Genetics and the Language Sciences

Simon E. Fisher and Sonja C. Vernes

Department of Language and Genetics, Max Planck Institute of Psycholinguistics, Nijmegen 6525 XD, Netherlands; email: Sonja.Vernes@mpi.nl

Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen 6525 EN, Netherlands

Abstract

Theories addressing the biological basis of language must be built on an appreciation of the ways that molecular and neurobiological substrates can contribute to aspects of human cognition. Here, we lay out the principles by which a genome could potentially encode the necessary information to produce a language-ready brain. We describe what genes are; how they are regulated; and how they affect the formation, function, and plasticity of neuronal circuits. At each step, we give examples of molecules implicated in pathways that are important for speech and language. Finally, we discuss technological advances in genomics that are revealing considerable genotypic variation in the human population, from rare mutations to common polymorphisms, with the potential to relate this variation to natural variability in speech and language skills. Moving forward, an interdisciplinary approach to the language sciences, integrating genetics, neurobiology, psychology, and linguistics, will be essential for a complete understanding of our unique human capacities.

1. BEYOND THE ABSTRACT GENE

A deeper understanding of genes and genomes has broad importance for scientists who study the bases and origins of natural language. Assumptions about the roles of genetics lie at the heart of controversial debates in the language sciences, concerning fundamental questions that have occupied generations of thinkers for decades. Yet, until recently, theories and discussions have been characterized largely by simple perspectives of genetics, typically treating genes as abstract entities with mysterious powers.

Consider the astonishing productivity of language, whether spoken or signed, in which we are able to take a finite set of elements and recombine them to create a potentially limitless number of meaningful utterances. One of the most influential ideas in linguistics posits that this ability depends on principles of combinatoriality (often referred to as rules), acting in the brains of every competent language user, for which we have no conscious awareness (Chomsky 1965). Where could such unconscious principles of language come from? If a child learning a language is exposed to a restricted number of examples, highly constrained by context, how is it that she or he successfully reconstructs the relevant rules, without formal teaching or explicit understanding of what these rules are? Some linguists conclude that a human infant must bring to this task an innate understanding of the nature of linguistic structures, for example in the form of pre-specified cognitive operations (Berwick et al. 2013). Although such theories routinely assume that a human gene (or a set of genes) somehow encodes the necessary information, they do so without reference to what genes are and how they work, neglecting the complex indirect routes by which genomes contribute to brain development and function. Over the years, unconstrained speculations about innately specified linguistic structure helped feed the myth of so-called grammar genes (Gopnik 1997).

Even for people who are skeptical about the existence of innate knowledge of linguistic structure and adopt broader perspectives on language acquisition, questions about genetic contributions remain pertinent. A typical human child acquires a vast lexicon of thousands of words; successfully maps them to their associated meanings in the world; and becomes adept at producing, perceiving, and understanding complex utterances that relate to the present, the future, the past, and the imagined. Exposure to linguistic input in the social environment is essential, but it is nonetheless remarkable that an infant masters these different aspects of language use without being explicitly taught any of them. The acquisition of language is in stark contrast to most other complex skills that we learn; as Charles Darwin noted almost 150 years ago, "[N]o child has an instinctive tendency to brew, bake, or write" (Darwin 1871, p. 55). Moreover, there is much interest in the concept of critical periods: that our capacities to acquire features of language (and/or to learn a second language) differ at various stages of development (Kuhl 2010). To reframe the issues from a biological perspective, one can ask which features of physiology, anatomy, and neurobiology underpin the acquisition of human linguistic prowess, and try to describe their ontogeny: how they develop during the lifetime of an organism. Inherited factors may play key roles in shaping the physiological, anatomical, and neurological structures that support language development, so principles of gene function can inform understanding of this area but have been largely ignored thus far.

This discussion leads to another related issue that has received substantial attention. Do certain linguistic functions depend on dedicated cognitive mechanisms, independent of other aspects of brain processing? To what extent is language the product of domain-general mechanisms sub-serving learning and memory? Early empirical claims of specificity drew from observations of putative cognitive dissociations in patients with acquired brain lesions or neurodegenerative disorders (Damasio & Geschwind 1984); it was argued that damage to particular cortical areas

impaired speech and/or language skills in selective ways while preserving other functions, pointing to modularity. As the field moved forward, the dissociation logic spilled over into studies of developmental disorders, albeit with some resistance (Bishop 1997, Paterson et al. 1999). There was optimism from certain quarters that dissociations of language function might be observable in neurodevelopmental disorders and that, if they were heritable, such observations would provide evidence not only of linguistic modules in the brain but also of the individual genes that create them (Gopnik 1997, van der Lely et al. 1998). Again, major theories in this area have depended on the model of the abstract gene, isolated from principles of molecular biology and developmental genetics. As we discuss in greater detail below, it is not possible for a gene to specify an individual cognitive process, and we should not expect neurodevelopmental disorders to carve up cognition into clear-cut dissociable packages, each directed by a different gene (Fisher 2006). Certainly there are developmental impairments in which various speech and language abilities are disproportionately disturbed compared with other cognitive skills, and studies of these impairments can be very valuable (Bishop et al. 2014). However, pure isolated impairments are seldom observed. For example, when children have unexplained difficulties acquiring language despite a rich linguistic environment, adequate intelligence, and normal hearing and in the absence of neurological disorder, they are often given a diagnosis of specific language impairment (SLI). But even when cases are referred to as SLI, they are usually not "specific" and involve a complicated cluster of problems with heterogeneity from one affected person to another (Bishop 2001). Interestingly, with the advent of neuroimaging, many researchers studying the neurobiology of language have moved away from the classic view (still found in leading textbooks), replacing it with more complex models involving distributed neural circuits with varying degrees of specialization (Poeppel & Hickok 2004).

A further area of intense discussion concerns the evolutionary origins of language capacities (Christiansen & Kirby 2003). Human language skills seem unmatched elsewhere in the animal kingdom. Several well-publicized theories account for this observation in terms of one mutation event occurring in our nonspeaking ancestors (e.g., Crow 1997, Klein & Edgar 2002, Chomsky 2011), a lone spark that was sufficient to trigger the sudden appearance of language and culture. This romantic notion is inconsistent with the messy mappings between genetics and cognitive processes, noted above. It is also incompatible with data from comparative genomics, which have catalogued the many small DNA changes that distinguish us from other hominins (Paabo 2014) and indicate that our origins cannot be explained by only a single genetic trigger (Fisher & Marcus 2006, Fisher & Ridley 2013).

Maybe we should not be surprised that genetics has often been subject to simplistic treatment in the language sciences. After all, little is understood at present about which genes contribute to speech and language skills and how they function. However, even without complete details of the crucial genes, there is a wealth of knowledge about core principles of gene action, the tortuous routes by which genomic information can ultimately contribute to brain development and function, and the ways that changes in genes and genomes are involved in the evolution of traits. We believe that theories about language capacities can benefit by being deeply rooted in these known biological constraints. Here, we provide a conceptual framework for considering potential connections between genomic information and aspects of speech and language, based on the state of the art in molecular neuroscience and developmental neurobiology. We provide examples from studies of relevant phenotypes when possible. This article is not intended to be a comprehensive review of genes and language (we refer the reader to other recent articles, e.g., Fisher 2013, Graham & Fisher 2013, Szalontai & Csiszar 2013), but rather a guide toward effective interdisciplinary integration of data on genes, neurons, circuits, and cognition for the language sciences.

DNA:
deoxyribonucleic acid; composed of four chemical nucleotides: adenine (A), guanine (G), cytosine (C), and thymine (T)
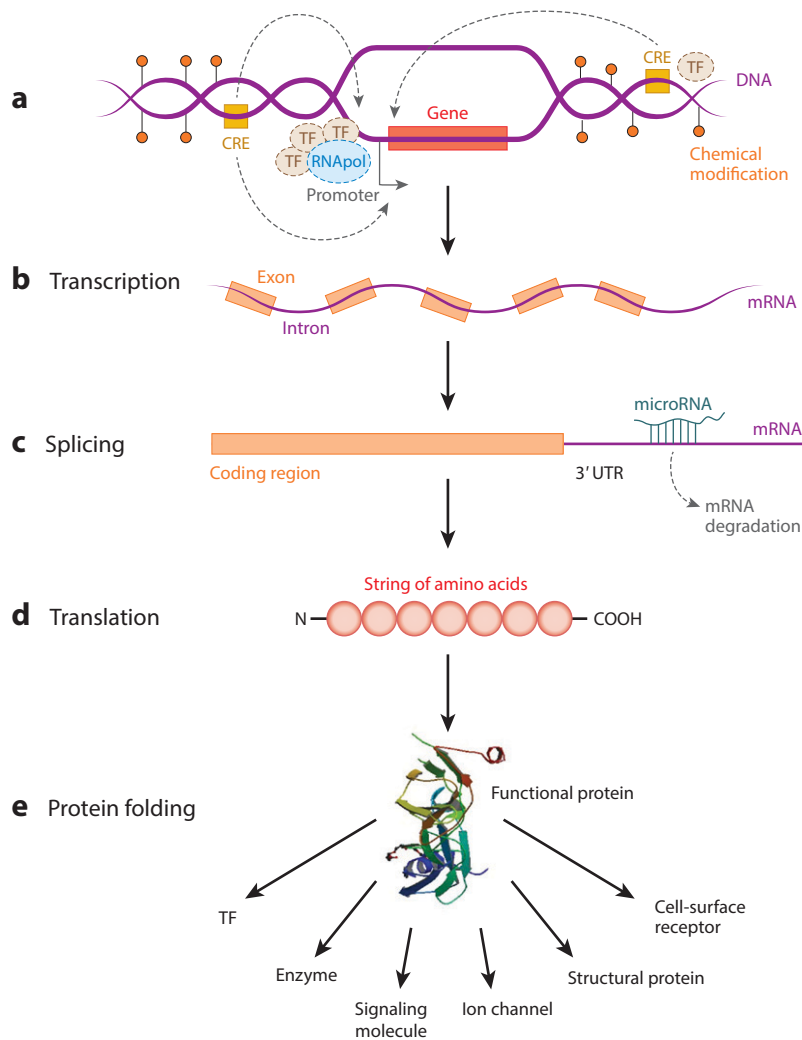
2. READING THE GENETIC CODE

