

## Review

Speech and Language:  
Translating the GenomePelagia Deriziotis<sup>1</sup> and Simon E. Fisher<sup>1,2,\*,@</sup>

Investigation of the biological basis of human speech and language is being transformed by developments in molecular technologies, including high-throughput genotyping and next-generation sequencing of whole genomes. These advances are shedding new light on the genetic architecture underlying language-related disorders (speech apraxia, specific language impairment, developmental dyslexia) as well as that contributing to variation in relevant skills in the general population. We discuss how state-of-the-art methods are uncovering a range of genetic mechanisms, from rare mutations of large effect to common polymorphisms that increase risk in a subtle way, while converging on neurogenetic pathways that are shared between distinct disorders. We consider the future of the field, highlighting the unusual challenges and opportunities associated with studying genomics of language-related traits.

## Genes As Entry Points into Language Neurobiology

Neurodevelopmental disorders with primary deficits in **speech** (see [Glossary](#)) and/or **language** provide a unique entry point into neurobiological pathways that facilitate these fascinating aspects of human development ([Table 1](#)) [1]. Advances in genomics are enhancing our understanding of the underlying genetic architecture. The success of this approach is perhaps best illustrated by the discovery that heterozygous *FOXP2* mutations cause a monogenic disorder, characterized by impaired sequencing of the mouth and face movements required for speech [childhood apraxia of speech (CAS)], along with deficits in expressive and receptive language (OMIM 602081) [2]. *FOXP2* provided the first robust evidence for molecular genetic contributions to language development [2]. Whilst *FOXP2* mutations are rare [3,4], research into its cellular and neurobiological functions served as a formidable gateway into neurogenetic mechanisms important for human communication ([Box 1](#)) [5].

Beyond *FOXP2*, discovery of novel risk genes underlying language-related disorders has been slow, partly because most cases involve complex genetic etiology. Moreover, researchers in this field face special challenges in phenotypic characterization that are not found for more traditional biomedical traits. Lacking well-defined diagnostic markers, disorders of speech, language, and/or reading are typically classified in terms of exclusionary criteria, based on observation of unexpected problems in those domains that cannot be explained by other factors such as deafness, neurological damage, or educational opportunity [6]. Individuals can present with multiple language-related disorders at a given time, the profile of impairments may change at different points of development, and linguistic deficits can occur in the context of broader cognitive syndromes, as in cases of intellectual disability (ID) and autism spectrum disorder (ASD) [7].

Despite these issues, clues into the genetic underpinnings of language-related phenotypes have emerged from classical linkage screens and targeted association studies (see reviews by [8–10]; [Table 2](#)). Several genomic loci that potentially harbor candidate genes have been

## Trends

Identification of genes involved in language-related disorders, including speech apraxia, stuttering, specific language impairment, and developmental dyslexia, provides molecular gateways for investigating the neurobiology of unusual human traits.

High-throughput genotyping and next-generation DNA sequencing offer novel opportunities for pinpointing genetic factors contributing to these disorders, as well as to normal variation in speech, language, and reading abilities.

Neurodevelopmental disorders that disrupt language involve complex genetic architectures; some shared risk factors act across diagnostic categories, while others have more specific effects, but they converge on common biological pathways.

Success in defining the genomic landscapes underlying language traits will depend on functional analyses of genetic variants in model systems, as well as on integration with complementary data from neuroimaging and comparative genomics.

<sup>1</sup>Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

<sup>2</sup>Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

@Twitter: @ProfSimonFisher

\*Correspondence: [simon.fisher@mpi.nl](mailto:simon.fisher@mpi.nl) (S.E. Fisher).

Table 1. Description of Neurodevelopmental Disorders Discussed in This Review

Disorder	Characteristics
Childhood apraxia of speech	Impairments with the learning and execution of coordinated movement sequences of the mouth, tongue, lips, and soft palate, yielding inconsistent speech errors that worsen with increased length and complexity of utterance. May be accompanied by additional deficits in expressive and receptive language. Also known as developmental verbal dyspraxia.
Stuttering	Disruptions in fluency of speech, characterized by involuntary prolongations and repetitions of syllables in addition to uncontrollable silent pauses during speech.
Specific language impairment	Diagnosis given to children with delayed or impaired ability for language acquisition in the absence of physical/neurological causes (i.e., cleft lip/palate, hearing loss, intellectual disability, autism) and despite growing up in a language-rich environment.
Developmental dyslexia	Unexplained problems in learning how to read and/or spell that are not accounted for by low intelligence quotient (IQ), visual acuity problems, or poor learning opportunities. Also known as specific reading disability.
Epilepsy-aphasia spectrum disorders	Collection of epilepsy syndromes with a range of severities, all characterized by seizures that occur in the centrottemporal areas of the brain during childhood but dissipate after adolescence. Seizures develop concurrently with behavioral, cognitive, and linguistic deficits. At the mild end of the spectrum, linguistic impairments disappear after remission of seizures, but they may persist in more severe syndromic forms.
Intellectual disability (ID)	Group of heterogeneous disorders characterized by impairments in general cognitive abilities. An IQ < 70 is required for diagnosis, as well as impairments in communication and self-care. Language delays are a common feature. ID is a feature of many different neurodevelopmental disorders.
Autism spectrum disorder (ASD)	Umbrella term to group a collection of disorders characterized by deficits in social communication and interaction skills alongside stereotyped, repetitive behaviors. Language impairments are often seen but may vary in type and severity. Individuals with ASD typically have problems with the use of language in a social context.

mapped in family cohorts of developmental dyslexia (reading disability), specific language impairment (SLI), and stuttering. Common **single-nucleotide polymorphisms (SNPs)** on multiple chromosomes have been suggested to increase susceptibility to dyslexia and/or SLI, whereas rare **single-nucleotide variants (SNVs)** on chromosomes 12q23, 15q21, and 16p13 have been proposed as risk factors in stuttering [8,9,11]. Follow-up investigations of putative risk factors for association with affection status or quantitative measures of performance have yielded mixed findings, with many failures to replicate, and much of the heritability of language-related disorders remains unaccounted for [8,9]. Together with improvements in molecular technologies, this is prompting a shift in strategies for identifying risk genes.

Recent advances in microarray, genotyping, and sequencing platforms are broadly transforming studies of neurodevelopmental traits. Such developments are fueling discovery of rare and **de novo variants** in disorders with complex phenotypes in which speech- or language-related deficits are prominent. High-throughput sequencing of thousands of cases of ID and ASD highlights a heterogeneous genetic landscape, encompassing hundreds of genes and implicating several cellular pathways, including chromatin remodeling and synaptic function [12–14]. Intriguingly, pathways identified in ID, ASD, epilepsy, and other neurodevelopmental disorders intersect with molecular networks known to go awry in forms of impairment with more selective effects on speech and language skills [15–17].

In this review, we provide a synthesis of molecular findings emerging from application of the latest gene discovery paradigms in this growing research area, describing how language

## Glossary

### 3' untranslated region:

untranslated region of mRNA transcript, following the termination codon.

**Cerebral organoid:** an artificially grown organ resembling a brain that is derived from pluripotent stem cells and can be used to model normal development as well as disorder biology.

### Chromatin immunoprecipitation:

a technique that detects the binding sites of endogenous proteins to DNA sequences by crosslinking the bound proteins to chromatin, followed by immunoprecipitation with a specific antibody and characterization of enriched genomic regions (usually by sequencing).

### Comparative genomic

**hybridization:** a molecular cytogenetic technique used to detect copy number changes in genomic DNA, by comparing a test sample with that of a reference.

### Compound heterozygote:

a genotype in which the two alleles of a gene carry different variants.

### Copy number variant (CNV):

structural variations in chromosomes involving the deletion or duplication of thousands or millions of nucleotides.

### De novo variant:

a genetic variant present in the genome of an individual that is not detected in the DNA of either parent and that may have arisen during oogenesis or spermatogenesis.

### Endophenotype:

in the context of neurogenetics, a heritable neurophysiological, neuroanatomical, or neuropsychological trait that can be measured, can act as a marker for a given condition, and may provide a closer link to the underlying biology.

