

REVIEW

Neurogenomics of speech and language disorders: the road ahead

Pelagia Deriziotis¹ and Simon E Fisher^{1,2*}

Abstract

Next-generation sequencing is set to transform the discovery of genes underlying neurodevelopmental disorders, and so offer important insights into the biological bases of spoken language. Success will depend on functional assessments in neuronal cell lines, animal models and humans themselves.

Keywords exome, genome, *FOXP2*, functional validation, language, next-generation sequencing, neurodevelopmental disorders, speech

Introduction

The human capacity for complex spoken language is unique [1]. Speech endows us with the ability to verbally express our ideas, opinions and feelings, using rapid precise control of the oral motor structures (larynx, mouth, tongue) to convert our thoughts into streams of sound that can be decoded by others. While vocal communication in other species sometimes exploits simple mappings between sound and meaning, the reach of human language extends far beyond this, most notably through its extraordinary generative power. A discrete number of individual units of language can be combined into a limitless number of utterances, giving us the potential to express and comprehend an infinite array of concepts. Moreover, when growing up in a language-rich environment, any normal human infant becomes highly proficient in his or her native language with astonishing ease, and without the need for explicit teaching.

It has been argued for many years that inherited factors must make a key contribution to the acquisition of spoken language [2]. It is only in the past decade or so, with the rise of molecular genetics, that biologists have been able to provide the first robust empirical evidence

regarding this issue. To begin investigating the pathways involved, research has focused on the roles of genes, proteins and cellular machinery in the etiology of language impairments, in which people mysteriously fail to develop normal skills despite adequate linguistic input and opportunity [3]. There is a diverse array of these language-related disorders, which usually appear in early childhood and often persist into later life, and they are common enough to have a major impact on modern society. Language problems are frequently observed co-occurring with other developmental disorders, such as autism and epilepsy [4,5].

Prior to the advent of molecular studies of language disorders, the importance of the genome was already evident from epidemiological analyses. These disorders typically cluster in families [6-9] and monozygotic twins display substantially higher rates of concordance than dizygotic twins [10-12]. Clearly, acquisition of fluent spoken language is also influenced by the environment and its interaction with our genes. However, beyond the obvious effects of impoverished language input (for example, due to hearing problems) there is little known regarding specific environmental risk factors that may disturb linguistic development [13].

Initial clues to the molecular bases of speech and language impairments came from low-density linkage screens [14], followed by targeted association studies of particular chromosome regions and/or focused mutation screens of candidate genes [15]. In addition, studies of chromosomal abnormalities are contributing to our understanding of such disorders, and genome-wide association scans using hundreds of thousands of single nucleotide polymorphisms (SNPs) are underway in several cohorts. However, it is evident that the future of gene discovery in language-related traits, as for many other complex phenotypes, lies in large-scale DNA sequencing of entire human genomes.

Traditional sequencing methods are slow, laborious and expensive; the original human genome sequencing project cost more than US\$3 billion and took more than a decade to finish [16]. Dramatic technological advances have transformed the ability to analyze our genetic make-up at single nucleotide resolution and commercialization

*Correspondence: simon.fisher@mpi.nl

¹Max Planck Institute for Psycholinguistics, 6525 XD Nijmegen, The Netherlands
Full list of author information is available at the end of the article

of these 'next-generation' platforms is growing fast. At the time of writing, a human genome can be entirely sequenced in a matter of days for only a few thousand dollars, and costs continue to fall at a remarkable rate. Nevertheless, excitement over the enormous potential of the new technologies must be tempered by acknowledging the associated analytical challenges. Already, our capacity to rapidly generate large swathes of sequence data from many individuals outstrips our capacity to infer the underlying biology of a trait using such information.

Here, we begin by summarizing approaches previously applied to identify and study the first genes implicated in speech and language disorders (Table 1). We go on to discuss the promise of next-generation sequencing (NGS) for uncovering the key genomic changes that affect our speech and language abilities, not only in relevant disorders, but also in the general population. We argue that it is essential to be able to assess the functional significance of identified variants if we are to understand their biological impact and elucidate their contributions to the human traits of interest. The success of such efforts will depend on synergies between diverse research techniques, including bioinformatics and experimental analyses using model systems, as well as integration of human genome sequences and functional gene network datasets (Figure 1).

Gene mapping in speech and language disorders

Speech apraxia

The first gene to be clearly implicated in a speech and language disorder was *FOXP2*. Disruptions of this gene cause a monogenic form of developmental verbal dyspraxia (DVD), also known as childhood apraxia of speech (CAS) [17], characterized by problems with the learning and execution of coordinated movement sequences of the mouth, tongue, lips and soft palate [18,19]. *FOXP2* was discovered through molecular studies of a large three-generational pedigree (the KE family) in which half the members have CAS, accompanied by wide-ranging deficits in both oral and written language, affecting not only production but also comprehension [17]. Linkage mapping in this family identified a region on chromosome 7q31 that co-segregated perfectly with the disorder [20]. An unrelated child with similar speech and language deficits was found to carry a *de novo* balanced translocation involving the same interval, which directly interrupted the coding region of a novel gene, *FOXP2* [17,21]. Screening of *FOXP2* in the KE family revealed that all affected members had inherited a heterozygous point mutation yielding an amino acid substitution at a key residue of the encoded protein [17]. Subsequent studies identified additional etiological *FOXP2* variants (nonsense mutations, translocations, deletions) in

individuals and families with speech and language problems, typically including CAS as a core feature (reviewed by Fisher and Scharff [22]). Although etiological mutations of *FOXP2* are rare [23,24], the gene provides a valuable molecular window into neurogenetic mechanisms contributing to human spoken language, as detailed elsewhere in this article.

Beyond *FOXP2*, additional loci that may contribute to CAS have emerged from cases of chromosomal abnormalities, identified using cytogenetic screening and/or comparative genomic hybridization (CGH). One report described a family in which three affected siblings all carry an unbalanced 4q;16q translocation [25]. Another study defined a small region on 12p13.3, containing the *ELKS/ERC1* gene, commonly deleted in nine unrelated patients with delayed speech development, most of whom had a formal diagnosis of CAS [26]. Interestingly, a key isoform encoded by *ELKS/ERC1* appears to be expressed specifically in the brain, where it binds to RIM proteins. In neurons, RIMs act within the presynaptic active zone, a site that integrates synaptic vesicle exo/endocytosis with intracellular signaling in the nerve terminal [27]. Certain copy number variant (CNV) syndromes with complex variable phenotypes have been linked to increased risk of CAS, including 16p11.2 microdeletions [28,29] and 7q11.23 microduplications [30]. The rare metabolic disorder, galactosemia, is also associated with elevated incidence of CAS [31].

Specific language impairment

When a child is delayed or impaired in acquiring language, without any obvious physical or neurological cause (cleft lip/palate, intellectual disability (ID), autism, deafness, and so on) he or she is usually diagnosed with specific language impairment (SLI). Since it is defined using exclusionary criteria, SLI encompasses a range of different cognitive and behavioral profiles. The most common forms involve deficits in expressive language, either in isolation or accompanied by receptive problems.

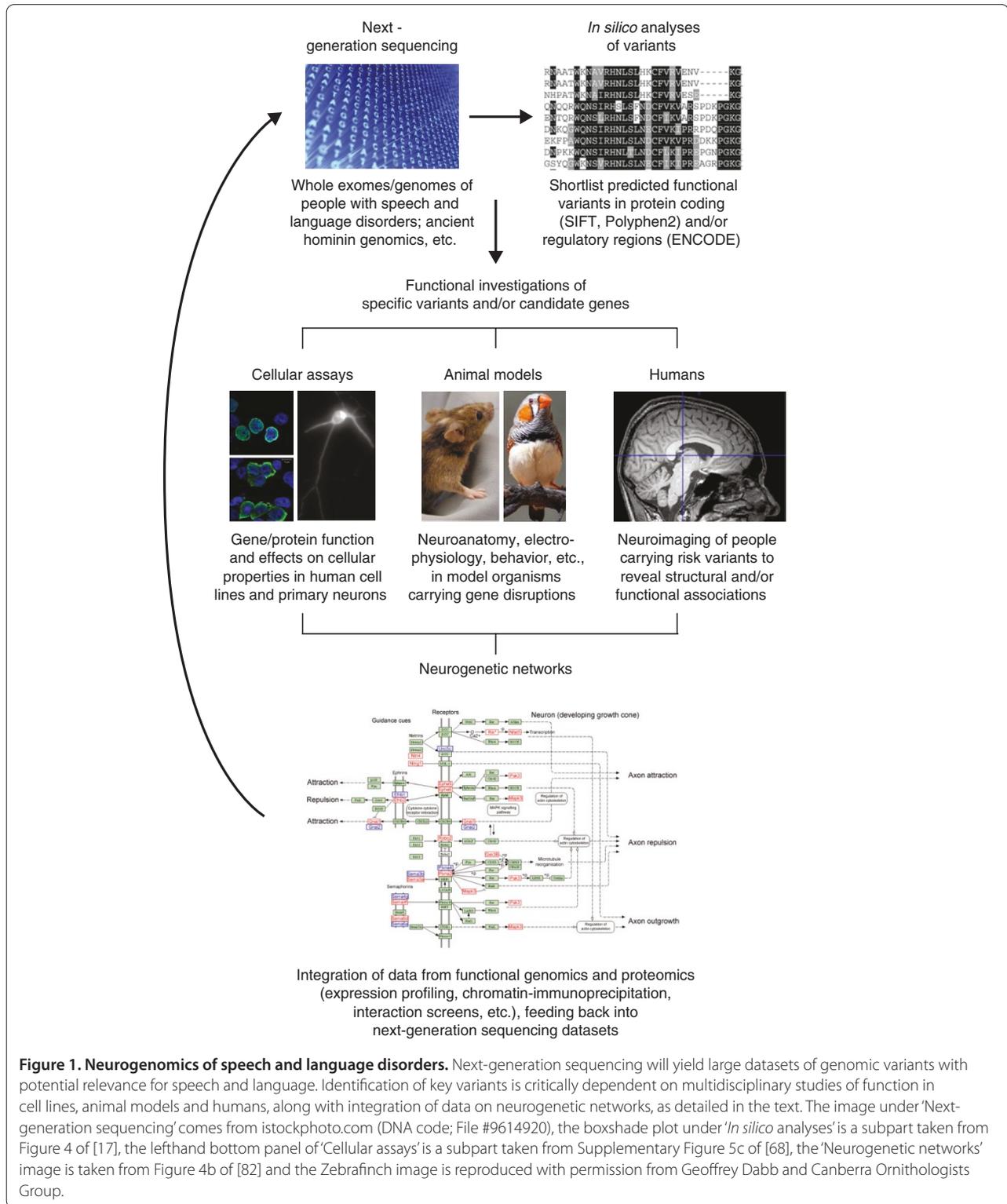
The estimated prevalence of SLI is up to 7% in kindergarten children [32] and it shows familial clustering; twin studies consistently indicate high heritability [10,11,33]. In contrast to the rare cases of monogenic CAS discussed above, typical forms of SLI have a complex multifactorial basis [34]. Genome-wide linkage mapping in families with SLI have suggested the existence of multiple risk loci, on chromosomes 16q and 19q [35-38], as well as 2p and 13q [39,40]. Targeted analysis of 16q identified variants in two genes, *ATP2C2* and *CMIP*, associated with deficits on a non-word repetition task, considered to be an index of impaired phonological short-term memory [15,41]. The *ATP2C2* gene encodes a single subunit integral membrane P-type ATPase that catalyzes the ATP-driven transport of cytosolic calcium