The human genome, like that of other organisms, encodes biological information in the form of DNA sequences using only four nucleotide "letters": adenine (A), cytosine (C), guanine (G), and thymine (T). Yet this deceptively simple code can guide the development of an incredibly complex and varied multicellular organism with highly sophisticated cognitive capacities, including the ability to acquire language. As we show in this section, the key lies in the hierarchies of control that determine how, where, and when those letters are read. We begin by describing the ways in which information is packaged and interpreted to produce the functional output of the genome: gene products in the form of RNA and protein. We then outline how these products control the function and activity of cells, with a particular focus on neurons, explaining how such genetically guided cellular mechanisms mediate the assembly of neuronal circuits that underlie complex behavior and cognition.

2.1. Regulating Genes

The human genome is spread across 24 different chromosomes (chromosomes 1–22, plus the sex chromosomes X and Y), each of which can be considered a long string of DNA, tightly packaged up by other molecules. An entire human genome sequence is composed of more than three billion letters, but the information encoded within these stretches of As, Gs, Cs, and Ts is organized into somewhat discrete functional units, known as genes (**Figure 1***a*). Although all the cells of the body carry an almost identical genome, they are able to differ considerably in the combination of genes that are switched on or off (the profile of gene expression). This feature of biological systems allows the same set of DNA instructions to produce a wide array of distinct cellular types and subtypes with remarkably different morphology, function, and specialization (e.g., neurons, muscle, skin, liver). Thus, changes in expression profiles obviously play crucial roles during development (Ooi & Wood 2008), but we emphasize that gene regulation remains a dynamic process; even in mature cells, expression of genes can increase or decrease to suit the demands on the system (Loebrich & Nedivi 2009).

What factors determine when and where a gene is switched on or off (i.e., its expression pattern)? For each gene there are DNA sequences in the surrounding parts of the genome that are not themselves genes but instead represent regulatory elements (**Figure 1***a*). Regulatory elements in the vicinity of a gene can undergo epigenetic modifications that either unwind or compact the structure of the DNA in the local area, making it easier or harder, respectively, to switch on the gene (see the sidebar, Epigenetics) (Ha 2013). Links between particular epigenetic modifications and aspects of language function have not yet been demonstrated. However, differences in patterns of modification or activity of enzymes that mediate such modifications have been clearly implicated in human cognitive disorders (van Bokhoven 2011), including broad syndromes that include speech/language disruptions. A well-known illustration of the impact of epigenetic disruptions on brain function is the *MeCP2* (methyl CpG–binding protein 2) gene (Chahrour & Zoghbi 2007). Mutations of this gene cause Rett syndrome, a disorder characterized by normal early milestones followed by a slowing of development, loss of purposeful use of the hands, development of distinctive hand stereotypies, regression of spoken language, problems with walking, cognitive impairments, autistic features, and sometimes seizures. Epigenetics is a rapidly expanding field, and given that the relevant mechanisms are known to be important for regulating gene activity in neurons, it is plausible that future work will uncover roles for epigenetic pathways in the development and/or function of language-related circuits.

Regulatory elements in the vicinity of a gene also act as binding sites for transcription factors (TFs), which significantly affect the activity of the gene (**Figure 1***a*). TFs are proteins that bind to

Figure 1

From DNA to language: reading the genetic code. (*a*) Information in the genome is packaged into genes, which are surrounded by regulatory elements (REs) that can interact with regulatory proteins, such as transcription factors (TFs) or enzymes that change the epigenetic state of the DNA, to influence how a gene is expressed. Complexes of regulatory proteins assemble near the start site of a gene and recruit cellular machinery that reads the DNA sequence into an RNA message (mRNA) in a process known as transcription. (*b*) Transcription produces a molecule that contains exons (the parts of the message coding for protein) and introns (intervening sequences that do not code for protein). A process known as splicing cuts out all the introns and assembles combinations of exons in an unbroken chain. (*c*) This "mature" mRNA molecule is regulated by microRNA molecules that control the amount of message that remains in the cell by degrading unwanted molecules. (*d*) The exons of the mRNA molecule are then translated by the cellular machinery to produce a long string of amino acids known as a polypeptide. (*e*) Polypeptides fold into a three-dimensional structure (based on the physical properties of the amino acids in the chain) to produce a functional protein. The structure of the protein is essential to the role that the protein will play in a cell. The structure shown here is that of the ROBO1 protein. Abbreviations: RNApol, RNA polymerase; UTR, untranslated region. Image in panel *e* modified from the RCSB Protein Data Bank (PDB) (**http://www.rcsb.org**; PDB identifier 4HLJ, by R. Barak and Y. Opatowsky).

DNA to regulate the transcription of a gene (Ooi & Wood 2008). Transcription can be considered the reading of the particular section of DNA code into a message [called messenger RNA (mRNA)] that can be interpreted by the cell (**Figure 1***b*). TFs often bind close to the start of the gene that they regulate (known as the promoter region) and recruit other proteins to this location. The formation of a protein complex at this location can switch on gene expression by recruiting RNA polymerase II (RNApol II), the enzyme responsible for transcribing the DNA code into mRNA, or switch off gene expression by preventing RNApol II binding (Shandilya & Roberts 2012). TFs can also interact with or alter the epigenetic pattern near a gene to make it more or less accessible to RNApol II (Rothbart & Strahl 2014).

Transcription: the process by which the DNA sequence of a gene is copied into messenger RNA

Messenger RNA (mRNA): a ribonucleic acid (RNA) transcribed from DNA that codes for proteins

Splicing: the removal of introns from an mRNA molecule to produce a protein coding message

TFs can target hundreds of genes spread throughout the genome and act as central hubs in regulating molecular networks. Thus, it is perhaps unsurprising that certain TFs are strongly implicated in neurodevelopmental phenotypes (Hong et al. 2005), including those related to language. Indeed, the best-studied gene in the language sciences is a TF gene known as *FOXP2* (Fisher & Scharff 2009). This gene was identified through studies of a large multigenerational family (the KE family), in which 15 people were affected with a monogenic form of a speech and language disorder (Lai et al. 2001). Different mutations or disruptions of *FOXP2* have now been found in multiple unrelated individuals or families affected with similar problems (e.g., Lai et al. 2001, MacDermot et al. 2005, Feuk et al. 2006, Shriberg et al. 2006, Rice et al. 2012, Turner et al. 2013). The disorder involves difficulties in learning to produce coordinated sequences of orofacial movements during speech (childhood apraxia of speech or developmental verbal dyspraxia), accompanied by expressive and receptive linguistic impairments affecting spoken and written modalities; effects on other cognitive aspects are less marked (Watkins et al. 2002). Hundreds of genes have been identified that are thought to be switched on or off by FOXP2 in the developing brain (Spiteri et al. 2007; Vernes et al. 2007, 2011). Through the regulation of these targets, FOXP2 ultimately can contribute to biological substrates involved in the development of speech and language skills.

Regulation of gene expression does not finish once DNA is transcribed into mRNA. Each mRNA must then undergo processing (**Figure 1***b*) to produce a mature molecule that can subsequently be translated into the amino acid sequence of a protein. Maturation of mRNAs involves a process known as splicing, which is rather like the editing of a movie. In splicing, cellular machinery cuts out any parts of a gene that do not code for protein (sections known as introns) so

EPIGENETICS

Epigenetics is the science of heritable, reversible chemical modifications that are present in the genome. Epigenetic modifications (also known as chromatin marks) do not directly change DNA sequence, but rather the way the DNA is read (Day & Sweatt 2011). These modifications affect whether nearby genes are switched on or off, largely by affecting whether regulatory proteins can interact with the DNA (Rothbart & Strahl 2014). Regions of DNA can display multiple different modifications; the combination of all these in the genome of a cell is called its epigenetic code. During development, different cell types acquire a characteristic set of epigenetic modifications, and the environment can influence the epigenetic code (Ha 2013)—one reason that identical twins look more different as they get older. Note that for epigenetic information to be passed to the next generation, it must be encoded within the gametes (germ cells) of an organism or affect the embryo during development. Thus, something like experiencing famine during pregnancy could dramatically affect the developing embryo (Heijmans et al. 2008). By contrast, epigenetic changes occurring in neurons in the brain (Day & Sweatt 2011), for example during language acquisition, are not likely to be passed to the next generation because this information is not transferred to the gametes.

that only the protein-coding regions (exons) remain (Zheng & Black 2013). After splicing, a processed mRNA usually contains the template for building a single coherent chain of amino acids. Often, the exons of a single gene can be joined together in different combinations. This process of alternative splicing enables the same gene to potentially encode multiple different protein products (Zheng & Black 2013). The various forms of the protein (isoforms) may have distinct properties and functions, act at different developmental time points, and/or be present in different tissues or cellular subpopulations. As a consequence, a single gene can have much greater diversity and flexibility in the functions that it encodes. Intriguingly, some proteins act to regulate alternative splicing in neurons (Grabowski 2011), and disruptions of these processes have been associated with human neurodevelopmental disorders. One notable example is RNA-binding protein, fox-1 homolog (encoded by the gene *RBFOX1*), which modulates splicing events implicated in neuronal development, maturation, and excitation (Fogel et al. 2012). *RBFOX1* has been independently implicated in autism spectrum disorders and in epilepsy (Gehman et al. 2011). Moreover, it is a direct target of FOXP2 and has been subject to Darwinian selection in recent human history (Ayub et al. 2013). *RBFOX2* (the most similar gene in the genome to *RBFOX1*) has shown suggestive association with quantitative measures of language and reading performance (Gialluisi et al. 2014).