### Expression quantitative trait

**locus:** a genomic region that carries variation that correlates with variability in expression levels of mRNAs.

**Indels:** sequence variants involving insertions or deletions of a small number of nucleotides.

**Language:** a structured system enabling the communication of thoughts, emotions, and concepts; involves the conventional use of arbitrary elements (spoken sounds, signs, or symbols), which are combined to form a potentially infinite

**Box 1. *FOXP2* As an Entry Point**

Over a decade and a half ago, investigation of a large multigenerational family and an independent translocation case uncovered the first mutations to be implicated in a monogenic speech and language disorder [2]. Since then, the *FOXP2* transcription factor gene has provided a paradigm for bridging genes, neurons, brains, and spoken language [5]. Molecular approaches in cellular systems have identified genes it regulates [15–17,84] and proteins it interacts with [88,89,110], implicating *FOXP2* in neuronal processes such as differentiation, neurite outgrowth, axon guidance, and synaptic plasticity. Intriguingly, a subset of its downstream targets and/or interaction partners has been associated with other neurodevelopmental phenotypes, such as SLI [84], epilepsy [69,86], schizophrenia [79], and ID/ASD [67,80,81,89–91]. Language is uniquely human, but *FOXP2* has a deep evolutionary history, and is highly similar among diverse vertebrates, with conserved expression, particularly in neuronal subtypes within the cortex, thalamus, basal ganglia, and cerebellum [5,108]. Thus, animal models offer clues to *FOXP2*'s neural roles [111,112]. For example, mice with mutations matching those that cause human speech disorder are reported to show impaired motor-skill learning [113] and altered sequencing of ultrasonic vocalizations [114]. Moreover, electrophysiological studies of these mutant mice have found effects on firing and plasticity of neural circuits where the gene is expressed [113,115]. Further clues have come from experiments in the zebra finch, an avian species in which males learn to sing a particular song from an adult tutor bird [112]. When RNA interference has been used to knockdown expression of the zebra finch ortholog of *FOXP2* in Area X, a key basal ganglia nucleus involved in song learning, this disturbs developmental and social modulation of song variability [116,117]. Taken together, studies in other species suggest that contributions of this gene to human speech and language are built on evolutionarily ancient functions in modulating the plasticity of a subset of brain circuits [108]. The roles of *FOXP2* are complex and diverse, even within the nervous system, and there is still a great deal left to learn about its fundamental functions, and why disruptions have disproportionate effects on speech and language abilities. Nonetheless, this story illustrates how a gene implicated in a rare form of disorder can provide informative entry points into the relevant neurobiological pathways [5].

genomics is being transformed by newer technologies. Based on a discussion of currently known neurogenetic pathways disturbed in primary forms of speech and language impairment, we argue for shared genetic etiology in different neurodevelopmental disorders involving language-related deficits. Finally, we consider the future challenges of the field, and discuss strategies for overcoming them.

### Contributions of Common Genetic Variation to Language-Related Phenotypes

As for other human traits, genetic analysis of language-related phenotypes has gained momentum with emergence of affordable high-throughput genotyping platforms. Genome-wide association scans (GWASs) provide a cost-effective way of systematically testing common genetic variation for association with disorders based on case–control comparisons, or with scores on quantitative measures of a phenotype of interest [18]. To achieve adequate statistical power, such efforts typically require many thousands of subjects, far exceeding the number of samples currently available in existing cohorts with language-related phenotypes. Despite this limitation, GWAS has begun to be applied to relevant traits in SLI, developmental dyslexia, and general population cohorts (Table 3) [19–24].

An early GWAS study in this area, involving two small dyslexia cohorts (discovery  $N = 200$ ; replication  $N = 186$ ), reported that an intergenic SNP on chromosome 4q32.1 was significantly associated with **mismatch negativity**, proposed as an electrophysiological **endophenotype** of the disorder. This SNP was suggested to have a potential trans-regulation effect on the expression of *SLC2A3*, a gene on chromosome 12 that encodes a neuronal glucose transporter [25]. Subsequently, an investigation of 133 165 SNPs in 101 families with dyslexic probands (718 individuals) found suggestive association with a marker 77 Kb downstream of fibroblast growth factor 18 (*FGF18*) [23]. A later GWAS study of SLI investigated parent–child trios (297 affected children from 278 families) and found paternal parent-of-origin effects on 14q12, meeting genome-wide significance [19]. The most significant SNP conferred a non-synonymous coding change within *NOP9*, a gene encoding an RNA-binding protein that has been found to be dysregulated in schizophrenia [26]. Suggestive evidence of maternal parent-of-origin effects was found on 5p13, in a region previously implicated in syndromic ID/ASD [27,28].

array of larger units of meaning (words and sentences).

**MicroRNAs:** highly conserved noncoding RNA molecules, typically 18–22 nucleotides long, that bind to the 3' untranslated region of specific mRNAs, acting as post-transcriptional regulators of expression.

**Mismatch negativity:** a component of the brain's electrophysiological response to an odd stimulus in a sequence of stimuli.

**Next-generation sequencing (NGS):** high-throughput technologies capable of massively parallel DNA sequencing applied to sequence protein-coding exons and noncoding RNAs in a genome (whole-exome sequencing) or all base pairs in a genome (whole-genome sequencing).

**Nonword repetition task:** a widely used endophenotype in studies of language and reading disorders, in which children are asked to repeat a pronounceable but meaningless string of speech sounds, of varied complexity and length.

**Single-nucleotide polymorphism (SNP):** a variant at a single base pair position in the genome that is present at an allele frequency of at least 1% in a population.

**Single-nucleotide variant (SNV):** a variant at a single base pair position in the genome that is not common in the population (allele frequency < 1%). If occurring at higher frequency, such a variant is more typically referred to as an SNP.

**Speech:** the human ability to convert thoughts into vocal sounds through coordinated movements of the larynx, tongue, lips, palate, and jaws (the articulators).

**Stop-gain variant:** a sequence variant that leads to a premature stop codon yielding a truncated protein and/or nonsense-mediated decay of the transcript.

Table 2. Examples of Initial Candidate Genes Implicated in Speech, Language, and Reading Disorders, As Identified through Traditional Linkage and Targeted Association Screens (reviewed by [8–11])

Phenotype	Gene	Chromosome	Protein name
Childhood apraxia of speech	<i>FOXP2</i>	7q31.1	Forkhead box protein P2
Stuttering	<i>GNPTAB</i>	12q23.2	<i>N</i> -Acetylglucosamine-1-phosphotransferase subunits $\alpha/\beta$
	<i>AP4E1</i>	15q21.2	Adaptor related protein complex 4 subunit $\epsilon$ 1
	<i>GNPTG</i>	16p13.3	<i>N</i> -Acetylglucosamine-1-phosphotransferase subunit $\gamma$
	<i>NAGPA</i>	16p13.3	<i>N</i> -Acetylglucosamine-1-phosphodiester Alpha- <i>N</i> -acetylglucosaminidase
Developmental dyslexia	<i>ROBO1</i>	3p12.3	Roundabout homolog 1
	<i>KIAA0319</i>	6p22.3	Dyslexia-associated protein KIAA0319
	<i>DCDC2</i>	6p22.3	Doublecortin domain-containing protein 2
	<i>DNAAF4</i>	15q21.3	Dynein axonemal assembly factor 4
Specific language impairment	<i>CNTNAP2</i>	7q35	Contactin-associated protein-like 2
	<i>CMIP</i>	16q23.2	C-Maf-inducing protein
	<i>ATP2C2</i>	16q24.1	Calcium-transporting ATPase type 2C member 2

High-throughput molecular approaches are also being used to characterize genetic architecture that is shared between different language-related phenotypes. For example, one report described a GWAS meta-analysis of quantitative scores on multiple reading- and language-related measures in two independent population samples, the first including Australian twins and siblings ( $N = 1177$  from 538 families), the second comprising children of the UK ALSPAC cohort (Avon Longitudinal Study of Parents and their Children, maximum  $N = 5472$ ). SNPs in *ABCC13* showed suggestive association with performance on a **nonword repetition task**, which has been well established as an endophenotype of SLI; however, while *ABCC13* in monkeys is involved in transporting ATP across membranes, the human version is a pseudogene incapable of encoding a functional protein [22]. Another study of ALSPAC selected individuals with low reading, writing, and language skills, performing case-control screens for reading problems ( $N = 353$ ), language impairment ( $N = 163$ ), and probands comorbid for both ( $N = 174$ ) [21]. Suggestive associations were reported for *NDST4* in language-impaired probands, and *ZNF385D* and *COL4A2* in comorbid probands [21]. More recently, a GWAS meta-analysis of three cohorts of individuals with dyslexia or SLI, as well as their siblings, used quantitative phenotypes on multiple measures to derive a principal component that captured much of the variance in language- and reading-related skills in these samples ( $N = 1862$ ). The authors reported suggestive associations with SNPs in *CCDC136/FLNC*, a gene encoding a structural protein with roles in cytoskeleton remodeling, and *RBFOX2*, which encodes a protein involved in alternative splicing in the nervous system [20]. Of interest, deletions of other *RBFOX* genes (*RBFOX1* and *RBFOX3*) have been reported in Rolandic epilepsy [29]; this form of epilepsy belongs to a group of epilepsy-aphasia spectrum disorders, in which speech and language impairments are predominant features (Table 1).