Table 1. Neurogenomics of speech and language: summary of key genes discussed in the article

Chromosome region	Candidate gene	Protein function	Neurodevelopmental disorders	Functional studies of specific risk variants
3p14	<i>FOXP1</i>	Forkhead-box transcriptional repressor [78]; can form heterodimers with FOXP2 protein	Rare point mutations, deletions and translocations reported in intellectual disability (ID) and autism spectrum disorder (ASD), accompanied by severe speech and language problems [68,100-102]	Cell-based analyses support a two-hit mechanism in a severely affected ASD case with two rare coding mutations; one in <i>FOXP1</i> , the other in <i>CNTNAP2</i> , disturbing a shared functional pathway [68]; reporter gene assays described for another FOXP1 variant, implicated in ID and ASD [100]
7q31	<i>FOXP2</i>	Forkhead-box transcriptional repressor [77]; can form heterodimers with FOXP1 protein	Rare point mutations, deletions and translocations reported in families and cases of developmental verbal dyspraxia (DVD)/childhood apraxia of speech (CAS) [17,22,24]	Known etiological point mutations disrupt function in cellular models [77,83,84] and mutant mice [82,117-120]; the latter studies suggested effects of risk variants on neurite outgrowth [82], neural plasticity and acquisition of motor skills [117,118], and auditory-motor association learning [120]
7q35	<i>CNTNAP2</i>	Transmembrane scaffolding protein; member of neurexin superfamily; clusters K ⁺ channels at nodes of Ranvier; implicated in neuronal migration, dendritic arborization and spine development [86-88]	Homozygous loss-of-function mutations cause cortical dysplasia with focal epilepsy [88]; common non-coding variants associated with various neuro-developmental disorders, for example, ASD [89-91], specific language impairment (SLI) [44], selective mutism [93]; also reports of links to schizophrenia [92] and Tourette syndrome [94]	Coding variants identified in ASD show impaired cellular trafficking [140]; neuroimaging genetics has suggested that common non-coding risk alleles have effects on brain structure/function in the general population (for example, [133-135]), although sample sizes were small and findings have been inconsistent between studies
12p13	<i>ELKS/ERC1</i>	Member of family of RIM-binding proteins; RIMs are active zone proteins that regulate neurotransmitter release [27]	Rare deletions reported in cases of DVD/CAS [26]	None reported to date
12q23	<i>GNPTAB</i>	Alpha and beta subunits of GlcNAc-phosphotransferase; catalyzes addition of mannose 6-phosphate tag to hydrolytic enzymes, allowing lysosomal targeting	Rare coding variants reported in cases of persistent stuttering [57]	None reported to date
16p13	<i>GNPTG</i>	Gamma subunits of GlcNAc-phosphotransferase (see above)	Rare coding variants reported in cases of persistent stuttering [57]	None reported to date
16p13	<i>NAGPA</i>	'Uncovering enzyme'; catalyzes second step in tagging hydrolytic enzymes for lysosomal targeting [103]	Rare coding variants reported in cases of persistent stuttering [57]	Coding mutations found to affect enzymatic activity, protein folding and proteasomal degradation in cell-based assays [103]
16q23	<i>CMIP</i>	Cytoskeletal adaptor protein; interacts with filamin A and RelA [43]	Common non-coding variants associated with non-word repetition deficits in families with SLI [41]	None reported to date
16q24	<i>ATP2C2</i>	Integral membrane P-type ATPase; catalyzes Ca ²⁺ /Mn ²⁺ transport into Golgi lumen [42]	Common non-coding variants associated with non-word repetition deficits in families with SLI [41]	None reported to date
18q12	<i>SETBP1</i>	Interacts with SET, an oncogene involved in DNA replication [64-66]	Haploinsufficiency reported in cases of expressive speech/language impairment (for example, [65]); dominant gain-of-function point mutations cause a distinct reproductively lethal disorder, Schinzel-Giedion syndrome [64]	None reported to date

and manganese into the Golgi lumen [42]. This cellular role makes it a plausible candidate for SLI susceptibility, since intracellular calcium levels are intimately linked to

multiple diverse aspects of neuronal function, ranging from migration to plasticity, while manganese dysregulation has been linked to neurodegenerative phenotypes.



The product of *CMIP* contains pleckstrin homology and leucine-rich repeat domains, and is hypothesized to be an adaptor protein of the actin cytoskeleton, interacting with filamin A and RelA (an NF-kappaB subunit) [43].

Although little is known about *CMIP* at this stage, it is again a credible candidate for involvement in nervous system function, since cytoskeletal reorganization makes essential contributions to processes like neuronal

migration and synapse formation/modification. Other candidate genes (such as *CNTNAP2*) have been implicated in SLI susceptibility through functional approaches [44], as highlighted elsewhere in this article.

Studies of isolated founder populations may also help pinpoint new genes contributing to language disorders. A notable example is Robinson Crusoe Island - an island of 633 residents lying west of Chile, South America - which was most recently colonized in the late 19th century [45]. Thirty-five percent of the colonizing children satisfy criteria for a diagnosis of SLI, substantially higher than the 4% prevalence rate for mainland Chile [45]. Initial molecular investigations identified several genomic regions of interest (on chromosomes 6, 7, 12, 13 and 17), but no specific risk genes have yet been discovered [46].

SLI has connections with another heritable neurodevelopmental trait, dyslexia, defined as specific significant impairments in reading and/or spelling that are not attributed to intelligence, visual acuity problems or inadequate learning opportunities. Although they do not display overt difficulties with speech or language, people with dyslexia often have subtle underlying deficits with aspects of linguistic processing [47]. Thus, genetic studies of dyslexia may be informative for understanding language pathways. We do not have space to discuss this here, and refer readers to other recent reviews [48,49].

Stuttering

Stuttering is a neurodevelopmental disorder that disturbs the flow of speech [50]. People who stutter are affected by uncontrollable repetitions and prolongations of syllables, and by involuntary silent pauses while speaking; these difficulties begin in childhood, persisting in about 20% of case referrals [51]. Most people who suffer from persistent stuttering nevertheless display normal linguistic proficiency [52]. Stuttering is thought to have a strong genetic basis [53]. Thus far, most genome-wide investigations of persistent familial stuttering have revealed only suggestive evidence of linkage, with loci distributed across at least ten chromosomes, and little overlap between different studies, indicating that this is a complex multifactorial trait [53-55].

One of the few reports of significant linkage focused on 46 consanguineous families from Pakistan, and highlighted chromosome 12q as a site of interest [56]. Subsequent analyses of the largest family from that study found that most affected relatives carried a coding variant in the 12q23.2 gene *GNPTAB*, which encodes two subunits of GlcNAc-phosphotransferase (GNPT) [57]. This putative risk variant (Q1200K), which altered a conserved residue of the protein, was identified in a number of other Pakistani cases, at higher frequency than Pakistani controls. GNPT is involved in addition of a mannose 6-phosphate tag to hydrolytic enzymes,

allowing them to be targeted to lysosomes. Further screening of *GNPTAB*, as well as *GNPTG* and *NAGPA*, two closely related genes in this metabolic pathway, identified several different coding variants that were only present in cases and not controls [57]. The proposed risk variants are rare even among people who stutter, so it is likely that there are other unknown genes involved in stuttering.

