Published in final edited form as:

J Dev Behav Pediatr. 2010 September ; 31(7): 555–563. doi:10.1097/DBP.0b013e3181ee3d9e.

Etiologies and molecular mechanisms of communication disorders

Shelley D. Smith, Ph.D.,

Department of Pediatrics, Munroe Meyer Institute for Genetics and Rehabilitation, University of Nebraska Medical Center

Elena Grigorenko, Ph.D.,

Child Study Center, Department of Psychology, Department of Epidemiology & Public Health, Yale University

Erik Willcutt, Ph.D.,

Department of Psychology and Neuroscience, Institute for Behavioral Genetics, University of Colorado

Bruce F. Pennington, Ph.D.,

Department of Psychology, University of Denver

Richard K. Olson, Ph.D., and

Department of Psychology and Neuroscience, Institute for Behavioral Genetics, University of Colorado

John C. DeFries, Ph.D.

Department of Psychology and Neuroscience, University of Colorado

Abstract

Quantitative behavioral genetic studies have made it clear that communication disorders such as reading disability (RD), language impairment (LI), and autism spectrum disorders (ASD) follow some basic principles: 1) Complex disorders have complex causes, in that each clinical disorder is influenced by a number of separate genes; and 2) at least some behaviorally related disorders are influenced by the same genes. Recent advances in molecular and statistical methods have confirmed these principles and are now leading to an understanding of the genes that may be involved in these disorders and how their disruption may affect the development of the brain. The prospect is that the genes involved in these disorders will define a network of interacting neurologic functions, and that perturbations of different elements of this network will produce susceptibilities for different disorders. Such knowledge would clarify the underlying deficits in these disorders and could lead to revised diagnostic conceptions. These goals are still in the future, however. Identifying the individual genes in such a network is painstaking, and there have been seemingly contradictory studies along the way. Improvements in study design and additional functional analysis of genes is gradually clarifying many of these issues. When combined with careful phenotypic studies, molecular genetic studies have the potential to refine the clinical definitions of communication disorders and influence their remediation.

Introduction

Reading Disability (RD), Language Impairment (LI), and Autism Spectrum Disorders (ASD) have distinct clinical definitions, but these definitions have components which overlap at the cognitive level and the disorders are often comorbid in children. They also share similarities at the genetic level, with comparable degrees of heritability and modes of inheritance.

Discovery of cognitive and genetic commonalities between these disorders could help redefine their diagnostic criteria and the optimal means of remediation based on etiology.

One method for determining the developmental mechanism behind a disorder is to identify the genes that influence it. Knowledge of a gene's function can help determine which aspects of brain development and operation are affected. An individual gene may participate in multiple activities in different parts of the brain, however, so pinpointing the function that influences a particular disorder becomes more apparent when multiple genes are identified, highlighting a network of genes forming a developmental or operational pathway. This type of approach has been effective in highlighting the importance of neuronal migration in the etiology of reading disability and of synaptic connections in autism spectrum disorders, as described below.

Disorders which share cognitive characteristics or which show comorbidity may also share common genetic pathways accounting for these overlaps. As these networks of genes and pathways are identified, there will be greater understanding of the neurodevelopmental functions and structures that are important for language and learning.

Genetic Methodology and Definitions

Identification of genes controlling the manifestation of a disorder is critically dependent upon careful definition of the disorder, a sufficient sample size, and appropriate genetic and statistical methods. This can be particularly challenging for complex disorders, where diagnosis and measurement may differ between studies and inheritance is due to multiple genes and environmental effects rather than a single gene dominant or recessive mechanism.

Phenotype definition

The definition of a disorder may be at several levels, including a clinical definition based on symptoms, such as DSM-IV criteria for diagnosis of ASD resulting in a designation of affected or unaffected¹, or a quantitative measure on a test or group of tests, such as a reading test or a composite factor score. Endophenotypes are defined as measures that reflect the underlying etiology of a disorder, such as an ability thought to be an essential to a process (eg phonological awareness for reading). The assumption is that endophenotypes are closer to the gene effect and are less likely to be influenced by environmental factors, and thus may make better phenotypes for genetic studies. In all cases, the phenotypes or endophenotypes need to be an accurate reflection of the genetic makeup. This can be determined by estimation of the heritability of the phenotype, symbolized as h^2 , which measures the proportion of variation in the phenotype that is due to additive genetic factors. In the specific case of disabilities, the heritability of the group deficit, h^2_g , measures the genetic contribution to the deficit in performance. Phenotypes in which most of the variation is due to genetic factors are better choices for gene identification studies. Since heritability is the proportion of variation that is due to genetic factors, that proportion may differ in different populations, however; for example, heritability may be higher in more homogeneous populations with better access to education, simply due to the decreased environmental influence of poor education. Other indicators of genetic influence include a high sibling relative risk (λ_s) which is the ratio of the risk of the disorder in siblings of a probands compared to the prevalence in the general population, or evidence for segregation of the disorder in families, particularly if segregation analysis shows a major gene effect or the involvement of only a few genes. Estimates of heritability are often done through twin studies, and other measures require family studies.

