The Genomic Landscape of Language: Insights into Evolution

Hayley S. Mountford and Dianne F. Newbury

Department of Biological and Medical Sciences, Faculty of Health and Life Sciences, Oxford Brookes University.

Abstract

Studies of severe, monogenic forms of language disorders have revealed important insights into the mechanisms that underpin language development and evolution. It is clear that monogenic mutations in genes such as *FOXP2* and *CNTNAP2* only account for a small proportion of language disorders seen in children, and the genetic basis of language in modern humans is highly complex and poorly understood. In this review, we examine why we understand so little of the genetic landscape of language disorders, and how the genetic background of an individual greatly affects the way in which a genetic change is expressed. We discuss how the underlying genetics of language disorders has informed our understanding of language evolution, and how recent advances may obtain a clearer picture of language capacity in ancient hominins.

Introduction

The ease with which most children acquire their native language has lead researchers to propose that language acquisition is innate, [1], and suggest that this reflects a genetically determined language-specific module [2]. Others argue that it simply reflects higher order processing in humans and is facilitated by their existing cognitive skills [3]. Major questions remain as to the evolutionary and genetic mechanisms that underpin these proposed models; did language evolution rely upon a small number of 'big-hit' mutations which rapidly changed cognition, or through a series of small-step changes where many variants were accumulated slowly over thousands of years? Did ancient hominins have the cognitive ability to use some form of language?

The study of genetic variation that underpins language ability in modern humans can provide insights into how higher language function evolved in our ancient ancestors. The application of next generation sequencing technology means that we are now able to generate a near-complete picture of genetic variation with relative ease. The discovery of genetic variants associated with language disorders results in the identification of the genes and molecular pathways necessary for the successful acquisition of language. Genetic studies of modern humans, therefore, have direct relevance to the study of how language evolved in our ancestors.

Discussion of the evolution of language in fields outside of genetics, still tend to consider 'a gene for language' as the principle driver of language evolution. While the consideration of single variants and genes has provided important insights, the field of human genetics has moved on. Here, we argue that in order to understand language evolution, we first need to consider the full genetic landscape in modern humans, then use this to inform our understanding of the forces that shaped language evolution in ancient hominins.

Language Disorders

When considering which genetic pathways contribute to language, researchers often choose to study the extremes of language ability - most often when a person's ability to speak is severely impaired. So

far, the greatest insights into the molecular biology of language have come from studying the genetics of families and individuals with persistent language disorders.

A recent study found that over 7% of British children (n=12,000, Surrey) at school entry had impaired language, either as part of a complex developmental disorder such as autism spectrum disorder (ASD), developmental delay or intellectual disability, or as a primary language disorder with no other explanatory features [4]. Previous smaller English speaking studies concluded similar rates [5; 6]. In real terms, this means that a staggering three children in every class have a language disorder [4]. Age appropriate language acquisition is so important to a child's development that receptive language ability at age 3 is a predictor of an individuals' future economic burden [7]. Despite educational intervention, over half of children with language disorders have lasting difficulties with language throughout their childhood [8]. This means that a child who struggles to understand or produce language, even from an early age, has an increased risk of behavioural disorders, unemployment and mental health issues later in life [9]. This importance is clearly demonstrated in a recent systematic review which found that there was a consistent strong association between young offenders and language disorders [10].From a genetics point-of-view, it is of particular interest when language disorder occurs in isolation (so-called primary language disorder), with no other features such as autism spectrum disorder or developmental delay that may confound difficulties with language. Primary language disorders may represent domain independent deficits and therefore provide an excellent opportunity to study the genetics that underpin speech.

Two such primary language disorders are childhood apraxia of speech (previously called developmental verbal dyspraxia) (CAS, OMIM #602081) and developmental language disorder (DLD) (also known as specific language impairment) (SLI, OMIM %606711, %606712, %607134, %612514). Although both conditions are primary language disorders, they are proposed to arise from different obstacles in language production pathways. CAS is primarily a motoric difficulty in which the brain cannot coordinate the fine muscles controlling the tongue, lips and mouth that are required to produce speech [6]. DLDs are a persistent difficulty with more generalised aspects of speech and

language, in the absence of any other explanatory medical condition such as hearing difficulties or developmental delay [11]. The diagnostic guidelines for DLDs are therefore less stringent than CAS and, accordingly, DLDs are an extremely common childhood developmental issue that can persist throughout the child's life. In this review we will focus on the primary language disorders DLD and CAS.

There is little doubt as to the impact of language disorders on children, but despite the frequency and impact on society, we still understand little of the underlying neurobiology. It is clear that the risk of speech and language disorder is increased if a parent or sibling has a speech disorder [12]. Many studies indicate that language ability is highly heritable, and that that genetic factors play a role in this familiality [12-14]. The identification of genetic variants or risk factors for DLDs may explain why some children struggle with language acquisition. It may also help explain why language ability is so often affected in related disorders such as ASD, developmental dyslexia, intellectual learning disability or attention deficit hyperactivity disorder (ADHD) and tease apart the phenotypic overlaps between these highly related disorders. Assuming that language impairments are at one end of a continuum of language ability, genetic studies are providing a better understanding of the molecular pathways that are important in language acquisition.