The next generation: uncovering novel risk variants

While it is clear that exciting progress has been made, many of the genetic risk factors underlying speech and language disorders and/or normal linguistic variation remain to be discovered. At the time of writing, no study had yet reported the use of NGS methodologies to specifically investigate language-related traits. However, the advent of NGS has transformed the identification of genetic variants in other important neurodevelopmental phenotypes that co-occur with language deficits, such as ID and autism spectrum disorders (ASDs). Thus far, most such research has focused on sequencing protein-coding regions of the genome (the exome) to detect *de novo* variants in rare and common forms of these disorders [58-60]. Since *de novo* mutations have highly deleterious effects and are subject to strong negative selection, it is hypothesized that they might be important explanations of sporadic occurrences of disorder.

Whole-exome sequencing first proved effective in detecting causal *de novo* variants in rare reproductively lethal neurodevelopmental disorders, such as Kabuki syndrome [61], Bohring-Opitz syndrome [62] and KBG syndrome [63]. The study that pioneered this approach assessed 13 cases of Schinzel-Giedion syndrome, which is characterized by severe ID and typical facial features, and revealed *de novo* gain-of-function mutations independently occurring in a single gene, *SETBP1* [64]. Interestingly, haploinsufficiency of *SETBP1* has been identified in some cases of expressive speech impairment [65]. *SETBP1* encodes a widely expressed nuclear protein that interacts with SET, an oncogene involved in DNA replication. Recent studies have shown that SET binding protein 1 (SETBP1) also includes three highly conserved AT-hooks (motifs that bind AT-rich DNA in a non-sequence-specific manner) and that it can act as a transcription factor, directly activating targets such as *Hoxa9* and *Hoxa10* [66]. Functional links between SETBP1 and brain development have yet to be explored.

NGS techniques are also shedding light on the roles of *de novo* changes in common non-syndromic disorders [59]. A pilot study of whole-exome sequencing in sporadic cases of non-syndromic ID and their parents (parent-child trios) reported nine non-synonymous *de novo* mutations in different genes in seven of ten probands [67]. Since then, multiple investigations have employed

similar approaches to screen trios or quads (trio plus unaffected siblings), including four large-scale whole-exome sequencing efforts across about 1,000 ASD families [68-72] (reviewed by Buxbaum *et al.* [60]). One conclusion of this work was that the rate of *de novo* mutations was higher in ASD probands than controls, and it pointed to six genes of particular interest that had recurrent loss-of-function mutations.

A major advantage of focusing on *de novo* mutations is that it dramatically reduces the search space for potential causative variants; it is estimated that an average of approximately one *de novo* coding variant arises per genome per generation [59]. Interpretation of NGS data becomes more difficult when the search criteria are broadened to encompass all potential etiological coding variants that a proband carries, and it is even more challenging if one also considers non-exonic variations throughout the entire genome. It is not currently known if the genetic architecture underlying specific speech and language disorders includes a significant role for *de novo* mutations. Thus, it will be important to develop alternative study designs and analytic strategies (for example, Yu *et al.* [73] and Lim *et al.* [74]) for pinpointing causative mutations in NGS data from cases and families with language impairments.

Bridging the gap from genetic variants to biology

In the near future, NGS methods will become standard tools in molecular studies of speech and language disorders. As noted above, gene discovery strategies will need to move beyond the *de novo* paradigms that have been so successful for ID and ASD. Researchers will be faced with the major challenge of discerning which of the many plausibly causal variants carried by each affected person are physiologically relevant to their speech and/or language impairments. Fortunately, distinct fields combining computational and experimental methods can help ascertain the biological roles of detected variants and ultimately highlight genes important for our unique capacity for spoken language.

When focusing on protein-coding sequences, after initial filtering of identified variants from NGS data, it is possible to use predictive algorithms such as SIFT [75] and PolyPhen2 [76] to flag the most promising mutations for subsequent analyses. Computational methods such as these use known information on protein sequence and evolutionary history to rank them as benign, possibly damaging or probably damaging. Nonetheless, as cellular pathways harbor some degree of redundancy, not all loss-of-function mutations will contribute to a given disorder and such predictions should be treated with caution. For example, sequencing of *FOXP2* in a cohort of CAS/DVD cases revealed a non-synonymous substitution near the N-terminus of the protein (Q17L) in one of the probands

[24], a variant that is predicted to be damaging by both SIFT and PolyPhen2. However, follow-up functional experiments of the Q17L substitution using cell models did not find adverse effects on protein characteristics, in contrast to observations for other proband mutations [77]. Together with the fact that the Q17L proband has an affected sibling who does not carry the substitution, it seems unlikely that this particular change is etiological. Thus, although bioinformatic approaches help narrow down the list of variants from ongoing high-throughput genetic screens of speech and language phenotypes, experimental analyses in model systems are often crucial for determining causality, as well as offering deeper insights into mechanisms.

The value of functional approaches is particularly apparent from studies of how *FOXP2* mutations lead to speech and language disorder [22]. *FOXP2* encodes a forkhead-box transcription factor. Following homo- or hetero-dimerization with other forkhead box P (FOXP) family members [78], the protein binds DNA and represses transcription of its target genes [79]. Human neuron-like cells have been used to assess two different mutant FOXP2 proteins that co-segregate with disorder in CAS/DVD families: pFOXP2.R553H [17] and pFOXP2.R328X [24]. The functional assays demonstrated that these mutations severely disrupt nuclear localization, DNA-binding ability and transactivation potential of the protein [77]. Investigations into downstream targets of FOXP2 highlighted several neuronal pathways that it regulates. Independent high-throughput studies of promoter occupancy in cells and human fetal brain reported that FOXP2 directly regulates genes involved in neurite outgrowth, synaptic plasticity and axon guidance [80, 81]. More recently, following genome-wide analyses of neural targets *in vivo* in mouse models, it has been shown that *Foxp2* mutations can alter neurite outgrowth and branching in primary neurons [82].

A subset of FOXP2 targets are implicated in neuro-developmental disorders that often co-occur with language deficits, such as the sushi repeat-containing protein X-linked 2 (SRPX2)-plasminogen activator receptor, urokinase-type (uPAR) complex in epilepsy and speech apraxia [83], *DISC1* in schizophrenia [84] and *MET* in ASD [85]. The most rigorously studied FOXP2 target is *CNTNAP2*, encoding contactin-associated protein-like 2 (CASPR2), a transmembrane scaffolding protein that clusters K⁺ channels in myelinated axons [86]. CASPR2 is a member of the neurexin superfamily and, in addition to its role in mature neurons, it has been implicated in neuronal migration, dendritic arborization and spine development [87]. Homozygous loss-of-function *CNTNAP2* mutations cause infant-onset epilepsy, learning deficits and language regression [88]. FOXP2 binds directly within the first intron of *CNTNAP2*

and is able to downregulate its expression [44]. Association analyses of quantitative phenotype data in 184 small SLI families identified a cluster of common intronic SNPs in *CNTNAP2* that correlated significantly with reduced performance on linguistic tests, most strongly for the non-word repetition endophenotype [44]. The identity of the precise functional variant(s) in this region is not yet determined, but it is hypothesized that they affect the way that *CNTNAP2* is regulated. Rare and common *CNTNAP2* variants have also been implicated independently in ASDs [89-91], consistent with prior hypotheses that SLI and ASDs may involve some degree of shared genetic etiology. Beyond SLI, ASD and epilepsy, contributions of *CNTNAP2* have been suggested for a range of other neurodevelopmental phenotypes, including schizophrenia [92], selective mutism [93] and Tourette syndrome [94].

A recent study of sporadic ASD demonstrates how the combination of NGS screens with functional experiments can shed light on language-related gene networks [68]. Whole-exome sequencing of parent-child trios identified a *de novo* frameshift mutation in an ASD proband, introducing a premature stop codon in *FOXP1* [68]. The child was severely affected, with regression and language delays. *FOXP1* is the most closely related gene to *FOXP2* in the human genome and they can act synergistically to regulate shared targets in regions of co-expression [78,95,96]. Remarkably, the proband with the *FOXP1* mutation also carried an extremely rare *CNTNAP2* missense variant, inherited from his unaffected mother [68]. In cell-based functional analyses, the aberrant *FOXP1* protein mislocalized to the cytoplasm and lost its transcriptional repressor properties; expression of the mutant *FOXP1* isoform in cells elevated *CNTNAP2* levels, unlike wild-type *FOXP1* [68]. These data were consistent with a two-hit mechanism in which abnormal *FOXP1* results in higher *CNTNAP2* levels, amplifying any potentially deleterious effects of the missense *CNTNAP2* variant of the proband [68]. Similar findings regarding multiple-hit mechanisms have emerged from independent studies of ASDs and other neurodevelopmental syndromes (for example, Leblond *et al.* [97]), suggesting that this may be an important model for genetic etiology of such disorders [98].

Previous screening of 49 children diagnosed with CAS/DVD did not detect any obviously etiological *FOXP1* mutations [99]. However, studies of patients with mild to moderate ID and language impairment have detected rare *de novo* deletions and a nonsense *FOXP1* variant [100,101]. High-throughput sequencing of balanced chromosomal abnormalities in neurodevelopmental disorders identified disruptions at the *FOXP1* locus [102].