There is yet another level of regulation by which gene expression can be modulated. Many parts of the genome are transcribed into RNA molecules that do not act as templates for building proteins (i.e., they are noncoding); their active form is as an RNA molecule. Some of these molecules, known as microRNAs, interact directly with mRNAs of protein-coding genes, via a range of different mechanisms, to prevent them being used for translation into protein (**Figure 1c**). Given the intimate relationship between the amount of mRNA expressed from a gene and the amount of the resulting protein that is made, the extra layer of microRNAs allows for rapid fine-tuning of protein production. As discussed in Section 3.3, this feature makes microRNA regulatory mechanisms critical for aspects of neural plasticity. MicroRNAs have been implicated in the etiology of several cognitive disorders, and it seems likely that future studies will link this class of molecules to aspects of language function and/or dysfunction. One particularly interesting microRNA, called Mir-137, is found at high levels in the brain, where it is known to influence development and connectivity of neurons (see Section 3; Smrt et al. 2010, Willemsen et al. 2011), and it has been associated with disorders including intellectual disability, schizophrenia, and autism (Cross-Disord. Group Psychiatr. Genomics Consort. 2013). Mir-137, itself a target of FOXP2 (Vernes et al. 2011), interacts with a number of different mRNAs to fine-tune the quantities of encoded proteins. Many of these regulated mRNAs encode proteins that have been independently implicated in brain development and language-related phenotypes (Kwon et al. 2013, Wright et al. 2013, Devanna & Vernes 2014).

2.2. From Genes to Proteins

Proteins are made up of strings of amino acids that naturally fold into a particular three-dimensional shape. Molecules of mRNA act as templates for building proteins through the process of translation, in which information is sequentially read in the form of triplets of letters, each coding for one of 20 different available amino acids (**Figure 1d**). The rules for matching triplets of nucleotide letters to corresponding amino acids represent one of the fundamental features of life on this planet. For example, an ATG codon at the DNA level specifies the amino acid methionine, whereas ACG corresponds to threonine, AAG to lysine, and so on (Szymanski & Barciszewski 2007). There are 64 possible triplet combinations of the 4 nucleotide letters ($4^3 = 64$), but only 20 amino acids are used as potential building blocks for proteins. This means that the genetic code

Amino acids: the chemical compounds of which proteins are made; they are encoded by triplets of nucleotides in mRNA

Alternative splicing: combining exons from a single gene to produce mRNA encoding different (but related) proteins

microRNA: a small RNA molecule that interacts with mRNA to regulate the levels of protein produced in a cell

Translation: the creation of polypeptides (strings of amino acids) from an mRNA template

has a degree of redundancy—sometimes multiple triplet combinations encode the same amino acid (e.g., CCA, CCG, CCT, and CCC all specify the amino acid proline). Thus, some DNA changes within a protein-coding gene may have no effect at the protein level, leaving the amino acid sequence unaltered—an important consideration when studying genome evolution and when tracing connections between genotypes and phenotypes.

Different amino acids have distinct chemical properties, so the precise sequence of amino acids in a protein determines the three-dimensional shape that it forms, and the shape of the protein determines the function that it performs. The set of proteins that are expressed in a cell defines the biological properties and behavior of that cell.

The 20 amino acid building blocks can produce a vast array of distinct proteins of different shapes and sizes with astonishingly diverse functions (**Figure 1***e*). For example, a protein can act as an enzyme to facilitate a chemical reaction, as a channel to allow external factors into the cell, as a signal to be sent to another cell, as a structural protein to give a cell its shape, and so on. As discussed above, some proteins also act to regulate gene transcription and splicing of genes, and such proteins are, of course, themselves encoded by genes. Thus, genes and proteins interact with each other to form complicated networks, which can be carefully traced out and studied through a variety of experimental methods (Fisher 2006, Vernes & Fisher 2009, Fogel et al. 2012, Carrion-Castillo et al. 2013, Rodenas-Cuadrado et al. 2014). But how might defined functions of genes and proteins at cellular levels affect the building of brain circuitry, or influence neural processing? This crucial bridge to cognition and behavior is the subject of our next section.

3. CONNECTING GENES TO NEURAL CIRCUITS

All cognitive abilities and behaviors, including the processes that underpin language, depend on the activities of sets of neurons that signal to each other in intricate patterns of connectivity. A number of different events need to occur for these neural circuits to form properly (Gao et al. 2013). Neuronal precursor cells must proliferate and differentiate into a particular subtype, taking on a specific neuronal identity (Tan & Shi 2013). Many types of neurons need to move past other cells to find their final location in the brain; for example, this process of neuronal migration is crucial for forming the six layers of cerebral cortex found in humans and all other mammals (Marin et al. 2010, Tan & Shi 2013). Each neuron has to wire up with other neurons, some of which might be close by, while others might be in rather distant brain regions (Kolodkin & Tessier-Lavigne 2011). To allow information to flow from one neuron to another in this circuit, connections known as synapses must form (Chia et al. 2013). Ongoing modification or refinement of these synapses changes the strength of connection and modulates the properties of the circuit (Holtmaat & Svoboda 2009). It is well established that the formation of neural circuits depends on interplay between the molecular factors that guide these processes of proliferation, differentiation, migration, and connectivity, together with experience-dependent contributions from the environment. Moreover, learning and memory rely on strengthening or weakening of sets of synaptic connections, and this process of plasticity is itself mediated by genes and proteins. The fundamental interweaving of genetics and experience in neural function is not widely appreciated outside the fields of molecular neuroscience and neurobiology.

If any of the above processes underlying circuit development or function is disrupted, it can influence the properties of the affected circuits and thereby yield disturbances in the cognitive processes that depend on such circuits. Often, the circuits underlying any given cognitive process are widely distributed throughout the brain, involve multiple different subtypes of neurons, and overlap with other circuits; in other words, one set of neurons may be embedded in multiple different circuits, with distinct wiring patterns. As such, molecular factors that affect neural

Synapse: a connection between two neurons that allows the exchange of information in the form of neurotransmitter molecules

circuits tend to be pleiotropic—that is, affect more than one aspect of behavior and cognition. From this perspective, it is unlikely that there could ever be a "gene for language" or even a single set of genes that are exclusive to language function. However, through a combination of genomic studies in humans and functional genetic analyses in cellular and animal models (White et al. 2006, Deriziotis & Fisher 2013, Vernes & Fisher 2013), scientists have identified genes that are related to the capacity for language and that contribute to the neural circuitry needed for normal language development and function. In the following sections, we consider how different types of gene products could affect the perception, processing, and production of language by contributing to the formation and activity of neural circuits in the brain.

3.1. On the Move: Neuronal Migration in Brain Development

At early stages of development, once a neuron has obtained its neuronal identity, it migrates to its final location to produce normal brain architecture (**Figure 2a**). Deficits in this process can result in severe disruptions of the gross anatomy of the brain, such as lissencephaly, a syndrome involving major abnormalities in layering and folding of the cortex (Manzini & Walsh 2011). However, disturbances in neuronal migration may also contribute to more subtle neurodevelopmental disorders that could affect language-related skills in the absence of widespread anatomical or phenotypic abnormalities (Carrion-Castillo et al. 2013). This idea comes from studies of developmental dyslexia (reading disability), a common childhood disorder involving unexplained problems with learning to read and spell, despite adequate intelligence and opportunity (Paracchini et al. 2007). Although children with dyslexia appear to have grossly normal spoken language skills, research has shown that a substantial proportion have underlying persistent difficulties with aspects of phonological processing; thus, many scientists in the field consider dyslexia a language-related disorder (Fisher & DeFries 2002). The possibility of aberrant neuronal migration in dyslexia was first proposed on the basis of neuroanatomical studies of human postmortem tissue, in which displaced neurons and glia were observed, tending to be located in cortical regions of the left hemisphere (Galaburda & Kemper 1979, Galaburda et al. 1985). Some years later, independent evidence emerged from an entirely novel direction. The first four candidate genes to be associated with dyslexia, *DYX1C1*, *KIAA0319*, *DCDC2*, and *ROBO1*, have all been implicated in neuronal migration mechanisms through studies of animal models (Paracchini et al. 2007, Scerri & Schulte-Körne 2010), although it is clear that this is not their only cellular function.