Genes Involved in Disorders of Language Development

When a language disorder recurs within multiple generations of a family, we often assume a strong genetic contribution. Such families have therefore traditionally been the obvious place to start when studying genetic inheritance. The principal insights into the genetics of DLDs have come from such family studies, and several genes have been identified using genetic linkage and candidate gene sequencing in related family members (Table 1). These genes were often identified from single families or a number of related individuals, using genetic linkage to look for regions of the genome shared by language impaired family members, or by testing for genetic association between large numbers of unrelated individuals with a similar phenotype (Table 1). Genetic linkage and association approaches have traditionally been the mainstay of neurodevelopment genetics, with much success.

The most successful study in this field, to date, has been the identification of an arginine to histidine mutation at amino acid position 553 (denoted as p.R553H) in the *FOXP2* gene, identified in a large, multigenerational family known as the KE family. Family members who carry this mutation have the CAS phenotype [15]. In genetic terminology, the p.R553H change is a dominant, fully penetrant mutation – one mutated copy of the gene is enough to result in a particular disorder. Fully penetrant cases are rare and presumably differ from more 'typical' cases of DLD, where one genetic change cannot be directly correlated with their disorder. While this remains the most studied and best characterised gene implicated in speech, mutations in *FOXP2* only account for about 2% of CAS cases [16], and as such, causative mutations in *FOXP2* are still considered a rare cause of language disorders.

FOXP2, dubbed a 'molecular window' into speech and language development, has been a leap-pad for the identification of other genes and mechanisms involved in language (for example, *CNTNAP2* [17], as described below). The discovery of *FOXP2* was hailed by the media as the 'speech gene' – suggesting that this single protein is responsible for language development in humans. This headline tag is an overly simplistic interpretation, which has endured in fields outside of genetics and language biology. More recently, investigation into the molecular function of FOXP2 has slowly built a more detailed picture of its role in language development [18-20]. The literature is clear – *FOXP2* is not the sole explanatory factor for presence of language.

Table 1 – Major genes implicated in language disorders, and associated overlapping phenotypes. The table shows genes from association or linkage of language disorders, and does not include a thorough review of other phenotypes (dyslexia, ASD etc). * indicates gene has been reported as monogenic.

Gene	Associated Disorder(s)	Key References
ABCC13	Language disorder	[21]

ARHGEF39	Language disorder	[22]
ATP2C2	Language disorder (short term memory)	[23; 24]
BCL11A	Language disorder (specifically CAS) with expressive language and mild intellectual delay	[25]
CMIP	Language disorder (short term memory) Language disorder and dyslexia Dyslexia	[23] [26] [27]
CNTNAP2	Language disorder Autism	[17] [28; 29]
DCDC2	Dyslexia Language disorder and dyslexia	[30; 31] [27; 32; 33]
ERC1	Language disorder (CAS)	[34; 35]
FLNC	Language disorder and reading difficulties	[36]
FOXP1*	Language disorder and intellectual delay	[37-41]
FOXP2 *	Language disorder (specifically CAS)	[15; 42-46]
GRIN2A	Focal epilepsy with speech disorder, with or without mental retardation	[35; 47-49]
KIAA0319	Dyslexia Language disorder	[27; 50] [26]
NDST4	Language disorder	[51]
NFXL1	Language disorder	[52]
NOP9	Language disorder	[53]
RBFOX2	Language disorder and reading difficulties	[36]
ROBO1	Dyslexia Language disorder and dyslexia	[54] [55]
ROBO2	Language disorder	[56]
SETBP1	Language disorder	[57-59]

SRPX2	Language disorder, rolandic seizures and intellectual delay	**35; 60**
TM4SF20 *	Language disorder	**61**

There are very few instances of monogenic inheritance, where the absence of a protein leads directly to language disorder. In Table 1, only *FOXP2, FOXP1* and *TM4SF20* have been described as monogenic drivers of language disorders. The remainder of the identified genes instead confer risk of language disorder through genetic variations that subtly alter the way in which genes and proteins work. The majority of genes have been implicated in language disorders through association with language-related phenotypes obtained from cohort studies. In contrast to *FOXP2,* where a mutation explains the observed language difficulties (monogenic model), these genes tend to play a role within a complex genetic model. Carrying a risk variant within these genes confers a 'susceptibility' to develop language disorder, however this remains difficult to quantify and is poorly understood. Nonetheless, the study of cases and their families has provided an important window into the underlying mechanisms of language disorders. At present, *FOXP2* and *FOXP1* remain the best characterised of the genes implicated in language disorders. Clinical diagnosis of the underlying molecular cause of a language disorder is not usually possible, unless the causative mutation is within *FOXP2, FOXP1* or *TM4SF20*. Mutations in these genes are rare, and therefore the majority of language disorder cases are unlikely to have an underlying molecular cause identified.

Large scale genome sequencing projects such as 1000 Genomes [62] and ExAC [63] have created a major shift in how we perceive human genetic variation and its contribution to disease. We have understood for decades that monogenic disorders usually involve rare mutations which impact upon the function of the protein. Such mutations usually lead to non-functional proteins which manifest in a disease phenotype. Access to large numbers of control genomes through 1000 Genomes and ExAC has enabled us to more accurately identify and assess genetic risk factors, which tend to be more common in the population, but may confer a modest risk of developing a phenotype.