A GWAS of early expressive vocabulary in infants from ALSPAC and the Dutch Generation-R study (15–18 months; total  $N = 8889$  of combined sample) found significant association for a polymorphism in *ROBO2*, which encodes an axon-guidance receptor and which is closely related to *ROBO1*, a prior candidate for dyslexia susceptibility [24]. However, these *ROBO* genes did not show evidence of association with expressive vocabulary in ALSPAC, Generation-R, and additional cohorts at later stages of language acquisition (24–30 months,  $N = 10\,819$ ), or with other

Table 3. Common Putative Risk Variants That Are Discussed in This Review and That Were Reported in Genome-Wide Association Studies of Language-Related Traits and Disorders

Phenotype	Study	Participants	Associations
Dyslexia	[25]	Children with dyslexia Discovery: 200 individuals Replication: 186 individuals Mismatch negativity as endophenotype	Significant association at 4q32.1 (intergenic)
	[23]	718 individuals from 101 families with dyslexia (at least two affected siblings)	Suggestive association at 5q35.1 (near <i>FGF18</i> )
SLI	[19]	Parent-child trios: 278 families including 297 probands Replication cohort: 313 children with SLI and their mothers from UK ALSPAC	Significant paternal parent-of-origin effects on 14q12 (most significant association within <i>NOP9</i> ) Suggestive maternal parent-of-origin effects on 5p13
Reading problems and language impairment	[21]	Children aged 7–9 from ALSPAC Language impairment only: 163 Reading problems only: 353 Both: 174 No disorder: 4177	Language impairment only: suggestive association at 4q26 ( <i>NDST4</i> ) Reading problems only: no associations Both: suggestive association at 3p24.3 ( <i>ZNF385D</i> ) and 13q34 ( <i>COL4A2</i> )
Reading and language abilities as continuous traits	[22]	Cohort 1: Australian siblings, including twins (1177 individuals aged 12–25 from 538 families) Cohort 2: ALSPAC (5472 children aged 8–9)	Suggestive association at 16q22.3 (intergenic) and 1p13.1 (intergenic) with word reading Suggestive association at 21q11.2 ( <i>ABCC13</i> ) and 16q23.3 (intergenic) with nonword repetition
	[20]	Cohort 1: children with dyslexia and their siblings (983 individuals, 608 families) Cohort 2: children with SLI and their siblings (548 individuals, 288 families) Cohort 3: children with dyslexia and/or attention deficit hyperactivity disorder and their siblings (749 individuals, 343 families)	Suggestive association at 7q32.1 ( <i>CCDC136/FLNC</i> ) and at 22q12.3 ( <i>RBFOX2</i> )
Expressive vocabulary in infancy	[24]	Younger age analyses: Discovery: 6851 individuals from ALSPAC (15-month old) Replication: 2038 individuals from GenR (18-month old) Older age analyses: Discovery: 6299 individuals from ALSPAC (24-month old) Replication: 4520 individuals from GenR, TEDS, and Raine (24–30-month old)	Younger age: Significant association at 3p12.3 ( <i>ROBO2</i> ) and suggestive association at 11p15.2 (near <i>INSC</i> ), 12q15 (near <i>CAND1</i> ) and 19p13.3 ( <i>DAPK3</i> ) in both discovery and follow-up samples Older age: Suggestive association at 5q22.1 ( <i>CAMK4</i> ) in discovery cohort; no association in replication sample

GenR, Generation Rotterdam; TEDS, Twins Early Development Study. Associations with  $P < 10^{-6}$  are listed.

language-related measures at different developmental time points. Thus, assuming the findings are true positives, association signals in studies of normal language acquisition may be highly sensitive to factors like developmental stage. Consistent with this idea, recent molecular epidemiology analyses of genome-wide SNP-chip data from ALSPAC, together with large independent cohorts for ASD (PGC-ASD: 5305 cases and 5305 pseudo-controls; iPSYCH-ASD: 7783 cases and 11 359 controls) and schizophrenia (PGC-SCZ2: 34 241 cases and 45 604 controls; 1235 trios) found that shared genetic influences between clinical disorders and social communication traits depend on the age at which the latter are measured, in line with the distinct developmental profiles of the different disorders [30]. Specifically, there were overlaps in genetic risk between ASD (an early onset disorder) and social communication difficulties in the general population; these overlaps emerged early in development but decreased with age. By contrast, overlaps in genetic risk between schizophrenia (a later-onset disorder) and social communication difficulties in the general population persisted through development, with increased magnitude when the quantitative traits were measured in late adolescence.



### Box 2. Neuroimaging Genomics: Promises and Pitfalls

Confronted with tiny effect sizes for associations of common genetic variants with behavioral measures, researchers have sought to move closer to biology by studying brain-based endophenotypes. The growing field of neuroimaging genomics looks for correlations between interindividual variations in aspects of brain structure and/or function (indexed by neuroimaging) and variability at the molecular level (indexed by genotyping) [118]. The initial hope was that genetic polymorphisms would have larger effects on brain structure and function than on distal behavioral/cognitive outputs, circumventing some of the issues of inadequate power faced by traditional behavior genetic association studies. However, meta-analyses of subcortical brain volumes in over 30 000 individuals demonstrate that structural measures derived from neuroimaging can show just as much complexity in genetic underpinnings as behavioral (endo)phenotypes, with effect sizes for individual genetic associations that are of a similar magnitude [119]. Thus, neuroimaging genomics methods are subject to similar constraints to classical association studies, with respect to study design, sample size, multiple-testing issues, and so on. Moreover, the increased complexity and richness of individual neuroimaging datasets, as compared with psychometric test data (for example), yield greater susceptibility to false positives, especially if exploratory work is performed in small samples without sufficient statistical adjustments or replication [120]. Analyses of *FOXP2* provide a cautionary tale. Neuroimaging of people with speech/language disorder due to rare high-penetrant *FOXP2* coding mutations revealed subtle but significant structural abnormalities affecting regions including the inferior frontal gyrus, striatum, and cerebellum [4,5]. Subsequently, a number of small-scale investigations proposed that common intronic polymorphisms of *FOXP2* (all of unknown functional significance) are associated with variations in brain structure, but the largest study to date, including 1300 people from the general population, found no evidence to support such claims [121]. Nonetheless, as sample sizes continue to increase for cohorts with robust measures of brain structure/function and matching genotype data [118], neuroimaging genomics could indeed help reveal genetic architecture underlying variability in language-related circuits. For instance, we can anticipate high-powered GWASs of individual variability in thickness, surface area, and volumes of key cortical regions involved in language processing, as well as of the white matter tracts that connect them, and of their functional properties. It will be of particular interest to integrate findings from such studies with those from GWASs of language-related behavioral endophenotypes [10]. Thus far, synergies have been limited to assessing specific language/reading-related GWAS signals for effects on brain structure (e.g., [122]).

Overall, emerging GWAS results in language-related phenotypes show potential, but there are very limited overlaps in association signals across the different investigations performed so far, likely reflecting small effect sizes of common variants coupled to the low power of most studies to date. Lack of consistency in GWAS efforts might also be due to inherent difficulties in applying matching diagnostic criteria and phenotypic definitions in the different studies, or perhaps even a lack of common variants contributing to the traits of interest. Nonetheless, the GWAS approach holds considerable potential for the future, through larger-scale meta-analyses, gathering of bigger datasets, and harmonization of phenotyping methods (as discussed later in this article). The expanding area of brain imaging genomics is also beginning to be used for understanding language neurobiology, and this has its own promises and pitfalls (Box 2).

### Insights into Neurodevelopment from Studying Rare Gene Variants

A complementary strategy focuses on rare genetic events of large effect. High-penetrance variants may only account for a minority of cases of speech and language impairment [3], but investigations of neurodevelopmental disorders such as ID and ASD illustrate the value of this approach for uncovering etiology [31–39]. Success in this area has been driven by dramatic technological advances matched with availability of well-phenotyped cohorts. **Comparative genomic hybridization** and **next-generation sequencing (NGS)** screens of thousands of sporadic ASD cases and their unaffected parents have highlighted hundreds of putative risk loci. These studies were based on the hypothesis that sporadic cases of severe neurodevelopmental disorders might be enriched for high-penetrant gene disruptions that arise *de novo* and do not spread due to intense negative selection. Indeed, contributing *de novo* **copy number variants (CNVs)**, SNVs, and insertion/deletion events (**indels**) are found in 10–20% of ASD cases [34,40,41].