3.2. Wiring the Brain: Neurite Outgrowth and Connectivity

To establish connectivity patterns in the brain, neurons extend long, thin processes, known as neurites, from their cell bodies (**Figure 2b**) (Kolodkin & Tessier-Lavigne 2011). From each neuron, one of the neurites develops into the axon, the main route by which the neuron transmits information (Polleux & Snider 2010). The axon must grow toward its correct synaptic targets, via a mechanism called axon guidance, sometimes navigating across great distances (Bashaw & Klein 2010, Chedotal & Richards 2010). The remaining neurites of each neuron become dendrites, which receive incoming signals from other neurons (Polleux & Snider 2010). The architecture of neural circuits, along with the balance of excitation and inhibition in the brain, strongly depends on these early wiring activities. Many genes have been identified that contribute to neurite growth, axon guidance, and connectivity (Kolodkin & Tessier-Lavigne 2011), and disruption of such pathways has been repeatedly implicated in neurodevelopmental disorders, most notably autism and schizophrenia (Piton et al. 2011). This finding also holds for primary disorders of speech and

Neurite: collective term for all the processes extending from the cell body of a developing neuron

Axon: a long projection from a neuron that makes connections to other neurons and sends signals via the synapse

Dendrite: long and often highly branched projections from neuronal cells that make connections with other neurons and receive signals via the synapse
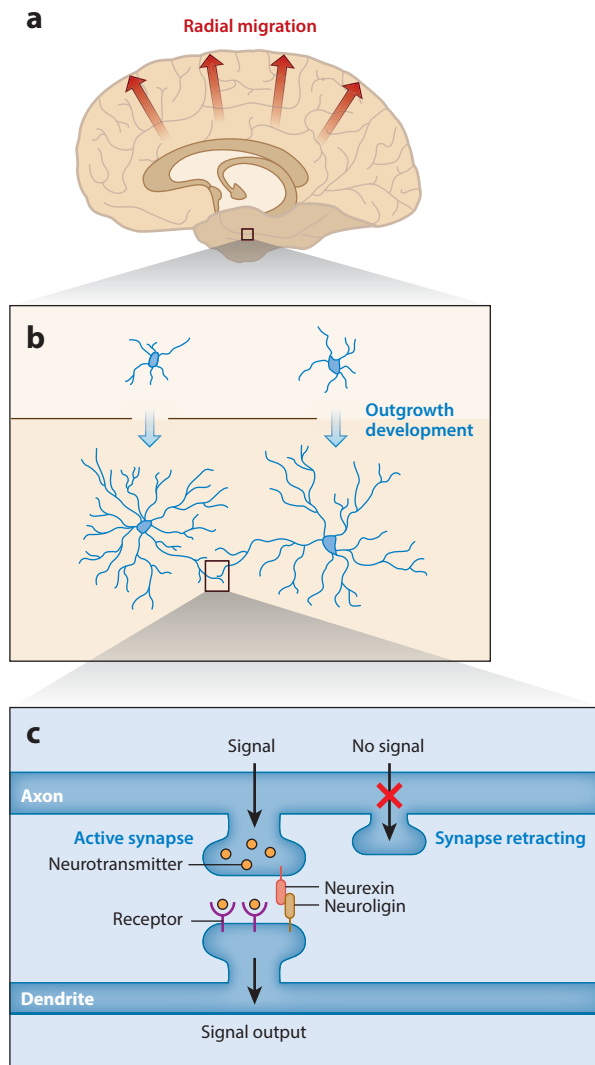
Figure 2

From DNA to language: wiring the brain. (*a*) Schematic view of the brain, sliced down the midline (sagittal view).
Many different neurons must migrate to a different part of the brain during development. They do so via a
combination of attractive and repulsive guidance cues given by other cells. The cortex represents an important
location for migration during development because cells must migrate from the base layers of the cortex into the
upper regions in order for the six-layer cortex to form (a process known as radial migration). Dyslexia-related genes
such as *ROBO1*, *KIAA0319*, *DYX1C1*, and *DCDC2* influence neuronal migration. (*b*) Neurons must develop
a complex network of axons and dendrites to form connected neuronal circuits. Many genes have been implicated
in the growth and guidance of axons and dendrites, including the language-related genes *CNTNAP2* and *ROBO1*.
(*c*) It is estimated that a single neuron can form on the order of tens of thousands of connections. During development
and throughout adult life, these connections, or synapses, allow information to flow from one neuron (via release
of neurotransmitter from its axon) to another (via uptake of neurotransmitter by receptors on the dendrite).
Synapses can be strengthened or retracted depending on the strength and frequency of signaling between the connected
neurons. MicroRNAs (such as miR-137) are located at the synapse and allow rapid changes in protein levels to
produce real-time conversion of neuronal activity into synaptic activity. Structural molecules such as neurexins (e.g.,
NRXN1) and neuroligins form a molecular bridge that contributes to the strength and plasticity of the synapse.

language; functional studies of *FOXP2* found key contributions to neurogenetic pathways that regulate neurite outgrowth (Vernes et al. 2011).

The extension of neurites to form connected networks is closely linked to neuronal migration, involving an elegant system of signals and receptors, along with structural changes within the cell (Bashaw & Klein 2010, Polleux & Snider 2010). Indeed, many of the molecular substrates underlying migration and outgrowth are shared (Marin et al. 2010). To give one specific example, we have already encountered *ROBO1*, a gene implicated in dyslexia and other language-related phenotypes (Paracchini et al. 2007, Carrion-Castillo et al. 2013). As noted in Section 3.1, this gene encodes a guidance molecule that helps neurons migrate to their correct location (Park et al. 2002, Gonda et al. 2013). However, the same molecule also becomes important in a different context, guiding the developing growth cones of axons as they extend to find the correct connectivity partners (Kim et al. 2011, Kolodkin & Tessier-Lavigne 2011). The successful wiring up of a human brain is an amazing feat, considering that it is estimated to comprise ∼80 billion neurons (Fonseca-Azevedo & Herculano-Houzel 2012) and that many of the connections are made between neurons located in distant regions. The higher-order cognitive processing that underpins language does not depend only on short-range local connections. It also involves the activity of distributed circuits that bridge the cortex with subcortical structures such as the striatum and thalamus. The growing axons of a microscopic neuron must interpret guidance cues to correctly navigate toward distant targets while ignoring inappropriate neuronal partners, allowing for the correct formation of long-range neural circuits.

3.3. Sending Signals: Synapse Development, Maintenance, and Plasticity

A synapse is a physical structure that forms where two neurons make contact, allowing signals to be passed from one neuron to the next (**Figure 2c**). Synapse formation is dynamic, ongoing, and essential to neural circuit development and function. Neurons develop new synapses, which can be strengthened if signals continue to be sent across them or retracted (even completely abolished) in the absence of such signals. Individual synapses strengthen and weaken in response to changes in activity of the circuit, such as those that occur when an organism interacts with its environment (Holtmaat & Svoboda 2009, Ebert & Greenberg 2013). This process of synaptic plasticity depends on multiple crucial molecular mechanisms, which provide the basis of learning and memory. For example, as you learn a new skill, the synapses in the circuitry involved in executing it become stronger as the behavior is repeated (Ebert & Greenberg 2013). Many genes have been identified that contribute to synapse development, maintenance, and plasticity.

To facilitate increases or decreases in signaling though a neural circuit, rapid, precise control of the molecular composition of individual synapses is paramount. Synapses are often located at long distances from the cell body, where most proteins are produced. Thus, neurons cannot rely on the relatively slow transport of proteins from cell body to synapse when responding to neuronal activity. Furthermore, because a single neuron may belong to multiple circuits, different synapses from the same cell must be able to respond to different inputs. One way that neurons deal with these challenges is by controlling protein levels directly at the synapse using microRNAs (Jung & Holt 2011; also see Section 2.1). In neurons, a pool of mRNA transcripts are transported out of the main part of the cell and maintained at individual synapses for translation into protein (Poon et al. 2006, Zivraj et al. 2010, Jung & Holt 2011). MicroRNAs located at the synapse act on this mRNA pool, mediating rapid changes in protein levels at individual synapses in response to inputs from different neuronal circuits (Goldie & Cairns 2012, McNeill & Van Vactor 2012). This process allows real-time conversion of neuronal activity into changes in the molecular, structural, and functional activity of individual synapses.

A number of structural proteins have also been identified that facilitate the transfer of information across synapses. Neurexins are a family of proteins with a distinctive position at the synapse; part of the protein is inside the cell, and another part juts out of the axon and into the synaptic cleft (Südhof 2008). Neurexins make contact with proteins that have a similar localization on the opposing synapse; some of the best-characterized partners of neurexins are molecules known as neuroligins. This neurexin–neuroligin partnership forms a molecular bridge across the synapse, maintaining the connection between two neurons and affecting the strength and plasticity of the synapse (Südhof 2008). Loss of neurexin or neuroligin activity reduces signal activity at the synapse, altering the activity of the relevant neural circuits (Zhang et al. 2005). The human genome contains three neurexin and four neuroligin genes, and the mRNAs from these can undergo extensive alternative splicing (see Section 2.1) to produce thousands of molecules (Ullrich et al. 1995, Boucard et al. 2005). The large number of possible combinations of such molecules at different synapses has been proposed to form a "synaptic code" that may influence neural circuit development and plasticity to produce different activity at different synapses (Boucard et al. 2005, Südhof 2008). The neurexin gene, *NRXN1*, has been implicated in complex neurological disorders such as schizophrenia and intellectual disability (Bena et al. 2013). Mutations of *NRXN1* have also been found in patients with language impairments, language delay, and absent language (Ching et al. 2010, Gregor et al. 2011, Zweier 2012, Bena et al. 2013). Moreover, mutations of neuroligin genes have been described in children with autism spectrum disorders (Bourgeron 2009).