Gene localization: Linkage analysis

When the genes influencing a disorder are unknown, linkage analysis provides a means of locating a region on a chromosome containing such a gene. Further studies can be done to identify the gene, determine its function, and find mutations that affect its function or degree

of expression. Linkage analysis takes advantage of the fact that genes are strung along DNA strands in a linear fashion, and DNA strands are packaged into individual chromosomes. Most genetic regions contain DNA variations, and the variations in a gene or a DNA position are termed alleles. All chromosomes come in pairs, with one inherited from one parent and the other from the other parent. Two genes that are close together on the same chromosome will tend to be inherited together from parent to child, although it is possible for genes on the same chromosome to swap positions with their partner genes on the paired chromosome. This process recombines the ancestral set of alleles of the linked genes. The incidence of recombination between genes is related to the distance between them, with smaller distances related to lower frequency of recombination. If genes are far apart or are on different chromosomes, their alleles will recombine randomly, following Mendel's law of individual assortment. In linkage studies, DNA locations (loci) across the chromosomes are "marked" at specific points by sites of DNA variation, and the inheritance of sets of marker alleles (generally microsatellite markers with of different numbers of repeated nucleotides) is compared to the inheritance of trait phenotypes to determine if any marker alleles are inherited along with the disorder phenotype. If a marker and a phenotype are inherited together in nonrandom fashion to a statistically significant extent, it is taken as evidence that a gene influencing the phenotype is located close to the marker. Traditionally, the level of statistical significance required for linkage analysis has been a LOD score (log of the odds of likelihood of linkage) of at least 3, roughly equivalent to a p-value of 0.001. As the number of marker loci genotyped across the chromosomes increases, however, the requirements for significance of linkage are increased to account for the greater number of tests. Linkage studies can be done with multigenerational families, which are particularly effective for major genes and in cases where the parameters of mode of inheritance and allele frequency are known or can be accurately estimated. If those parameters are unknown, studies with pairs of siblings can be used, with or without parental DNA.

The regions defined by linkage analysis are generally fairly large, containing many possible candidate genes. The region can be narrowed by a second round of linkage analysis with an additional, closely spaced markers in the region, but ultimately linkage studies will be limited by the decreasing frequency of recombination as markers are tested which are closer to the trait locus. In that case, further evaluation of candidate genes in a region can be done with association analysis or through direct DNA sequencing.

Gene localization: Association analysis

In linkage analysis, the marker allele that is linked to the trait allele (the causal mutation) may be different across families but is consistent within families. Association is a special case of linkage in which the marker alleles linked to the causal mutation are the same across families. This implies that there have been very few founder mutations in a population and that there have been few instances of recombination between the mutation and the marker (linkage disequilibrium), or the marker may itself be causal. The genetic markers that are used for these studies are generally single nucleotide polymorphisms (SNPs) since they occur much more frequently across the chromosomes and thus are more likely to be tightly linked to a mutation. Association can be detected through case/control studies, thus decreasing the cost of gathering and genotyping relatives, as long as the cases and controls are drawn from the same genetic background. Ethnic stratification, in which cases and controls come from different backgrounds, can produce spurious results based on differences in allele frequencies that are not due to the differences in the trait phenotype. Genetic isolates are ideal for these studies because of the limited number of founders and the homogeneous genetic background. Genes discovered in isolated populations may not reflect the causes in other populations, however, and may not be generalizable to populations other than the population in which they were identified. Family-based association analyses have been used in more diverse populations to

try to mitigate the problems of ethnic stratification. In these studies, the transmission of parental alleles to an affected child (or children) is tested across families.

The ease of genotyping large numbers of SNPs and the development of statistical analyses that can handle large amounts of data have led to Genome-Wide Association Studies (GWAS) using a million or more markers and thousands of affected individuals matched with unaffected controls. These studies have successfully identified genes influencing complex traits such as Type 2 diabetes and schizophrenia.

While linkage and association analysis have been very successful in identifying some loci for complex traits, large sample sizes with hundreds or thousands of subjects are often necessary to obtain sufficient statistical power for reliable results. Genetic heterogeneity can seriously weaken a study, since it will be more difficult to detect multiple genes with smaller effect sizes vs. a few genes with greater influence. This is particularly true in genome-wide studies. The use of a dense panel of SNPs in a GWAS has the advantage of covering nearly all of the genome, but the corresponding disadvantage is in the high false positive rate for multiple testing at that level. For some disorders, very large sample sizes are needed to obtain statistically significant results that withstand corrections for multiple testing. Since these are sometimes impractical, the field has had to rely on replication studies to sift out false positive results. The replication studies also need to be carefully designed and have sufficient power, since there can be additional reasons for lack of replication, including genetic differences between populations, differences in phenotypes, or difference in analysis methodology. This can result in an uncomfortable state of limbo while an initial finding undergoes further evaluation.

Identification of Candidate Genes

Since SNP markers are so frequent across the genome, they can target individual genes. If necessary, further fine mapping with an even more dense panel of SNP markers covering a gene and its regulatory regions can be used to identify associations within a candidate gene.

DNA sequencing can be used to identify mutations in the gene that disrupt its function or regulation. Mutations that disrupt function are generally in the coding regions of a gene, the regions that determine the structure of the gene product, and are usually rare in the general population. These mutations may be obvious, in that they cause changes that would prematurely terminate a protein or interfere with its function. In general, mutations in a candidate gene that significantly disrupt a gene product are more likely to have a major phenotypic effect, and the phenotypes they produce are more likely to be transmitted in a Mendelian (dominant or recessive) manner, but mutations which cause subtle protein changes or changes in the amount of protein produced may have milder effects. In other cases it can be difficult to distinguish a deleterious mutation from a benign variation in DNA code, and further functional studies in cells or in animal studies are needed to determine if the sequence change has an effect. Benign changes are usually more common, and are termed polymorphisms if they occur in greater than 1% of chromosomes in the general population; otherwise, they are termed rare variants. The frequency of a variant does not automatically determine its pathogenicity, however; some polymorphisms may turn out to contribute to susceptibility to a disorder, and some rare variants may be benign. In general, however, deleterious variants are more likely to be rare if there is evolutionary pressure against them. In complex conditions such as communication disorders, deleterious mutations may not be found in coding regions of genes; in those cases, they are presumed to be in regulatory regions surrounding the gene. These regions are not always well characterized, so it can be particularly difficult to determine if sequence changes are deleterious or benign. A current collaborative mass sequencing effort called the 1000 Genomes Project (<http://www.1000genomes.org>) will sequence the entire genomes of 1000 individuals from different populations around the world, which should give a better idea of common genetic variation in genes and regulatory regions.