While this issue has been extensively studied with respect to ASD and ID, it is not yet clear how much of the genetic architecture underlying primary speech and language impairments involves

rare or *de novo* causal variants of large effect. To the extent that such disorders are not as severe as ASD/ID, the sporadic *de novo* hypothesis may be less applicable. Rare etiological *FOXP2* coding variants (private to particular families) were estimated in one study to account for approximately 2% of CAS referrals [3] and there have also been reports of causative *de novo* variants in this gene [42,43]. Similarly, rare coding variants in genes involved in lysosomal targeting (*NAGPA*, *GNPTAB*, and *GNPTG*) and intracellular trafficking (*AP4E1*) have been suggested as risk factors in stuttering [44,45], but are likely to explain only a small proportion of cases, especially once background mutation rates in unaffected individuals are taken into account. Recent comparative genomic hybridization and SNP-based analyses of independent cohorts with dyslexia indicate that rare large CNVs (>50 Kb) do not confer a significant burden compared with controls [46,47]. A study of 127 independent cases of SLI concluded that most of the risk conferred by CNVs on this disorder involves common, inherited events under a 'common disorder–common variant' model [48]. By contrast, case studies support a role for rare CNVs at the severe end of the spectrum of speech and language disorders. For instance, rare and *de novo* multigene microdeletions in 16p11.2 and 12p13.3 (spanning the *ELKS/ERC1* gene) have been reported as risk factors for disorders involving CAS as a prominent feature [49–52]. Rare microduplications in 16p11.2 have also been reported in epilepsy–aphasia spectrum disorders [53].

There are interesting cases where deletions or duplications of the equivalent chromosomal region have contrasting consequences for language-related phenotypes. For instance, 7q11.23 deletions cause Williams–Beuren syndrome, involving mild-to-moderate ID, with severe problems in visuospatial construction against a background of relatively spared verbal short-term memory and language, while duplications of the same interval lead to impaired speech and expressive language, with relative strengths in visuospatial construction [54]. In another example, a terminal deletion of 22q was found in a girl with autism and severe language impairments, while a partial trisomy of the region was found in her brother, who had Asperger syndrome, with fluent speech and precocious language development [55].

In some cases, a rare CNV points to involvement of a single gene. A hemizygous *de novo* deletion of *BCL11A*, encoding a transcription factor with roles in regulation of hemoglobin expression and in neurite outgrowth [56], was identified in a proband with CAS, expressive language delay, dysarthria, mild ID, hypotonia, and general oral and gross motor dyspraxia [57]. Subsequently, *de novo* heterozygous missense, nonsense, and frameshift variants in *BCL11A* were found in several cases of an ID syndrome characterized by global delay in developmental milestones, including speech and language delay [58]. Similarly, rare inherited and *de novo* disruptions of *GRIN2A*, which encodes the NR2A subunit of the *N*-methyl-D-aspartate glutamate receptor, are estimated to account for 9–20% of epilepsy–aphasia spectrum disorders [59–61]. SNPs in the genes *ATP2C2* and *CMIP* were previously associated with nonword repetition deficits in SLI [62] and *de novo* deletions involving these genes have also been reported in other SLI cases [63], as well as in an autistic proband with speech/language impairments [64].

Speech and/or language deficits observed in individuals carrying rare or *de novo* CNVs may sometimes be due to deletion or duplication of functional elements that regulate the expression of genes beyond the actual site of the CNV itself. For example, one report described an individual with CAS who carries a 2-Mb *de novo* deletion on 7q31, encompassing a noncoding RNA (*EIF3IP1*) and two protein-coding genes, *LRRN3* and *IMMP2L* [65]. Although the site of the deletion mapped 3 Mb away from *FOXP2*, the authors demonstrated that it impacted on *FOXP2* expression *in cis*, presumably due to regulatory elements mapping in the deleted interval.

Very few studies have assessed contributions of rare and *de novo* SNVs to primary speech and language disorders. One study carried out exome sequencing in 10 participants diagnosed with idiopathic CAS [66]. However, although data were obtained exome wide, the authors focused primarily on a subset of known candidate genes from prior literature on communication disorders, with limited consideration of potential contributions from elsewhere in the exomes. They reported nonsynonymous heterozygous coding changes in six candidate genes: *FOXP1*, *CNTNAP1*, *CNTNAP2*, *KIAA0319*, *ATP13A4*, and *SETX*. Because parents were not tested, it was not known whether these variants occurred *de novo*, but a few were rare variants predicted to be deleterious based on available bioinformatics methods. The study did not include experimental validation, and a subsequent independent investigation of the *FOXP1* missense variant (p.I107T) indicated that it does not alter protein function in cells [67]. This illustrates the importance of performing functional characterization of missense variants before assigning causality, particularly when the inheritance status is unknown, and even when they are found in known disorder genes.

A recent study reported rare and *de novo* variants identified by exome sequencing in 43 unrelated probands with severe SLI [68]. The first stage of analysis focused on a predefined set of 19 known candidate genes from prior literature. However, in contrast to the exome-based candidate-gene study of 10 CAS cases [66] described earlier, this SLI exome sequencing study of 43 probands [68] went on to perform more systematic analyses to identify rare/novel variants of potential high risk from throughout the exome, by defining **stop-gain** variants and searching for cases of **compound heterozygotes** for rare disruptive variants. Moreover, they evaluated segregation patterns of all these high-risk variants in the available parents and siblings. Thus, the SLI exome-wide screen not only uncovered variants of interest in known candidates (such as *ERC1*, *GRIN2A*, and *SRPX2*) but also highlighted several novel genes that might be relevant for SLI, including *OXR1*, *SCN9A*, and *KMT2D* [68]. The study also identified that some of the probands carried multiple rare coding variants at different loci, consistent with prior proposals that neurodevelopmental problems may sometimes involve a ‘double-hit’ model in which combinations of variants interact to increase disorder risk [38,46]. Support from recurrent mutations in larger cohorts and experimental evidence of functional impact will be important for future validation of all these findings, but they represent an initial step toward understanding the contributions of rare exonic SNVs to typical forms of SLI.

### A Shared Genetic Landscape Underlying Distinct Disorders

As noted earlier, recent SLI exome-sequencing efforts identified potential disruptive variants in genes (including *ERC1*, *GRIN2A*, and *SRPX2*) that have been implicated in epilepsy-aphasia spectrum disorders and/or speech motor problems [51,60,68,69]. Indeed, an emerging theme is that variants of the same candidate gene may contribute to multiple different language-related disorders (Table 4). Thus, fundamental genetic overlaps may link neurodevelopmental phenotypes that are traditionally deemed as clinically distinct. Such overlaps are perhaps not surprising, given that speech and language impairments often coexist with disorders such as ASD, developmental delays, schizophrenia, epilepsy, and attention deficit hyperactivity disorder. Rare CNVs that have been identified in cases of speech/language impairment have also been argued as causal in these other disorders. For example, deletion or duplication of 16p11.2 has been associated with CAS [49,50,52], dyslexia [70], ASD [70], schizophrenia [71], developmental delay [72], and epilepsy [73]. Similarly, microdeletions of 12p13.3 that encompass *ERC1* have been reported in cases of both CAS [51] and ID [74].