The processes underlying neuronal circuit formation do not occur independently of one another, and genes have been identified that affect these different but related elements of migration, connectivity, and signaling. *CNTNAP2* is a gene that has been implicated in SLI and associated with a number of other disorders involving language disruptions, such as autism and epilepsy (Alarcón et al. 2008, Vernes et al. 2008, Rodenas-Cuadrado et al. 2014). *CNTNAP2* expression is tightly controlled, and it has been found at high levels in brain circuits involved in higher-order cognitive functions. The protein produced from the *CNTNAP2* gene is known as CASPR2 (contactin associated protein–like 2). CASPR2 plays multiple roles in the developing and adult brain, influencing how neurons migrate, how they connect to one another, and how they signal to each other once connected (Rodenas-Cuadrado et al. 2014). In this way, *CNTNAP2* can influence the development of neural circuits as well as their ongoing functional activity, and the phenotypic outcomes of disruption of this gene vary depending on when and how severely the activity of CASPR2 is impaired. Interestingly, *CNTNAP2* expression is regulated by TFs that have themselves been implicated in cognitive disorders involving language, including FOXP2 (discussed in Section 2.1), FOXP1 (a closely related TF), and TCF4 (O'Roak et al. 2011, Forrest et al. 2012). Mutations of *TCF4* and *FOXP1* can cause phenotypes encompassing autism spectrum disorder, intellectual disability, and speech impairments/delay (Bacon & Rappold 2012, Hamdan et al. 2010, Horn 2011, O'Roak et al. 2011, Navarrete et al. 2013, Palumbo et al. 2013, Sweatt 2013). Thus, scientists are beginning to trace networks of functionally related genes that jointly contribute to neural pathways that are important for the normal development of speech and language.

4. UNIVERSALITY, VARIABILITY, AND THE ERA OF PERSONAL GENOMICS

As a human-specific trait, the capacity to acquire complex speech and language presents special challenges to geneticists, given that discovery of the critical genes must rely on studies of our own species. In such cases, one can treat the human population as a kind of natural experiment by searching for correlations between variation at the genotypic and phenotypic levels. As is evident from the examples of genetic disruptions discussed in Sections 2 and 3, this approach is particularly effective at the extremes; it is used to study developmental pathologies ranging from primary

Synaptic cleft: the space between two synapses, into which neurotransmitter molecules are released

speech and language disorders (such as childhood apraxia of speech and SLI) to cognitive syndromes that sometimes disturb these skills within a broader syndrome (such as autism). Crucially, the same overall approach can be used to investigate variability in speech and language skills in the normal range. For some linguists, embedded in a strong tradition that focuses on language universals and that treats linguistic competence as a fixed unvarying feature of *Homo sapiens*, this endeavor might seem futile. Nonetheless, the study of variation represents a cornerstone of biology, a powerful empirical tool for revealing pathways and mechanisms at multiple levels. In this final section, we consider some ways in which analyses of genetic and phenotypic variability are opening up new research avenues for the language sciences.

Due to the emphasis placed on universals, although language researchers have embraced the idea of studying pathology, little work has been done on other sources of within-population differences in linguistic skills. Therefore, we do not yet have a full picture of the degree of relevant phenotypic variation. In contrast, we know a great deal about diversity at the genetic level, which has turned out to be greater than previously realized (Abecasis et al. 2012, Lappalainen et al. 2013). Modern human populations harbor an array of genomic variations, including large-scale rearrangements of chromosomes, altered numbers of copies of genes, and changes of single letters of DNA. Variants range in frequency from very common polymorphisms to extremely rare mutations that might even be unique to a particular family or individual (known as private mutations). Many genomic variants are effectively silent (see Section 2.2); they have no apparent consequences for gene function and thus are unlikely to have biological effects on distal phenotypes. However, if a variant changes the coding sequence of a gene, this change may alter the amino acid sequence of the encoded protein in a way that affects the protein's function. Even a variant that does not alter protein sequences can have functional consequences; for example, if it is located in a regulatory element it can change the way that a protein-coding gene is expressed, when and where it is switched on in the developing and/or mature organism, how much protein is made, and so on. As described in Section 3, these alterations can lead to changes in development and function of neural circuits via effects on processes such as proliferation, migration, connectivity, and plasticity. Although different loci can show distinct patterns and frequencies, virtually every gene in the genome shows some type of potentially functional variation in modern human populations (Abecasis et al. 2012, Lappalainen et al. 2013).

The wider impact of human genetics on biology and medicine stems largely from the use of existing variation, of one kind or another, to trace causal links between functional variants in our genomes and phenotypic outcomes. Although early successes concerned mainly single-gene disorders and rare syndromes arising from major chromosomal abnormalities, work has progressed to tackle complex multifactorial traits and has even extended the reach beyond pathology into the general population. These efforts are not about documenting trivial aspects of variation—after relevant genes are identified, they are used as windows into the biology of the trait of interest. Most importantly, these principles, methods, and technologies are not exclusive to biomedical traits, but are applicable to features of human cognition and behavior, including speech and language.

Our first glimpse of the molecular underpinnings of developmental speech and language disorders has already revealed a mixed genetic architecture with a spectrum of effects (Graham & Fisher 2013). Researchers have documented rare, highly disruptive gene variants with large consequences, as epitomized by the various private *FOXP2* mutations that cause childhood apraxia of speech in different families and individuals (e.g., Lai et al. 2001, MacDermot et al. 2005, Feuk et al. 2006, Shriberg et al. 2006, Rice et al. 2012, Turner et al. 2013). A half-dosage of functional FOXP2 protein appears sufficient, by itself, to derail a person's speech and language skills (Fisher & Scharff 2009). At the same time, studies of hundreds of families with typical forms

Private mutation:
a novel and rare DNA variant found only in a single individual or family

of SLI have uncovered complex relationships between language impairments and common genetic variation. For instance, when children from SLI families carry particular variants of *ATP2C2* and/or *CMIP* (two neighboring genes on chromosome 16), they show reduced performance on language measures, such as the ability to repeat pronounceable nonsense words (Newbury et al. 2009). In this case, the putative risk variants are common polymorphisms that are found at appreciable frequency in unaffected people; they affect language skills in a quantitative manner, accounting for a relatively small proportion of total trait variance, and so must be acting together with other contributory factors in a multifactorial context (Newbury et al. 2009). Moreover, as is often observed for common risk variants, the polymorphisms lie outside the coding regions of the genes and are instead hypothesized to alter gene regulation in some way; the precise mechanisms have yet to be experimentally determined.

In studies of speech and language disorders, genes harboring rare etiological mutations and those carrying common risk polymorphisms converge on shared functional pathways at the molecular level, as demonstrated by regulation of *CNTNAP2* expression by the FOXP proteins (Vernes et al. 2008, O'Roak et al. 2011). Sometimes, the same gene can show independent evidence of rare and common variants affecting related disorders. Again, *CNTNAP2* is a prime example (Rodenas-Cuadrado et al. 2014), given that rare coding mutations cause cortical dysplasia–focal epilepsy syndromes involving language regression and autistic features, and are also found in cases of intellectual disability (Gregor et al. 2011), whereas common noncoding variants at this locus are implicated in language deficits in autism and SLI (Alarcón et al. 2008, Vernes et al. 2008). Efforts are under way to trace the functional effects of these networks of known risk genes on the key features of neuronal biology, discussed in Section 3.

As in biomedical traits, advances in molecular technologies are accelerating the pace of research on genotype–phenotype correlations in language-related disorders (Deriziotis & Fisher 2013). One key development has been the production of DNA chips for high-throughput genotyping. For a few hundred dollars per individual, one can now simultaneously screen millions of different genetic markers across the genome, targeting all the known sites of variation within every gene. Researchers can thereby carry out large-scale screens for genome-wide association, with the potential to reveal novel connections between particular genes and phenotypes of interest (Visscher et al. 2012). Given the degree of multiple testing involved in such screens and the potentially small effect sizes of the variants being tested, such studies require hundreds (even thousands) of people to achieve adequate statistical power. Genome-wide association scans are being carried out for common language-related disorders, but the sample sizes are relatively small compared with those in other fields in human genetics, so clearly we are still scratching the surface (Roeske et al. 2011, Gialluisi et al. 2014, Nudel et al. 2014).

Another innovation that has transformed genetics and genomics is the rise of new platforms for fast, affordable DNA sequencing (Metzker 2010). We are rapidly approaching a time when an individual's entire genome can be fully sequenced in a matter of hours for a cost of less than US$1,000. (As of June 2014, the cost is only around two to three times this amount, so this target is not far off.) Indeed, sequencing of all the coding parts of the genome (the exome) has already begun to be used to study neurodevelopmental disorders such as autism and childhood apraxia of speech (e.g., O'Roak et al. 2011, Worthey et al. 2013). In principle, the latest technologies will enable researchers to catalog virtually every gene variant (from private mutations to common polymorphisms) that each study participant carries, providing comprehensive genomic data in which to search for correlations with speech and language disorders. Nevertheless, this opportunity raises major challenges with regard to study design, analyses, and interpretation. Perhaps the most substantial will be to determine, against a background of many potentially interesting changes,

which genetic variants are functionally connected with speech and language skills (Deriziotis & Fisher 2013; also see Future Issues, below).

