The Genetics of Language Acquisition

Hayley S. Mountford and Dianne F. Newbury, Oxford Brookes University, UK

Introduction

Language acquisition in early childhood is remarkably robust. Even in cases of severe neurological disorders, a child's ability to understand and use language is often preserved to some degree. Such findings demonstrate the human drive to communicate and underpin the idea that language is innate (Pinker, 1994), and therefore is likely to be genetically driven. There has been considerable debate in linguistics, philosophy, psychology and genetics as to whether language acquisition is innate or taught, a blend of both, or perhaps a balance which shifts at different points in a child's development. This chapter will focus on the current understanding of the role of genetics in language. When thinking about the role of genetics in language acquisition, it is important to consider it within a framework of other external influences discussed elsewhere in this book.

The extent to which language ability is inherited, or heritable, is well established in the field (Barry, Yasin, & Bishop, 2007; Bishop, Adams, & Norbury, 2006; Stromswold, 1998). Genetic factors play a key role in an individual's ability to successfully acquire and use language. Many genetic factors are usually involved, and they interact both with each other and environmental factors in a complex way. In this chapter, we will explore how genetics contribute to language, and how new genetic techniques can offer inroads into the molecular basis of language acquisition. We will also discuss the limitations of genetic studies, what they cannot tell us, and how this fits with our current understanding.

Genes do not act in isolation, and it is clear that environmental factors also play an important role in the acquisition of language. For example, maternal education level, and by proxy socio-economic status, is associated with academic and language abilities of their children (Reilly et al., 2010). A similar story is seen in measures of general cognition; that the mother's socio-economic status is associated with a child's IQ (Hanscombe et al., 2012). We discuss some of the key findings of gene x environment studies and provide a snapshot of the current understanding in the field, considering some of the limitations of this type of study design.

The relationships between language and other aspects of cognition are equally difficult to disentangle. As Vygotskiĭ noted "A word devoid of thought is a dead thing, and a thought unembodied in words remains a shadow" (Vygotskiĭ, 2012). Language ability shows a strong association with generalised cognitive ability: a high functioning child will generally develop more advanced language at an earlier age, and conversely, a cognitively delayed child will be more likely to struggle to meet language milestones. Language ability is also strongly linked to other developmental domains such as reading, which closely relate and influence each other. As such, these related domains need to be considered as part of a spectrum, and not

considered as independent phenotypes. In terms of scientific study, this can be extremely challenging. At a physiological level, language and speech are separate faculties (a mental representation and the realisation of that representation, accordingly). Nonetheless, these two attributes are intricately intertwined and can be difficult to disentangle at the behavioural level. For this reason, we have chosen to use "language" as an umbrella term encompassing both.

In this chapter we explore the current field of play in the genetics of language acquisition. We begin by discussing the heritability of language and the role of family and twin studies in the understanding of language. We go on to explore the inheritance mechanisms that are implicated in language development. Finally, we look forward by considering how modern DNA sequencing approaches are revolutionising the field of language genetics.

The Heritability of Language

There have been many studies that use twin pairs to evaluate the relative importance of genetics in language ability. These studies compare specific traits between monozygotic (or identical) twins, who share their entire DNA sequence, and dizygotic (or non-identical) twins who, on average, share half of their segregating DNA, that is, a similar level to ordinary siblings. Heritability studies look at the variation of a trait between and within twin pairs. They estimate the proportion of variation in a given trait that can be explained by genetic factors. Heritability estimates range from 0.0, for a trait that is entirely environmental, to 1.0, for a trait that is entirely genetic.

Environmental effects are often unique to any given individual (e.g. bumping your head as a child, or getting ill from a virus). This means that relationships between environmental factors and outcomes can be hard to predict unless we have a very detailed history across lots of individuals. Because twins are born and reared together, however, we can assume (rightly or wrongly) that they have a higher level of shared environmental factors (e.g. number of books in the home, number of days of education, parent's socioeconomic status) than two members of the general population (or even two members of the same family). This assumption provides a powerful mechanism by which we can estimate the importance of shared environmental influences upon a trait. If a given environment is shared between twins and if this environment directly affects an outcome, then any given twin pair should be more similar to each other than to any other member of the study population in terms of this outcome. This is true for all twin pairs, regardless of whether they are identical or non-identical. Similarly, by considering similarities between twin pairs, we can estimate the relative importance of genetic effects. Identical twins share their entire segregating DNA sequence. Thus, if genetic factors influence a given outcome, then identical twins should be similar to each other in terms of that outcome. However, this tenet does not hold true for non-identical twins who, on average, share only half of their DNA sequence.

Given that shared environment and genetic influences show different patterns of correlation between and within twin pairs, we can use these patterns of variation to estimate the relative importance of each of these factors. If a trait is purely genetic then monozygotic twins will always have the same outcome and dizygotic twins will have the same outcome half of the time. If a trait is purely down to shared environment then both monozygotic and dizygotic twins will always have the same outcome as their twin pairs. If a trait is purely due to unique environment then there will be a random pattern between twin type and outcome and, on average, both types of twin pairs will have the same outcome about half of the time. In reality, no trait can be clearly characterised into one of these three boxes, but by looking at patterns of variation within a trait, we can estimate the importance of genetics and environment. A major benefit to performing studies of this type is that these methods do not necessarily require us to actually look at the genetic information to estimate heritability and can be performed on large longitudinal study data sets.

An important finding in behavioural genetics is that the heritability of intelligence increases steadily throughout life. In infancy, genetic influences account for about 20% of variation in intelligence. In adulthood, they account for about 60% of variation. This, so-called, "Wilson effect" is thought to reflect a "genetic-amplification" procedure in which children select differential environments that compound their genetic propensities across a life-span (Plomin & Deary, 2015). Interestingly, Finkel et al. (1998) found that language ability is strongly and stably heritable across the adult life-span, and, perhaps unexpectedly that in this particular study, language was independent of general cognition. This finding is supported by more recent studies which suggest that gene x gene and gene x environment interactions become increasingly important in verbal ability in later life (Reynolds & Finkel, 2015).

The Wilson effect is primarily reported in adult samples and general intelligence. However the trends described are echoed in studies looking at language acquisition. In their study of children spanning 2 to 12 years old, Hayiou-Thomas, Dale, and Plomin (2012) found that environmental factors accounted for most of the differences seen in the 2 to 4 year old group, while heritability increased between the 2 to 4 and 7 to 10 year old groups. Their work suggested that genetic factors may become increasingly important as a child develops. Tosto et al. (2017) refined this even further, and showed that the heritability of language increases from age 7, 12 and 16. They also showed a strong genetic correlation between oral language and reading comprehension at the ages 12 and 16, and to a lesser extent between oral language and reading fluency at 7, 12 and 16. These findings again reflect studies of cognition where it is found that although the influence of genetic factors increased over a life-span, it is the same set of genetic factors that influence cognition over time. Furthermore, it is found that the same set of genes influence multiple different aspects of intelligence, from vocabulary to spatial memory to processing speed and executive function. These studies

clearly show that genetics plays an important role in the acquisition and maintenance of language throughout life but also that these effects can be moderated by life course.

What is more challenging to delineate is the specific role of genetic factors across generalised cognition, language and reading ability. The overlap of these developmental domains remains one of the biggest challenges in behavioural genetics.

Heritability in Language Disorders

In addition to looking at heritability of language ability in the general population, there have been several studies which specifically investigate the heritability of language disorders. The term language disorder is the diagnosis given to children with persistent difficulties in the use or understanding of language (Bishop et al., 2017). Within this broad umbrella term for all children who struggle with language, there are several more specifically defined sub-groups. These include developmental language disorders (DLDs), which overlap heavily with the previously used category of specific language impairment (SLI), in which a primary language difficulty exists without any other explanatory factor such as autism spectrum disorder (ASD), hearing loss, or developmental delay. Childhood apraxia of speech (CAS), also known as developmental verbal dyspraxia, is a specific fine motor control disorder affecting speech production, and ultimately language ability, frequently due to mutations in the *FOXP2* gene described later in this section. Specific subtypes and diagnosis of language impairments are discussed in detail in section III of this book.