Gene expression and function

Confirmation that a gene is involved in the etiology of a genetic condition requires that statistically significant association is confirmed in an independent population, a causal mutation in the gene or regulatory region is found that affects the phenotype, and/or disruption of the gene in a model system which produces a phenotype that is consistent with the disorder. Several methods are available to examine gene function by manipulating the sequence of the gene or the degree of expression in a model system, which may be an *in vitro* system such as tissue culture or an *in vivo* system such as a genetically engineered mouse. An *in vivo* model is more difficult, but it is also a more accurate representation of the genetic and developmental surroundings that a gene finds itself in. The expression of the gene can be eliminated completely in a “knock out” procedure which introduces a mutation that completely disrupts the gene, and then the resulting phenotype can be examined. Total elimination of a gene in all tissues at all times may be too drastic, so knock outs can be made to be tissue-specific through the inclusion of a promoter sequence that is only active in that tissue. Conditional knock-outs can be structured so that genes are turned on or off at specific points in development by the administration of a trigger substance such as tetracycline. More realistic representations of human mutations can be obtained by introduction of a transgene, a mutant form of the gene that is injected into the model system, or the expression of a gene can be decreased in a “knock down” procedure where small interfering RNAs are introduced which decrease but do not eliminate the production of the gene product.

For a communication disorder, the use of model systems such as a mouse model can be problematic, since the behavioral phenotype cannot be duplicated. As more is learned about the neurobiological characteristics of these conditions, however, phenotypes such as specific changes in brain development can be examined. Despite the similarities between humans and animal models, though, a gene may not produce the same phenotype in a mouse as it does in humans. The presence of other genes which modify the expression of the gene may explain some of these discrepancies, so it can be important to test gene function in several different mouse strains or in other animal models before deciding that a gene is irrelevant to the human phenotype.

Copy Number Variation

Copy number variants (CNVs) are a form of DNA structural variation in which stretches of DNA are deleted or duplicated. These stretches can range in length from a thousand to several million base pairs and the larger ones can contain several genes and regulatory regions. CNV discovery studies have estimated that there are around 8000-9,000 regions in the human genome subject to copy number variation, with a median number per individual around 900-1000². They are too small to be visible under the microscope, but can be detected through Comparative Genome Hybridization (CGH) arrays or through analysis of stretches of missing or extra SNPs. As with other types of DNA variation, many CNVs are benign, but some may disrupt a gene or regulatory region and contribute to a disorder. Interestingly, there seems to be preferential involvement of CNVs in genes involving processes of communication between cells². Accordingly, CNVs have been detected in studies of autism, with increased numbers of CNVs occurring around genes involved in synaptogenesis and postsynaptic densities^{3,4}, and CNVs associated with ADHD were primarily involved in neurodevelopment⁵. There are important differences among CNVs that affect the ability to generalize about their effects, however; differences in size, inheritance, frequency, and the mechanism of generation of CNVs can be related to the probability that CNVs are pathogenic. Copy number variations can be duplications or deletions, and in some cases, it appears that deletions are more deleterious than duplications^{5,6}. The presence of short stretches of repeated sequences appear to predispose an area to copy number variation^{2,7} by promoting mispairing of the homologous chromosomes in meiosis. These abnormal recombination events primarily result in deletions, but also

duplications⁸. Some (but not all) longer CNVs appear to be due to different mechanisms based on errors in DNA repair or replication rather than recombination; such variants are more likely to be rare and pathogenic^{8,9}. *De novo* CNVs, representing a new mutation, are also presumed to be more pathogenic. Conversely, some CNVs are inherited and fairly common, and even show linkage disequilibrium with adjacent SNPs, indicating stable inheritance. These CNVs are more likely to be benign or related to less deleterious disorders¹⁰ and since they are associated with common SNPs, common CNVs are unlikely to add new information about possible genes².

Further complicating the interpretation of CNVs is the finding that CNVs with the same breakpoints can be associated with multiple phenotypes including autism, schizophrenia, and intellectual disability. This variation may be due to the presence of genetic modifiers or environmental influences^{3,6,11}. Moreover, studies of phenotypically concordant and discordant MZ twins have pointed out that somatic mutation may be an additional source of phenotypic variation. In studies of a neurodegenerative phenotype, phenotypically discordant MZ twins were shown to be discordant for a CNV, demonstrating that the variation in copy number occurred somatically after the split of the twins (or, conversely, there was a spontaneous normalization of a CNV in one twin). The discordant CNV thus points to a genetic region which could be important to the etiology of the phenotype in the affected twin. Lack of a CNV in one twin of a pair of phenotypically concordant MZ twins further implies that somatic mosaicism can exist; the CNV was not present in the DNA from the blood sample from both of the twins, but presumably was present in the appropriate tissues of the brain in both¹². The proportion of cells which carry a CNV may also contribute to variability in the severity or expression of the resulting phenotype. As with any sequence variation that does not obviously disrupt a gene product, it can be difficult to determine if a CNV is causal, coincidental, or in linkage disequilibrium with a nearby variation that is the true cause. Merikangas et al. (2009) have also pointed out problems in studies of CNVs in disease, including overlapping clinical samples, incomplete phenotypic information, inappropriate controls, variation in CNV detection methods, and unknown baseline frequencies of specific CNVs in the general population. Overall, CNVs are an additional source of genetic variation and appear to affect the types of processes that are important in neurodevelopmental disorders, but additional careful studies are needed to define the effects of specific CNVs on particular phenotypes.

Identification of networks

It is assumed that most of the genes that are being identified for communication disorders contribute to susceptibility rather than being a sole cause, so that each gene individually may have a relatively small effect. The value of identification of these genes is magnified if they can define a developmental process, thus getting at the etiology of the disorder itself. The ability to tie genes together into networks is dependent upon the existing knowledge of their functions and pathways, and computerized algorithms are essential in sifting through this information. At this time, functional information is incomplete for most genes, but these methods are still able to point out potential gene interactions^{13,14,15}. As this is a very active area of research, these capabilities can be expected to improve as genetic and statistical methods improve. In addition, comprehensive phenotypic assessments will enhance the ability to define gene functions and thus define the relationships between communication disorders¹⁶.