The impact of affordable rapid high-throughput genotyping and DNA sequencing extends beyond language-related disorders and into the normal range of speech and language skills in unaffected people. Studies have already targeted candidate genes implicated in SLI and dyslexia and have shown that common variation of some of these genes have effects in the general population (Carrion-Castillo et al. 2013, Graham & Fisher 2013). For example, the same *CNTNAP2* risk variants that are associated with reduced language performance in SLI and with "age at first word" in autism correlate with measures of early language acquisition, as assessed at age 2 years, in a study of more than 1,000 children from an unselected Australian birth cohort (Whitehouse et al. 2011). Genome-wide association screens are being carried out for reading and language skills in general-population cohorts (e.g., Luciano et al. 2013). It may not be long before psychologists and neuroscientists collect saliva samples from all their study participants as a matter of routine for subsequent analyses using genome-wide genotyping or sequencing. The success of future work in this area will require expertise at the genetic level, with respect to study design (including considerations of sample size and power), analyses, and interpretation. Just as crucially, it will need sophisticated approaches for characterizing the phenotypes of interest in a way that is both practical and meaningful for biological studies. Language is not a simple unitary phenomenon that can be captured with a single test. One might envisage useful cognitive markers being developed from aspects of phonology, morphology, syntax, semantics, and so on, which will depend on close communication between psychologists, linguists, and biologists.

One emerging area with a great deal of potential is that of brain-imaging genetics. With this approach, scientists test for associations between genomic variants and variations in aspects of brain structure or function, assessed with noninvasive neuroimaging techniques. Both genomics and neuroimaging, in isolation, can generate extremely rich, complex data sets. When these techniques are combined for a brain-imaging genetics study, it is essential to constrain the analyses, for example by targeting particular candidate genes of interest and/or defining specific aspects of brain structure or function to focus on. Moreover, it has become clear that, just as observed for other measures of cognition and behavior, the effect sizes of common genomic variants are small and require large samples to achieve sufficient power (Thompson et al. 2014). Imaging genetic studies of language are beginning to be performed, with intriguing results; again, to take the example of *CNTNAP2*, common variants of this gene are reported to be associated with language-related brain activations (Whalley et al. 2011), event-related potentials (Kos et al. 2012), and altered structural connectivity (Dennis et al. 2011) in the healthy brain. However, note that most studies in this area have used small sample sizes, yielding low power and high risk of type I error (false-positive findings). A recent investigation into the impact of common *FOXP2* variants on human brain structure, using a sample of 1,300 people, did not detect any significant associations, contrasting with observations from earlier studies that had used much smaller samples (Hoogman et al 2014).

We are at a turning point in the evolution of the language sciences. We have access to extraordinary technologies and can generate new types of data that we could only have dreamed of before. More importantly, we have exciting opportunities to synthesize approaches, expertise, hypotheses, and findings from different fields can begin to address the big questions about language. By identifying genomic variations that are associated with language-related phenotypes and studying their impacts on neurobiology, we are in a unique position to empirically approach the major issues that we raised at the start of this article, concerning the development, specificity, and origins of our fascinating human capacities. As such, interdisciplinary dialogues between geneticists, neuroscientists, psychologists, and linguists hold enormous promise for the future.

SUMMARY POINTS

1. A basic understanding of molecular biology is essential for researchers interested in the biological basis of language because genetics lies at the heart of questions about ontogeny, specificity, and evolutionary origins.

2. The human genome encodes a large amount of information packaged into genes. Expression of these genes is intricately regulated at multiple levels, increasing the complexity of the system.

3. Diverse gene products (RNA and proteins) influence the fundamental properties of cells in the brain, affecting how those cells develop and function.

4. Gene products mediate aspects of neuronal development and function, including proliferation, migration, neurite outgrowth, and axon guidance, as well as development, maintenance, and plasticity of synapses.

5. Complex behaviors such as language are underpinned by activities of neuronal circuits. The development, connectivity, and plasticity of circuits are driven by genetic factors in a complex interplay with environmental input.

6. Gene disruptions are implicated in developmental disorders involving severe speech and language problems. There is also considerable genetic variability in the population, which is likely to contribute to more subtle differences in ability.

7. Existing genomic variation in the population, from common polymorphisms to rare mutations, can guide us to molecular substrates underlying language, particularly given novel high-throughput genetic technologies.

8. Interaction between disciplines is crucial to reliably connect genotypes and phenotypes. One example is neuroimaging genetics, the success of which depends on robust study design, adequate power, and careful analyses.

FUTURE ISSUES

1. As the field develops, high-throughput genomic technologies will reveal many gene variants that could be candidates for involvement in phenotypic variations related to language. However, it will be highly challenging to robustly identify which genes and variants have real causal links to the traits of interest.

2. Introducing gene variants of interest into human cell lines will allow rapid testing to reveal effects on protein expression, protein function, and cellular phenotypes (e.g., O'Roak et al. 2011). Moreover, skin or blood cells obtained from patients carrying mutations of interest can be induced to look and act like neurons from different parts of the brain (cortical layers, basal ganglia, etc.) with characteristic profiles of gene expression. These cells can then be used to investigate the impact of variants on the function of the protein, as well as the activity and connectivity of neurons (e.g., the ability to form synapses or send/receive signals) (Chailangkarn et al. 2012).

3. Animal models will become increasingly important. Although only humans display the capacity for language, animal models have still proven useful for investigating phenotypic outcomes in the context of the whole brain or whole organism. Functional investigations of *FOXP2* provide a good illustration of this approach. Studies of mice

carrying an equivalent mutation to that observed in the KE family described effects on neurite outgrowth and synaptic plasticity, which affect the brain regions implicated in humans with *FOXP2*-related speech disorders (Groszer et al. 2008, Vernes et al. 2011, French et al. 2012). Furthermore, research with songbirds is proving highly informative in this area, as they display complex learned vocalizations. For example, experimental manipulations of the songbird version of FOXP2 in song-related brain areas of zebra finches severely disturb the structure of the learned song (Haesler et al. 2007).

4. Investigations of language origins will depend on uniting diverse data sets, including those from evolutionary genomics and neuroimaging genetics (see Section 4). For example, screening the genome for relationships between genomic variants and speech/language phenotypes in modern human populations will yield lists of candidate genes that may have mechanistic connections to linguistic pathways. These findings can be integrated with emerging information on evolutionary differences between our species, Neanderthals, Denisovans, and other primates (Paabo 2014). Cross-referencing between these different types of data might be valuable, for example, for determining which of the candidate genes show interesting patterns of change in our ancestors, in the context of multifactorial models of language evolution (Fisher & Marcus 2006, Fisher & Ridley 2013).

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56–65

Alarcón M, Abrahams BS, Stone JL, Duvall JA, Perederiy JV, et al. 2008. Linkage, association, and gene-expression analyses identify *CNTNAP2* as an autism-susceptibility gene. *Am. J. Hum. Genet.* 82:150–59

Ayub Q, Yngvadottir B, Chen Y, Xue Y, Hu M, et al. 2013. *FOXP2* targets show evidence of positive selection in European populations. *Am. J. Hum. Genet.* 92:696–706

Bacon C, Rappold GA. 2012. The distinct and overlapping phenotypic spectra of *FOXP1* and *FOXP2* in cognitive disorders. *Hum. Genet.* 131:1687–98

Bashaw GJ, Klein R. 2010. Signaling from axon guidance receptors. *Cold Spring Harb. Perspect. Biol.* 2:a001941

Bena F, Bruno DL, Eriksson M, van Ravenswaaij–Arts C, Stark Z, et al. 2013. Molecular and clinical characterization of 25 individuals with exonic deletions of *NRXN1* and comprehensive review of the literature. *Am. J. Med. Genet. B* 162:388–403

Berwick RC, Friederici AD, Chomsky N, Bolhuis JJ. 2013. Evolution, brain, and the nature of language. *Trends Cogn. Sci.* 17:89–98

Bishop DV. 1997. Cognitive neuropsychology and developmental disorders: uncomfortable bedfellows. *Q. J. Exp. Psychol. A* 50:899–923

Bishop DV. 2001. Genetic and environmental risks for specific language impairment in children. *Philos. Trans. R. Soc. Lond. B* 356:369–80

Bishop DV, Nation K, Patterson K. 2014. When words fail us: insights into language processing from developmental and acquired disorders. *Philos. Trans. R. Soc. Lond. B* 369:20120403

Boucard AA, Chubykin AA, Comoletti D, Taylor P, Südhof TC. 2005. A splice code for trans–synaptic cell adhesion mediated by binding of neuroligin 1 to α- and β-neurexins. *Neuron* 48:229–36

Bourgeron T. 2009. A synaptic trek to autism. *Curr. Opin. Neurobiol.* 19:231–34

Carrion-Castillo A, Franke B, Fisher SE. 2013. Molecular genetics of dyslexia: an overview. *Dyslexia* 19:214–40

Chahrour M, Zoghbi HY. 2007. The story of Rett syndrome: from clinic to neurobiology. *Neuron* 56:422–37

Chailangkarn T, Acab A, Muotri AR. 2012. Modeling neurodevelopmental disorders using human neurons. *Curr. Opin. Neurobiol.* 22:785–90

Chedotal A, Richards LJ. 2010. Wiring the brain: the biology of neuronal guidance. *Cold Spring Harb. Perspect. Biol.* 2:a001917

Chia PH, Li P, Shen K. 2013. Cell biology in neuroscience: cellular and molecular mechanisms underlying presynapse formation. *J. Cell Biol.* 203:11–22

Ching MS, Shen Y, Tan WH, Jeste SS, Morrow EM, et al. 2010. Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders. *Am. J. Med. Genet. B* 153:937–47

Chomsky N. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press. 251 pp.