Risk factors shared between distinct language-related phenotypes also include disruptions at the single-gene level. *De novo* mutations in *BCL11A* have been described in CAS [57], ASD [34], and a defined ID syndrome including developmental and language delays as well as dysmorphic features [58]. Rare disruptions in *GRIN2A* can lead to epilepsy-aphasia spectrum



Table 4. Examples of Rare Risk Variants Shared across Distinct Language-Related Phenotypes and Discussed in This Review

Chromosome	Variants	Associated phenotypes	Refs
2p16.1	CNVs, SNVs, and indels affecting <i>BCL11A</i>	Childhood apraxia of speech Autism spectrum disorders Intellectual disability syndrome	[34,57,58]
12p13.33	CNVs affecting <i>ERC1</i>	Childhood apraxia of speech Intellectual disability	[51,74]
16p11.2	CNVs (deletions)	Childhood apraxia of speech Developmental dyslexia Autism spectrum disorders Global developmental delay	[49,50,52,70,72]
	CNVs (duplications)	Autism spectrum disorders Schizophrenia Global developmental delay Epilepsy	[70–73]
16p13.2	CNVs, SNVs, and indels affecting <i>GRIN2A</i>	Epilepsy–aphasia spectrum disorders Speech deficits in absence of seizures Intellectual disability Autism spectrum disorders	[59–61,75–77]

disorders [59,60], speech deficits in absence of seizures [75], mild-to-severe ID [59,61], or ASD [76,77]. Recurrent *de novo* mutations affecting the closely related *GRIN2B* and *GRIN2D* genes have been found in severe ASD [32,36,37] and in epileptic encephalopathy [78], respectively. Overall, neurodevelopmental disorders involve complex genetic architectures, with some shared risk factors acting across different diagnostic categories, while other risk factors may have more specific effects. We still have only a poor understanding of how environmental and/or stochastic influences modulate the phenotypic presentation.

Studies of the functions of *FOXP2*, mutated in CAS (Box 1), have underscored molecular intersections between networks involved in spoken language and pathways implicated in ID/ASD [5]. *FOXP2* encodes a transcription factor that interacts with multiple other proteins and regulates hundreds of genes [5]. Functional investigations of its downstream targets in cellular and animal models point to genetic overlaps in pathways known to go awry in ID/ASD, such as *Wnt* signaling, protein turnover, synaptic plasticity, axon guidance, neurite outgrowth, and cell motility and migration [15–17]. Furthermore, genes known to be regulated by *FOXP2* have been implicated in disorders such as schizophrenia (e.g., *DISC1*) [79] and ASD (e.g., *MET* and *MEF2C*) [80,81]. One of *FOXP2*'s target genes, *TCF4* [17], has been linked to a diverse array of neurodevelopmental phenotypes including Pitt–Hopkins syndrome [82], ASD [36], and schizophrenia [83].

Perhaps the most intensively characterized *FOXP2* target to date is *CNTNAP2* (*contactin-associated protein-like 2*). In early work on *FOXP2* downstream pathways, **chromatin immunoprecipitation** in human neuronal cell models, coupled to shotgun sequencing, revealed that the protein directly binds the first intron of *CNTNAP2* [84], a gene encoding a neurexin cell-adhesion molecule with key roles in several aspects of neurodevelopment [85]. Homozygous loss-of-function *CNTNAP2* mutations result in cortical dysplasia, early onset focal epilepsy, language regression, and cognitive impairment [86,87], whereas heterozygous disruptions give rise to milder phenotypes that can include speech and language problems [84,85]. Common intronic SNPs in *CNTNAP2* showed association with quantitative measures of language skills in a cohort of 184 SLI families, especially with reduced performance on the nonword repetition endophenotype [84].

Molecular links between clinically distinct neurodevelopmental disorders are also evident from investigations of FOXP2 protein–protein interactions. For instance, FOXP2 interacts with FOXP1 and TBR1 [88,89], two other transcription factors involved in brain development. Rare and *de novo* mutations disrupting either *FOXP1* or *TBR1* result in recognizable syndromes characterized by global developmental delay, ID, autistic traits, and moderate-to-severe speech deficits [37,67,90,91]. Interactions between FOXP2, FOXP1, and TBR1 in brain areas where they are coexpressed (including the striatum in the case of FOXP2–FOXP1 and the deep layers of the cortex in the case of FOXP2–TBR1) may be important for coregulating downstream targets relevant to language development, such as *CNTNAP2* [32]. Disruptions of these interactions could yield dysregulation of targets and contribute to the speech/language impairments seen in the distinct phenotypes. This hypothesis is supported by cell-based assays of protein function; *de novo* mutations of *TBR1* found in ASD disturb the interaction of the encoded protein with FOXP2 [89]. Intriguingly, pathogenic *FOXP2* mutations implicated in CAS also disrupt FOXP2–TBR1 interactions [89].

## Concluding Remarks

### The Future: Challenges and Opportunities

Clearly, the genetic architecture shaping language proficiency within each individual and across the population is heterogeneous, involving both common and rare variation. To uncover novel language-related genes and to better understand the molecular basis of the relevant suite of skills, state-of-the-art methods for genomic characterization must be coupled to data integration from complementary approaches in multiple disciplines, from molecules to the clinic.

To date, GWAS studies of speech, language, and reading phenotypes, whether for disorders or normal variation, have revealed few significant associations, and none have yet been convincingly replicated. The limited success may largely reflect a lack of power, due to insufficient sample sizes for detecting SNPs with anticipated small effects, against a background of massive multiple testing across the genome. The experiences of similar research efforts for other brain-related phenotypes, such as schizophrenia, suggest that numbers of genotyped/phenotyped participants in the order of tens of thousands will be required for robust associations [92]. One strategy for achieving the necessary sample sizes is to target independent cohorts with existing genome-wide genotype information, and which have already collected data on speech, language, and/or reading measures, and attempt to harmonize the available phenotypes across cohorts for meta-analyses (or even mega-analyses). A new network, the GenLang consortium (<http://genlang.org/>), was recently established to facilitate these kinds of coordinated efforts.

Here it is worth noting some special challenges for language-related GWAS designs. First, many large-scale GWAS studies in biomedical fields take advantage of cohorts collected through routine clinical or neuropsychiatric screening, but people with developmental speech, language, and/or reading disorders do not typically enter these kinds of medical routes. Second, ‘language’ is not a unitary dimension but a complex multifaceted phenotype, involving aspects ranging from vocabulary size and speech fluency to understanding of grammar and pragmatic inferences. How such aspects relate to each other within an individual, how they vary between individuals, and how profiles change at different points of development are all still poorly understood. Third, the distinct properties of languages spoken in different parts of the world impose their own limitations on phenotypic characterization. Some of the aforementioned phenotyping issues will be addressed in coming years with development of standardized Web-/app-based test batteries for tapping into variation in language-related skills in different languages (not only in disorders but also in the normal range of abilities).

## Outstanding Questions

Are sporadic cases of severe developmental speech and language impairments enriched for *de novo* causative variants?

What proportion of the genomic architecture that underlies language-related disorders comprises common risk factors of small effect size, versus rare mutations with large effects?

Moving beyond germ-line mutations, how do somatic mutations and epigenetic effects contribute to language-related disorders?

Are there genes which harbor both rare mutations and common risk variants that contribute to related phenotypes, or do the different types of variation have distinct consequences for development of brain and behavior?

Which of the genes operating at the disorder end of the spectrum of language abilities have effects also in the normal range? Considering the other end of the spectrum, is there a genetic basis for exceptional speech, language, or reading skills, and which genes are involved?

How might similar mutations involved in speech and language deficits manifest their profiles of impairment in different cultures with highly distinct languages?

How do genetic factors involved in speech and language skills affect brain development? How do they influence the structure and function of particular human brain circuits, as captured through neuroimaging methods?

Do genes implicated in interindividual variability in speech and language skills in modern humans overlap with the genes that were involved in the emergence of human linguistic capacities during evolution of our lineage?

What are the deeper evolutionary histories of language-related genes? What functions do they have in other species and/or in non-neural tissues of the body?

Complementary to studying common variation, identification of novel genetic contributions to speech/language phenotypes will greatly benefit from increased use of NGS platforms to discover rare variants with large effects. One strategy is to apply whole-exome or whole-genome sequencing in large families showing potential monogenic transmission of a language-related disorder [93,94], although there are no guarantees that such approaches will pinpoint a single causal variant in the family being studied. Useful insights may also emerge from GWAS and/or NGS in geographical isolates with dramatically increased prevalence of language disorders, as shown by studies of SLI implicating variants in *SETBP1* and *MEF2*-regulated genes in a remote cluster of Russian villages [95], and a variant of *NFXL1* in a founder population from Robinson Crusoe Island (Chile) [96]. Another means for identifying high-penetrance variants is to sequence entire exomes/genomes of parent-child trios, in which there is a proband with a severe sporadic disorder. This method has been successful in identifying *de novo* causative variants in ID and ASD, even with small sample sizes (10–20 trios) [32,97], and is in the midst of being applied to severe cases of speech/language disorder (see Outstanding Questions).