Overview: Genes Affecting Communication Disorders

Reading Disability

Reading Disability is defined as difficulty learning to read and spell despite adequate instruction and neurosensory abilities, and is generally thought to be due to underlying defects in phonological decoding and phonemic awareness. It affects 8-10% of US school children,

and estimates of heritability of the group deficit in reading range from 0.54 to 0.84^{17,18,19,20}. Segregation analyses have supported a major gene effect (as opposed to the effect of many genes with individually small effects) for variation in reading ability^{21,22}, and studies of specific reading phenotypes have estimated that several major genes may be responsible^{23,24}, confirming the impression that RD is a complex genetic trait due to a combination of genetic and environmental influences. The findings of linkage and association analyses are consistent with the idea that RD is influenced by multiple genes; these studies have localized genetic effects to at least 9 chromosomal regions (designated by the genetic symbols DYX1-9) and have highlighted several genes, including DYX1C1 (Dyslexia 1 candidate 1) on chromosome 15²⁵, DCDC2 (doublecortin 2)²⁶ and KIAA0391²⁷ on chromosome 6, ROBO1 (roundabout 1) on chromosome 3²⁸, C2orf3 and MRPL2 (mitochondrial ribosomal protein L2)²⁹ on chromosome 2, KIAA0319L on chromosome 1³⁰, and SLC2A3 (solute carrier 2 A3) on chromosome 12³¹. In all of these genes, mutations were not found in the DNA that codes for the proteins produced by these genes, but associations were seen with variation in adjacent noncoding DNA. These regions include regulatory elements that control the degree to which the gene is expressed, thus affecting the amount of protein that is produced. At least 4 of these genes are involved in neuronal and axonal migration in the brain. Knockdown experiments which decreased the expression of the genes KIAA0319, DCDC2, or DYX1C1 resulted in impairment of migration of neurons to the cortex in developing rat brains^{32,33,34}, and ROBO1 is known to affect the migration of axons across the midline in mice, rats, and fruitflies³⁵. Galaburda et al. (2006) have proposed a mechanism by which these 4 genes interact to promote neuronal and axonal migration, with the transmembrane molecules KIAA0319 and ROBO1 receiving signals and guidance for migration, and DCDC2 affecting the cell's microtubular structure and mobility. DYX1C1 also produces an intracellular protein which may affect the cell's mobility, but also has an estrogen receptor domain, and it is interesting to speculate that this may be responsive to estrogen-dependent growth factors³⁷.

Language Impairment

Language Impairment (LI) is defined as a deficit in the development of expressive or receptive language in the face of adequate intelligence and opportunity, and without sensory or motor handicaps. The definition excludes ASD. It can involve any 5 domains: phonology, morphology, syntax, semantics, and pragmatics^{38,39,40}. A phonologic language deficit involving articulation is sometimes considered separately and termed Phonologic Disorder or Speech Sound Disorder. The incidence of LI is around 8% of pre-school children, and the heritability of the group deficit ranges from 0.45 to 0.73^{41,42}. Recent studies have suggested that most of the heritability of LI is found when the population is restricted to those children who have a history of treatment by a speech pathologist for an articulation defect⁴³; in this group, heritabilities for the group deficit were as high as 0.97. This could mean that LI+SSD is etiologically different from LI or SSD alone, but more detailed studies of each disorder alone and in combination are needed to fully account for the overlapping cognitive and genetic contributions to RD, SSD, and LI.⁷⁶

Like RD, the mode of inheritance of LI is complex. Linkage and association analyses for Specific Language Impairment (SLI, which requires a discrepancy between language ability and intelligence) has identified 2 genetic regions, SLI1 on chromosome 16 and SLI2 on chromosome 19^{44,45,46}, and a separate study of families with language impairment identified a location on chromosome 13⁴⁷.

Overlap between RD, SSD, and LI

Although genome-wide linkage studies of LI have not shown overlap with genetic regions containing RD loci, more detailed studies of specific regions have indicated that there may be several areas of overlap between RD and LI/SSD linkage regions. In the study of LI and

chromosome 13 markers⁴⁷, the linked phenotype was actually a reading measure. The SLI2 locus on chromosome 19 also influenced reading measures; however, the SLI1 locus did not⁴⁶. Linkage studies of Speech Sound Disorder have confirmed localization of SSD to RD regions on chromosome 1, 3, 6, and 15^{48,49,50,51}. Similarly, linkage with the RD regions of chromosomes 1, 6, and 15 has been replicated with reading and language phenotypes in families with children with LI⁵². These same families showed association with the same SNPs that have shown association with RD in the candidate gene KIAA0319. Association was not seen with SNPs in the adjacent DCDC2 gene, but this gene has been found to be associated with ADHD⁵³, suggesting that it may have a slightly different mechanism.

The overlapping linkage results between LI, SSD, and RD are particularly interesting given longitudinal studies demonstrating that children with SSD and LI are more likely to develop reading disability than children with SSD alone⁵⁴. If these overlapping linkages reflect the presence of single genes affecting different phenotypes, the next steps will be to determine why some families have predominantly language problems while others show more literacy problems. Possible explanations include different genetic variations at a gene, different environmental factors, and/or the interactions of additional genes. These studies also point out the importance of complete phenotypic assessments to adequately capture the phenotypic range of a gene within and between individuals, families, and disorders.