Developmental language disorders are extremely common and occur in 7% of school age children, meaning we would expect to find two children with DLD in every primary school class of thirty children (Norbury et al., 2016). The long-term effects of language disorders present a significant burden on individuals and society alike. Children who struggle with language have a greater chance of developing anxiety and depression in adolescence (Conti-Ramsden & Botting, 2008), and are more likely to have lower socio-economic status in adulthood (Ruben, 2000). Despite the clear social and economic importance of understanding DLDs, we still do not understand much of the underlying aetiology.

DLDs have been shown by a number of studies to be highly heritable (Bishop, North, & Donlan, 1995; Hayiou-Thomas, Oliver, & Plomin, 2005; Lewis & Thompson, 1992) and it is clear that the risk of language disorder is increased if a first degree relative has a language disorder (Stromswold, 1998). Interestingly, studies have shown that language shows a higher degree of heritability in the lower range of language ability children than in the general population (Dale et al., 1998; Spinath et al., 2004). This finding may suggest that shared

genetics factors are more relevant to the extremes of the ability spectrum, compared to the typical range.

The understanding that environment is important in general cognitive ability (Finkel et al., 1998) and that interactions between genetic and environmental factors can influence the heritability of traits across time is crucial for the way in which we conceptualise both language acquisition and language disorders (Bishop & Hayiou-Thomas, 2008). Differences in testing procedures, how and when a child received a referral to speech and language clinics, and IQ can all affect the clinical diagnoses, which in turn, can affect heritability estimates.

Limitations of Heritability Studies

Heritability studies have provided many key insights into the genetics of both language acquisition and language disorders. There are, however, some important limitations to this methodology that should be considered. Heritability studies measure a particular trait, in a particular population, and a specific time. The way in which a trait is measured, the time at which the measurements are taken, and the study population can affect heritability estimates. As discussed below, genetic and heritability studies usually only consider additive genetic effects, rather than gene x gene and gene x environment interactions. This is an important point given that we know genetic effects themselves can change over time, responding to differing environmental pressures. As heritability studies consider one (or a small number) of traits, they cannot capture the multitude of factors that may be relevant at any one time. Similarly, differences in inclusion criteria or how a trait is measured may have a major impact on how heritable a trait is found to be (Bishop & Hayiou-Thomas, 2008). This last point is especially pertinent to language disorders which currently lack gold-standard testing procedures.

Twin studies suggest that heritability may vary for different aspects of language, for example heritability estimates of speech production tend to be higher than those for vocabulary (Bishop & Hayiou-Thomas, 2008; Hayiou-Thomas et al., 2006). This adds a further challenge in determining heritability, as difference exist between sub-phenotypes of global language ability and it is not always easy to distinguish between these subtypes.

There are also challenges with dealing with differences between monozygotic and dizygotic twin groupings when it comes to the consideration of language-related traits. For example, simply being born a twin is associated with delayed language, particularly for monozygotic twins. Termed the “twinning effect” (Rice et al., 2014), this may act to decrease estimates of genetic heritability by over-estimating shared environment influences. Rice et al. (2017)

recently showed that the twinning effect detected in younger children becomes less pronounced between 4 and 6 years old as twins successfully catch up to their single child peers. Similarly, they showed that the monozygotic twin pairs caught up to their dizygotic and single child peers, and from four years old the twinning and zygosity effects are no longer detectable.

Heritability studies have provided some interesting insights into language acquisition and allow us to model changes in relative influences over the life-span. While they have shown unequivocally that genetic factors play an important role in language, as in all genetic studies, they rely upon the binary concepts of genetic and environment. This oversimplifies the biological complexities of the genome, and how it interacts with conditions in a molecular pathway, in a particular cell type, at a specific point in development, at the mercy of a plethora of environmental factors. Heritability studies indicate that genetics may be important, but they do not allow us to identify which specific genetic variants, molecular pathways or cell types are involved. To begin to understand the underlying neuromolecular pathways involved in language acquisition, we need to undertake molecular studies which look at specific genes.

The Study of Language Through Extreme Traits

Language ability is thought to be a continuous and normally distributed trait, with most people falling around the mean of the bell curve of ability, and fewer people at the extreme ends of the range. As mentioned above, heritability studies indicate that the same genetic effects contribute across the distribution of ability but that heritability increases at the extremes. To a geneticist, therefore, the extremes of this spectrum (that is people who are incredibly good at language or those who struggle with language) have the potential to be most informative. These extremes have been the traditional place to begin exploring the underlying molecular genetics of language. The underlying premise of studying participants who sit at the extremes is that the neurological pathways or mechanisms that underpin their disorder are also important to language development in the general population. These extreme traits, known in genetics as phenotypes, are the key to understanding language acquisition. Most commonly, these studies focus upon individuals with DLD. Since DLD is characterised by “unexpected” language difficulties, they are presumed to represent a “pure” form of language difficulties reducing confounding developmental issues that may confuse an analysis.

Studying families in which several members share the same disorder has been a fruitful place to begin looking for genes and molecular pathways that may be involved in language. This approach assumes that family members with the same phenotype share the same underlying

genetic cause which results in their disorder. By investigating the regions shared between affected family members, we can identify the potential cause of their disorder.

Mendelian Genetics

Genetic studies typically focus upon large families in which a severe form of disorder is inherited in a characteristic way. These types of family-based disorders are termed Mendelian disorders because they show similar inheritance patterns (recessive and dominant) to those described by Gregor Mendel in his famous pea crossing experiment in the 1800's. Mendel showed that when he crossed purple flowering pea plants with white flowering pea plants, it resulted in all purple flowering offspring. When he crossed these purple offspring to create the next generation, he always ended up with the ratio of one white to three purple flowering pea plants, indicating that the genetic factor determining white flowers was present in the first generation even though it was not visible. Mendel explained this observation by proposing that genetic factors are inherited in pairs (one from each parent). Within any given pair, some outcomes can be genetically "dominant" (such as the purple flower) and some genetically "recessive" (such as the white flower). Thus, he discovered it is possible for a pea plant to carry a genetic variant (allele) that determines purple flowers and an allele that determines white flowers. In this case, the flowers will appear purple, and the white allele will be masked but still present. While we now understand that the picture is much more complicated than this, the basic principle of dominant and recessive model still applies to many genetic disorders and traits. Mendelian forms of language disorders are collectively rare, and may represent the most severe forms. They have, however, been extremely important in our understanding of language genetics.

To understand how Mendelian disorders can be used to understand language, we must first cover some basic genetics principles.

The human genome carries the instructions for how to build a human. It can be thought of as a book, with two identical copies carried in each cell in the body (Figure 1). Each book is broken up into chapters which represent the chromosomes - individual chunks of genome packaged into discreet units. We each have twenty two chapters and two sex chromosomes (either XX or XY), making a total of 46 chromosomes. Each chapter can be further divided into thousands of paragraphs, each representing a gene. Each paragraph contains a discrete set of instructions to build a single protein, in the form of combinations of letters arranged into words that are read by the molecular machinery within the cell. When there is a spelling mistake, or mutation, in one of these genes, it can lead to a word becoming unreadable and the meaning being completely changed. In some cases, these mutations can render the entire paragraph nonsensical as the true meaning of the sentence can no longer be interpreted by the

cellular machinery. The protein may not be produced at all or may be unable to carry out its function properly. These deficiencies may result in the person having a genetic disorder.

