a characteristic that is thought to be biological, but is yet to be understood
Monozygotic Twins: identical twins who share 100% percent of their DNA
Phenotype: An observable set of characteristics resulting from the expression of genetic variants and interaction with environment
Pleiotropy: when variants in a gene result in different disorders
Rare Variant: a rare genetic spelling mistake found in less than 1% of the population
Recessive Disorder: when two copies of causative variants are needed to be inherited in order for a genetic disorder to occur
Trait: a particular characteristic or quantitative measure
Variant, genetic: a spelling change carried in the DNA, also known as an allele or mutation

List of items for index:

Allele
Childhood Apraxia of Speech
Complex Genetic Disorder
Copy Number Change (CNV

De novo genetic model
Dizygotic Twins
Dominant Inheritance
Dyslexia
Epigenetics
Gene-gene interaction
Gene-environment interaction
Genome-Wide Association Study
Fully penetrant
Heritability
Incomplete penetrance
Linkage Study
Mendelian Inheritance
Missing heritability
Monozygotic Twins
Phenotype
Pleiotropy
Rare Variant
Recessive Disorder
Specific Language Impairment
Trait
Variant, genetic