There has been little reported to date on functional analyses of other genes (such as *ATP2C2* and *CMIP*)

associated with speech and language disorders, in part because no protein-coding variants have been pinpointed. As noted above, some cases of persistent stuttering carry coding variants in genes (*GNPTAB*, *GNPTG* and *NAGPA*) involved in lysosomal targeting of hydrolase enzymes. Interestingly, loss-of-function mutations of this pathway cause mucopolysaccharidosis disorders, which involve severe abnormalities affecting multiple systems, including skeletal, respiratory and cardiovascular tissues. Cell-based assays were recently used to analyze Mannose 6-phosphate-uncovering enzyme variants found in people who stutter, and were reported to yield incorrect protein folding, decreased enzymatic activity and degradation by the proteasome [103].

It is not always feasible to carry out experimental assessments of putative risk variants. The nature of assessment is highly dependent on the type of gene product; it is difficult to test protein function if there are no known measurable properties. In contrast to NGS technologies, functional experiments typically remain high cost, time-consuming and laborious, and are less amenable for up-scaling. Nevertheless, as NGS reveals additional variants potentially implicated in language impairments and other neurodevelopmental traits, we will inevitably need access to high-throughput techniques for simultaneous mutation testing to define disease-causing variants across the genome [104]. Indeed, several multiplex approaches for characterizing the functional effects of genetic variation in proteins [105], mammalian regulatory elements [106,107] and RNA [108] have recently been developed. More and more emphasis will be placed on possible functional variants that lie outside protein-coding regions. Various efforts are underway to facilitate this transition, most notably the ENCODE project, which aims to characterize all functional elements at a genome-wide scale, including non-coding RNA and *cis*-regulatory elements [109]. RegulomeDB is of particular interest, as it combines data from the ENCODE project, GEO and published literature into a single, integrated database that can be used to query the functional significance of variants in both coding and non-coding regions of the genome [110].

Integrating data networks

Beyond establishing causality, functional characterization of candidate risk variants in model organisms may also help highlight pathways implicated in the origins and bases of language. For example, studies of *FOXP2* across different species (mouse, bird, human) have given us initial clues into neurogenetic networks facilitating human spoken language [22,111]. *FOXP2* expression is enriched in several brain areas, including the basal ganglia, deep cortical layers, thalamus and cerebellum [112], some of which display subtle structural and functional abnormalities

in people carrying *FOXP2* mutations [19,112-114]. From an evolutionary perspective, this is a highly conserved gene with regard to both the amino acid sequence of the encoded protein and the neural sites where it is expressed [95,115]. These data suggest that ancestral forms of *FOXP2* were involved in important aspects of brain development long before the emergence of spoken language. There is evidence that the functions of the gene may have been modified during human evolution ([116]; also see below), but it remains clear that its roles in the human brain are built on evolutionarily ancient pathways [1].

Extensive characterization of rodent models carrying etiological *Foxp2* variants indicates roles in synaptic plasticity, motor-skill learning, and processing and integration of auditory information [117-120]. When mice are heterozygous for the mutation that causes speech problems in the human KE family, they display decreased synaptic plasticity in corticostriatal circuits and motor-skill learning deficits [117]. These mouse findings are intriguing given that affected humans have problems learning to master the rapid coordinated orofacial movements underlying speech [121]. *In vivo* electrophysiology recordings in awake-behaving mice revealed more about the impacts of *Foxp2* on corticostriatal circuitry; mice heterozygous for the KE mutation displayed higher basal striatal activity than wild-type controls, and medium spiny neurons showed aberrant negative modulation of their firing rates during motor-skill learning [118]. Separate studies used mouse models to explore whether impairments in auditory processing and auditory-motor integration might also be relevant to *FOXP2*-related disorders [119,120]. Mice carrying the KE mutation were reported to have altered auditory brainstem responses to sound, although this finding was not replicated in mice carrying a different mutation associated with speech/language problems in another family [119]. Mice carrying either etiological mutation have deficits in learning to associate auditory stimuli with motor outputs [120].

Songbirds carry their own version of *FOXP2*, referred to as *FoxP2*, and it appears to make important contributions to the functions of a striatal nucleus called Area X [122]. In zebra finches, Area X is critical for auditory-guided vocal learning, a process in which young male birds learn their song by imitating an adult tutor. Vocal learning is also a key component of human speech acquisition. *FoxP2* mRNA levels in Area X are enriched in young birds during the critical song-learning period [123] and show rapid downregulation when adult birds practice their songs outside the context of courtship [124-126]. Furthermore, selective knockdown of *FoxP2* in Area X disrupts the song-learning process [127] and alters dendritic spine density in this region [128].

Functional studies of genes implicated in language-related disorders may also give us entry points into

mechanisms involved in language function in the general population. As discussed above, variants of *CNTNAP2*, a direct target of *FOXP2*, were associated with linguistic deficits in clinically distinct neurodevelopmental disorders [44,88,89,129-131]. Subsequent studies revealed that *CNTNAP2* may contribute to language processing in healthy individuals [132-134]. The cluster of *CNTNAP2* SNPs that is associated with language phenotypes in SLI and ASDs has also been reported to correlate with assessments of early language development in general population samples [132]. Neuroimaging genetics studies of common *CNTNAP2* SNPs in healthy samples have proposed associations with functional brain measures related to language [133,134] and with altered structural connectivity patterns [135]. However, imaging genetics of language is a field that is only in its infancy; reports thus far involved small sample sizes with limited power, as well as a substantial multiple-testing burden, and results of different studies have been largely inconsistent. Additional analyses are required to elucidate how *FOXP2*, *CNTNAP2* and other language-related genes influence brain circuits at multiple levels of description - molecular, cellular, structural and functional.

Insights from ancient genomes

The reach of NGS technologies extends well beyond living species. These innovations have allowed molecular anthropologists to reconstruct large portions of nuclear genomes from extinct hominins that co-existed with our ancestors, such as Neanderthals [136] and Denisovans [137]. By comparing modern human sequences to ancient hominin genomes, as well as to our closest extant relatives, chimpanzees, it is possible to identify molecular variants that arose during human evolution, and roughly date them with regard to branches of the primate phylogenetic tree. As for other NGS projects, our capacity to generate large amounts of sequence data exceeds our ability to interpret it. So although scientists have successfully catalogued many of the DNA changes that occurred on our lineage, an extraordinary feat in itself, it is still a major challenge to determine which of these evolutionary events were relevant for the emergence of traits such as speech and language acquisition [1]. Here, success may depend on the integration of findings from evolutionary genomics with data from molecular studies of language-related disorders.

The best illustration of this approach comes again from work on the *FOXP2* gene, which was targeted for evolutionary investigations, based on its prior link to a severe speech and language disorder. Comparative primate genomics suggests that *FOXP2* probably underwent at least two interesting evolutionary events on the lineage that led to modern humans. After splitting from the chimpanzee (several million years ago) there were

changes in the coding region of the locus that yielded two amino acid substitutions in the encoded protein [138]. Although these are minor changes outside the known functional domains, when such substitutions are inserted into the endogenous *Foxp2* gene of a mouse, they have subtle detectable effects on brain structure and function, including altered connectivity and plasticity of corticostriatal circuits [116]. NGS approaches indicate that these amino acid substitutions are shared by Neanderthals [136] and Denisovans [137]. (It is worth emphasizing here that status of a single gene is not enough to determine whether or not a species can speak.) Researchers went on to identify a number of non-coding variants in intronic regions of *FOXP2* that had occurred more recently on the human lineage, after splitting from Neanderthal/Denisovan a few hundred thousand years ago [139]. One of these changes lies in a region that underwent a recent selective sweep, and alters a putative binding site for the POU class 3 homeobox 2 (POU3F2) transcription factor, such that it may have affected regulation of *FOXP2* expression; cell-based analyses are consistent with this hypothesis [139]. Thus, just like sequence-based analyses of language-related disorders, evaluation of the biological significance of interesting variants from ancient genomics requires functional studies using model systems.

Conclusion

The advent of whole genome NGS means that data generation will no longer be the limiting factor in understanding how genetic factors contribute to mechanisms underlying complex neurodevelopmental traits. Coupling NGS approaches to functional validation in model systems will facilitate network mapping and pathway investigation in speech and language disorders, and ultimately in normal linguistic development.

Abbreviations

ASD, autism spectrum disorder; CAS, childhood apraxia of speech; DVD, developmental verbal dyspraxia; FOXP, forkhead box P; ID, intellectual disability; NGS, next-generation sequencing; SLI, specific language impairment.

Competing interests

The authors declare that they have no competing interests

Author details

¹Max Planck Institute for Psycholinguistics, 6525 XD Nijmegen, The Netherlands. ²Donders Institute for Brain, Cognition and Behaviour, Radboud University, 6525 EN Nijmegen, The Netherlands.

Published: 18 April 2013

References

1. Fisher SE, Marcus GF: **The eloquent ape: genes, brains and the evolution of language.** *Nat Rev Genet* 2006, **7**:9-20.
2. Pinker S, Bloom P: **Natural language and natural selection.** *Behav Brain Sci* 1990, **13**:707-784.
3. Bishop DV: **Genes, cognition, and communication: insights from neurodevelopmental disorders.** *Ann NY Acad Sci* 2009, **1156**:1-18.