These databases also provide unprecedented power to inform our understanding of gene function in modern humans, and by proxy, our ancestors. It is well established that Neanderthals and Denisovans shared the 'humanised' version of *FOXP2*, which differs from ancestral *FOXP2* at two positions; chromosome 7, base-pair 114,282,597 (denoted as chr7:114,282,597) resulting in an arginine rather than the ancestral threonine at position 303 (denoted as p.N303) and chromosome 7, base-pair 114,282,663 (denoted as chr7:114,282,663) resulting in a serine at amino acid position 325 rather than the ancestral arginine (denoted as p.S325) (variant 1, hg19) [64]. This important finding gave rise to the idea that Neanderthals may have had a sophisticated level of cognitive processing to support some form of language [64].

Interestingly, the 'humanised' *FOXP2* amino acid at position 325 is somewhat called into question by the presence of two apparently healthy controls in the ExAC database. These two individuals carry one copy (heterozygous) of a T>G change at neighbouring position (chr7:114,282,664), essentially reverting the amino acid sequence to the ancestral form, resulting in a serine to arginine change (p.S325N). This change is extremely rare (allele frequency=0.00001648) and only seen in 2 of more than 60,000 individuals, but it poses the question - did these apparently healthy individuals have language difficulties? Although ExAC participants were not specifically screened for cognitive function or language ability, it is unlikely that they had an overt phenotype as this would have excluded them from the study. This presents an interesting line of thought, that if these two amino acids are the hominin form of *FOXP2*, then there are at least two functioning humans out there who do not have a fully 'humanised' version of *FOXP2*. The presence of a non-human *FOXP2* amino acid change in these two healthy individuals shows the power of these databases to identify extremely rare occurrences of a variant carried in less than 0.0016% of the population. It provides a more accurate snapshot of human variation with which we can more effectively predict which variants are likely to be important.

Even in monogenic disorders, when it is clear that the trait is directly caused by a dominant mutation, we still observe a high degree of variability between individuals (incomplete penetrance). Such

phenotypic variability is even present within the KE family who have a 'fully' penetrant dominant *FOXP2* mutation with a clear-cut phenotype [15; 65]. It is widely reported that some individuals of the KE family present with non-verbal difficulties. The performance IQ scores of five affected KE family members are varied - on male affected (age 10) scored 112 compared to a second 10 year old affected male who scores 66. These individuals carry the p.R553H mutation which explains their CAS phenotype, but the differences in performance IQ are likely due to genetic modifiers, and not directly related to *FOXP2*. , For the majority of language disorder loci discovered to date, it is likely that they explain only part of the risk and the modifier, and additional variants have yet to be identified. We are only just beginning to understand the actions of modifiers and risk factors, but this concept underlies a shift from the traditional genetic model, in which phenotypes are truly dominant or recessive. Instead we now understand the importance of considering all variation on a genetic background.

Complex Inheritance and Genetic Risk

The power of familial studies are a proven method to identify contributory genes, but increasingly molecular genetics is focussing on the role of modifiers and risk factors in DLDs. The majority of genes listed in Table 1 that have been associated with language disorders fall into this category. An example is an asparagine to lysine change at amino acid position 150 (denoted as p.N150K) in the *NFXL1* gene. This variant (rs144169475), identified by sequencing five affected Islanders, was found to be associated with language impairment on Robinson Crusoe Island, an isolated Chilean population with an exceptionally high rate of language disorders [52]. This variant likely forms a key part of a complex inheritance model where a single variant only explains part of the DLD risk. The variant is seen in 4.1% in South American control genomes, and is therefore considered common in Latin America, suggesting that it may confer susceptibility to DLD when inherited in combination with other variants that are yet to be identified.

The study of complex genetic factors is primarily performed using large numbers of unrelated cases specifically selected to have a high degree of phenotypic similarity. Large scale genome-wide association studies (GWAS) with several thousands of participants may be able to successfully

identify common risk variants involved in DLDs, however a large scale study of this nature has not yet been attempted. A recent GWAS into the genetic basis of schizophrenia successfully identified more than 100 associated loci using 37,000 schizophrenia patients and 113,000 controls [66]. The application of these methods in clinical traits such as schizophrenia, have shown that enormous sample sizes are required to enable the consistent replication of associated loci.

A major limiting factor in performing a large scale GWAS for language disorders remains the systematic phenotyping of enough participants to gain the statistical power required to detect contributory variants. This challenge is common to most large complex genetics studies, but is particularly pronounced for the field of language disorders where there is little consensus on what constitutes a speech and language disorder, or how it should be diagnosed and classified. A recent report by the CATALISE consortium aims to do exactly that [11]. Even the terminology used to describe language disorders and DLDs required standardisation across disciplines, and although these are the current approved terms, they are taking time to become standard in research and education. Establishing consistent terminology is the keystone to developing standardised diagnostic criteria. Once these definitions are consistent within and across disciplines, then a large scale study could be successfully developed. It would likely lead to the identification of a novel pathways and gene networks involved in language production.