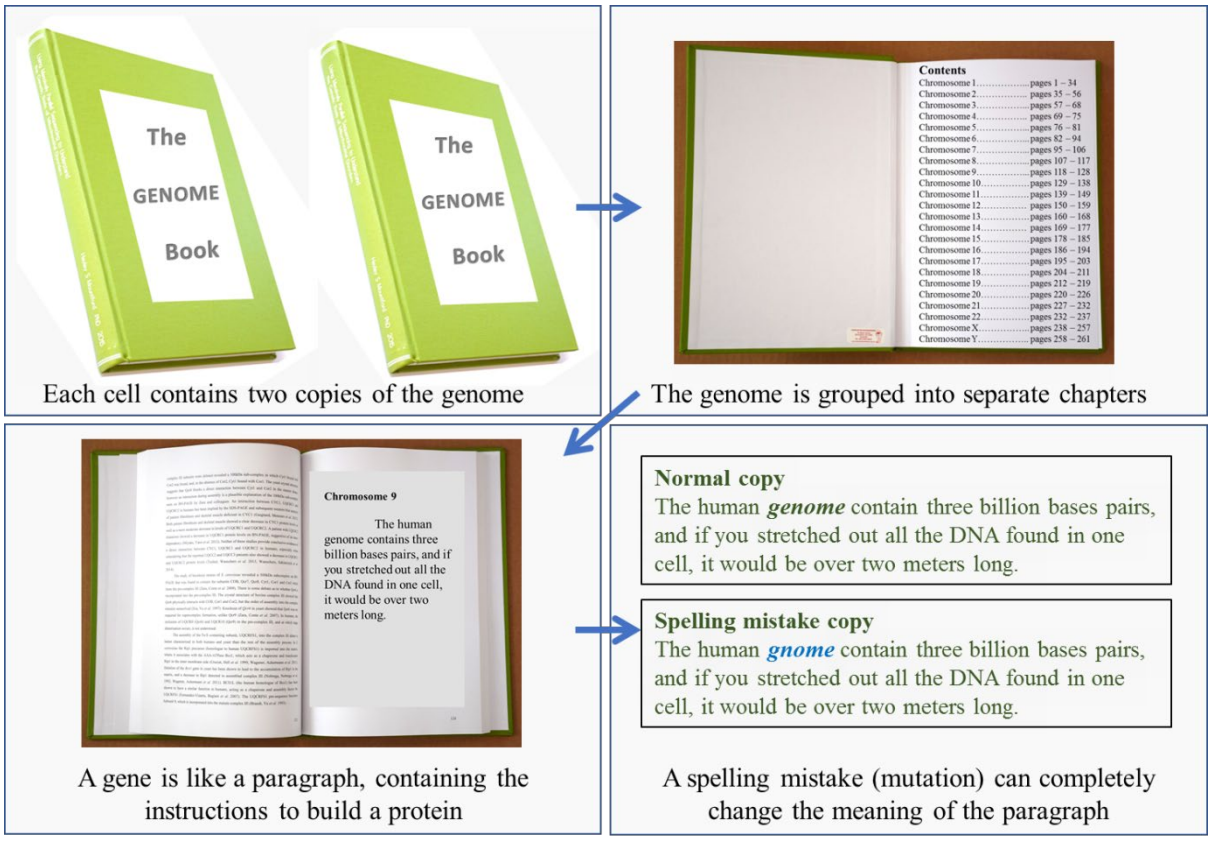

Figure 1 *- The human genome represented as a book.*

Examples of the patterns typical of dominant and recessive inherited disorders are shown in figure 2. Figure 2A shows a dominant inheritance pattern in a theoretical family of six. An affected family member is indicated by a shaded shape, and males are represented by a square and females by a circle. Figure 2A shows an affected Father (black square), unaffected Mother (white circle), and half of the children (n=2/4) are also affected. This ratio of affected offspring indicates that the underlying cause is likely to be dominantly inherited, that is, one copy of the mutation is enough for the trait to be visible. In this instance the mutation is shared by Father, and children 1 and 4. Dominant disorders are a special case where one functional copy of the protein is not enough to do the job properly, and the cell cannot function. In practice, a person still has two copies of the gene, but the dominant mutation masks the second wild-type copy. The wild-type copy is still present, and can be passed onto offspring but its function is overwritten by the mutant copy when both are present together. An example of a dominantly inherited genetic disorder is Huntington's disease, where affected family members develop severe progressive neurological deterioration.

In contrast, recessive disorders require both copies of the gene to be disrupted before a trait is visible. Figure 2B shows a recessive pedigree where both Father and Mother are unaffected. One of their children is affected while the other three remain unaffected. This pattern is typical of a recessive disorder. Both Father and Mother carry the mutation (termed carriers), but one mutant copy is not enough to cause disorder. By chance, one in four children will inherit two copies of the mutation (one from each parent) and so will be affected. The other three children will all be unaffected but, on average, two of them will be carriers (one copy of the mutation inherited) and so could pass a mutant allele onto their children. If their partner also happens to be a carrier then their children may also be affected. In recessive disorders, an affected person needs to inherit a mutation from both parents in the same gene. This combination renders the protein completely dysfunctional leading to disease. In contrast to a dominant disorder, these proteins are not dosage sensitive; as long as you can produce some protein, the function of the cell will not be disrupted, so one copy of a mutation is tolerated. Some examples of recessive genetic disorders include cystic fibrosis, Tay-Sachs disease, and sickle cell anaemia.

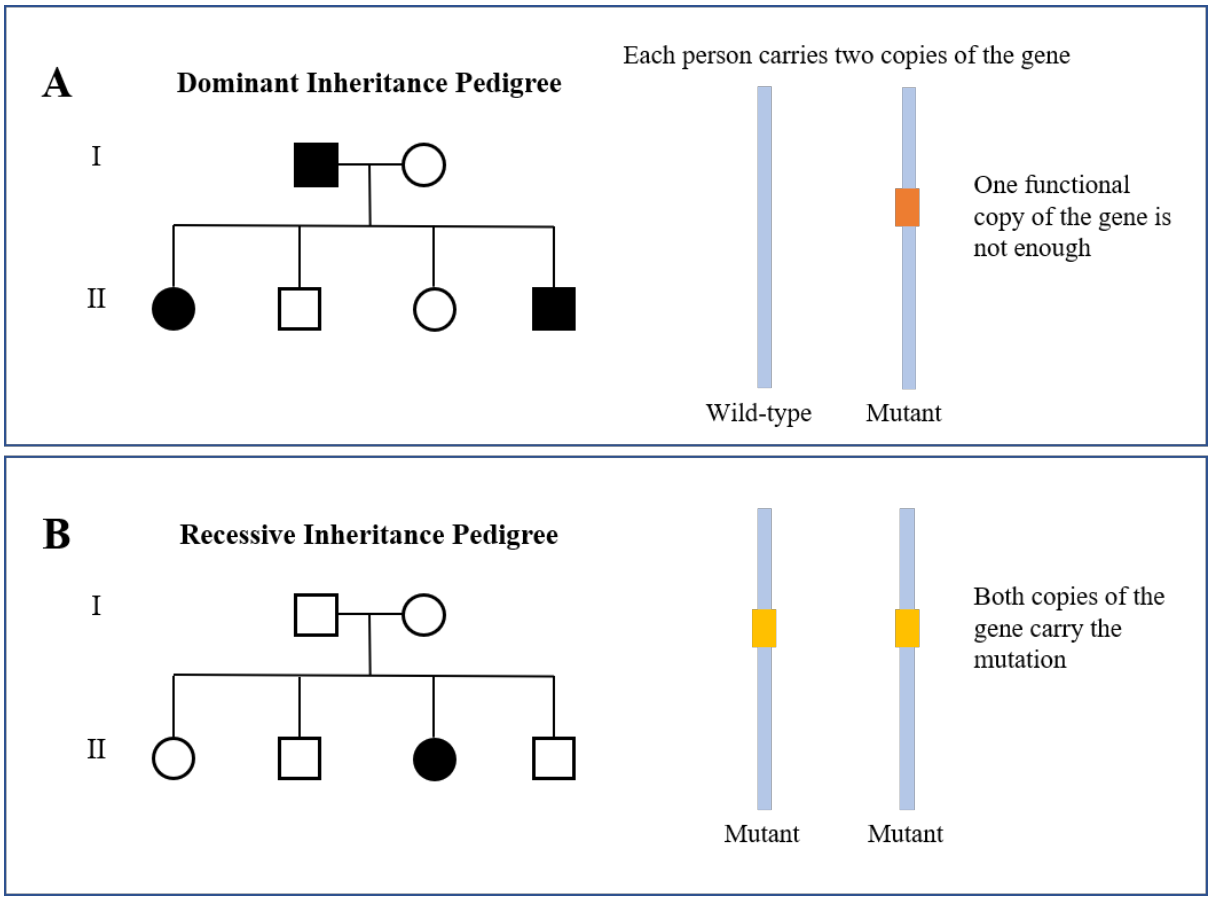

Figure 2A *shows an example dominant pedigree. Males are indicated by a square, females by a circle, affected family members are indicated by a black shape, and white indicates unaffected. In the example, the father is affected with disorder but the mother is not. Two of four children have inherited the mutation from the father and have the disease, and two are unaffected. Because the protein containing the mutation is dosage sensitive, inheriting only one working copy is not sufficient to produce enough functional protein.* ***Figure 2B*** *shows an*

example recessive pedigree. In the example, neither father nor mother is affected but both are carriers. One child is affected and carries two mutant copies, two children are carriers (one copy) and one child is unaffected with no mutations inherited.