Chomsky N. 2011. Language and other cognitive systems. What is special about language? *Lang. Learn. Dev.* 7:263–78

Christiansen MH, Kirby S. 2003. *Language Evolution.* Oxford, UK/New York: Oxford Univ. Press. 395 pp.

Cross-Disord. Group Psychiatr. Genomics Consort. 2013. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 381:1371–79

Crow TJ. 1997. Is schizophrenia the price that *Homo sapiens* pays for language? *Schizophr. Res.* 28:127–41

Damasio AR, Geschwind N. 1984. The neural basis of language. *Annu. Rev. Neurosci.* 7:127–47

Darwin C 1871. *The Descent of Man, and Selection in Relation to Sex*. New York: D. Appleton & Co.

Day JJ, Sweatt JD. 2011. Epigenetic mechanisms in cognition. *Neuron* 70:813–29

Dennis EL, Jahanshad N, Rudie JD, Brown JA, Johnson K, et al. 2011. Altered structural brain connectivity in healthy carriers of the autism risk gene, *CNTNAP2. Brain Connect.* 1:447–59

Deriziotis P, Fisher SE. 2013. Neurogenomics of speech and language disorders: the road ahead. *Genome Biol.* 14:204

Devanna P, Vernes SC. 2014. A direct molecular link between the autism candidate gene *RORa* and the schizophrenia candidate *MIR137. Sci. Rep.* 4:3994

Ebert DH, Greenberg ME. 2013. Activity-dependent neuronal signalling and autism spectrum disorder. *Nature* 493:327–37

Feuk L, Kalervo A, Lipsanen-Nyman M, Skaug J, Nakabayashi K, et al. 2006. Absence of a paternally inherited *FOXP2* gene in developmental verbal dyspraxia. *Am. J. Hum. Genet.* 79:965–72

Fisher SE. 2006. Tangled webs: tracing the connections between genes and cognition. *Cognition* 101:270–97

Fisher SE. 2013. Building bridges between genes, brains and language. In *Birdsong, Speech, and Language: Exploring the Evolution of Mind and Brain*, ed. JJ Bolhuis, M Everaert, pp. 425–54. Cambridge, MA:MIT Press

Fisher SE, DeFries JC. 2002. Developmental dyslexia: genetic dissection of a complex cognitive trait. *Nat. Rev. Neurosci.* 3:767–80

Fisher SE, Marcus GF. 2006. The eloquent ape: genes, brains and the evolution of language. *Nat. Rev. Genet.* 7:9–20

Fisher SE, Ridley M. 2013. Culture, genes, and the human revolution. *Science* 340:929–30

Fisher SE, Scharff C. 2009. *FOXP2* as a molecular window into speech and language. *Trends Genet.* 25:166–77

Fogel BL, Wexler E, Wahnich A, Friedrich T, Vijayendran C, et al. 2012. *RBFOX1* regulates both splicing and transcriptional networks in human neuronal development. *Hum. Mol. Genet.* 21:4171–86

Fonseca-Azevedo K, Herculano-Houzel S. 2012. Metabolic constraint imposes tradeoff between body size and number of brain neurons in human evolution. *Proc. Natl. Acad. Sci. USA* 109:18571–76

Forrest M, Chapman RM, Doyle AM, Tinsley CL, Waite A, Blake DJ. 2012. Functional analysis of TCF4 missense mutations that cause Pitt–Hopkins syndrome. *Hum. Mutat.* 33:1676–86

French CA, Jin X, Campbell TG, Gerfen E, Groszer M, et al. 2012. An aetiological *Foxp2* mutation causes aberrant striatal activity and alters plasticity during skill learning. *Mol. Psychiatry* 17:1077–85

Galaburda AM, Kemper TL. 1979. Cytoarchitectonic abnormalities in developmental dyslexia: a case study. *Ann. Neurol.* 6:94–100

Galaburda AM, Sherman GF, Rosen GD, Aboitiz F, Geschwind N. 1985. Developmental dyslexia: four consecutive patients with cortical anomalies. *Ann. Neurol.* 18:222–33

Gao P, Sultan KT, Zhang XJ, Shi SH. 2013. Lineage-dependent circuit assembly in the neocortex. *Development* 140:2645–55

Gehman LT, Stoilov P, Maguire J, Damianov A, Lin CH, et al. 2011. The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. *Nat. Genet.* 43:706–11

Gialluisi A, Newbury DF, Wilcutt EG, Olson RK, DeFries JC, et al. 2014. Genome-wide screening DNA variants associated with reading and language traits. *Genes Brain Behav* 13:686–701

Goldie BJ, Cairns MJ. 2012. Post-transcriptional trafficking and regulation of neuronal gene expression. *Mol. Neurobiol.* 45:99–108

Gonda Y, Andrews WD, Tabata H, Namba T, Parnavelas JG, et al. 2013. Robo1 regulates the migration and laminar distribution of upper-layer pyramidal neurons of the cerebral cortex. *Cereb. Cortex* 23:1495–508

Gopnik M. 1997. Language deficits and genetic factors. *Trends Cogn. Sci.* 1:5–9

Grabowski P. 2011. Alternative splicing takes shape during neuronal development. *Curr. Opin. Genet. Dev.* 21:388–94

Graham SA, Fisher SE. 2013. Decoding the genetics of speech and language. *Curr. Opin. Neurobiol.* 23:43–51

Gregor A, Albrecht B, Bader I, Bijlsma EK, Ekici AB, et al. 2011. Expanding the clinical spectrum associated with defects in *CNTNAP2* and *NRXN1*. *BMC Med. Genet.* 12:106

Groszer M, Keays DA, Deacon RM, de Bono JP, Prasad-Mulcare S, et al. 2008. Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits. *Curr. Biol.* 18:354–62

Ha M. 2013. Understanding the chromatin remodeling code. *Plant Sci.* 211:137–45

Haesler S, Rochefort C, Georgi B, Licznerski P, Osten P, Scharff C. 2007. Incomplete and inaccurate vocal imitation after knockdown of *FoxP2* in songbird basal ganglia nucleus Area X. *PLOS Biol.* 5:e321

Hamdan FF, Daoud H, Rochefort D, Piton A, Gauthier J, et al. 2010. De novo mutations in *FOXP1* in cases with intellectual disability, autism, and language impairment. *Am. J. Hum. Genet.* 87:671–78

Heijmans BT, Tobi EW, Stein AD, Putter H, Blauw GJ, et al. 2008. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc. Natl. Acad. Sci. USA* 105:17046–49

Holtmaat A, Svoboda K. 2009. Experience-dependent structural synaptic plasticity in the mammalian brain. *Nat. Rev. Neurosci.* 10:647–58

Hong EJ, West AE, Greenberg ME. 2005. Transcriptional control of cognitive development. *Curr. Opin. Neurobiol.* 15:21–28

Hoogman M, Guadalupe T, Zwiers MP, Klarenbeek P, Francks C, Fisher SE. 2014. Assessing the effects of common variation in the FOXP2 gene on human brain structure. *Front. Hum. Neurosci.* 8:473

Horn, D. 2011. Mild to moderate intellectual disability and significant speech and language deficits in patients with FOXP1 deletions and mutations. *Mol. Syndromol.* 2:213–16

Jung H, Holt CE. 2011. Local translation of mRNAs in neural development. *Wiley Interdiscip. Rev. RNA* 2:153–65

Kim M, Roesener AP, Mendonca PR, Mastick GS. 2011. Robo1 and Robo2 have distinct roles in pioneer longitudinal axon guidance. *Dev. Biol.* 358:181–88

Klein RG, Edgar B. 2002. *The Dawn of Human Culture*. New York: Wiley. 288 pp.

Kolodkin AL, Tessier-Lavigne M. 2011. Mechanisms and molecules of neuronal wiring: a primer. *Cold Spring Harb. Perspect. Biol.* 3:a001727

Kos M, van den Brink D, Snijders TM, Rijpkema M, Franke B, et al. 2012. *CNTNAP2* and language processing in healthy individuals as measured with ERPs. *PLOS ONE* 7:e46995

Kuhl PK. 2010. Brain mechanisms in early language acquisition. *Neuron* 67:713–27

Kwon E, Wang W, Tsai LH. 2013. Validation of schizophrenia-associated genes *CSMD1*, *C10orf26*, *CACNA1C* and *TCF4* as miR-137 targets. *Mol. Psychiatry* 18:11–12

Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. 2001. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413:519–23

Lappalainen T, Sammeth M, Friedländer MR, 't Hoen PA, Monlong J, et al. 2013. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 501:506–11

Loebrich S, Nedivi E. 2009. The function of activity-regulated genes in the nervous system. *Physiol. Rev.* 89:1079–103

Luciano M, Evans DM, Hansell NK, Medland SE, Montgomery GW, et al. 2013. A genome-wide association study for reading and language abilities in two population cohorts. *Genes Brain Behav.* 12:645–52

MacDermot KD, Bonora E, Sykes N, Coupe AM, Lai CS, et al. 2005. Identification of *FOXP2* truncation as a novel cause of developmental speech and language deficits. *Am. J. Hum. Genet.* 76:1074–80

Manzini MC, Walsh CA. 2011. What disorders of cortical development tell us about the cortex: One plus one does not always make two. *Curr. Opin. Genet. Dev.* 21:333–39

Marin O, Valiente M, Ge X, Tsai LH. 2010. Guiding neuronal cell migrations. *Cold Spring Harb. Perspect. Biol.* 2:a001834

McNeill E, Van Vactor D. 2012. MicroRNAs shape the neuronal landscape. *Neuron* 75:363–79