Table 1 reveals the striking number of genes implicated in DLDs which are also implicated in other, closely related neurodevelopmental disorders. Vernes and colleagues identified an association between variants in the contactin-associated protein like 2 gene *CNTNAP2* and DLDs through its interaction with the transcription factor *FOXP2* [17]. Variants in *CNTNAP2* are also associated with ASD [28; 67], cortical dysplasia focal epilepsy syndrome (OMIM #610042) [68], and Pitt-Hopkins-like syndrome (OMIM #610042) [69]. Another example of genes implicated in language overlapping with related disorders is the axon guidance receptor protein ROBO1. It was first implicated as a candidate gene for dyslexia in a patient with a translocation involving the *ROBO1* region [54], and was subsequently found to be associated with short term memory of words, a key feature of DLD [55]. Other

examples of genes involved in language disorders that overlap with a dyslexia phenotype, include *DCDC2, KIAA0319* and *CMIP* [27; 30] .

This observation suggests the documented phenotypic overlap between developmental disorders like DLD, ASD, and dyslexia may be driven by shared genetic aetiology. We should note, however that the level of shared aetiology is hard to objectively ascertain without genome-wide data. Technical and financial limitations mean that many studies of DLDs to date are limited to candidate genes, leading to substantial ascertainment bias.

The factors that determine how a given genetic variant manifests to become one phenotype over another is not fully understood, but they are likely to involve interactions between genetic variants. This emphasises the need to consider the genetic background of an individual within any candidate gene analyses. These multiple layers of complexity partly explain why genetic studies have so far struggled to elucidate the genetic basis of many neurodevelopmental disorders.

Limitations of current genomic studies

There are a number of reasons why we do not have a better picture of the genetics of speech and language disorders. As discussed above, the majority of studies have used relatively low resolution mapping methods within small sample sizes with inconsistent characterisation between studies. Recent advances in DNA sequencing technology allow us to generate a more complete picture of genetic variation across the entire genome (whole genome sequencing) or across all known genes in the genome (whole exome sequencing). Whilst such technologies afford better resolution and, to some extent, offset these problems, the identification of risk variants, which only have a small effect size, remain difficult.

The average human genome contains between 4 and 5 million variants that differ from published reference sequences. Only about 1% of the human genome actually encodes genes, and these gene encoding regions will contain about 150 coding mutations which result in the loss-of-function of the

protein. They will also contain around 10,000 ‘silent’ mutations that fall within genes but do not alter the amino sequence. Each person’s genome will contain about 120 novel coding variants which have not previously been reported [62]. The vast majority of variation we see in the human genome does not directly change the protein, and is non-coding. Once we consider that these non-coding changes may have a function affecting gene expression (how much of each protein is made), the list of potential variants can be vast, and extremely challenging.

Exome sequencing studies investigate just the coding regions of the genes of individuals affected by DLD [35; 59] or CAS [16]. These preliminary, small-scale investigations confirm the complexity of the underlying genetics in the majority of cases and reinforce the need for larger-scale screening studies.

Even though they lie outside of gene sequences, non-coding variants can change gene functions, for example by increasing or decreasing expression. It is highly likely that these non-coding variants will be involved in neurodevelopmental disorders. These variants represent a far greater challenge than coding variants. They are often not captured by whole exome sequencing meaning that we may simply be missing important mutations. Whole genome sequencing is becoming more commonly used, but cost is often prohibitively expensive. Even when these variants are captured, their categorisation is difficult. A recent study demonstrated a role for variations within non-coding regulatory regions in DLD and other neurodevelopmental conditions underlining the importance of this route of investigation [22]. The use of whole genome sequencing produces vastly more data, and analysis can be more computationally expensive, and requires a much greater level of analytical expertise. Since the effects of these variants are often indirect, their characterisation usually involves complex functional validation steps that are challenging to complete for a high number of variants.

Genetic studies tend to be performed on European or American cohorts. Findings in these participants may not be relevant in other populations as some variants can be more or less common in a different population, and different groups may need their own specific studies to gain a better global understanding. For example, the *NFXL1* variant found to increase risk of DLD on Robinson Crusoe

Island was found in 4.1% of Latin Americans, but 0% of Europeans [52]. Similarly, investigations of an isolated Russian population have yielded novel loci in relation to DLD [59]. The availability of 1000 Genomes data has improved power to detect variants that differ in allele frequencies between populations however, these are still limited to relatively small numbers of individuals from a restricted set of countries.

Another degree of complexity is added by tissue specificity; while present in the genomic DNA of every cell, some mutations may only have a detectable effect in a specific tissue, at a particular time in development. The function of a gene can vary between cell types and conditions, and many genes have multiple, and often surprisingly different, functions. *FOXP2* is highly expressed in the brain, but is also highly expressed in the lungs and many other tissue types all of which will carry the mutation at a DNA level [70]. The brain appears particularly sensitive to this particularly change and, as far as we are aware, the lungs of the KE family are unaffected [15]. It is therefore important to remember that although genomic technologies can give us a window into what is happening in a particular individual, it is far more challenging to predict the cellular context in which it will become important.