Both these example pedigrees show the odds of inheriting a Mendelian mutation under both recessive and dominant models. However, in practice, patterns can vary from this theoretical model. A point of interest is that inheriting mutations is not always deleterious. We can inherit single copies of recessive mutations without adverse effects, and some changes do not affect protein function at all. Most mutations are inherited but on average, 100 mutations occur new (termed *de novo*) in every generation. Some of these will be deleterious, some may be advantageous and many will be neutral. To return to the Genome book analogy, non-damaging changes may alter how a word is spelled in a sentence, but the meaning can still be clearly interpreted. These types of tolerated changes are termed genetic variants, and are relatively common in the human genome. In general, a mutation eliminates the meaning of a critical word but may also alter the meaning to be more efficient. This is analogous to text speak where the spelling of words is altered but the meaning retained in a more efficient manner. These changes to the genome represent evolutionary changes that spread through families and populations over time. On average, the Human Genome carries around 2,000 protein altering changes (The 1000 Genomes Project Consortium, 2015), and over 1,000,000 genetic variants. The challenge in any genetic study is to figure out which of these are relevant to the disorder of interest.

Families provide a unique opportunity to study genetic contributions to disorders and traits. Children inherit half of their DNA from each parent. This means they share half of their segregating DNA with their Mother, half with their Father, and, on average, half with their sibling(s). We can use these shared regions to narrow down our search for relevant mutations. For example, if both children in a given family are affected, we can look for regions of the genome that are shared between them reducing our search space by 50%. Additional siblings and relatives allow us to further narrow our focus. Large families therefore represent a particularly useful tool for gene identification, particularly if they have multiple affected and unaffected family members. This approach is relatively straightforward if one mutation is involved and individuals can be classified as affected or unaffected. For example, in Figure 2A, we know that the Father, child 1 and child 4 all share the same mutation, whereas Mother, child 2 and child 3 do not. This principle forms the basis of Mendelian genetics studies, and has been extremely powerful in understanding the underlying cause of severe language disorders in families. We will explore some of these in the next section.

Insights from Mendelian Studies

The first, and most well-known example of a gene to be associated with a language disorder involved the KE family (Lai et al., 2001). Several members in this large multigenerational family present with a form of language disorder known as childhood apraxia of speech (CAS) (Figure 3). This disorder, described in detail in Chapter 19, manifests with a specific difficulty in the fine motor control movements required to produce speech sounds. In their studies of the KE family, Lai *et al.* identified a dominant mutation in the gene *FOXP2*. This mutation was found to be present in a single copy in all affected family members, but never in unaffected individuals. This gene encodes an important protein involved in the regulation of many developmental genes (Carlsson & Mahlapuu, 2002). Subsequent studies identified additional mutations in other unrelated individuals with CAS, and mutations in *FOXP2* are now a well characterised cause of CAS (MacDermot et al., 2005; Moralli et al., 2015; Reuter et al., 2017; Tomblin et al., 2009; Turner et al., 2013). Because it regulates other genes, changes in the levels of the FOXP2 protein can have big knock-on effects upon many pathways. They are therefore usually dominant – one mutation is enough to cause disorder. However, mutations in *FOXP2* are still extremely rare, and represent the genetic cause for only about 2% of CAS cases (Worthey et al., 2013). The KE family is an extremely unusual and special case. Cases of *FOXP2* mutations are collectively rare, and to find a large multigenerational family with a clearly dominant form of language disorder is exceptional.

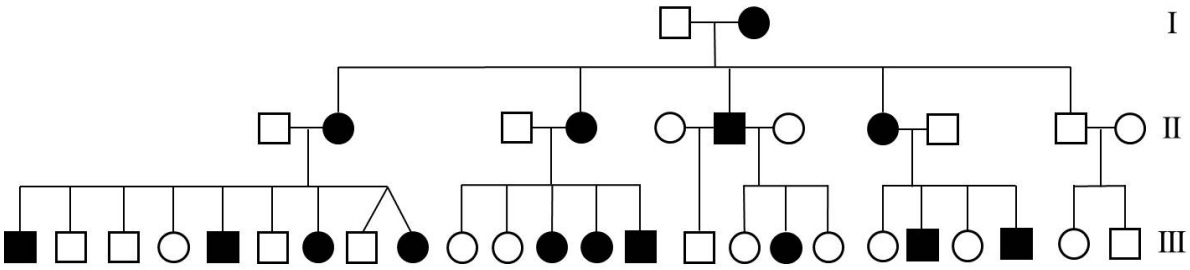

Figure 3 *Showing the family pedigree from the KE family. Affected family members are coloured in black, and unaffected family members are in white. Males are represented by squares and females by circles. Adapted from Watkins et al. (2002).*

FOXP2's most important and best characterised function is its role in the growth and control of neurons in the developing brain (Chabout et al., 2016; Spiteri et al., 2007; Vernes et al., 2011). This role has been shown to be important in neuroplasticity and is thought to be key to how we learn language. Studies in song-birds and mice show that FOXP2 also regulates neuroplasticity and vocalisations in other species. Perhaps somewhat surprisingly, FOXP2 is not just a brain protein, but it plays a role in the regulation of hundreds of genes in many parts of the human body that are unaffected by the mutation that causes CAS. For some reason, these mutations seem to particularly affect the brain.

FOXP2 paved the way for our current understanding of the molecular pathways involved in language development. Because of this, *FOXP2* is often referred to as a 'molecular window' into speech and language (Fisher & Scharff, 2009). It has formed the basis for an entirely new field of research, and has led to a wealth of discoveries that aid our understanding. One such example is the identification of the *CNTNAP2* gene (Vernes et al., 2008), a synaptic protein, which is regulated by *FOXP2*, and has also been associated with both language disorders (Devanna et al., 2017), ASD (Alarcon et al., 2008; Arking et al., 2008) and a wide range of neuropsychiatric disorders and functions. Such studies indicate that FOXP2-regulated pathways are likely to be important in brain development and relevant to many cognitive processes beyond language acquisition.

The identification and characterisation of *FOXP2* has facilitated important insights into how the human brain develops language. However current understanding is far from comprehensive, and many questions remain. Each identified pathway can open up additional questions and networks that need exploring. For example, we know that immature neurons require guiding to find their way to the final location where they then fully mature. These represent important steps in brain development (Vernes et al., 2011; Vernes et al., 2007). Retinoic acid, a metabolite of vitamin A, has been shown to play a vital role in cellular differentiation of neurons (Devanna, Middelbeek, & Vernes, 2014) and is driven by *FOXP2*. These neurodevelopmental pathways are critical for brain function, and it appears that their role in language may be extremely important and understudied.

Aside from *FOXP2*, genetic studies have also identified some unlikely pathways in the quest to understand language disorders. One particularly example is the gene *GNPTAB* which was identified across large, multigenerational Pakistani families with persistent stuttering (Kang et al., 2010). The *GNPTAB* gene encodes a lysosomal protein that is important for the recycling of proteins within cells. This represents a basic cellular process that is relevant to all cells and, as with *FOXP2*, appears to affect brain cells in particular. Disruption of lysosomal pathways can lead to a variety of symptoms from developmental delay to seizures, deafness and blindness. Exactly how this mechanism is linked to stuttering and language is yet to be uncovered, but it is clear that there is still a lot to be investigated, and discovered.