Autism Spectrum Disorders

The DSM-IV definition of Autism Spectrum Disorders (ASDs) requires deficits in 3 areas: socialization, such as poor interaction with peers, lack of reciprocity and absent social judgment; communication, including delayed expressive and pragmatic language; and behavior, including repetitive or stereotypic actions, restricted range of interests, and resistance to change (APA, 2000). While some cases of ASD are related to genetic syndromes such as Fragile X or tuberous sclerosis, genetic factors are involved in the etiology of nonsyndromic ASD as well, as evidenced by heritabilities as high as 90% for the diagnosis of autism⁵⁵, and 57% for continuous measures of autistic characteristics in the general population⁵⁶. There is evidence for heterogeneity in the mode of inheritance, however, meaning that the recurrence risks for siblings in families may be quite different. One study has found evidence for two subtypes with different male recurrence risks; one subgroup had a recurrence risk close to 50% for male siblings and 20% for female siblings, suggesting inherited dominant genes with reduced penetrance in females. Another, larger subgroup was defined with a low recurrence risk, reflecting multifactorial inheritance or *de novo* mutations in the probands, accounting for sporadic cases⁵⁷. The influence of gender-neutral new dominant mutations was further supported by Anello et al. (2009) who hypothesized that increased new mutation in older fathers should affect the gender ratio in their offspring. This hypothesis was supported by the finding that the male:female ratio, which is generally cited as being around 4:1, was actually closer to 6:1 for fathers under 30 years of age, and then decreased to 1.2:1 with advancing paternal age. The existence of two subtypes with different genetic mechanisms would presumably complicate the search for responsible genes and suggests that familial cases would need to be considered separately from sporadic cases.

Genome scans with SNPs and CNVs and mutation analyses have highlighted many possible candidate genes and regions; for a comprehensive review, see Abrahams and Geschwind, 2008. Some of the genes in which mutations have been identified interact in pathways underlying synaptic function, including neuroligin 3, (NLGN3), neuroligin 4 (NLGN4X), SHANK3, neurexin 1 (NRXN1), and CNTNAP1⁵⁹, and genes associated with CNVs also tend to be involved in synaptogenesis⁶⁰. There appears to be increasing evidence from genetic and imaging studies implicating deficits in synaptic maturation and connectivity in the etiology of

autism^{59,61,62,63}, and particularly genes that are expressed in response to neuronal activity⁶⁴, indicating influence on postnatal developmental processes.

Overlap between ASD and LI

Given the commonality of language problems in ASD and LI, it has been hypothesized that there may be phenotypic and genetic overlap. Children with LI have a higher frequency of ASD^{65,66} and children with ASD have features of LI, particularly pragmatic speech deficits^{67,68}. Several studies have argued against an etiological connection between the two, however; for example, Whitehouse et al. (2008) did not find support for similar etiologies of nonword repetition deficits in ASD and LI, and Lindgren et al. (2009) also did not find evidence of similar language problems in sibs of probands with autism vs. sibs of probands with LI. On the other hand, genetic linkage studies show some regions of overlap (reviewed in Smith, 2004 and Abrahams and Geschwind, 2008), and in particular, the gene CNTNAP2 has been reported to be associated with both ASD^{72,73} and LI⁷⁴. Thus, although the language characteristics of the two disorders appear to be dissimilar, it may be that there are some common genetic factors that contribute to abnormal language development in both conditions.

Summary: Genetic commonalities in communication disorders

By identifying interacting networks of genes influencing communication disorders, genetic studies have the potential to define the developmental causes of these conditions. This is most evident in the overlapping genetic influences on RD, LI, and SSD, and the contrasting distance between these disorders and ASD. As these genetic factors have been defined, a common theme has been the necessity to look beyond current clinical definitions and consider quantitative endophenotypes. The genotype/endophenotype correspondences that are found—or rejected—can identify sources of shared liability and redefine the boundaries of clinical conditions based on their etiology^{59,71,75,76,77}. These endeavors are clearly dependent upon the collaboration of specialists from different fields to rigorously define the disorders using accepted standardized measures. With that basis, further research can be done to accurately define cognitive endophenotypes which contribute to the etiology of the disorders. Molecular genetic tests can then be applied to identify the genes that influence the disorders and endophenotypes and determine any overlaps between them.

The results of these collaborations could refine diagnostic evaluations so that they cover the cognitive endophenotypes rather than depending solely upon achievement testing. Moreover, since referrals to therapists and the therapies that are applied are often tied to clinical diagnosis and clinical specialties, a change in the conceptualization of these conditions and their boundaries could promote changes in therapies and inclusion of multiple specialties. For example, if children with RD and LI have common cognitive liabilities in phonologic processing, this should be considered in designing therapy regardless of whether the predominant deficit is in reading or language. Greater understanding of comorbidities would ensure that assessments cover appropriate cognitive problems; that is, a child with LI and SSD would be followed more carefully for later reading problems, and appropriate therapies would be drawn from the fields of learning disability as well as speech pathology.

The definition of endophenotypes and their genetic relationships has a synergistic effect on further genetic and phenotypic studies. With more appropriate endophenotypes, genetic results may be more reliable, and gene networks can be built, leading to still more candidate genes and greater understanding of cognitive development and the endophenotypes that best reflect disorders of development. This further improves diagnosis at the educational level, and assessment of endophenotypes may also lead to earlier identification of susceptibility for a disorder; for example, current longitudinal studies of children of dyslexic parents demonstrate that early recognition of risk status is possible through an electrophysiologic

endophenotype⁷⁸. As genes are identified, direct genetic testing may also be beneficial. Since multiple genes are involved in these disorders, screening would likely be array-based and may not predict a specific outcome, but a pattern of susceptibility alleles might highlight a potential problem in the development of particular processes important for communication. Regardless of the methods for early identification, children at risk could be followed carefully to identify and remediate problems before they reach the level of academic failure required to qualify for services.