Paleogenetics and Language

In evolutionary terms, the window to understand genetic effects on cognitive function and language ability in hominins is even narrower than in modern humans, and must be interpreted with extreme caution. The humanised version of *FOXP2* is thought to have become fixed in the population around 500 KYA, prior to the last shared common ancestor (370-450 KYA) [71] and the presence of this version in Neanderthals supports the notion of cognitive function sophisticated enough to support language. More recently, a regulatory region of *FOXP2* was identified exclusively in modern humans at a binding site of the transcription factor *POUF3F2* which is absent in Neanderthals [72]. This suggests that differences in gene regulation and expression may be involved in cognitive function, and that species differences are due to far more than just two variants in a single gene. We must be cautious when interpreting such information as it is extremely unlikely that these *FOXP2* changes are

solely responsible for the presence (or absence) of language function, and any observations, modern or otherwise, should consider the entire genetic background. [73]

To further complicate the underlying assumptions in evolutionary studies, the small numbers of Neanderthals sequenced heavily biases the study findings. The difficulties in obtaining ancient DNA of suitable quality for sequencing, means that sequenced individuals are not representative of time periods or geographical locations, and are from a small number of sites where preservation conditions were optimal. As discussed above, genome sequence studies clearly illustrate that small numbers of individuals from one or two geographical locations do not represent the entire population. This is the modern genetic equivalent of sequencing one family and assuming that everyone else is the same – this is not genetically plausible. There is not enough available population data to be able to accurately predict genetic affects, particularly with respect to complex cognitive processes like language function.

Paleogenetics researchers are slowly building a broader and more accurate picture of ancient hominin genetics through sequencing larger numbers from a range of geographical locations. A larger sample size will greatly improve the statistical significance of findings, and increase confidence in their implication for language and higher cognitive function. Genes that are implicated in language disorders in modern can inform investigation of language in ancient hominins, and there have been several efforts to investigate the impact of language associated genes more broadly [73]. Through the expansion of genetic technologies and a greater understanding of their application and limitation, we will continue to build a more accurate picture of both modern and ancient language cognition slowly, piece by piece, applying the scientific rigour and multiple lines of evidence of molecular biology.

Discussion

The study of language disorders has been fruitful in implicating genes, and subsequent molecular pathways that are involved in the mechanisms of language. While there have been many exciting discoveries spanning the past two decades, there remains much more to understand. We still do not

fully understand the underlying causes of DLDs, and what makes some children are more susceptible. Family studies can still provide novel insights into the underlying mechanisms of DLDs. There is strong potential for using a familial shared genetics-based approach, particularly when combined with recent advances in sequencing technologies that can investigate more of the genome than ever before.. We increasingly recognise that genetic risk plays a key role in language disorders and many current approaches are investigating a genetic background of susceptibility. To be statistically sound, these studies require much larger sample sizes and more consistently phenotyped datasets to generate sufficient statistical power. The reality is that DLDs are likely to involve some high impact rare mutations, genetic rearrangements and common sequence variations, all of which create a background of susceptibility. Family based and association studies are still uncovering some unlikely pathways which play a role in language disorders, and it is clear that it will not be a simple story.

The idea that a single gene has a distinct role or confers a single trait is an outdated concept. Similarly, the idea that a gene will have a single role in the cell has been dispelled. We understand that non-coding variants can play a crucial role in gene regulation, and are highly likely to have an important function in DLDs, and other neurodevelopmental disorders. The genetic background and regulation of gene expression and function is dynamic, and depends greatly on individual cell types. While this is still poorly understood, methods for detecting and experimentally validating such context dependent states are in development. The function of a gene in a particular cellular circumstance can, and will be validated by molecular biology in model animal or cellular systems. Genetic control is no longer beyond our testing capability, and we have a range of technologies to characterise gene function and expression across different cell types and under different conditions.

Environmental factors clearly play a role in language development, and poor life circumstance may impact the DLD phenotype. Nature versus nurture is a falsely binary concept, and the underlying genetics plays a key role within an environmental (nurture) context.

The theory that the presence of 'humanised' *FOXP2* gene in Neanderthals drove language ability is naive and overly simplistic. *FOXP2* clearly plays an important role in speech evolution and production, however, we must be cautious to avoid making over-inflated statements about language in Neanderthals based on a single gene [19]. We are only just beginning to unravel the highly complex developmental processes that underlie speech in modern humans, and should be extremely cautious in extrapolating any findings into hominins. The identification of risk factors for DLDs in modern humans will inform our understanding of capacity for language in ancient hominins. We may be able to build a far clearer picture of how language evolved once we increase our understanding of the neuromolecular pathways involved language development in modern humans.

References

1. Chomsky, N. (1998). On the nature, use and acquisition of language. J Toribio & A Cíark (éds), Language and meaning in cognitíoe science: cognitioe issues and semantic theory, 1-20.

2. Pinker, S. (1994). The Language Instinct 1994 New York. NY Harper Perennial Modern Classics http://dx doi org/101037/e412952005-009.

3. Locke, J. (1836). An essay concerning human understanding.(T. Tegg and Son).

4. Norbury, C.F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., and 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, 1247-1257.