4. Bishop DV: **Overlaps between autism and language impairment: phenomimicry or shared etiology?** *Behav Genet* 2010, **40**:618-629.
5. Pal DK: **Epilepsy and neurodevelopmental disorders of language.** *Curr Opin Neurol* 2011, **24**:126-131.
6. Neils J, Aram DM: **Family history of children with developmental language disorders.** *Percept Mot Skills* 1986, **63**:655-658.
7. Tomblin JB: **Familial concentration of developmental language impairment.** *J Speech Hear Disord* 1989, **54**:287-295.
8. Barry JG, Yasin I, Bishop DV: **Heritable risk factors associated with language impairments.** *Genes Brain Behav* 2007, **6**:66-76.
9. Lewis BA, Freebairn LA, Hansen AJ, Miscimarra L, Iyengar SK, Taylor HG: **Speech and language skills of parents of children with speech sound disorders.** *Am J Speech Lang Pathol* 2007, **16**:108-118.
10. Lewis BA, Thompson LA: **A study of developmental speech and language disorders in twins.** *J Speech Hear Res* 1992, **35**:1086-1094.
11. Bishop DV, North T, Donlan C: **Genetic basis of specific language impairment: evidence from a twin study.** *Dev Med Child Neurol* 1995, **37**:56-71.
12. Bishop DV, Hayiou-Thomas ME: **Heritability of specific language impairment depends on diagnostic criteria.** *Genes Brain Behav* 2008, **7**:365-372.
13. Bishop DV: **Genetic and environmental risks for specific language impairment in children.** *Philos Trans R Soc Lond B Biol Sci* 2001, **356**:369-380.
14. Fisher SE, Lai CS, Monaco AP: **Deciphering the genetic basis of speech and language disorders.** *Annu Rev Neurosci* 2003, **26**:57-80.
15. Newbury DF, Monaco AP: **Genetic advances in the study of speech and language disorders.** *Neuron* 2010, **68**:309-320.
16. International Human Genome Sequencing Consortium: **Finishing the euchromatic sequence of the human genome.** *Nature* 2004, **431**:931-945.
17. Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP: **A forkhead-domain gene is mutated in a severe speech and language disorder.** *Nature* 2001, **413**:519-523.
18. Alcock KJ, Passingham RE, Watkins KE, Vargha-Khadem F: **Oral dyspraxia in inherited speech and language impairment and acquired dysphasia.** *Brain Lang* 2000, **75**:17-33.
19. Watkins KE, Vargha-Khadem F, Ashburner J, Passingham RE, Connelly A, Friston KJ, Frackowiak RSJ, Mishkin M, Gadian DG: **MRI analysis of an inherited speech and language disorder: structural brain abnormalities.** *Brain* 2002, **125**:465-478.
20. Fisher SE, Vargha-Khadem F, Watkins KE, Monaco AP, Pembrey ME: **Localisation of a gene implicated in a severe speech and language disorder.** *Nat Genet* 1998, **18**:168-170.
21. Lai CS, Fisher SE, Hurst JA, Levy ER, Hodgson S, Fox M, Jeremiah S, Povey S, Jamison DC, Green ED, Vargha-Khadem F, Monaco AP: **The SPCH1 region on human 7q31: genomic characterization of the critical interval and localization of translocations associated with speech and language disorder.** *Am J Hum Genet* 2000, **67**:357-368.
22. Fisher SE, Scharff C: **FOXP2 as a molecular window into speech and language.** *Trends Genet* 2009, **25**:166-177.
23. Newbury DF, Bonora E, Lamb JA, Fisher SE, Lai CS, Baird G, Jannoun L, Slonims V, Stott CM, Merricks MJ, Bolton PF, Bailey AJ, Monaco AP; International Molecular Genetic Study of Autism Consortium: **FOXP2 is not a major susceptibility gene for autism or specific language impairment.** *Am J Hum Genet* 2002, **70**:1318-1327.
24. MacDermot KD, Bonora E, Sykes N, Coupe AM, Lai CS, Vernes SC, Vargha-Khadem F, McKenzie F, Smith RL, Monaco AP, Fisher SE: **Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits.** *Am J Hum Genet* 2005, **76**:1074-1080.
25. Shriberg LD, Jakielski KJ, El-Shanti H: **Breakpoint localization using array-CGH in three siblings with an unbalanced 4q;16q translocation and childhood apraxia of speech (CAS).** *Am J Med Genet* 2008, **146A**:2227-2233.
26. Thevenon J, Callier P, Andrieux J, Delobel B, David A, Sukno S, Minot D, Mosca Anne L, Marle N, Sanlaville D, Bonnet M, Masurel-Paulet A, Levy F, Gaunt L, Farrell S, Le Caignec C, Toutain A, Carmignac V, Mugneret F, Clayton-Smith J, Thauvin-Robinet C, Faivre L: **12p13.33 microdeletion including ELKS/ERC1, a new locus associated with childhood apraxia of speech.** *Eur J Hum Genet* 2013, **21**:82-88.
27. Wang Y, Liu X, Biederer T, Sudhof TC: **A family of RIM-binding proteins regulated by alternative splicing: Implications for the genesis of synaptic active zones.** *Proc Natl Acad Sci U S A* 2002, **99**:14464-14469.
28. Newbury DF, Mari F, Sadighi Akha E, Macdermot KD, Canitano R, Monaco AP, Taylor JC, Renieri A, Fisher SE, Knight SJ: **Dual copy number variants**