Acknowledgments

This work was funded by NIH-NICHD 5P50HD027802: Differential Diagnosis in Learning Disabilities (R.K Olson, P.I)

References

1. Diagnostic and statistical manual of mental disorders: DSM-IV-TR. Washington, DC: American Psychological Association (APA); 2000.
2. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. *Nature*. oct 7;2009 epub ahead of print.
3. Merikangas AK, Corvin AP, Gallagher L. Copy-number variants in neurodevelopmental disorders: promises and challenges. *Trends Genet* 2009;25(12):536–44. [PubMed: 19910074]
4. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, et al. Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* 2008;82(2):477–488. [PubMed: 18252227]
5. Elia J, Gai X, Xie HM, Perin JC, Geiger E, Glessner JT, et al. Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. *Mol Psychiatry*. 2009 June 29; epub ahead of print.
6. Itsara A, Cooper GM, Baker C, Girirajan S, Li J, Absher D, et al. Population analysis of large copy number variants and hotspots of human genetic disease. *Am J Hum Genet* 2009;84(2):148–161. [PubMed: 19166990]
7. de Smith AJ, Walters RG, Coin LJ, Steinfeld I, Yakhini Z, Sladek R, et al. Small deletion variants have stable breakpoints commonly associated with alu elements. *PLoS One* 2008;3(8):e3104. [PubMed: 18769679]
8. Vissers LE, Bhatt SS, Janssen IM, Xia Z, Lalani SR, Pfundt R, et al. Rare pathogenic microdeletions and tandem duplications are microhomology-mediated and stimulated by local genomic architecture. *Hum Mol Genet* 2009;18(19):3579–3593. [PubMed: 19578123]
9. Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. *Annu Rev Med* 2010;61:437–455. [PubMed: 20059347]
10. McCarroll SA. Extending genome-wide association studies to copy-number variation. *Hum Mol Genet* 2008;17(R2):R135–142. [PubMed: 18852202]
11. O'Donovan MC, Kirov G, Owen MJ. Phenotypic variations on the theme of CNVs. *Nat Genet* 2008;40(12):1392–1393. [PubMed: 19029974]
12. Bruder C, Piotrowski A, Gijsbers AA, Andersson R, Erickson S, de Stahl TD, et al. Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *Am J Hum Genet* 2008;82(3):763–771. [PubMed: 18304490]
13. Tao Y, Sam L, Li J, Friedman C, Lussier YA. Information theory applied to the sparse gene ontology annotation network to predict novel gene function. *Bioinformatics* 2007;23(13):i529–538. [PubMed: 17646340]
14. Chen Y, Zhu J, Lum PY, Yang X, Pinto S, MacNeil DJ, Zhang C, Lamb J, Edwards S, Sieberts SK, Leonardson A, Castellin LW, Wang S, Champy MF, Zhang B, Emilsson V, Doss S, Ghazalpour A, Horvath S, Drake TA, Lusis AJ, Schadt EE. Variations in DNA elucidate molecular networks that cause disease. *Nature* 2008;452(7186):429–435. [PubMed: 18344982]
15. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;4(1):44–57. [PubMed: 19131956]

16. Grigorenko EL. Speaking genes or genes for speaking? Deciphering the genetics of speech and language. *J Child Psychol Psychiatry* 2009;50(1-2):116–125. [PubMed: 19220595]
17. DeFries, JC.; Gillis, JJ. Genetics of reading disability. In: Plomin, R.; McClearn, G., editors. *Nature, Nature, and Psychology*. Washington, DC: APA Press; 1993.
18. DeFries JC, Alarcón M. Genetics of specific reading disability. *Mental Retard Devel Disab Res Rev* 1996;2:39–47.
19. Hawke JL, Wadsworth SJ, DeFries JC. Genetic influences on reading difficulties in boys and girls: the Colorado twin study. *Dyslexia* 2006;12(1):21–29. [PubMed: 16512171]
20. Astrom RL, Wadsworth SJ, DeFries JC. Etiology of the stability of reading difficulties: the longitudinal twin study of reading disabilities. *Twin Res Hum Genet* 2007;10(3):434–439. [PubMed: 17564501]
21. Pennington BF, Gilger JW, Pauls D, Smith SA, Smith SD, DeFries JC. Evidence for major gene transmission of developmental dyslexia. *JAMA* 1991;266(11):1527–1534. [PubMed: 1880884]
22. Gilger JW, Borecki IB, DeFries JC, Pennington BF. Commingling and segregation analysis of reading performance in families of normal reading probands. *Behav Genet* 1994;24(4):345–355. [PubMed: 7993313]
23. Wijsman EM, Peterson D, Leutenegger AL, Thomson JB, Goddard KA, Hsu L, et al. Segregation analysis of phenotypic components of learning disabilities. I. Nonword memory and digit span. *Am J Hum Genet* 2000;67(3):631–646. [PubMed: 10924405]
24. Chapman NH, Raskind WH, Thomson JB, Berninger VW, Wijsman EM. Segregation analysis of phenotypic components of learning disabilities. II. Phonological decoding. *Am J Med Genet B Neuropsychiatr Genet* 2003;121B(1):60–70. [PubMed: 12898577]
25. Taipale M, Kaminen N, Nopola-Hemmi J, Haltia T, Myllyluoma B, Lyytinen H, et al. A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. *Proc Natl Acad Sci U S A* 2003;100(20):11553–11558. [PubMed: 12954984]
26. Schumacher J, Anthoni H, Dahdouh F, Konig IR, Hillmer AM, Kluck N, et al. Strong genetic evidence of DCDC2 as a susceptibility gene for dyslexia. *Am J Hum Genet* 2006;78(1):52–62. [PubMed: 16385449]
27. Cope N, Harold D, Hill G, Moskvina V, Stevenson J, Holmans P, et al. Strong evidence that KIAA0319 on chromosome 6p is a susceptibility gene for developmental dyslexia. *Am J Hum Genet* 2005;76(4):581–591. [PubMed: 15717286]
28. Hannula-Jouppi K, Kaminen-Ahola N, Taipale M, Eklund R, Nopola-Hemmi J, Kaariainen H, et al. The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. *PLoS Genet* 2005;1(4):e50. [PubMed: 16254601]
29. Anthoni H, Zucchelli M, Matsson H, Muller-Myhsok B, Fransson I, Schumacher J, et al. A locus on 2p12 containing the co-regulated MRPL19 and C2ORF3 genes is associated to dyslexia. *Hum Mol Genet* 16(6):667–677. [PubMed: 17309879]
30. Couto JM, Gomez L, Wigg K, Cate-Carter T, Archibald J, Anderson B, et al. The KIAA0319-like (KIAA0319L) Gene on chromosome 1p34 as a candidate for reading disabilities. *J Neurogenet* 2008;22(4):295–313. [PubMed: 19085271]
31. Roeske D, Ludwig KU, Neuhoﬀ N, Becker J, Bartling J, Bruder J, et al. First genome-wide association scan on neurophysiological endophenotypes points to trans-regulation effects on SLC2A3 in dyslexic children. *Mol Psychiatry*. 2009 Sept 29; epub ahead of print.
32. Paracchini S, Thomas A, Castro S, Lai C, Paramasivam M, Wang Y, et al. The chromosome 6p22 haplotype associated with dyslexia reduces the expression of KIAA0319, a novel gene involved in neuronal migration. *Hum Mol Genet* 2006;15(10):1659–1666. [PubMed: 16600991]
33. Meng H, Smith SD, Hager K, Held M, Liu J, Olson RK, et al. DCDC2 is associated with reading disability and modulates neuronal development in the brain. *Proc Natl Acad Sci U S A* 2005;102(47):17053–17058. [PubMed: 16278297]
34. Wang Y, Paramasivam M, Thomas A, Bai J, Kaminen-Ahola N, Kere J, et al. DYX1C1 functions in neuronal migration in developing neocortex. *Neuroscience* 2006;143(2):515–522. [PubMed: 16989952]