Insights into mechanism can also come from the opposite end of the language ability spectrum. A recent study identified a rare variant in a Serbian family with the ability to reverse words with incredible speed and precision (Prekovic et al., 2016). The variant lies in the gene *RIC3*, which is involved in establishing communication systems between neural cells. Prekovic and colleagues showed that the family members able to backwards speak rapidly have exceptional working memory, allowing words to be reversed. The mechanisms involved in this pseudo-language may shed light on natural language-acquisition processes. This study shows that novel genes and pathways can be identified not just from language

disorders, but that families with super language abilities can be just as useful in discovering new mechanisms underlying language acquisition.

Limitations to Mendelian Studies

Each of the above cases provides an example of when a family or a small number of individuals led to the identification of a gene. Each gene is part of a specific molecular pathway which leads to an improved understanding of individual influences upon language. Each new gene identified may only directly affect a small number of individuals, but it can, as in the case of *FOXP2*, open up a "molecular window" facilitating major insights in neurodevelopment and revolutionising the field (Fisher & Scharff, 2009). Each finding is a small part of a highly complex picture, with each new insight providing a gateway into specific processes and pathways of importance. Language acquisition is clearly an extremely complex process involving many aspects of neurodevelopmental and a plethora of genetic and environmental factors. It is likely that a large numbers of molecular pathways play a role in the successful acquisition of language, and to gain a clear picture of how this works will be extremely challenging. Nonetheless, the more steps we can uncover, the more links we can make and the clearer the picture will become.

In each of the three cited examples, the variant responsible has fallen within the coding region of the genome, creating a clearly and easily interpretable mutation. However, the majority of the Human Genome does not directly code for proteins. Variations in these regions are far less well understood, but are likely to play a role in how genes are controlled. Interpreting these non-coding variants can be extremely challenging, and as a result, clear links between genetic changes and protein function can be few and far between. As the current understanding of these non-coding variants increases, this will improve.

Complex genetic models

Mendelian disorders represent a special case in which a single mutation is necessary and sufficient to cause a disorder that is easily identifiable and distinct from typical development. However, we know that in the majority of cases, these principles are unlikely to hold. The diagnosis of DLD is not always clear-cut. Individuals can have borderline diagnoses, simply because the inclusion criteria are arbitrarily assigned, and will often miss children who sit on this border. Even in cases where children are clearly affected, the nature of disorder will vary between individuals and over time. DLD diagnoses represent a spectrum of severity and dimensions and it is unclear whether the same factors contribute across these dimensions. In the majority of cases, Mendelian dominant and recessive models will not be appropriate. We have come to understand that there is not one clear genetic cause for the majority of DLD cases, but that the model instead involves many variants that interact to contribute to genetic

susceptibility or risk. This is termed a "complex genetic model", meaning that each variant contributes to an individual's overall level of risk of developing a language disorder. Coming back to our analogy, this is akin to many spelling errors spread across a paragraph that obscure the meaning. While "text-speak" can make communication more efficient, the combination of contracted variations can sometimes make it difficult to understand. The knowledge of the surrounding context and external factors (such as exposure to technology) may affect the ability of the individual to cope with these combined variations. Similarly, combinations of genetic variations can put an individual at an increased risk of disorder. A familiar example of a complex disorder is type 2 diabetes – it is a common disorder in which a balance of genetic susceptibility and lifestyle factors combine to modulate individual risk (Fuchsberger et al., 2016). Each individual genetic variant and lifestyle factor explains a proportion of the risk matrix but the combinations of factors can differ between individuals and no one factor alone explains the majority of risk.

One successful method for identifying genetic risk variants is to perform a genome wide association study (GWAs). These look for shared variants that are over-represented in a large cohort of unrelated patients. GWAs are excellent at identifying common variants in common disease but rely upon careful diagnosis of affection status and explain only part of the genetic risk model. By definition, complex genetic disorders involve hundreds of genetic variants, many of which will only have a small impact upon risk. These variants are much more challenging to identify than the high impact variants associated with Mendelian disorder.

DLDs present several difficult challenges to gene identification approaches. Although some candidate genes (*CNTNAP2*, *CMIP*, *ATP2C2*) and chromosome regions (chromosome 15q deletions) have been identified, large-scale studies are still few and far between. Several GWAs studies have been completed but no variants have been found to confer an increased risk of language disorders across investigations (Carrion-Castillo et al., 2016; Eicher et al., 2013; Gialluisi et al., 2014; Harlaar et al., 2014; Luciano et al., 2013; Nudel et al., 2014; St Pourcain et al., 2014). For a recent review of candidate genes and GWAs, we refer the reader to Carrion-Castillo et al. (2016). The lack of replication is characteristic of complex genetic disorders and is likely to reflect the low effect size of the contributory variants. In order to detect these variants, a GWAs on a huge scale would need to be performed. The goal post for number of participants required to detect moderate impact variants is demonstrated by a recent study into the genetics of schizophrenia which tested 37,000 schizophrenia patients and an enormous 113,000 controls, and were able to identify over 100 new regions of the genome (Ripke et al., 2014). This number shows that while previous studies have provided some useful insights, they are woefully underpowered to detect moderate impact variants. Most GWAs studies to date have had hundreds to thousands of participants, but as evidenced by the success of the schizophrenia study, larger numbers provide the power to detect regions with a lower effect size and are therefore harder to detect. GenLang (http://genlang.org/) is a

global initiative set up to address the challenges of study size and standardisation of phenotyping.

To conduct a "successful" GWAs, all participants must be characterised in the same way and all cases must share an underlying genetic aetiology. This can be particularly problematic for a disorder like DLD, in which we understand little of the contributing biology. When studying Mendelian inherited disorders that run in families, it is relatively easy to precisely split phenotypes into subsets, drawing clear boundaries around what constitutes DLD, and what constitutes an overlapping phenotype such as dyslexia or ASD. In unrelated individuals and a trait that differs between individuals and over time, this can be a daunting task. If two individuals are both classified as affected but, in fact, have different conditions, the assumptions underlying GWAs will be violated and the power to detect underlying factors will be reduced.

A ray of light in this sobering reality is offered by the heritability studies described at the beginning of this chapter. Recall that these studies suggest that the genetic contributions to verbal ability are relatively stable over time and overlap with those influencing other cognitive domains. If this is true, some genetic contributions will be shared between individuals even when clinical presentations differ slightly. Thus the GWAs approach can be used to detect these related and overlapping factors by lumping cases into broad phenotypes with shared related pathways. This premise is supported at the molecular level where we find that some genes contribute to multiple neurodevelopmental traits (e.g. *CNTNAP2* in Rodenas-Cuadrado, Ho, and Vernes (2014)) while others may contribute to particular subsets of individuals and disorders (e.g. *DCDC2* in Scerri et al. (2011). These overlaps confound the identification of disorder-specific variants but may actually assist in the identification of variants that affect general cognition and language. It may not be possible to separate out those that impact language only from those that affect cognitive processes in general. These scenarios make performing a large-scale GWAs of DLD an intimidating and exciting prospect.

Interactive Effects and Epigenetics

The complexity of language acquisition is evident from the neurological mechanisms of brain development. The genes encoded in our DNA play a vital role which we are only just beginning to understand, but they only tell part of the story. It is evident that environment plays perhaps an equally important role in language development. But how do our genes interact with our environment to make us better or worse at language? A relatively novel area of science attempts to address questions such as this by considering gene x gene, gene x environment interactions and epigenetic effects.

As explained above, we know that some genes play a general role in development while others have a more specific effect. Similarly, the role of a given genetic variants may be modulated by other variants or in particular environments. Genetic studies tend to consider specific genes or variants as independent factors but we know that genes act within dynamic interacting pathways that are responsive to environmental factors. The serotonin transporter gene *SLC6A4* has been associated with depression. A common variation within this gene results in two alternative protein forms - a short form (s) or a long form (l). The short form has been shown to increase the risk of negative developmental outcomes but this effect only becomes apparent when environmental stress is encountered (van Ijzendoorn, Belsky, & Bakermans-Kranenburg, 2012). Conversely, the short form is associated with positive developmental outcomes when the environment is positive (van Ijzendoorn et al., 2012). This fascinating study highlights how the same genetic variant can be associated with differential outcomes in response to a particular environment. If the environment was not considered within a study then the outcomes associated with the short form would be mixed and no clear pattern would be seen. It is only by this consideration that the patterns become clear. Although environment is clearly important in language and neurodevelopment, the number of dimensions associated with gene x gene or gene x environment screens currently prohibit the analysis of realistic models unless these studies are hypothesis driven.