- involving 16p11 and 6q22 in a case of childhood apraxia of speech and pervasive developmental disorder. *Eur J Hum Genet* 2012, **21**:361-365.
29. Raca G, Baas BS, Kirmani S, Laffin JJ, Jackson CA, Strand EA, Jakielski KJ, Shriberg LD: **Childhood Apraxia of Speech (CAS) in two patients with 16p11.2 microdeletion syndrome.** *Eur J Hum Genet* 2013, **21**:455-459.
30. Van der Aa N, Rooms L, Vandeweyer G, van den Ende J, Reyniers E, Fichera M, Romano C, Delle Chiaie B, Mortier G, Menten B, Destrée A, Maystadt I, Männik K, Kurg A, Reimand T, McMullan D, Oley C, Brueton L, Bongers EM, van Bon BW, Pfund R, Jacquemont S, Ferrarini A, Martinet D, Schrandt-Stumpel C, Stegmann AP, Frints SG, de Vries BB, Ceulemans B, Kooy RF: **Fourteen new cases contribute to the characterization of the 7q11.23 microduplication syndrome.** *Eur J Med Genet* 2009, **52**:94-100.
31. Shriberg LD, Potter NL, Strand EA: **Prevalence and phenotype of childhood apraxia of speech in youth with galactosemia.** *J Speech Lang Hear Res* 2011, **54**:487-519.
32. Tomblin JB, Records NL, Buckwalter P, Zhang X, Smith E, O'Brien M: **Prevalence of specific language impairment in kindergarten children.** *J Speech Lang Hear Res* 1997, **40**:1245-1260.
33. Tomblin JB, Buckwalter PR: **Heritability of poor language achievement among twins.** *J Speech Lang Hear Res* 1998, **41**:188-199.
34. Newbury DF, Bishop DV, Monaco AP: **Genetic influences on language impairment and phonological short-term memory.** *Trends Cogn Sci* 2005, **9**:528-534.
35. SLI Consortium: **A genome-wide scan identifies two novel loci involved in specific language impairment.** *Am J Hum Genet* 2002, **70**:384-398.
36. SLI Consortium: **Highly significant linkage to the SLI1 locus in an expanded sample of individuals affected by specific language impairment.** *Am J Hum Genet* 2004, **74**:1225-1238.
37. Monaco AP: **Multivariate linkage analysis of specific language impairment (SLI).** *Ann Hum Genet* 2007, **71**:660-673.
38. Falcaro M, Pickles A, Newbury DF, Addis L, Banfield E, Fisher SE, Monaco AP, Simkin Z, Conti-Ramsden G: **Genetic and phenotypic effects of phonological short-term memory and grammatical morphology in specific language impairment.** *Genes Brain Behav* 2008, **7**:393-402.
39. Bartlett CW, Flax JF, Logue MW, Vieland VJ, Bassett AS, Tallal P, Brzustowicz LM: **A major susceptibility locus for specific language impairment is located on 13q21.** *Am J Hum Genet* 2002, **71**:45-55.
40. Bartlett CW, Flax JF, Logue MW, Smith BJ, Vieland VJ, Tallal P, Brzustowicz LM: **Examination of potential overlap in autism and language loci on chromosomes 2, 7, and 13 in two independent samples ascertained for specific language impairment.** *Hum Hered* 2004, **57**:10-20.
41. Newbury DF, Winchester L, Addis L, Paracchini S, Buckingham LL, Clark A, Cohen W, Cowie H, Dworzynski K, Everitt A, Goodyer IM, Hennessy E, Kindley AD, Miller LL, Nasir J, O'Hare A, Shaw D, Simkin Z, Simonoff E, Slonims V, Watson J, Ragoussis J, Fisher SE, Seckl JR, Helms PJ, Bolton PF, Pickles A, Conti-Ramsden G, Baird G, Bishop DV, Monaco AP: **CMIP and ATP2C2 modulate phonological short-term memory in language impairment.** *Am J Hum Genet* 2009, **85**:264-272.
42. Xiang M, Mohamalawari D, Rao R: **A novel isoform of the secretory pathway Ca2+,Mn(2+)-ATPase, hSPCA2, has unusual properties and is expressed in the brain.** *J Biol Chem* 2005, **280**:11608-11614.
43. Kamal M, Valanciate A, Dahan K, Ory V, Pawlak A, Lang P, Guellaen G, Sahali D: **C-mip interacts physically with RelA and inhibits nuclear factor kappa B activity.** *Mol Immunol* 2009, **46**:991-998.
44. Vernes SC, Newbury DF, Abrahams BS, Winchester L, Nicod J, Groszer M, Alarcón M, Oliver PL, Davies KE, Geschwind DH, Monaco AP, Fisher SE: **A functional genetic link between distinct developmental language disorders.** *N Engl J Med* 2008, **359**:2337-2345.
45. Villanueva P, de Barbieri Z, Palomino HM, Palomino H: **[High prevalence of specific language impairment in Robinson Crusoe Island. A possible founder effect].** *Rev Med Chil* 2008, **136**:186-192.
46. Villanueva P, Newbury DF, Jara L, De Barbieri Z, Mirza G, Palomino HM, Fernandez MA, Cazier JB, Monaco AP, Palomino H: **Genome-wide analysis of genetic susceptibility to language impairment in an isolated Chilean population.** *Eur J Hum Genet* 2011, **19**:687-695.
47. Fisher SE, DeFries JC: **Developmental dyslexia: genetic dissection of a complex cognitive trait.** *Nat Rev Neurosci* 2002, **3**:767-780.
48. Raskind WH, Peter B, Richards T, Eckert MM, Berninger VW: **The genetics of reading disabilities: from phenotypes to candidate genes.** *Front Psychol* 2012, **3**:601.
49. Graham SA, Fisher SE: **Decoding the genetics of speech and language.** *Curr Opin Neurobiol* 2013, **23**:43-51.
50. Prasse JE, Kikano GE: **Stuttering: an overview.** *Am Fam Physician* 2008, **77**:1271-1276.
51. Drayna D, Kang C: **Genetic approaches to understanding the causes of stuttering.** *J Neurodev Disord* 2011, **3**:374-380.
52. Nippold MA: **Stuttering and language ability in children: questioning the connection.** *Am J Speech Lang Pathol* 2012, **21**:183-196.
53. Kang C, Drayna D: **Genetics of speech and language disorders.** *Annu Rev Genomics Hum Genet* 2011, **12**:145-164.
54. Raza MH, Amjad R, Riazuddin S, Drayna D: **Studies in a consanguineous family reveal a novel locus for stuttering on chromosome 16q.** *Hum Genet* 2012, **131**:311-313.
55. Raza MH, Gertz EM, Mundorff J, Lukong J, Kuster J, Schaffer AA, Drayna D: **Linkage analysis of a large African family segregating stuttering suggests polygenic inheritance.** *Hum Genet* 2012, **132**:385-396.
56. Riaz N, Steinberg S, Ahmad J, Pluzhnikov A, Riazuddin S, Cox NJ, Drayna D: **Genomewide significant linkage to stuttering on chromosome 12.** *Am J Hum Genet* 2005, **76**:647-651.
57. Kang C, Riazuddin S, Mundorff J, Krasnewich D, Friedman P, Mullikin JC, Drayna D: **Mutations in the lysosomal enzyme-targeting pathway and persistent stuttering.** *N Engl J Med* 2010, **362**:677-685.
58. Gilissen C, Hoischen A, Brunner HG, Veltman JA: **Unlocking Mendelian disease using exome sequencing.** *Genome Biol* 2011, **12**:228.
59. Veltman JA, Brunner HG: **De novo mutations in human genetic disease.** *Nat Rev Genet* 2012, **13**:565-575.
60. Buxbaum JD, Daly MJ, Devlin B, Lehner T, Roeder K, State MW: **The autism sequencing consortium: large-scale, high-throughput sequencing in autism spectrum disorders.** *Neuron* 2012, **76**:1052-1056.
61. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J: **Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome.** *Nat Genet* 2010, **42**:790-793.
62. Hoischen A, van Bon BW, Rodriguez-Santiago B, Gilissen C, Vissers LE, de Vries P, Janssen I, van Lier B, Hastings R, Smithson SF, Newbury-Ecob R, Kjaergaard S, Goodship J, McGowan R, Bartholdi D, Rauch A, Peippo M, Cobben JM, Wiczorek D, Gillessen-Kaesbach G, Veltman JA, Brunner HG, de Vries BB: **De novo nonsense mutations in ASXL1 cause Bohring-Opitz syndrome.** *Nat Genet* 2011, **43**:729-731.
63. Sirmaci A, Spiliopoulos M, Brancati F, Powell E, Duman D, Abrams A, Bademci G, Agolini E, Guo S, Konuk B, Kavaz A, Blanton S, Digilio MC, Dallapiccola B, Young J, Zuchner S, Tekin M: **Mutations in ANKRD11 cause KBG syndrome, characterized by intellectual disability, skeletal malformations, and macrodontia.** *Am J Hum Genet* 2011, **89**:289-294.
64. Hoischen A, van Bon BW, Gilissen C, Arts P, van Lier B, Steehouwer M, de Vries P, de Reuver R, Wieskamp N, Mortier G, Devriendt K, Amorim MZ, Revencu N, Kidd A, Barbosa M, Turner A, Smith J, Oley C, Henderson A, Hayes IM, Thompson EM, Brunner HG, de Vries BB, Veltman JA: **De novo mutations of SETBP1 cause Schinzel-Giedion syndrome.** *Nat Genet* 2010, **42**:483-485.
65. Marsaglia G, Scordo MR, Pescucci C, Nannetti G, Biagini E, Scandurra V, Gerundino F, Magi A, Benelli M, Torricelli F: **372 kb Microdeletion in 18q12.3 causing SETBP1 haploinsufficiency associated with mild mental retardation and expressive speech impairment.** *Eur J Med Genet* 2012, **55**:216-221.
66. Oakley K, Han Y, Vishwakarma BA, Chu S, Bhatia R, Gudmundsson KO, Keller J, Chen X, Vasko V, Jenkins NA, Copeland NG, Du Y: **Setbp1 promotes the self-renewal of murine myeloid progenitors via activation of Hoxa9 and Hoxa10.** *Blood* 2012, **119**:6099-6108.
67. Vissers LE, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, van Lier B, Arts P, Wieskamp N, del Rosario M, van Bon BW, Hoischen A, de Vries BB, Brunner HG, Veltman JA: **A de novo paradigm for mental retardation.** *Nat Genet* 2010, **42**:1109-1112.
68. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, Karakoc E, Mackenzie AP, Ng SB, Baker C, Rieder MJ, Nickerson DA, Bernier R, Fisher SE, Shendure J, Eichler EE: **Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations.** *Nat Genet* 2011, **43**:585-589.
69. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, Levy R, Ko A, Lee C, Smith JD, Turner EH, Stanaway IB, Vernot B, Malig M, Baker C, Reilly B, Akey JM, Borenstein E, Rieder MJ, Nickerson DA, Bernier R, Shendure J, Eichler EE: **Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations.** *Nature* 2012, **485**:246-250.

70. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL, Walker MF, Ober GT, Teran NA, Song Y, El-Fishawy P, Murtha RC, Choi M, Overton JD, Bjornson RD, Carriero NJ, Meyer KA, Bilguvar K, Mane SM, Sestan N, Lifton RP, Günel M, Roeder K, Geschwind DH, Devlin B, State MW: **De novo mutations revealed by whole-exome sequencing are strongly associated with autism.** *Nature* 2012, **485**:237-241.
71. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, Lin CF, Stevens C, Wang LS, Makarov V, Polak P, Yoon S, Maguire J, Crawford EL, Campbell NG, Geller ET, Valladares O, Schafer C, Liu H, Zhao T, Cai G, Lihm J, Dannenfelser R, Jabado O, Peralta Z, Nagaswamy U, Muzny D, Reid JG, Newsham I, Wu Y, et al.: **Patterns and rates of exonic de novo mutations in autism spectrum disorders.** *Nature* 2012, **485**:242-245.
72. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, Yamrom B, Lee YH, Narzisi G, Leotta A, Kendall J, Grabowska E, Ma B, Marks S, Rodgers L, Stepansky A, Troge J, Andrews P, Bekirsky M, Pradhan K, Ghiban E, Kramer M, Parla J, Demeter R, Fulton LL, Fulton RS, Magrini VJ, Ye K, Darnell JC, Darnell RB, Mardis ER, Wilson RK, Schatz MC, McCombie WR, Wigler M: **De novo gene disruptions in children on the autistic spectrum.** *Neuron* 2012, **74**:285-299.
73. Yu TW, Chahrouh MH, Coulter ME, Jiralerspong S, Okamura-Ikeda K, Ataman B, Schmitz-Abe K, Harmin DA, Adli M, Malik AN, D'Gama AM, Lim ET, Sanders SJ, Mochida GH, Partlow JN, Sunu CM, Felie JM, Rodriguez J, Nasir RH, Ware J, Joseph RM, Hill RS, Kwan BY, Al-Saffar M, Mukaddes NM, Hashmi A, Balkhy S, Gascon GG, Hisama FM, LeClair E, et al.: **Using whole-exome sequencing to identify inherited causes of autism.** *Neuron* 2013, **77**:259-273.
74. Lim ET, Raychaudhuri S, Sanders SJ, Stevens C, Sabo A, MacArthur DG, Neale BM, Kirby A, Ruderfer DM, Fromer M, Lek M, Liu L, Flannick J, Ripke S, Nagaswamy U, Muzny D, Reid JG, Hawes A, Newsham I, Wu Y, Lewis L, Dinh H, Goss S, Wang LS, Lin CF, Valladares O, Gabriel SB, dePristo M, Altshuler DM, Purcell SM, et al.: **Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders.** *Neuron* 2013, **77**:235-242.
75. Ng PC, Henikoff S: **SIFT: Predicting amino acid changes that affect protein function.** *Nucleic Acids Res* 2003, **31**:3812-3814.
76. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR: **A method and server for predicting damaging missense mutations.** *Nat Methods* 2010, **7**:248-249.
77. Vernes SC, Nicod J, Elahi FM, Coventry JA, Kenny N, Coupe AM, Bird LE, Davies KE, Fisher SE: **Functional genetic analysis of mutations implicated in a human speech and language disorder.** *Hum Mol Genet* 2006, **15**:3154-3167.
78. Li S, Weidenfeld J, Morrissy EE: **Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions.** *Mol Cell Biol* 2004, **24**:809-822.
79. Shu W, Yang H, Zhang L, Lu MM, Morrissy EE: **Characterization of a new subfamily of winged-helix/forkhead (Fox) genes that are expressed in the lung and act as transcriptional repressors.** *J Biol Chem* 2001, **276**:27488-27497.
80. Spiteri E, Konopka G, Coppola G, Bomar J, Oldham M, Ou J, Vernes SC, Fisher SE, Ren B, Geschwind DH: **Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain.** *Am J Hum Genet* 2007, **81**:1144-1157.
81. Vernes SC, Spiteri E, Nicod J, Groszer M, Taylor JM, Davies KE, Geschwind DH, Fisher SE: **High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders.** *Am J Hum Genet* 2007, **81**:1232-1250.
82. Vernes SC, Oliver PL, Spiteri E, Lockstone HE, Puliyadi R, Taylor JM, Ho J, Mombereau C, Brewer A, Lowy E, Nicod J, Groszer M, Baban D, Sahgal N, Cazier JB, Ragoussis J, Davies KE, Geschwind DH, Fisher SE: **Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain.** *PLoS Genet* 2011, **7**:e1002145.
83. Roll P, Vernes SC, Bruneau N, Cillario J, Ponsole-Lenfant M, Massacrier A, Rudolf G, Khalife M, Hirsch E, Fisher SE, Szepeetowski P: **Molecular networks implicated in speech-related disorders: FOXP2 regulates the SRPX2/uPAR complex.** *Hum Mol Genet* 2010, **19**:4848-4860.
84. Walker RM, Hill AE, Newman AC, Hamilton G, Torrance HS, Anderson SM, Ogawa F, Derizioti P, Nicod J, Vernes SC, Fisher SE, Thomson PA, Porteous DJ, Evans KL: **The DISC1 promoter: characterization and regulation by FOXP2.** *Hum Mol Genet* 2012, **21**:2862-2872.
85. Mukamel Z, Konopka G, Wexler E, Osborn GE, Dong H, Bergman MY, Levitt P, Geschwind DH: **Regulation of MET by FOXP2, genes implicated in higher cognitive dysfunction and autism risk.** *J Neurosci* 2011, **31**:11437-11442.
86. Poliak S, Salomon D, Elhanany H, Sabanay H, Kiernan B, Pevny L, Stewart CL, Xu X, Chiu SY, Shrager P, Furley AJ, Peles E: **Juxtaparanodal clustering of Shaker-like K⁺ channels in myelinated axons depends on Caspr2 and TAG-1.** *J Cell Biol* 2003, **162**:1149-1160.
87. Anderson GR, Galfin T, Xu W, Aoto J, Malenka RC, Sudhof TC: **Candidate autism gene screen identifies critical role for cell-adhesion molecule CASPR2 in dendritic arborization and spine development.** *Proc Natl Acad Sci U S A* 2012, **109**:18120-18125.
88. Strauss KA, Puffenberger EG, Huentelman MJ, Gottlieb S, Dobrin SE, Parod JM, Stephan DA, Morton DH: **Recessive symptomatic focal epilepsy and mutant contactin-associated protein-like 2.** *N Engl J Med* 2006, **354**:1370-1377.
89. Alarcón M, Abrahams BS, Stone JL, Duvall JA, Perederiy JV, Bomar JM, Sebat J, Wigler M, Martin CL, Ledbetter DH, Nelson SF, Cantor RM, Geschwind DH: **Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene.** *Am J Hum Genet* 2008, **82**:150-159.
90. Arking DE, Cutler DJ, Brune CW, Teslovich TM, West K, Ikeda M, Rea A, Guy M, Lin S, Cook EH, Chakravarti A: **A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism.** *Am J Hum Genet* 2008, **82**:160-164.
91. Bakkaloglu B, O'Roak BJ, Louvi A, Gupta AR, Abelson JF, Morgan TM, Chawarska K, Klin A, Ercan-Sencicek AG, Stillman AA, Tanriover G, Abrahams BS, Duvall JA, Robbins EM, Geschwind DH, Biederer T, Gunel M, Lifton RP, State MW: **Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders.** *Am J Hum Genet* 2008, **82**:165-173.
92. Friedman JI, Vrijenhoek T, Markx S, Janssen IM, van der Vliet WA, Faas BH, Knoers NV, Cahn W, Kahn RS, Edelman L, Davis KL, Silverman JM, Brunner HG, van Kessel AG, Wijmenga C, Ophoff RA, Veltman JA: **CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy.** *Mol Psychiatry* 2008, **13**:261-266.
93. Stein MB, Yang BZ, Chavira DA, Hitchcock CA, Sung SC, Shipon-Blum E, Gelernter J: **A common genetic variant in the neurexin superfamily member CNTNAP2 is associated with increased risk for selective mutism and social anxiety-related traits.** *Biol Psychiatry* 2011, **69**:825-831.
94. Verkerk AJ, Mathews CA, Joosse M, Eussen BH, Heutink P, Oostra BA: **CNTNAP2 is disrupted in a family with Gilles de la Tourette syndrome and obsessive compulsive disorder.** *Genomics* 2003, **82**:1-9.
95. Teramitsu I, Kudo LC, London SE, Geschwind DH, White SA: **Parallel FoxP1 and FoxP2 expression in songbird and human brain predicts functional interaction.** *J Neurosci* 2004, **24**:3152-3163.
96. Shu W, Lu MM, Zhang Y, Tucker PW, Zhou D, Morrissy EE: **Foxp2 and Foxp1 cooperatively regulate lung and esophagus development.** *Development* 2007, **134**:1991-2000.
97. Leblond CS, Heinrich J, Delorme R, Proepper C, Betancur C, Huguet G, Konyukh M, Chaste P, Ey E, Rastam M, Anckarsäter H, Nygren G, Gillberg IC, Melke J, Toro R, Regnault B, Fauchereau F, Mercati O, Lemière N, Skuse D, Poot M, Holt R, Monaco AP, Järvelä I, Kantojärvi K, Vanhala R, Curran S, Collier DA, Bolton P, Chiochetti A, et al.: **Genetic and functional analyses of SHANK2 mutations suggest a multiple hit model of autism spectrum disorders.** *PLoS Genet* 2012, **8**:e1002521.
98. Ruzzo EK, Pappas AL, Goldstein DB: **Modifier genetics in neuropsychiatric disease: challenges and opportunities.** *Genome Biol* 2012, **13**:150.
99. Vernes SC, MacDermot KD, Monaco AP, Fisher SE: **Assessing the impact of FOXP1 mutations on developmental verbal dyspraxia.** *Eur J Hum Genet* 2009, **17**:1354-1358.
100. Hamdan FF, Daoud H, Rochefort D, Piton A, Gauthier J, Langlois M, Foomani G, Dobrzyńska S, Krebs MO, Joobar R, Lafrenière RG, Lacaille JC, Mottron L, Drapeau P, Beauchamp MH, Phillips MS, Fombonne E, Rouleau GA, Michaud JL: **De novo mutations in FOXP1 in cases with intellectual disability, autism, and language impairment.** *Am J Hum Genet* 2010, **87**:671-678.
101. Horn D, Kapeller J, Rivera-Brugués N, Moog U, Lorenz-Depiereux B, Eck S, Hempel M, Wagenstaller J, Gawthroppe A, Monaco AP, Bonin M, Riess O, Wohlleber E, Illig T, Bezzina CR, Franke A, Spranger S, Villavicencio-Lorini P, Seifert W, Rosenfeld J, Klopocki E, Rappold GA, Strom TM: **Identification of FOXP1 deletions in three unrelated patients with mental retardation and significant speech and language deficits.** *Hum Mutat* 2010, **31**:E1851-1860.
102. Talkowski ME, Rosenfeld JA, Blumenthal I, Pillalamarri V, Chiang C, Heilbut A, Ernst C, Hanscom C, Rossin E, Lindgren AM, Pereira S, Ruderfer D, Kirby A, Ripke S, Harris DJ, Lee JH, Ha K, Kim HG, Solomon BD, Gropman AL, Lucente D, Sims K, Ohsumi TK, Borowsky ML, Loranger S, Quade B, Lage K, Miles J, Wu BL, Shen Y, et al.: **Sequencing chromosomal abnormalities reveals**