5. Tomblin, J.B., Records, N.L., Buckwalter, P., Zhang, X., Smith, E., and O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. J Speech Lang Hear Res 40, 1245-1260.

6. Shriberg, L.D., Tomblin, J.B., and McSweeny, J.L. (1999). Prevalence of speech delay in 6-year-old children and comorbidity with language impairment. J Speech Lang Hear Res 42, 1461-1481.

7. Caspi, A., Houts, R.M., Belsky, D.W., Harrington, H., Hogan, S., Ramrakha, S., Poulton, R., and Moffitt, T.E. (2016). Childhood forecasting of a small segment of the population with large economic burden. Nature Human Behaviour 1, 0005.

8. Hulme, C., and Snowling, M.J. (2009). Developmental disorders of language learning and cognition.(John Wiley & Sons).

9. Conti-Ramsden, G., and Botting, N. (2008). Emotional health in adolescents with and without a history of specific language impairment (SLI). J Child Psychol Psychiatry 49, 516-525.

10. Anderson, S.A., Hawes, D.J., and Snow, P.C. (2016). Language impairments among youth offenders: A systematic review. Children and Youth Services Review 65, 195-203.

11. Bishop, D.V.M., Snowling, M.J., Thompson, P.A., Greenhalgh, T., and and the, C.-c. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. Journal of Child Psychology and Psychiatry, n/a-n/a.

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

13. Bishop, D.V., Adams, C.V., and 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, 158-169.

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

15. Lai, C.S., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F., and Monaco, A.P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. Nature 413, 519-523.

16. Worthey, E.A., Raca, G., Laffin, J.J., Wilk, B.M., Harris, J.M., Jakielski, K.J., Dimmock, D.P., Strand, E.A., and Shriberg, L.D. (2013). Whole-exome sequencing supports genetic heterogeneity in childhood apraxia of speech. J Neurodev Disord 5, 29.

17. 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., et al. (2008). A functional genetic link between distinct developmental language disorders. N Engl J Med 359, 2337-2345.

18. Dediu, D., and Christiansen, M.H. (2016). Language Evolution: Constraints and Opportunities From Modern Genetics. Top Cogn Sci 8, 361-370.

19. Fitch, W.T. (2017). Empirical approaches to the study of language evolution. Psychon Bull Rev 24, 3-33.

20. Fisher, S.E. (2017). Evolution of language: Lessons from the genome. Psychon Bull Rev 24, 34-40.

21. 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 Behav 12, 645-652.

22. 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. Mol Psychiatry.

23. Newbury, D.F., Winchester, L., Addis, L., Paracchini, S., Buckingham, L.L., Clark, A., Cohen, W., Cowie, H., Dworzynski, K., Everitt, A., et al. (2009). CMIP and ATP2C2 modulate phonological short-term memory in language impairment. Am J Hum Genet 85, 264-272.

24. Smith, A.W., Holden, K.R., Dwivedi, A., Dupont, B.R., and Lyons, M.J. (2015). Deletion of 16q24. 1 supports a role for the ATP2C2 gene in specific language impairment. Journal of child neurology 30, 517-521.

25. Peter, B., Matsushita, M., Oda, K., and Raskind, W. (2014). De novo microdeletion of BCL11A is associated with severe speech sound disorder. Am J Med Genet A 164A, 2091-2096.

26. Newbury, D.F., Paracchini, S., Scerri, T.S., Winchester, L., Addis, L., Richardson, A.J., Walter, J., Stein, J.F., Talcott, J.B., and Monaco, A.P. (2011). Investigation of dyslexia and SLI risk variants in reading- and language-impaired subjects. Behav Genet 41, 90-104.

27. 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. Biol Psychiatry 70, 237-245.

28. 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., et al. (2008). A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism. Am J Hum Genet 82, 160-164.

29. Bakkaloglu, B., O'Roak, B.J., Louvi, A., Gupta, A.R., Abelson, J.F., Morgan, T.M., Chawarska, K., Klin, A., Ercan-Sencicek, A.G., and Stillman, A.A. (2008). Molecular cytogenetic

analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. The American Journal of Human Genetics 82, 165-173.

30. Schumacher, J., Anthoni, H., Dahdouh, F., Konig, I.R., Hillmer, A.M., Kluck, N., Manthey, M., Plume, E., Warnke, A., Remschmidt, H., et al. (2006). Strong genetic evidence of DCDC2 as a susceptibility gene for dyslexia. Am J Hum Genet 78, 52-62.

31. Marino, C., Meng, H., Mascheretti, S., Rusconi, M., Cope, N., Giorda, R., Molteni, M., and Gruen, J.R. (2012). DCDC2 genetic variants and susceptibility to developmental dyslexia. Psychiatric genetics 22, 25.

32. Marino, C., Mascheretti, S., Riva, V., Cattaneo, F., Rigoletto, C., Rusconi, M., Gruen, J.R., Giorda, R., Lazazzera, C., and Molteni, M. (2011). Pleiotropic effects of DCDC2 and DYX1C1 genes on language and mathematics traits in nuclear families of developmental dyslexia. Behavior genetics 41, 67-76.