To date, there have not been any studies published investigating the molecular gene x environment specific to language. However it represents an exciting future direction for the field. As noted above, it is sometimes difficult to delineate genetic effects from environmental effects, particularly when considering parents and children, who often share both genetics and environment. Gene x environment correlation is the term given to traits which influence each other, and then become part of the shared environment. This is particularly relevant to behavioural genetics where a genetic predisposition may alter the home environment, which then affects the phenotypic outcomes. For example, a recent study found that latent factors of maternal language significantly predict storybook exposure in children (Puglisi et al., 2017). It is important to bear in mind that correlation does not equate to causality, and that directionality cannot be assumed. Studies of gene x gene (Asbury, Wachs, & Plomin, 2005) and gene x environment correlation (Dale et al., 2015) have been attempted using twins, and while they are not the optimum way to tackle this question, they do represent the first studies of this exciting new approach.

The role of environmental factors in genetic susceptibility of DLDs offers an interesting research route. It may be that genetic susceptibility interacts with environmental factors such as lower socioeconomic status, that put a child at a higher risk of developing a language disorder given the “perfect storm” of conditions (Virgin & Todd, 2011). That is to say, each susceptibility variant may only contribute a small amount of risk, but when inherited together, under the right conditions, they may be enough to result in a language disorder.

Another new field of research that considers differential effects is epigenetics. It refers to alterations in gene regulation in response to environmental stimuli, without changing the

underlying DNA sequence. Some of these epigenetic modifications are heritable (inherited by the next generation) and some are not. Two of the most commonly referred to epigenetic mechanisms are DNA methylation and histone modification, but in general any extrinsic mechanism by which a gene is modulated can be referred to as epigenetics. DNA methylation and histone modifications refer to two different ways in which chemical modifications can modulate gene expression and both have been related to neurological dysfunction. For example, changes in DNA methylation profile have been associated with neurological disorders, particularly neurodegenerative disorders such as late onset Alzheimer's disease (Nicolia et al., 2017). Increased histone modifications in in dopaminergic neurons are associated with Parkinson's disease (Park et al., 2016). These modifications have long been recognised as methods of gene expression regulation but recent research indicates that epigenetic changes are much more dynamic than previously thought, and can alter in response to external factors such as diet, presence of toxins, or the environment in the womb, setting up new axes of epigenetic x environment interactions. In addition, we now understand that such modifications may be inherited leading to trans-generational changes in gene expression in response to environmental factors. Epigenetic changes therefore represent an intriguing mechanism which may partially explain the complex interplay between genetics and environmental factors, and why certain conditions results in a specific change in some people.

Epigenetic factors have been proposed as a possible contributory mechanism for DLDs by several groups (Kraft & DeThorne, 2014; Rice, 2012; Smith, 2011). Although no direct associations between language disorders and epigenetic control have been reported in the literature, recent studies do suggest a role for these mechanisms in ASD. Increased methylation of the oxytocin receptor gene (*OXTR*) and the reelin gene (*RELN*) have both been reported in individuals affected by ASD (reviewed by Loke, Hannan, and Craig (2015)). This system therefore represents a highly plausible mechanism which may be of relevance to language disorder.

Future Directions for the Genetics of Language Acquisition

Language disorders are incredibly common but have proved difficult to study due to the complexity of the trait and contributory mechanisms and influences. The underlying genetic mechanisms remain far from solved. Many valuable insights have come from studying families at the phenotypic extremes, and from studies looking at the population, as whole. There is still much to understand about how genetics plays a role in language acquisition.

The advent of Next Generation Sequencing (NGS) has opened up exciting possibilities for studying the role of genetics in language disorders. NGS allows for the rapid sequencing of the entire genome simultaneously, providing more complete genetic information than previously possible. The current challenge is the computational analysis that is required to make sense of such enormous volumes of data, and remains a bottle neck for genomic investigations. Similarly, NGS can be cost prohibitive, both in terms of the performing the sequencing itself, but also the cost of the computational resources and expertise required to

use it. NGS has revolutionised genetics, and will certainly be key to future advances in the genetics of language. Such studies (Chen et al., 2017; Devanna et al., 2017; Eising et al., 2018) are starting to appear in the literature and have facilitated the identification of new genes and pathways although sample sizes remain small.

In the future, genetic screening could be used to identify children at an increased risk of developing DLD and help them to access the most appropriate treatment, early. Intervention may well be effective, even if a child's particular disorder is largely genetic in origin, and access to the most appropriate educational support may improve the child's outcome. It may even be possible to use a genetic risk profile to guide the most effective form of educational support, using this to insure a child receives the most appropriate type of intervention.

Several challenging areas are currently being addressed by researchers around the world. The most fundamentally important being the efforts to develop better diagnostic terminology for DLDs (Bishop et al., 2016; Bishop et al., 2017). These will form a solid foundation for future studies. Larger studies, with systematic phenotyping of thousands of DLD cases and controls will allow for the identification of large and moderate impact genetic variants as risk factors for DLD. These will implicate new genes and molecular mechanisms in language acquisition, developing our understanding of how the brain learns to use language to communicate.

References

Alarcon, M., Abrahams, B. S., Stone, J. L., Duvall, J. A., Perederiy, J. V., Bomar, J. M., . . . Geschwind, D. H. (2008). Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. *Am J Hum Genet, 82*(1), 150-159. doi:10.1016/j.ajhg.2007.09.005

Arking, D. E., Cutler, D. J., Brune, C. W., Teslovich, T. M., West, K., Ikeda, M., . . . Chakravarti, A. (2008). A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism. *Am J Hum Genet, 82*(1), 160-164. doi:10.1016/j.ajhg.2007.09.015

Asbury, K., Wachs, T. D., & Plomin, R. (2005). Environmental moderators of genetic influence on verbal and nonverbal abilities in early childhood. *Intelligence, 33*(6), 643-661. doi:10.1016/j.intell.2005.03.008

Barry, J. G., Yasin, I., & Bishop, D. V. M. (2007). Heritable risk factors associated with language impairments. *Genes, Brain and Behavior, 6*(1), 66-76. doi:10.1111/j.1601-183X.2006.00232.x

Bishop, D., & Hayiou-Thomas, M. (2008). Heritability of specific language impairment depends on diagnostic criteria. *Genes, Brain and Behavior, 7*(3), 365-372. doi:10.1111/j.1601-183X.2007.00360.x

Bishop, D. V., North, T., & Donlan, C. (1995). Genetic basis of specific language impairment: Evidence from a twin study. *Developmental Medicine & Child Neurology, 37*(1), 56-71.

Bishop, D. V., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & consortium, C. (2016). CATALISE: A Multinational and Multidisciplinary Delphi Consensus Study. Identifying Language Impairments in Children. *PLoS One, 11*(7), e0158753. doi:10.1371/journal.pone.0158753

Bishop, D. V. M., Adams, C. V., & Norbury, C. F. (2006). Distinct genetic influences on grammar and phonological short-term memory deficits: evidence from 6-year-old twins. *Genes, Brain and Behavior, 5*(2), 158-169. doi:10.1111/j.1601-183X.2005.00148.x

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & Catalise-consortium, 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, 58*(10), 1068-1080. doi:10.1111/jcpp.12721

Carlsson, P., & Mahlapuu, M. (2002). Forkhead transcription factors: key players in development and metabolism. *Developmental biology, 250*(1), 1-23. doi:10.1006/dbio.2002.0780

Carrion-Castillo, A., van Bergen, E., Vino, A., van Zuijen, T., de Jong, P. F., Francks, C., & Fisher, S. E. (2016). Evaluation of results from genome-wide studies of language and reading in a novel independent dataset. *Genes, Brain and Behavior, 15*(6), 531-541. doi:10.1111/gbb.12299

Chabout, J., Sarkar, A., Patel, S. R., Radden, T., Dunson, D. B., Fisher, S. E., & Jarvis, E. D. (2016). A Foxp2 Mutation Implicated in Human Speech Deficits Alters Sequencing of Ultrasonic Vocalizations in Adult Male Mice. *Front Behav Neurosci, 10*, 197. doi:10.3389/fnbeh.2016.00197

Chen, X. S., Reader, R. H., Hoischen, A., Veltman, J. A., Simpson, N. H., Francks, C., . . . Fisher, S. E. (2017). Next-generation DNA sequencing identifies novel gene variants and pathways involved in specific language impairment. *Sci Rep, 7*, 46105. doi:10.1038/srep46105

Conti-Ramsden, G., & Botting, N. (2008). Emotional health in adolescents with and without a history of specific language impairment (SLI). *J Child Psychol Psychiatry, 49*(5), 516-525. doi:10.1111/j.1469-7610.2007.01858.x

Dale, P., Simonoff, E., Bishop, D., Eley, T., Oliver, B., Price, T., . . . Plomin, R. (1998). Genetic influence on language delay in two-year-old children. *Nature neuroscience, 1*(4), 324.