To properly translate NGS findings, it is essential to get a clear understanding of the biological impact of the sequence variants that are identified. Initial data filtering is done using computational tools that infer functional importance of, for example, single amino acid substitutions based on sequence conservation only [98–101] or on sequence- and structure-based features [102,103]. However, empirical assessments of functional significance in experimental systems will be crucial for establishing whether variants are truly pathogenic and for understanding their biological impact [104]. Cellular assays are already being used to compare effects conferred by rare variants in known language-related genes, emerging from NGS screening [67,91,105]. Two recent examples include a functional study of seven rare *FOXP2* variants, including missense and frameshift variants, found in cases of primary speech/language disorder [105], and an experimental assessment of 11 different *FOXP1* variants, including *de novo* events, identified in children with neurodevelopmental disorder [67]. Although these reports confirmed functional impact for multiple variants, both studies found examples of variants that had been predicted as causal, but did not affect protein functions in experimental assays and are most probably incidental to the phenotypes observed in the children that carry them [67,105].

While tests of protein function help identify likely disease-causing mutations against the high background of incidental variation found in each person's genome, it is not always feasible to experimentally assess putative risk variants, especially if they occur in genes of unknown function or in noncoding regions. Nonetheless, already in the context of speech/language phenotypes, progress is being made in assessing biological roles of variants falling outside coding regions. A recent study analyzed **3' untranslated regions** in exome sequencing data from 43 children with language impairments (the same SLI cohort as that studied by [68]) and identified a functional SLI-associated variant affecting a **microRNA** binding site in the gene *ARHGEF39* [106]. Functional assays in cellular models demonstrated that this variant affected regulation of *ARHGEF39*, which was consistently supported by **expression quantitative trait locus** data from postmortem human brains. As well as identifying an interesting new candidate gene for SLI susceptibility, this work demonstrates the potential for systematically evaluating functional impacts of noncoding variants identified in NGS screens of neurodevelopmental disorders, especially as whole-genome sequencing becomes routine.

Translating the genome for understanding the biology of human language will ultimately depend on integrating findings from diverse fields, only a subset of which we have discussed here. Close interactions between clinical geneticists and experts in speech/language phenotyping are needed to improve characterization of the impacts of different gene variants, and to aid in

### Box 3. Learning from Archaic Hominin Genomes

Molecular data from ancient hominins offer a complementary source for gaining insights into the biology of language [107,123]. Until a few years ago, comparative genomic approaches faced certain limitations in explaining the emergence of human traits, because these methods had to largely focus on extant species. Comparisons of the human nuclear genome sequence with that of non-human primates enabled researchers to assemble a fairly comprehensive catalog of derived genomic features, that is, virtually every change that occurred on our lineage after splitting from the common ancestor with the chimpanzee lineage, several million years ago [108]. However, the number of such features is high (millions of single-nucleotide changes and indels) and it is no trivial task to identify which of these many different genomic changes were functionally relevant for evolution of linguistic skills [108]. Analyses of variation in human populations can help narrow regions of evolutionary interest, for example, by identifying signatures of selection in the genome [123]. Crucially, advances in NGS have now enabled molecular anthropologists to determine relatively complete genome sequences from Neandertals and Denisovans, archaic hominins that diverged from our own lineage 400–500 000 years ago [124]. Such data make it possible to date evolutionary events more precisely, identifying subsets of genomic changes that occurred following the splitting of anatomically modern humans from the Neandertal/Denisovan lineages. A study of Neandertal DNA highlighted approximately 31 000 single-nucleotide changes and approximately 4000 indels that arose on the human lineage after this split, including 96 amino acid substitutions and approximately 3000 potential regulatory variants that became fixed in modern humans [124]. Moreover, this kind of genomic analyses revealed the existence of interbreeding between early modern humans and archaic hominins, such that up to 4% of the genomes of non-Africans constitute introgressed fragments from Neandertals [123]. Putting aside debates about the linguistic prowess (or lack thereof) of Neandertals, it will be interesting in the future to integrate data from comparative genomics with findings from gene mapping studies of language-related phenotypes [107]. As with other areas of work, *FOXP2* provides a useful illustration of the principles. Against a background of strong conservation across diverse vertebrates (Box 1), *FOXP2* underwent multiple interesting evolutionary events on the lineage that led to modern humans, including two amino acid substitutions after splitting from the chimpanzee lineage [125], and at least one putative regulatory change after splitting from the Neandertal/Denisovan lineages [126], and it lies in a gene desert that has resisted introgression of archaic alleles [127]. Crucially, the functional significance of these evolutionary changes is open to empirical investigation in model systems. For example, when the *FOXP2* amino acid substitutions that occurred on the human lineage are inserted into a mouse model, they yield changes in plasticity of a subset of neural circuits that express the gene, in ways that differ from consequences of gene knockout [125].

developing novel targeted therapies. Only humans have the capacity to acquire proficient spoken language, and there is much we can learn from molecular evolutionary data (Box 3), especially comparative genomics of archaic hominins [107]. If, as has been argued [108], neurogenetic pathways involved in human language built on mechanisms with deeper evolutionary histories, valuable insights will also come from investigating language-related genes in animal and bird models (see also Box 1). At the same time, we should bring to our studies of genes and language more sophisticated systems for modeling human brain development and function in the laboratory, including reprogramming of human cells to produce neurons in a dish, along with **cerebral organoids** [109]. Through the synergies offered by these diverse approaches, we will eventually bridge the gaps between genes, neurons, brain circuits, and language.

### References

- Graham, S.A. *et al.* (2015) Insights into the genetic foundations of human communication. *Neuropsychol. Rev.* 25, 3–26
- Lai, C.S. *et al.* (2001) A Forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413, 519–523
- MacDermot, K.D. *et al.* (2005) Identification of *FOXP2* truncation as a novel cause of developmental speech and language deficits. *Am. J. Hum. Genet.* 76, 1074–1080
- Morgan, A. *et al.* (2016) *FOXP2*-related speech and language disorders. In *GeneReviews* (Pagon, R.A., ed.), pp. 1993–2017, University of Washington
- Fisher, S.E. and Scharff, C. (2009) *FOXP2* as a molecular window into speech and language. *Trends Genet.* 25, 166–177
- Bishop, D.V. (2014) Ten questions about terminology for children with unexplained language problems. *Int. J. Lang. Commun. Disord.* 49, 381–415
- Tomblin, B. (2011) Co-morbidity of autism and SLI: kinds, kin and complexity. *Int. J. Lang. Commun. Disord.* 46, 127–137
- Carrion-Castillo, A. *et al.* (2013) Molecular genetics of dyslexia: an overview. *Dyslexia* 19, 214–240
- Reader, R.H. *et al.* (2014) Genome-wide studies of specific language impairment. *Curr. Behav. Neurosci. Rep.* 1, 242–250
- Graham, S.A. and Fisher, S.E. (2015) Understanding language from a genomic perspective. *Annu. Rev. Genet.* 49, 131–160
- Kang, C. and Drayna, D. (2011) Genetics of speech and language disorders. *Annu. Rev. Genomics Hum. Genet.* 12, 145–164
- van Bokhoven, H. (2011) Genetic and epigenetic networks in intellectual disabilities. *Annu. Rev. Genet.* 45, 81–104
- Krumm, N. *et al.* (2014) A *de novo* convergence of autism genetics and molecular neuroscience. *Trends Neurosci.* 37, 95–105
- Huguet, G. *et al.* (2013) The genetic landscapes of autism spectrum disorders. *Annu. Rev. Genomics Hum. Genet.* 14, 191–213
- Spiteri, E. *et al.* (2007) Identification of the transcriptional targets of *FOXP2*, a gene linked to speech and language, in developing human brain. *Am. J. Hum. Genet.* 81, 1144–1157
- Vernes, S.C. *et al.* (2007) High-throughput analysis of promoter occupancy reveals direct neural targets of *FOXP2*, a gene