35. Kidd T, Brose K, Mitchell KJ, Fetter RD, Tessier-Lavigne M, Goodman CS, et al. Roundabout controls axon crossing of the CNS midline and defines a novel subfamily of evolutionarily conserved guidance receptors. *Cell* 1998;92(2):205–215. [PubMed: 9458045]
36. Galaburda AM, LoTurco J, Ramus F, Fitch RH, Rosen GD. From genes to behavior in developmental dyslexia. *Nat Neurosci* 2006;9(10):1213–1217. [PubMed: 17001339]
37. Massinen S, Tammimies K, Tapia-Paez I, Matsson H, Hokkanen ME, Soderberg O, et al. Functional interaction of DYX1C1 with estrogen receptors suggests involvement of hormonal pathways in dyslexia. *Hum Mol Genet* 2009;18(15):2802–2812. [PubMed: 19423554]
38. Tallal, P. Developmental dysphasia. In: Adelman, G., editor. *Encyclopedia of Neuroscience*. Vol. 1. Boston: Birkhauser; 1987.
39. Leonard LB. Language learnability and specific language impairment in children. *Appl Psycholinguistics* 1989;10:179–202.
40. Tomblin JB, Records NL, Buckwalter P, et al. Prevalence of specific language impairment in kindergarten children. *J Speech Lang Hear Res* 1997;40:1245–1260. [PubMed: 9430746]
41. Tomblin JB, Buckwalter PR. Heritability of poor language achievement among twins. *J Speech Lang Hear Res* 1998;41(1):188–199. [PubMed: 9493744]
42. Dale PS, Simonoff E, Bishop DV, Eley TC, Oliver B, Price TS, Purcell S, Stevenson J, Plomin R. Genetic influence on language delay in two-year-old children. *Nat Neurosci* 1998;1(4):324–328. [PubMed: 10195167]
43. Bishop DV, Hayiou-Thomas ME. Heritability of specific language impairment depends on diagnostic criteria. *Genes Brain Behav* 2008;7(3):365–372. [PubMed: 17919296]
44. SLI Consortium. A genomewide scan identifies two novel loci involved in specific language impairment. *Am J Hum Genet* 2002;70(2):384–398. [PubMed: 11791209]
45. 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(6):1225–1238. [PubMed: 15133743]
46. Falcár 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(4):393–402. [PubMed: 18005161]
47. 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(1):45–55. [PubMed: 12048648]
48. Stein CM, Schick JH, Gerry Taylor H, Shriberg LD, Millard C, Kundtz-Kluge A, Russo K, Minich N, Hansen A, Freebairn LA, Elston RC, Lewis BA, Iyengar SK. Pleiotropic effects of a chromosome 3 locus on speech-sound disorder and reading. *Am J Hum Genet* 2004;74(2):283–297. [PubMed: 14740317]
49. Smith SD, Pennington BF, Boada R, Shriberg LD. Linkage of speech sound disorder to reading disability loci. *J Child Psychol Psychiatry* 2005;46(10):1057–1066. [PubMed: 16178929]
50. Stein CM, Millard C, Kluge A, Miscimarra LE, Cartier KC, Freebairn LA, Hansen AJ, Shriberg LD, Taylor HG, Lewis BA, Iyengar SK. Speech Sound Disorder Influenced by a Locus in 15q14 Region. *Behav Genet* 2006;36(6):858–868. [PubMed: 16786424]
51. Miscimarra L, Stein C, Millard C, Kluge A, Cartier K, Freebairn L, Hansen A, Shriberg L, Taylor HG, Lewis B, Iyengar SK. Further evidence of pleiotropy influencing speech and language: analysis of the DYX8 region. *Hum Hered* 2007;63(1):47–58. [PubMed: 17230025]
52. Rice ML, Smith SD, Gayán J. Convergent genetic linkage and associations to language, speech and reading measures in families of probands with Specific Language Impairment. *J Neurodev Disord* 2009;1(4):264–282. [PubMed: 19997522]
53. Couto JM, Gomez L, Wigg K, Ickowicz A, Pathare T, Malone M, Kennedy JL, Schachar R, Barr CL. Association of Attention-Deficit/Hyperactivity Disorder with a Candidate Region for Reading Disabilities on Chromosome 6p. *Biol Psychiatry* 2009;66(4):368–75. [PubMed: 19362708]
54. Peterson RL, Pennington BF, Shriberg LD, Boada R. What influences literacy outcome in children with speech sound disorder? *J Speech Lang Hear Res* 2009;52(5):1175–1188. [PubMed: 19403946]