Dale, P. S., Tosto, M. G., Hayiou-Thomas, M. E., & Plomin, R. (2015). Why does parental language input style predict child language development? A twin study of gene–environment correlation. *Journal of communication disorders, 57*, 106-117. doi:10.1016/j.jcomdis.2015.07.004

Devanna, P., Chen, X. S., Ho, J., Gajewski, D., Smith, S. D., Gialluisi, A., . . . Vernes, S. C. (2017). Next-gen sequencing identifies non-coding variation disrupting miRNA-binding sites in neurological disorders. *Mol Psychiatry*. doi:10.1038/mp.2017.30

Devanna, P., Middelbeek, J., & Vernes, S. C. (2014). FOXP2 drives neuronal differentiation by interacting with retinoic acid signaling pathways. *Frontiers in Cellular Neuroscience, 8*(305). doi:10.3389/fncel.2014.00305

Eicher, J. D., Powers, N. R., Miller, L. L., Akshoomoff, N., Amaral, D. G., Bloss, C. S., . . . Genetics, S. (2013). Genome-wide association study of shared components of reading disability and language impairment. *Genes Brain Behav, 12*(8), 792-801. doi:10.1111/gbb.12085

Eising, E., Carrion-Castillo, A., Vino, A., Strand, E. A., Jakielski, K. J., Scerri, T. S., . . . Mazoyer, B. (2018). A set of regulatory genes co-expressed in embryonic human brain is implicated in disrupted speech development. *Molecular Psychiatry*, 1. doi:10.1038/s41380-018-0020-x

Finkel, D., Pedersen, N. L., Plomin, R., & McClearn, G. E. (1998). Longitudinal and cross-sectional twin data on cognitive abilities in adulthood: The Swedish Adoption/Twin Study of Aging. *Developmental psychology, 34*(6), 1400. doi:0012-1649/98/

Fisher, S. E., & Scharff, C. (2009). FOXP2 as a molecular window into speech and language. *Trends in Genetics, 25*(4), 166-177. doi:10.1016/j.tig.2009.03.002

Fuchsberger, C., Flannick, J., Teslovich, T. M., Mahajan, A., Agarwala, V., Gaulton, K. J., . . . McCarthy, D. J. (2016). The genetic architecture of type 2 diabetes. *Nature, 536*(7614), 41. doi:10.1038/nature18642

Gialluisi, A., Newbury, D. F., Wilcutt, E. G., Olson, R. K., DeFries, J. C., Brandler, W. M., . . . Fisher, S. E. (2014). Genome-wide screening for DNA variants associated with reading and language traits. *Genes Brain Behav, 13*(7), 686-701. doi:10.1111/gbb.12158

Hanscombe, K. B., Trzaskowski, M., Haworth, C. M., Davis, O. S., Dale, P. S., & Plomin, R. (2012). Socioeconomic status (SES) and children's intelligence (IQ): In a UK-representative sample SES moderates the environmental, not genetic, effect on IQ. *PLoS One, 7*(2), e30320. doi:10.1371/journal.pone.0030320

Harlaar, N., Meaburn, E. L., Hayiou-Thomas, M. E., Davis, O. S., Docherty, S., Hanscombe, K. B., . . . Dale, P. S. (2014). Genome-wide association study of receptive language ability of 12-year-olds. *Journal of Speech, Language, and Hearing Research, 57*(1), 96-105. doi:10.1044/1092-4388(2013/12-0303

Hayiou-Thomas, M. E., Oliver, B., & Plomin, R. (2005). Genetic influences on specific versus nonspecific language impairment in 4-year-old twins. *Journal of Learning Disabilities, 38*(3), 222-232. doi:10.1177/00222194050380030401

Hayiou-Thomas, M. E., Dale, P. S., & Plomin, R. (2012). The etiology of variation in language skills changes with development: A longitudinal twin study of language

from 2 to 12 years. *Developmental science, 15*(2), 233-249. doi:10.1111/j.1467-7687.2011.01119.x

Hayiou-Thomas, M. E., Harlaar, N., Dale, P. S., & Plomin, R. (2006). Genetic and environmental mediation of the prediction from preschool language and nonverbal ability to 7-year reading. *Journal of Research in Reading, 29*(1), 50-74. doi:10.1111/j.1467-9817.2006.00292.x

Kang, C., Riazuddin, S., Mundorff, J., Krasnewich, D., Friedman, P., Mullikin, J. C., & Drayna, D. (2010). Mutations in the lysosomal enzyme–targeting pathway and persistent stuttering. *New England Journal of Medicine, 362*(8), 677-685. doi:10.1056/NEJMoa0902630

Kraft, S. J., & DeThorne, L. S. (2014). The Brave New World of Epigenetics: Embracing Complexity in the Study of Speech and Language Disorders. *Current Developmental Disorders Reports, 1*(3), 207-214. doi:10.1007/s40474-014-0024-4

Lai, C. S., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., & Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature, 413*(6855), 519-523. doi:10.1038/35097076

Lewis, B. A., & Thompson, L. A. (1992). A study of developmental speech and language disorders in twins. *Journal of Speech, Language, and Hearing Research, 35*(5), 1086-1094.

Loke, Y. J., Hannan, A. J., & Craig, J. M. (2015). The Role of Epigenetic Change in Autism Spectrum Disorders. *Frontiers in Neurology, 6*, 107. doi:10.3389/fneur.2015.00107

Luciano, M., Evans, D. M., Hansell, N. K., Medland, S. E., Montgomery, G. W., Martin, N. G., . . . Bates, T. C. (2013). A genome-wide association study for reading and language abilities in two population cohorts. *Genes Brain Behav, 12*(6), 645-652. doi:10.1111/gbb.12053

MacDermot, K. D., Bonora, E., Sykes, N., Coupe, A. M., Lai, C. S., Vernes, S. C., . . . Fisher, S. E. (2005). Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. *Am J Hum Genet, 76*(6), 1074-1080. doi:10.1086/430841

Moralli, D., Nudel, R., Chan, M. T. M., Green, C. M., Volpi, E. V., Benítez-Burraco, A., . . . García-Bellido, P. (2015). Language impairment in a case of a complex chromosomal rearrangement with a breakpoint downstream of FOXP2. *Molecular cytogenetics, 8*(1), 36. doi:10.1186/s13039-015-0148-1

Nicolia, V., Cavallaro, R. A., López-González, I., Maccarrone, M., Scarpa, S., Ferrer, I., & Fuso, A. (2017). DNA Methylation Profiles of Selected Pro-Inflammatory Cytokines in Alzheimer Disease. *Journal of Neuropathology & Experimental Neurology, 76*(1), 27-31. doi:10.1093/jnen/nlw099

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., . . . Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study. *J Child Psychol Psychiatry, 57*(11), 1247-1257. doi:10.1111/jcpp.12573

Nudel, R., Simpson, N. H., Baird, G., O'Hare, A., Conti-Ramsden, G., Bolton, P. F., . . . Newbury, D. F. (2014). Genome-wide association analyses of child genotype effects and parent-of-origin effects in specific language impairment. *Genes Brain Behav, 13*(4), 418-429. doi:10.1111/gbb.12127