33. Powers, N.R., Eicher, J.D., Butter, F., Kong, Y., Miller, L.L., Ring, S.M., Mann, M., and Gruen, J.R. (2013). Alleles of a polymorphic ETV6 binding site in DCDC2 confer risk of reading and language impairment. Am J Hum Genet 93, 19-28.

34. Thevenon, J., Callier, P., Andrieux, J., Delobel, B., David, A., Sukno, S., Minot, D., Anne, L.M., Marle, N., and Sanlaville, D. (2013). 12p13. 33 microdeletion including ELKS/ERC1, a new locus associated with childhood apraxia of speech. European Journal of Human Genetics 21, 82.

35. 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. Sci Rep 7, 46105.

36. Gialluisi, A., Newbury, D.F., Wilcutt, E.G., Olson, R.K., DeFries, J.C., Brandler, W.M., Pennington, B.F., Smith, S.D., Scerri, T.S., Simpson, N.H., et al. (2014). Genome-wide screening for DNA variants associated with reading and language traits. Genes Brain Behav 13, 686-701.

37. Horn, D., Kapeller, J., Rivera-Brugues, N., Moog, U., Lorenz-Depiereux, B., Eck, S., Hempel, M., Wagenstaller, J., Gawthrope, A., Monaco, A.P., et al. (2010). Identification of FOXP1 deletions in three unrelated patients with mental retardation and significant speech and language deficits. Hum Mutat 31, E1851-1860.

38. Hamdan, F.F., Daoud, H., Rochefort, D., Piton, A., Gauthier, J., Langlois, M., Foomani, G., Dobrzeniecka, S., Krebs, M.-O., and Joober, R. (2010). De novo mutations in FOXP1 in cases with intellectual disability, autism, and language impairment. The American Journal of Human Genetics 87, 671-678.

39. Le Fevre, A.K., Taylor, S., Malek, N.H., Horn, D., Carr, C.W., Abdul-Rahman, O.A., O'donnell, S., Burgess, T., Shaw, M., and Gecz, J. (2013). FOXP1 mutations cause intellectual disability and a recognizable phenotype. American journal of medical genetics Part A 161, 3166-3175.

40. Srivastava, S., Cohen, J.S., Vernon, H., Barañano, K., McClellan, R., Jamal, L., Naidu, S., and Fatemi, A. (2014). Clinical whole exome sequencing in child neurology practice. Annals of neurology 76, 473-483.

41. Sollis, E., Graham, S.A., Vino, A., Froehlich, H., Vreeburg, M., Dimitropoulou, D., Gilissen, C., Pfundt, R., Rappold, G.A., and Brunner, H.G. (2015). Identification and functional characterization of de novo FOXP1 variants provides novel insights into the etiology of neurodevelopmental disorder. Human molecular genetics 25, 546-557.

42. 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., et al. (2005). Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. Am J Hum Genet 76, 1074-1080.

43. 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 52, 1157-1174.

44. 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 161, 2321-2326.

45. Moralli, D., Nudel, R., Chan, M.T., 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 8, 36.

46. Reuter, M.S., Riess, A., Moog, U., Briggs, T.A., Chandler, K.E., Rauch, A., Stampfer, M., Steindl, K., Gläser, D., and 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, 64-72.

47. Endele, S., Rosenberger, G., Geider, K., Popp, B., Tamer, C., Stefanova, I., Milh, M., Kortüm, F., Fritsch, A., and Pientka, F.K. (2010). Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. Nature genetics 42, 1021-1026.

48. De Ligt, J., Willemsen, M.H., Van Bon, B.W., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., De Vries, P., and Gilissen, C. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. New England Journal of Medicine 367, 1921-1929.

49. Carvill, G.L., Regan, B.M., Yendle, S.C., O'Roak, B.J., Lozovaya, N., Bruneau, N., Burnashev, N., Khan, A., Cook, J., and Geraghty, E. (2013). GRIN2A mutations cause epilepsy-aphasia spectrum disorders. Nature genetics 45, 1073-1076.

50. Kirsten, H., Wilcke, A., Ligges, C., Boltze, J., and Ahnert, P. (2012). Association study of a functional genetic variant in KIAA0319 in German dyslexics. Psychiatric genetics 22, 216-217.

51. Eicher, J.D., Powers, N.R., Miller, L.L., Akshoomoff, N., Amaral, D.G., Bloss, C.S., Libiger, O., Schork, N.J., Darst, B.F., Casey, B.J., et al. (2013). Genome-wide association study of shared components of reading disability and language impairment. Genes Brain Behav 12, 792-801.

52. Villanueva, P., Nudel, R., Hoischen, A., Fernandez, M.A., Simpson, N.H., Gilissen, C., Reader, R.H., Jara, L., Echeverry, M.M., Francks, C., et al. (2015). Exome sequencing in an admixed isolated population indicates NFXL1 variants confer a risk for specific language impairment. PLoS Genet 11, e1004925.

53. Nudel, R., Simpson, N.H., Baird, G., O'Hare, A., Conti-Ramsden, G., Bolton, P.F., Hennessy, E.R., Ring, S.M., Davey Smith, G., Francks, C., et al. (2014). Genome-wide association analyses of child genotype effects and parent-of-origin effects in specific language impairment. Genes Brain Behav 13, 418-429.