55. Rutter M. Genetic studies of autism: from the 1970s into the millennium. *J Abnorm Child Psychol* 2000;28:3–14. [PubMed: 10772346]
56. Hoekstra RA, Bartels M, Verweij CJ, Boomsma DI. Heritability of autistic traits in the general population. *Arch Pediatr Adolesc Med* 2007;161(4):372–377. [PubMed: 17404134]
57. Zhao X, Leotta A, Kustanovich V, Lajonchere C, Geschwind DH, Law K, et al. A unified genetic theory for sporadic and inherited autism. *Proc Natl Acad Sci U S A* 2007;104(31):12831–12836. [PubMed: 17652511]
58. Anello A, Reichenberg A, Luo X, Schmeidler J, Hollander E, Smith CJ, et al. Brief report: parental age and the sex ratio in autism. *J Autism Dev Disord* 2009;39(10):1487–1492. [PubMed: 19452267]
59. Abrahams BS, Geschwind DH. Advances in autism genetics: on the threshold of a new neurobiology. *Nat Rev Genet* 2008;9(5):341–355. [PubMed: 18414403]
60. Merikangas AK, Corvin AP, Gallagher L. Copy-number variants in neurodevelopmental disorders: promises and challenges. *Trends Genet* 2009;25(12):536–544. [PubMed: 19910074]
61. The Autism Genome Project Consortium. Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu XQ, et al. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* 2007;39(3):319–328. [PubMed: 17322880]
62. Sutcliffe JS. Genetics. Insights into the pathogenesis of autism. *Science* 2008;321(5886):208–209. [PubMed: 18621658]
63. Levy SE, Mandell DS, Schultz RT. Autism. *Lancet* 2009;374(9701):1627–1638. [PubMed: 19819542]
64. Morrow EM, Yoo SY, Flavell SW, Kim TK, Lin Y, Hill RS, et al. Identifying autism loci and genes by tracing recent shared ancestry. *Science* 2008;321(5886):218–223. [PubMed: 18621663]
65. Conti-Ramsden G, Simkin Z, Botting N. The prevalence of autistic spectrum disorders in adolescents with a history of specific language impairment (SLI). *J Child Psychol Psychiatry* 2006;47(6):621–628. [PubMed: 16712639]
66. Leyfer OT, Tager-Flusberg H, Dowd M, Tomblin JB, Folstein SE. Overlap between autism and specific language impairment: comparison of Autism Diagnostic Interview and Autism Diagnostic Observation Schedule scores. *Autism Res* 2008;1(5):284–296. [PubMed: 19360680]
67. Kjelgaard MM, Tager-Flusberg H. An Investigation of Language Impairment in Autism: Implications for Genetic Subgroups. *Lang Cogn Process* 2001;16(2-3):287–308. [PubMed: 16703115]
68. Bishop DV, Norbury CF. Exploring the borderlands of autistic disorder and specific language impairment: a study using standardised diagnostic instruments. *J Child Psychol Psychiatry* 2002;43(7):917–929. [PubMed: 12405479]
69. Whitehouse AJ, Barry JG, Bishop DV. The broader language phenotype of autism: a comparison with specific language impairment. *J Child Psychol Psychiatry* 2007;48(8):822–830. [PubMed: 17683454]
70. Lindgren KA, Folstein SE, Tomblin JB, Tager-Flusberg H. Language and reading abilities of children with autism spectrum disorders and specific language impairment and their first-degree relatives. *Autism Res* 2009;2(1):22–38. [PubMed: 19358305]
71. Smith, SD. Localization and identification of genes influencing language and learning disorders. In: Rice, ML.; Warren, SF., editors. *Developmental Language Disorder: From Phenotype to Etiologies*. Mahwah, New Jersey: Lawrence Erlbaum Associates; 2004. p. 329-354.
72. Arking DE, Cutler DJ, Brune CW, Teslovich TM, West K, Ikeda M, et al. A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism. *Am J Hum Genet* 2008;82(1):160–164. [PubMed: 18179894]
73. Alarcon M, Abrahams BS, Stone JL, Duvall JA, Perederiy JV, Bomar JM, et al. Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. *Am J Hum Genet* 2008;82(1):150–159. [PubMed: 18179893]
74. Vernes SC, Newbury DF, Abrahams BS, Winchester L, Nicod J, Groszer M, et al. A functional genetic link between distinct developmental language disorders. *N Engl J Med* 2008;359(22):2337–2345. [PubMed: 18987363]
75. Levy Y, Ebstein RP. Research review: crossing syndrome boundaries in the search for brain endophenotypes. *J Child Psychol Psychiatry* 2009;50(6):657–668. [PubMed: 19175806]

76. Pennington BF, Bishop DV. Relations among speech, language, and reading disorders. *Annu Rev Psychol* 2009;60:283–306. [PubMed: 18652545]
77. Grigorenko EL. Speaking genes or genes for speaking? Deciphering the genetics of speech and language. *J Child Psychol Psychiatry* 2009;50(1-2):116–125. [PubMed: 19220595]
78. Guttorm TK, Leppanen PH, Hamalainen JA, Eklund KM, Lyytinen HJ. Newborn Event-Related Potentials Predict Poorer Pre-Reading Skills in Children at Risk for Dyslexia. *J Learn Disabil*. Nov; 2009 ePub ahead of print.

Table 1

Abbreviations

SNP	Single Nucleotide Polymorphism
CNV	Copy Number Variant
GWAS	Genome-Wide Association Study
MZ	Monozygotic
RD	Reading Disability
LI	Language Impairment
ASD	Autism Spectrum Disorder
SSD	Speech Sound Disorder