Park, G., Tan, J., Garcia, G., Kang, Y., Salvesen, G., & Zhang, Z. (2016). Regulation of Histone Acetylation by Autophagy in Parkinson Disease. *The Journal of Biological Chemistry, 291*(7), 3531-3540. doi:10.1074/jbc.M115.675488

Pinker, S. (1994). The Language Instinct 1994 New York. *NY Harper Perennial Modern Classics*. doi:10.1037/e412952005-009

Plomin, R., & Deary, I. J. (2015). Genetics and intelligence differences: five special findings. *Molecular Psychiatry, 20*(1), 98. doi:10.1038/mp.2014.105

Prekovic, S., Durdevic, D. F., Csifcsak, G., Sveljo, O., Stojkovic, O., Jankovic, M., . . . Newbury, D. F. (2016). Multidisciplinary investigation links backward-speech trait and working memory through genetic mutation. *Sci Rep, 6*, 20369. doi:10.1038/srep20369

Puglisi, M. L., Hulme, C., Hamilton, L. G., & Snowling, M. J. (2017). The home literacy environment is a correlate, but perhaps not a cause, of variations in children's language and literacy development. *Scientific Studies of Reading, 21*(6), 498-514. doi:10.1080/10888438.2017.1346660

Reilly, S., Wake, M., Ukoumunne, O. C., Bavin, E., Prior, M., Cini, E., . . . Bretherton, L. (2010). Predicting language outcomes at 4 years of age: findings from Early Language in Victoria Study. *Pediatrics, 126*(6), e1530-e1537. doi:10.1542/peds.2010-0254

Reuter, M. S., Riess, A., Moog, U., Briggs, T. A., Chandler, K. E., Rauch, A., . . . Joset, P. (2017). FOXP2 variants in 14 individuals with developmental speech and language disorders broaden the mutational and clinical spectrum. *Journal of medical genetics, 54*(1), 64-72. doi:10.1136/jmedgenet-2016-104094

Reynolds, C. A., & Finkel, D. G. (2015). A meta-analysis of heritability of cognitive aging: minding the "missing heritability" gap. *Neuropsychology review, 25*(1), 97-112. doi:10.1007/s11065-015-9280-2

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, 4*(1), 27. doi:10.1186/1866-1955-4-27

Rice, M. L., Zubrick, S. R., Taylor, C. L., Gayán, J., & Bontempo, D. E. (2014). Late language emergence in 24-month-old twins: Heritable and increased risk for late language emergence in twins. *Journal of Speech, Language, and Hearing Research, 57*(3), 917-928. doi:10.1044/1092-4388(2013/12-0350)

Rice, M. L., Zubrick, S. R., Taylor, C. L., Hoffman, L., & Gayán, J. (2017). Longitudinal Study of Language and Speech of Twins at 4 and 6 Years: Twinning Effects Decrease, Zygosity Effects Disappear, and Heritability Increases. *Journal of Speech, Language, and Hearing Research*, 1-15. doi:10.1044/2017_JSLHR-L-16-0366

Ripke, S., Neale, B. M., Corvin, A., Walters, J. T. R., Farh, K., H., Holmans, P. A., . . . O'Donovan, M. C. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature, 511*(7510), 421-427. doi:10.1038/nature13595

Rodenas-Cuadrado, P., Ho, J., & Vernes, S. C. (2014). Shining a light on CNTNAP2: complex functions to complex disorders. *Eur J Hum Genet, 22*(2), 171-178. doi:10.1038/ejhg.2013.100

Ruben, R. J. (2000). Redefining the survival of the fittest: communication disorders in the 21st century. *Laryngoscope, 110*(2 Pt 1), 241-245. doi:10.1097/00005537-200002010-00010

Scerri, T. S., Morris, A. P., Buckingham, L. L., Newbury, D. F., Miller, L. L., Monaco, A. P., . . . Paracchini, S. (2011). DCDC2, KIAA0319 and CMIP are associated with reading-related traits. *Biol Psychiatry, 70*(3), 237-245. doi:10.1016/j.biopsych.2011.02.005

Smith, S. D. (2011). Approach to epigenetic analysis in language disorders. *Journal of Neurodevelopmental Disorders, 3*(4), 356-364. doi:10.1007/s11689-011-9099-y

Spinath, F. M., Price, T. S., Dale, P. S., & Plomin, R. (2004). The genetic and environmental origins of language disability and ability. *Child Development, 75*(2), 445-454. doi:10.1111/j.1467-8624.2004.00685.x

Spiteri, E., Konopka, G., Coppola, G., Bomar, J., Oldham, M., Ou, J., . . . Geschwind, D. H. (2007). Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. *The American Journal of Human Genetics, 81*(6), 1144-1157. doi:10.1086/522237

St Pourcain, B., Cents, R. A., Whitehouse, A. J., Haworth, C. M., Davis, O. S., O'Reilly, P. F., . . . Davey Smith, G. (2014). Common variation near ROBO2 is associated with expressive vocabulary in infancy. *Nat Commun, 5*, 4831. doi:10.1038/ncomms5831

Stromswold, K. (1998). Genetics of spoken language disorders. *Hum Biol, 70*(2), 297-324.

The 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature, 526*(7571), 68-74. doi:10.1038/nature15393

Tomblin, J. B., O'Brien, M., Shriberg, L. D., Williams, C., Murray, J., Patil, S., . . . Ballard, K. (2009). Language features in a mother and daughter of a chromosome 7;13 translocation involving FOXP2. *J Speech Lang Hear Res, 52*(5), 1157-1174. doi:10.1044/1092-4388(2009/07-0162)

Tosto, M. G., Hayiou-Thomas, M. E., Harlaar, N., Prom-Wormley, E., Dale, P. S., & Plomin, R. (2017). The genetic architecture of oral language, reading fluency, and reading comprehension: A twin study from 7 to 16 years. *Developmental psychology, 53*(6), 1115. doi:10.1037/dev0000297

Turner, S. J., Hildebrand, M. S., Block, S., Damiano, J., Fahey, M., Reilly, S., . . . Morgan, A. T. (2013). Small intragenic deletion in FOXP2 associated with childhood apraxia of speech and dysarthria. *American journal of medical genetics Part A, 161*(9), 2321-2326. doi:10.1002/ajmg.a.36055

van Ijzendoorn, M. H., Belsky, J., & 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. *Transl Psychiatry, 2*, e147. doi:10.1038/tp.2012.73

Vernes, S. C., Newbury, D. F., Abrahams, B. S., Winchester, L., Nicod, J., Groszer, M., . . . Fisher, S. E. (2008). A functional genetic link between distinct developmental language disorders. *N Engl J Med, 359*(22), 2337-2345. doi:10.1056/NEJMoa0802828

Vernes, S. C., Oliver, P. L., Spiteri, E., Lockstone, H. E., Puliyadi, R., Taylor, J. M., . . . Fisher, S. E. (2011). Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. *PLoS Genet, 7*(7), e1002145. doi:10.1371/journal.pgen.1002145

Vernes, S. C., Spiteri, E., Nicod, J., Groszer, M., Taylor, J. M., Davies, K. E., . . . Fisher, S. E. (2007). High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders. *The American Journal of Human Genetics, 81*(6), 1232-1250. doi:10.1086/522238

Virgin, H. W., & Todd, J. A. (2011). Metagenomics and personalized medicine. *Cell, 147*(1), 44-56. doi:10.1016/j.cell.2011.09.009

Vygotskiĭ, L. S. (2012). *Thought and language*: MIT press.

Watkins, K. E., Vargha-Khadem, F., Ashburner, J., Passingham, R. E., Connelly, A., Friston, K. J., . . . Gadian, D. G. (2002). MRI analysis of an inherited speech and language disorder: structural brain abnormalities. *Brain, 125*(3), 465-478. doi:10.1093/brain/awf057

Worthey, E. A., Raca, G., Laffin, J. J., Wilk, B. M., Harris, J. M., Jakielski, K. J., . . . Shriberg, L. D. (2013). Whole-exome sequencing supports genetic heterogeneity in childhood apraxia of speech. *J Neurodev Disord, 5*(1), 29. doi:10.1186/1866-1955-5-29