- mutated in speech and language disorders. *Am. J. Hum. Genet.* 81, 1232–1250
17. Vernes, S.C. *et al.* (2011) Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. *PLoS Genet.* 7, e1002145
  18. Visscher, P.M. *et al.* (2012) Five years of GWAS discovery. *Am. J. Hum. Genet.* 90, 7–24
  19. Nudel, R. *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
  20. Gialluisi, A. *et al.* (2014) Genome-wide screening for DNA variants associated with reading and language traits. *Genes Brain Behav.* 13, 686–701
  21. Eicher, J.D. *et al.* (2013) Genome-wide association study of shared components of reading disability and language impairment. *Genes Brain Behav.* 12, 792–801
  22. Luciano, M. *et al.* (2013) A genome-wide association study for reading and language abilities in two population cohorts. *Genes Brain Behav.* 12, 645–652
  23. Field, L.L. *et al.* (2013) Dense-map genome scan for dyslexia supports loci at 4q13, 16p12, 17q22; suggests novel locus at 7q36. *Genes Brain Behav.* 12, 56–69
  24. St Pourcain, B. *et al.* (2014) Common variation near ROBO2 is associated with expressive vocabulary in infancy. *Nat. Commun.* 5, 4831
  25. Roeske, D. *et al.* (2011) First genome-wide association scan on neurophysiological endophenotypes points to trans-regulation effects on SLC2A3 in dyslexic children. *Mol. Psychiatry* 16, 97–107
  26. Glatt, S.J. *et al.* (2011) Similarities and differences in peripheral blood gene-expression signatures of individuals with schizophrenia and their first-degree biological relatives. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 156B, 869–887
  27. Krantz, I.D. *et al.* (2004) Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of *Drosophila melanogaster* Nipped-B. *Nat. Genet.* 36, 631–635
  28. Novara, F. *et al.* (2013) 5p13 microduplication syndrome: a new case and better clinical definition of the syndrome. *Eur. J. Med. Genet.* 56, 54–58
  29. Lal, D. *et al.* (2013) RBFOX1 and RBFOX3 mutations in Rolandic epilepsy. *PLoS One* 8, e73323
  30. St Pourcain, B. *et al.* (2017) ASD and schizophrenia show distinct developmental profiles in common genetic overlap with population-based social communication difficulties. *Mol. Psychiatry* Published online January 3, 2017. <http://dx.doi.org/10.1038/mp.2016.198>
  31. Gilissen, C. *et al.* (2014) Genome sequencing identifies major causes of severe intellectual disability. *Nature* 511, 344–347
  32. O’Roak, B.J. *et al.* (2011) Exome sequencing in sporadic autism spectrum disorders identifies severe *de novo* mutations. *Nat. Genet.* 43, 585–589
  33. Sanders, S.J. *et al.* (2012) *De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 485, 237–241
  34. De Rubeis, S. *et al.* (2014) Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 515, 209–215
  35. Neale, B.M. *et al.* (2012) Patterns and rates of exonic *de novo* mutations in autism spectrum disorders. *Nature* 485, 242–245
  36. Talkowski, M.E. *et al.* (2012) Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. *Cell* 149, 525–537
  37. O’Roak, B.J. *et al.* (2012) Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* 338, 1619–1622
  38. O’Roak, B.J. *et al.* (2012) Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* 485, 246–250
  39. de Ligt, J. *et al.* (2012) Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.* 367, 1921–1929
  40. Iossifov, I. *et al.* (2014) The contribution of *de novo* coding mutations to autism spectrum disorder. *Nature* 515, 216–221
  41. Gaugler, T. *et al.* (2014) Most genetic risk for autism resides with common variation. *Nat. Genet.* 46, 881–885
  42. Turner, S.J. *et al.* (2013) Small intragenic deletion in FOXP2 associated with childhood apraxia of speech and dysarthria. *Am. J. Med. Genet. A* 161A, 2321–2326
  43. Reuter, M.S. *et al.* (2017) FOXP2 variants in 14 individuals with developmental speech and language disorders broaden the mutational and clinical spectrum. *J. Med. Genet.* 54, 64–72
  44. Kang, C. *et al.* (2010) Mutations in the lysosomal enzyme-targeting pathway and persistent stuttering. *N. Engl. J. Med.* 362, 677–685
  45. Raza, M.H. *et al.* (2015) Association between rare variants in AP4E1, a component of intracellular trafficking, and persistent stuttering. *Am. J. Hum. Genet.* 97, 715–725
  46. Girirajan, S. *et al.* (2011) Relative burden of large CNVs on a range of neurodevelopmental phenotypes. *PLoS Genet.* 7, e1002334
  47. Gialluisi, A. *et al.* (2016) Investigating the effects of copy number variants on reading and language performance. *J. Neurodev. Disord.* 8, 17
  48. Simpson, N.H. *et al.* (2015) Genome-wide analysis identifies a role for common copy number variants in specific language impairment. *Eur. J. Hum. Genet.* 23, 1370–1377
  49. Newbury, D.F. *et al.* (2013) Dual copy number variants involving 16p11 and 6q22 in a case of childhood apraxia of speech and pervasive developmental disorder. *Eur. J. Hum. Genet.* 21, 361–365
  50. Raca, G. *et al.* (2013) Childhood apraxia of speech (CAS) in two patients with 16p11.2 microdeletion syndrome. *Eur. J. Hum. Genet.* 21, 455–459
  51. Thevenon, J. *et al.* (2013) 12p13.33 microdeletion including ELKS/ERC1, a new locus associated with childhood apraxia of speech. *Eur. J. Hum. Genet.* 21, 82–88
  52. Fedorenko, E. *et al.* (2016) A highly penetrant form of childhood apraxia of speech due to deletion of 16p11.2. *Eur. J. Hum. Genet.* 24, 302–306
  53. Dimassi, S. *et al.* (2014) A subset of genomic alterations detected in Rolandic epilepsies contains candidate or known epilepsy genes including GRIN2A and PRRT2. *Epilepsia* 55, 370–378
  54. Osborne, L.R. and Mervis, C.B. (2007) Rearrangements of the Williams–Beuren syndrome locus: molecular basis and implications for speech and language development. *Expert Rev. Mol. Med.* 9, 1–16
  55. Durand, C.M. *et al.* (2007) Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat. Genet.* 39, 25–27
  56. Kuo, T.Y. *et al.* (2010) X-linked mental retardation gene CASK interacts with Bcl11A/CTIP1 and regulates axon branching and outgrowth. *J. Neurosci. Res.* 88, 2364–2373
  57. Peter, B. *et al.* (2014) *De novo* microdeletion of BCL11A is associated with severe speech sound disorder. *Am. J. Med. Genet. A* 164A, 2091–2096
  58. Dias, C. *et al.* (2016) BCL11A haploinsufficiency causes an intellectual disability syndrome and dysregulates transcription. *Am. J. Hum. Genet.* 99, 253–274
  59. Carvill, G.L. *et al.* (2013) GRIN2A mutations cause epilepsy-aphasia spectrum disorders. *Nat. Genet.* 45, 1073–1076
  60. Lesca, G. *et al.* (2013) GRIN2A mutations in acquired epileptic aphasia and related childhood focal epilepsies and encephalopathies with speech and language dysfunction. *Nat. Genet.* 45, 1061–1066
  61. Lemke, J.R. *et al.* (2013) Mutations in GRIN2A cause idiopathic focal epilepsy with Rolandic spikes. *Nat. Genet.* 45, 1067–1072
  62. Newbury, D.F. *et al.* (2009) CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *Am. J. Hum. Genet.* 85, 264–272
  63. Smith, A.W. *et al.* (2015) Deletion of 16q24.1 supports a role for the ATP2C2 gene in specific language impairment. *J. Child Neurol.* 30, 517–521
  64. Van der Aa, N. *et al.* (2012) Haploinsufficiency of CMIP in a girl with autism spectrum disorder and developmental delay due to