- neurodevelopmental loci that confer risk across diagnostic boundaries. *Cell* 2012, **149**:525-537.
103. Lee WS, Kang C, Drayna D, Kornfeld S: **Analysis of mannose 6-phosphate uncovering enzyme mutations associated with persistent stuttering.** *J Biol Chem* 2011, **286**:39786-39793.
 104. Cooper GM, Shendure J: **Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data.** *Nat Rev Genet* 2011, **12**:628-640.
 105. Fowler DM, Araya CL, Fleishman SJ, Kellogg, EH, Stephany, JJ, Baker D, Fields, S: **High resolution mapping of protein sequence-function relationships.** *Nat Methods* 2010, **7**:741-746.
 106. Patwardhan RP, Lee C, Litvin O, Young DL, Pe'er D, Shendure J: **High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis.** *Nat Biotech* 2009, **27**:1173-1175.
 107. Patwardhan RP, Hiatt JB, Witten DM, Kim MJ, Smith RP, May D, Lee C, Andrie JM, Lee SI, Cooper GM, Ahituv N, Pennacchio LA, Shendure J: **Massively parallel functional dissection of mammalian enhancers in vivo.** *Nat Biotechnol* 2012, **30**:265-270.
 108. Pitt JN, Ferré-D'Amaré AR: **Rapid construction of empirical RNA fitness landscapes.** *Science* 2010, **330**:376-379.
 109. ENCODE Project Consortium, Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Fretz S, Harrow J, Kaul R, Khatun J, Lajoie BR, Landt SG, Lee BK, Pauli F, Rosenbloom KR, Sabo P, Safi A, Sanyal A, Shores N, Simon JM, Song L, Trinklein ND, Altshuler RC, Birney E, Brown JB, Cheng C, Djebali S, Dong X, et al.: **An integrated encyclopedia of DNA elements in the human genome.** *Nature* 2012, **489**:57-74.
 110. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M: **Annotation of functional variation in personal genomes using RegulomeDB.** *Genome Res* 2012, **22**:1790-1797.
 111. White SA, Fisher SE, Geschwind DH, Scharff C, Holy TE: **Singing mice, songbirds, and more: models for FOXP2 function and dysfunction in human speech and language.** *J Neurosci* 2006, **26**:10376-10379.
 112. Lai CS, Gerrelli D, Monaco AP, Fisher SE, Copp AJ: **FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder.** *Brain* 2003, **126**:2455-2462.
 113. Belton E, Salmond CH, Watkins KE, Vargha-Khadem F, Gadian DG: **Bilateral brain abnormalities associated with dominantly inherited verbal and orofacial dyspraxia.** *Hum Brain Mapp* 2003, **18**:194-200.
 114. Liegeois F, Baldeweg T, Connelly A, Gadian DG, Mishkin M, Vargha-Khadem F: **Language fMRI abnormalities associated with FOXP2 gene mutation.** *Nat Neurosci* 2003, **6**:1230-1237.
 115. Ferland RJ, Cherry TJ, Preware PO, Morrissey EE, Walsh CA: **Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain.** *J Comp Neurol* 2003, **460**:266-279.
 116. Enard W: **FOXP2 and the role of cortico-basal ganglia circuits in speech and language evolution.** *Curr Opin Neurobiol* 2011, **21**:415-424.
 117. Groszer M, Keays DA, Deacon RM, de Bono JP, Prasad-Mulcare S, Gaub S, Baum MG, French CA, Nicod J, Coventry JA, Enard W, Fray M, Brown SD, Nolan PM, Pääbo S, Channon KM, Costa RM, Eilers J, Ehret G, Rawlins JN, Fisher SE: **Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits.** *Curr Biol* 2008, **18**:354-362.
 118. French CA, Jin X, Campbell TG, Gerfen E, Groszer M, Fisher SE, Costa RM: **An aetiological Foxp2 mutation causes aberrant striatal activity and alters plasticity during skill learning.** *Mol Psychiatry* 2011, **17**:1077-1085.
 119. Kurt S, Groszer M, Fisher SE, Ehret G: **Modified sound-evoked brainstem potentials in Foxp2 mutant mice.** *Brain Res* 2009, **1289**:30-36.
 120. Kurt S, Fisher SE, Ehret G: **Foxp2 mutations impair auditory-motor association learning.** *PLoS One* 2012, **7**:e33130.
 121. Watkins KE, Dronkers NF, Vargha-Khadem F: **Behavioural analysis of an inherited speech and language disorder: comparison with acquired aphasia.** *Brain* 2002, **125**:452-464.
 122. Scharff C, Adam I: **Neurogenetics of birdsong.** *Curr Opin Neurobiol* 2013, **23**:29-36.
 123. Haesler S, Wada K, Nshdejan A, Morrissey EE, Lints T, Jarvis ED, Scharff C: **FoxP2 expression in avian vocal learners and non-learners.** *J Neurosci* 2004, **24**:3164-3175.
 124. Miller JE, Spiteri E, Condro MC, Dosumu-Johnson RT, Geschwind DH, White SA: **Birdsong decreases protein levels of FoxP2, a molecule required for human speech.** *J Neurophysiol* 2008, **100**:2015-2025.
 125. Teramitsu I, Poopatanapong A, Torrisi S, White SA: **Striatal FoxP2 is actively regulated during songbird sensorimotor learning.** *PLoS One* 2010, **5**:e8548.
 126. Teramitsu I, White SA: **FoxP2 regulation during undirected singing in adult songbirds.** *J Neurosci* 2006, **26**:7390-7394.
 127. Haesler S, Rochefort C, Georgi B, Licznerns P, Osten P, Scharff C: **Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X.** *PLoS Biol* 2007, **5**:e321.
 128. Schulz SB, Haesler S, Scharff C, Rochefort C: **Knockdown of FoxP2 alters spine density in Area X of the zebra finch.** *Genes Brain Behav* 2010, **9**:732-740.
 129. Zweier C, de Jong EK, Zweier M, Orrico A, Ousager LB, Collins AL, Bijlsma EK, Oortveld MA, Ekici AB, Reis A, Schenck A, Rauch A: **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* 2009, **85**:655-666.
 130. Gregor A, Albrecht B, Bader I, Bijlsma EK, Ekici AB, Engels H, Hackmann K, Horn D, Hoyer J, Klapecki J, Kohlhasse J, Maystadt I, Nagl S, Prott E, Tinschert S, Ullmann R, Wohlleber E, Woods G, Reis A, Rauch A, Zweier C: **Expanding the clinical spectrum associated with defects in CNTNAP2 and NRXN1.** *BMC Med Genet* 2011, **12**:106.
 131. Newbury DF, Paracchini S, Scerri TS, Winchester L, Addis L, Richardson AJ, Walter J, Stein JF, Talcott JB, Monaco AP: **Investigation of dyslexia and SLI risk variants in reading- and language-impaired subjects.** *Behav Genet* 2011, **41**:90-104.
 132. Whitehouse AJ, Bishop DV, Ang QW, Pennell CE, Fisher SE: **CNTNAP2 variants affect early language development in the general population.** *Genes Brain Behav* 2011, **10**:451-456.
 133. Whalley HC, O'Connell G, Sussmann JE, Peel A, Stanfield AC, Hayiou-Thomas ME, Johnstone EC, Lawrie SM, McIntosh AM, Hall J: **Genetic variation in CNTNAP2 alters brain function during linguistic processing in healthy individuals.** *Am J Med Genet B Neuropsychiatr Genet* 2011, **156B**:941-948.
 134. Kos M, van den Brink D, Snijders TM, Rijpkema M, Franke B, Fernandez G, Hagoort P: **CNTNAP2 and language processing in healthy individuals as measured with ERPs.** *PLoS One* 2012, **7**:e46995.
 135. Dennis EL, Jahanshad N, Rudie JD, Brown JA, Johnson K, McMahon KL, de Zubicaray GI, Montgomery G, Martin NG, Wright MJ, Bookheimer SY, Dapretto M, Toga AW, Thompson PM: **Altered structural brain connectivity in healthy carriers of the autism risk gene, CNTNAP2.** *Brain Connect* 2011, **1**:447-459.
 136. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH, Hansen NF, Durand EY, Malaspina AS, Jensen JD, Marques-Bonet T, Alkan C, Prüfer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Höber B, Höffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, et al.: **A draft sequence of the Neandertal genome.** *Science* 2010, **328**:710-722.
 137. Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prüfer K, de Filippo C, Sudmant PH, Alkan C, Fu Q, Do R, Rohland N, Tandon A, Siebauer M, Green RE, Bryc K, Briggs AW, Stenzel U, Dabney J, Shendure J, Kitman J, Hammer MF, Shunkov MV, Derevianko AP, Patterson N, Andrés AM, Eichler EE, et al.: **A high-coverage genome sequence from an archaic Denisovan individual.** *Science* 2012, **338**:222-226.
 138. Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T, Monaco AP, Pääbo S: **Molecular evolution of FOXP2, a gene involved in speech and language.** *Nature* 2002, **418**:869-872.
 139. Maricic T, Günther V, Georgiev O, Gehre S, Culin M, Schreiwies C, Naumann R, Burbano HA, Meyer M, Lalueza-Fox C, de la Rasilla M, Rosas A, Gajovic S, Kelso J, Enard W, Schaffner W, Pääbo S: **A recent evolutionary change affects a regulatory element in the human FOXP2 gene.** *Mol Biol Evol* 2012, **30**:844-852.
 140. Falivelli G, De Jacobo A, Favaloro FL, Kim H, Wilson J, Dubi N, Ellisman MH, Abrahams BS, Taylor P, Comolletti D: **Inherited genetic variants in autism-related CNTNAP2 show perturbed trafficking and ATF6 activation.** *Hum Mol Genet* 2012, **21**:4761-4773.

doi:10.1186/gb-2013-14-4-204

Cite this article as: Deriziotis P, Fisher SE: **Neurogenomics of speech and language disorders: the road ahead.** *Genome Biology* 2013, **14**:204.