Metzker ML. 2010. Sequencing technologies—the next generation. *Nat. Rev. Genet.* 11:31–46

Navarrete K, Pedroso I, De Jong S, Stefansson H, Steinberg S, et al. 2013. *TCF4* (*e2-2; ITF2*): a schizophrenia-associated gene with pleiotropic effects on human disease. *Am. J. Med. Genet. B* 162:1–16

Newbury DF, Winchester L, Addis L, Paracchini S, Buckingham LL, et al. 2009. *CMIP* and *ATP2C2* modulate phonological short-term memory in language impairment. *Am. J. Hum. Genet.* 85:264–72

Nudel R, Simpson NH, Baird G, O'Hare A, Conti-Ramsden G, et al. 2014. Genome-wide association analyses of child genotype effects and parent-of-origin effects in specific language impairment. *Genes Brain Behav.* 13:418–29

O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, et al. 2011. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat. Genet.* 43:585–89

Ooi L, Wood IC. 2008. Regulation of gene expression in the nervous system. *Biochem. J.* 414:327–41

Paabo S. 2014. The human condition—a molecular approach. *Cell* 157:216–26

Palumbo O, D'Agruma L, Minenna AF, Palumbo P, Stallone R, et al. 2013. 3p14.1 de novo microdeletion involving the *FOXP1* gene in an adult patient with autism, severe speech delay and deficit of motor coordination. *Gene* 516:107–13

Paracchini S, Scerri T, Monaco AP. 2007. The genetic lexicon of dyslexia. *Annu. Rev. Genomics Hum. Genet.* 8:57–79

Park HT, Wu J, Rao Y. 2002. Molecular control of neuronal migration. *BioEssays* 24:821–27

Paterson SJ, Brown JH, Gsödl MK, Johnson MH, Karmiloff-Smith A. 1999. Cognitive modularity and genetic disorders. *Science* 286:2355–58

Piton A, Gauthier J, Hamdan FF, Lafrenière RG, Yang Y, et al. 2011. Systematic resequencing of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia. *Mol. Psychiatry* 16:867–80

Poeppel D, Hickok G. 2004. Towards a new functional anatomy of language. *Cognition* 92:1–12

Polleux F, Snider W. 2010. Initiating and growing an axon. *Cold Spring Harb. Perspect. Biol.* 2:a001925

Poon MM, Choi SH, Jamieson CA, Geschwind DH, Martin KC. 2006. Identification of process-localized mRNAs from cultured rodent hippocampal neurons. *J. Neurosci.* 26:13390–99

Rice GM, Raca G, Jakielski KJ, Laffin JJ, Iyama-Kurtycz CM, et al. 2012. Phenotype of *FOXP2* haploinsufficiency in a mother and son. *Am. J. Med. Genet. A* 158:174–81

Rodenas-Cuadrado P, Ho J, Vernes SC. 2014. Shining a light on *CNTNAP2*: complex functions to complex disorders. *Eur. J. Hum. Genet.* 22:171–78

Roeske D, Ludwig KU, Neuhoff N, Becker J, Bartling J, 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

Rothbart SB, Strahl BD. 2014. Interpreting the language of histone and DNA modifications. *Biochim. Biophys. Acta* 1839:627–43

Scerri TS, Schulte-Körne G. 2010. Genetics of developmental dyslexia. *Eur. Child Adolesc. Psychiatry* 19:179–97

Shandilya J, Roberts SG. 2012. The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochim. Biophys. Acta* 1819:391–400

Shriberg LD, Ballard KJ, Tomblin JB, Duffy JR, Odell KH, Williams CA. 2006. Speech, prosody, and voice characteristics of a mother and daughter with a 7;13 translocation affecting *FOXP2*. *J. Speech Lang. Hear. Res.* 49:500–25

Smrt RD, Szulwach KE, Pfeiffer RL, Li X, Guo W, et al. 2010. MicroRNA miR-137 regulates neuronal maturation by targeting ubiquitin ligase mind bomb-1. *Stem Cells* 28:1060–70

Spiteri E, Konopka G, Coppola G, Bomar J, Oldham M, 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.* 81:1144–57

Südhof TC. 2008. Neuroligins and neurexins link synaptic function to cognitive disease. *Nature* 455:903–11

Sweatt JD. 2013. Pitt–Hopkins syndrome: intellectual disability due to loss of TCF4-regulated gene transcription. *Exp. Mol. Med.* 45:e21

Szalontai A, Csiszar K. 2013. Genetic insights into the functional elements of language. *Hum. Genet.* 132:959–86

Szymanski M, Barciszewski J. 2007. The genetic code—40 years on. *Acta Biochim. Pol.* 54:51–54

Tan X, Shi SH. 2013. Neocortical neurogenesis and neuronal migration. *Wiley Interdiscip. Rev. Dev. Biol.* 2:443–59

Thompson PM, Stein JL, Medland SE, Hibar DP, Vasquez AA, et al. 2014. The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav.* 8:153–82

Turner SJ, Hildebrand MS, Block S, Damiano J, Fahey M, et al. 2013. Small intragenic deletion in *FOXP2* associated with childhood apraxia of speech and dysarthria. *Am. J. Med. Genet. A* 161:2321–26

Ullrich B, Ushkaryov YA, Südhof TC. 1995. Cartography of neurexins: more than 1,000 isoforms generated by alternative splicing and expressed in distinct subsets of neurons. *Neuron* 14:497–507

van Bokhoven H. 2011. Genetic and epigenetic networks in intellectual disabilities. *Annu. Rev. Genet.* 45:81–104

van der Lely HK, Rosen S, McClelland A. 1998. Evidence for a grammar-specific deficit in children. *Curr. Biol.* 8:1253–58

Vernes SC, Fisher SE. 2009. Unravelling neurogenetic networks implicated in developmental language disorders. *Biochem. Soc. Trans.* 37:1263–69

Vernes SC, Fisher SE. 2013. Genetic pathways implicated in speech and language. In *Animal Models of Speech and Language Disorders*, ed. S Helekar, pp. 13–40. New York: Springer

Vernes SC, Newbury DF, Abrahams BS, Winchester L, Nicod J, et al. 2008. A functional genetic link between distinct developmental language disorders. *N. Engl. J. Med.* 359:2337–45

Vernes SC, Oliver PL, Spiteri E, Lockstone HE, Puliyadi R, et al. 2011. *Foxp2* regulates gene networks implicated in neurite outgrowth in the developing brain. *PLOS Genet.* 7:e1002145

Vernes SC, Spiteri E, Nicod J, Groszer M, Taylor JM, 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.* 81:1232–50

Visscher PM, Brown MA, McCarthy MI, Yang J. 2012. Five years of GWAS discovery. *Am. J. Hum. Genet.* 90:7–24

Watkins KE, Dronkers NF, Vargha-Khadem F. 2002. Behavioural analysis of an inherited speech and language disorder: comparison with acquired aphasia. *Brain* 125:452–64

Whalley HC, O'Connell G, Sussmann JE, Peel A, Stanfield AC, et al. 2011. Genetic variation in *CNTNAP2* alters brain function during linguistic processing in healthy individuals. *Am. J. Med. Genet. B* 156:941–48

White SA, Fisher SE, Geschwind DH, Scharff C, Holy TE. 2006. Singing mice, songbirds, and more: models for *FOXP2* function and dysfunction in human speech and language. *J. Neurosci.* 26:10376–79

Whitehouse AJ, Bishop DV, Ang QW, Pennell CE, Fisher SE. 2011. *CNTNAP2* variants affect early language development in the general population. *Genes Brain Behav.* 10:451–56

Willemsen MH, Vallès A, Kirkels LA, Mastebroek M, Olde Loohuis N, et al. 2011. Chromosome 1p21.3 microdeletions comprising *DPYD* and *MIR137* are associated with intellectual disability. *J. Med. Genet.* 48:810–18

Worthey EA, Raca G, Laffin JJ, Wilk BM, Harris JM, et al. 2013. Whole-exome sequencing supports genetic heterogeneity in childhood apraxia of speech. *J. Neurodev. Disord.* 5:29

Wright C, Turner JA, Calhoun VD, Perrone-Bizzozero N. 2013. Potential impact of miR-137 and its targets in schizophrenia. *Front. Genet.* 4:58

Zhang W, Rohlmann A, Sargsyan V, Aramuni G, Hammer RE, et al. 2005. Extracellular domains of α-neurexins participate in regulating synaptic transmission by selectively affecting N- and P/Q-type Ca^{2+} channels. *J. Neurosci.* 25:4330–42

Zheng S, Black DL. 2013. Alternative pre-mRNA splicing in neurons: growing up and extending its reach. *Trends Genet.* 29:442–48

Zivraj KH, Tung YC, Piper M, Gumy L, Fawcett JW, et al. 2010. Subcellular profiling reveals distinct and developmentally regulated repertoire of growth cone mRNAs. *J. Neurosci.* 30:15464–78

Zweier C. 2012. Severe intellectual disability associated with recessive defects in *CNTNAP2* and *NRXN1*. *Mol. Syndromol.* 2:181–85

RELATED RESOURCES

Griffiths AJF, Miller JH, Suzuki DT, Lewontin RC, Gelbart WM. 2000. *An Introduction to Genetic Analysis.* New York: W.H. Freeman. 7th ed.

Strachan T, Read AP. 2012. *Human Molecular Genetics.* New York: Wiley-Liss. 4th ed.

Contents

Errata

An online log of corrections to *Annual Review of Linguistics* articles may be found
at http://www.annualreviews.org/errata/linguistics