- a *de novo* deletion on chromosome 16q23.2. *Autism Res.* 5, 277–281
65. Adegbola, A.A. *et al.* (2015) Monoallelic expression of the human FOXP2 speech gene. *Proc. Natl. Acad. Sci. U. S. A.* 112, 6848–6854
  66. Worthey, E.A. *et al.* (2013) Whole-exome sequencing supports genetic heterogeneity in childhood apraxia of speech. *J. Neurodev. Disord.* 5, 29
  67. Sollis, E. *et al.* (2016) Identification and functional characterization of *de novo* FOXP1 variants provides novel insights into the etiology of neurodevelopmental disorder. *Hum. Mol. Genet.* 25, 546–557
  68. Chen, X.S. *et al.* (2017) Next-generation DNA sequencing identifies novel gene variants and pathways involved in specific language impairment. *Sci. Rep.* 7, 46105
  69. Roll, P. *et al.* (2010) Molecular networks implicated in speech-related disorders: FOXP2 regulates the SRPX2/uPAR complex. *Hum. Mol. Genet.* 19, 4848–4860
  70. Weiss, L.A. *et al.* (2008) Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* 358, 667–675
  71. McCarthy, S.E. *et al.* (2009) Microduplications of 16p11.2 are associated with schizophrenia. *Nat. Genet.* 41, 1223–1227
  72. Shinawi, M. *et al.* (2010) Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. *J. Med. Genet.* 47, 332–341
  73. Mefford, H.C. *et al.* (2011) Rare copy number variants are an important cause of epileptic encephalopathies. *Ann. Neurol.* 70, 974–985
  74. Abdelmoity, A.T. *et al.* (2011) 1.39 Mb inherited interstitial deletion in 12p13.33 associated with developmental delay. *Eur. J. Med. Genet.* 54, 198–203
  75. Turner, S.J. *et al.* (2015) GRIN2A: an aptly named gene for speech dysfunction. *Neurology* 84, 586–593
  76. Girirajan, S. *et al.* (2013) Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. *Am. J. Hum. Genet.* 92, 221–237
  77. D’Gama, A.M. *et al.* (2015) Targeted DNA sequencing from autism spectrum disorder brains implicates multiple genetic mechanisms. *Neuron* 88, 910–917
  78. Li, D. *et al.* (2016) GRIN2D recurrent *de novo* dominant mutation causes a severe epileptic encephalopathy treatable with NMDA receptor channel blockers. *Am. J. Hum. Genet.* 99, 802–816
  79. Walker, R.M. *et al.* (2012) The DISC1 promoter: characterization and regulation by FOXP2. *Hum. Mol. Genet.* 21, 2862–2872
  80. Mukamel, Z. *et al.* (2011) Regulation of MET by FOXP2, genes implicated in higher cognitive dysfunction and autism risk. *J. Neurosci.* 31, 11437–11442
  81. Chen, Y.C. *et al.* (2016) Foxp2 controls synaptic wiring of corticostriatal circuits and vocal communication by opposing Mef2c. *Nat. Neurosci.* 19, 1513–1522
  82. Rosenfeld, J.A. *et al.* (2009) Genotype-phenotype analysis of TCF4 mutations causing Pitt–Hopkins syndrome shows increased seizure activity with missense mutations. *Genet. Med.* 11, 797–805
  83. Cross-Disorder Group of the Psychiatric Genomics Consortium *et al.* (2013) Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* 45, 984–994
  84. Vernes, S.C. *et al.* (2008) A functional genetic link between distinct developmental language disorders. *N. Engl. J. Med.* 359, 2337–2345
  85. Rodenas-Cuadrado, P. *et al.* (2014) Shining a light on CNTNAP2: complex functions to complex disorders. *Eur. J. Hum. Genet.* 22, 171–178
  86. Strauss, K.A. *et al.* (2006) Recessive symptomatic focal epilepsy and mutant contactin-associated protein-like 2. *N. Engl. J. Med.* 354, 1370–1377
  87. Rodenas-Cuadrado, P. *et al.* (2016) Characterisation of CASPR2 deficiency disorder – a syndrome involving autism, epilepsy and language impairment. *BMC Med. Genet.* 17, 8
  88. Li, S. *et al.* (2004) Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions. *Mol. Cell Biol.* 24, 809–822
  89. Deriziotis, P. *et al.* (2014) *De novo* TBR1 mutations in sporadic autism disrupt protein functions. *Nat. Commun.* 5, 4954
  90. Le Fevre, A.K. *et al.* (2013) FOXP1 mutations cause intellectual disability and a recognizable phenotype. *Am. J. Med. Genet. A* 161A, 3166–3175
  91. Lozano, R. *et al.* (2015) A *de novo* FOXP1 variant in a patient with autism, intellectual disability and severe speech and language impairment. *Eur. J. Hum. Genet.* 23, 1702–1707
  92. Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511, 421–427
  93. Briscoe, J. *et al.* (2012) A specific cognitive deficit within semantic cognition across a multi-generational family. *Proc. Biol. Sci.* 279, 3652–3661
  94. Peter, B. *et al.* (2016) Genetic candidate variants in two multi-generational families with childhood apraxia of speech. *PLoS One* 11, e0153864
  95. Kornilov, S.A. *et al.* (2016) Genome-wide association and exome sequencing study of language disorder in an isolated population. *Pediatrics* 137, e20152469
  96. Villanueva, P. *et al.* (2015) Exome sequencing in an admixed isolated population indicates NFXL1 variants confer a risk for specific language impairment. *PLoS Genet.* 11, e1004925
  97. Vissers, L.E. *et al.* (2010) A *de novo* paradigm for mental retardation. *Nat. Genet.* 42, 1109–1112
  98. Kumar, P. *et al.* (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 4, 1073–1081
  99. Stone, E.A. and Sidow, A. (2005) Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Res.* 15, 978–986
  100. Thomas, P.D. and Kejariwal, A. (2004) Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. *Proc. Natl. Acad. Sci. U. S. A.* 101, 15398–15403
  101. Cooper, G.M. *et al.* (2005) Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* 15, 901–913
  102. Adzhubei, I.A. *et al.* (2010) A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249
  103. Kircher, M. *et al.* (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* 46, 310–315
  104. Deriziotis, P. and Fisher, S.E. (2013) Neurogenomics of speech and language disorders: the road ahead. *Genome Biol.* 14, 204
  105. Estruch, S.B. *et al.* (2016) Functional characterization of rare FOXP2 variants in neurodevelopmental disorder. *J. Neurodev. Disord.* 8, 44
  106. Devanna, P. *et al.* (2017) Next-gen sequencing identifies non-coding variation disrupting miRNA-binding sites in neurological disorders. *Mol. Psychiatry* Published online March 14, 2017. <http://dx.doi.org/10.1038/mp.2017.30>
  107. Fisher, S.E. (2017) Evolution of language: lessons from the genome. *Psychon. Bull. Rev.* 24, 34–40
  108. Fisher, S.E. and Marcus, G.F. (2006) The eloquent ape: genes, brains and the evolution of language. *Nat. Rev. Genet.* 7, 9–20
  109. Quadrato, G. *et al.* (2016) The promises and challenges of human brain organoids as models of neuropsychiatric disease. *Nat. Med.* 22, 1220–1228
  110. Estruch, S.B. *et al.* (2016) The language-related transcription factor FOXP2 is post-translationally modified with small ubiquitin-like modifiers. *Sci. Rep.* 6, 20911
  111. French, C.A. and Fisher, S.E. (2014) What can mice tell us about Foxp2 function? *Curr. Opin. Neurobiol.* 28, 72–79
  112. Wohlgenuth, S. *et al.* (2014) FoxP2 in songbirds. *Curr. Opin. Neurobiol.* 28, 86–93
  113. Groszer, M. *et al.* (2008) Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits. *Curr. Biol.* 18, 354–362

114. Chabout, J. *et al.* (2016) A Foxp2 mutation implicated in human speech deficits alters sequencing of ultrasonic vocalizations in adult male mice. *Front. Behav. Neurosci.* 10, 197
115. French, C.A. *et al.* (2012) An aetiological Foxp2 mutation causes aberrant striatal activity and alters plasticity during skill learning. *Mol. Psychiatry* 17, 1077–1085
116. Haesler, S. *et al.* (2007) Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. *PLoS Biol.* 5, e321
117. Murugan, M. *et al.* (2013) Diminished FoxP2 levels affect dopaminergic modulation of corticostriatal signaling important to song variability. *Neuron* 80, 1464–1476
118. Bearden, C.E. and Thompson, P.M. (2017) Emerging global initiatives in neurogenetics: The Enhancing Neuroimaging Genetics through Meta-analysis (ENIGMA) Consortium. *Neuron* 94, 232–236
119. Hibar, D.P. *et al.* (2015) Common genetic variants influence human subcortical brain structures. *Nature* 520, 224–229
120. Button, K.S. *et al.* (2013) Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376
121. Hoogman, M. *et al.* (2014) Assessing the effects of common variation in the FOXP2 gene on human brain structure. *Front. Hum. Neurosci.* 8, 473
122. Gialluisi, A. *et al.* (2016) Neuroimaging genetic analyses of novel candidate genes associated with reading and language. *Brain Lang.* Published online July 27, 2017. <http://dx.doi.org/10.1016/j.bandl.2016.07.002>
123. Paabo, S. (2014) The human condition – a molecular approach. *Cell* 157, 216–226
124. Prufer, K. *et al.* (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505, 43–49
125. Enard, W. *et al.* (2009) A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. *Cell* 137, 961–971
126. Maricic, T. *et al.* (2013) A recent evolutionary change affects a regulatory element in the human FOXP2 gene. *Mol. Biol. Evol.* 30, 844–852
127. Sankararaman, S. *et al.* (2016) The combined landscape of Denisovan and Neanderthal ancestry in present-day humans. *Curr. Biol.* 26, 1241–1247