54. Hannula-Jouppi, K., Kaminen-Ahola, N., Taipale, M., Eklund, R., Nopola-Hemmi, J., Kaariainen, H., and Kere, J. (2005). The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. PLoS Genet 1, e50.

55. Bates, T.C., Luciano, M., Medland, S.E., Montgomery, G.W., Wright, M.J., and Martin, N.G. (2011). Genetic variance in a component of the language acquisition device: ROBO1 polymorphisms associated with phonological buffer deficits. Behav Genet 41, 50-57.

56. St Pourcain, B., Cents, R.A., Whitehouse, A.J., Haworth, C.M., Davis, O.S., O'Reilly, P.F., Roulstone, S., Wren, Y., Ang, Q.W., Velders, F.P., et al. (2014). Common variation near ROBO2 is associated with expressive vocabulary in infancy. Nat Commun 5, 4831.

57. Filges, I., Shimojima, K., Okamoto, N., Röthlisberger, B., Weber, P., Huber, A.R., Nishizawa, T., Datta, A.N., Miny, P., and Yamamoto, T. (2010). Reduced expression by SETBP1 haploinsufficiency causes developmental and expressive language delay indicating a phenotype distinct from Schinzel–Giedion syndrome. Journal of medical genetics, jmg. 2010.084582.

58. Marseglia, G., Scordo, M.R., Pescucci, C., Nannetti, G., Biagini, E., Scandurra, V., Gerundino, F., Magi, A., Benelli, M., and Torricelli, F. (2012). 372 kb Microdeletion in 18q12. 3 causing SETBP1 haploinsufficiency associated with mild mental retardation and expressive speech impairment. European Journal of Medical Genetics 55, 216-221.

59. 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 137.

60. Roll, P., Rudolf, G., Pereira, S., Royer, B., Scheffer, I.E., Massacrier, A., Valenti, M.P., Roeckel-Trevisiol, N., Jamali, S., Beclin, C., et al. (2006). SRPX2 mutations in disorders of language cortex and cognition. Hum Mol Genet 15, 1195-1207.

61. Wiszniewski, W., Hunter, J.V., Hanchard, N.A., Willer, J.R., Shaw, C., Tian, Q., Illner, A., Wang, X., Cheung, S.W., Patel, A., et al. (2013). TM4SF20 ancestral deletion and susceptibility to a pediatric disorder of early language delay and cerebral white matter hyperintensities. Am J Hum Genet 93, 197-210.

62. Genomes Project Consortium. (2015). A global reference for human genetic variation. Nature 526, 68-74.

63. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285-291.

64. Krause, J., Lalueza-Fox, C., Orlando, L., Enard, W., Green, R.E., Burbano, H.A., Hublin, J.J., Hanni, C., Fortea, J., de la Rasilla, M., et al. (2007). The derived FOXP2 variant of modern humans was shared with Neandertals. Curr Biol 17, 1908-1912.

65. Watkins, K.E., Dronkers, N.F., and Vargha-Khadem, F. (2002). Behavioural analysis of an inherited speech and language disorder: comparison with acquired aphasia. Brain 125, 452-464.

66. Schizophrenia Working Group of the Psychiatric Genomics, C. (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421-427.

67. 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., et al. (2008). Linkage, association, and gene-

expression analyses identify CNTNAP2 as an autism-susceptibility gene. Am J Hum Genet 82, 150-159.

68. Strauss, K.A., Puffenberger, E.G., Huentelman, M.J., Gottlieb, S., Dobrin, S.E., Parod, J.M., Stephan, D.A., and Morton, D.H. (2006). Recessive symptomatic focal epilepsy and mutant contactin-associated protein-like 2. N Engl J Med 354, 1370-1377.

69. Zweier, C., de Jong, E.K., Zweier, M., Orrico, A., Ousager, L.B., Collins, A.L., Bijlsma, E.K., Oortveld, M.A., Ekici, A.B., Reis, A., et al. (2009). CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic protein in Drosophila. Am J Hum Genet 85, 655-666.

70. Shu, W., Lu, M.M., Zhang, Y., Tucker, P.W., Zhou, D., and Morrisey, E.E. (2007). Foxp2 and Foxp1 cooperatively regulate lung and esophagus development. Development 134, 1991-2000.

71. Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H., et al. (2010). A draft sequence of the Neandertal genome. Science 328, 710-722.

72. Maricic, T., Gunther, V., Georgiev, O., Gehre, S., Curlin, M., Schreiweis, C., Naumann, R., Burbano, H.A., Meyer, M., Lalueza-Fox, C., et al. (2013). A recent evolutionary change affects a regulatory element in the human FOXP2 gene. Mol Biol Evol 30, 844-852.

73. Mozzi, A., Forni, D., Clerici, M., Pozzoli, U., Mascheretti, S., Guerini, F.R., Riva, S., Bresolin, N., Cagliani, R., and Sironi, M. (2016). The evolutionary history of genes involved in spoken and written language: beyond FOXP2. Sci Rep 6, 22157.