Understanding the Etiology of Complex Traits: Symbiotic Relationships Between Psychology and Genetics

Elena L. Grigorenko[1,2]

ABSTRACT—The present article offers comments on the infusion of methodologies, approaches, reasoning strategies, and findings from the fields of genetics and genomics into studies of complex human behaviors (hereafter, complex phenotypes). Specifically, I discuss issues of generality and specificity, causality, and replicability as they pertain to molecular genetic studies of human phenotypes. These issues are illustrated with findings from genetic linkage and association studies investigating the etiology of disorders of spoken and written language—an area of inquiry that has been consistently referenced as one of the most successful in terms of its progress in understanding the genetic bases of human behaviors. I complete this discussion with comments on how the stronger presence of genetics and genomics in psychology is changing the conceptualization and investigation of research questions and affecting the next generation of interdisciplinary research.

The completion of the Human Genome Project in 2003 (http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml) and ensuing research has resulted in the discoveries of genes and gene variants associated with typical and atypical human development. Subsequently, recent years have seen an explosion of DNA markers that can be genotyped (or measured) at ever-increasing speed and decreasing expense. As a result, psychologists and educators interested in the etiology of complex human traits and skills have encountered yet another source of information allowing them to correlate variation in behavior with variation in DNA. Publications on these correlations are rampant and only show signs of increase. Here I discuss the ongoing infusion of methodologies, approaches, reasoning strategies, and findings from genetics and genomics into psychology and illustrate this infusion with molecular genetic linkage and association studies of etiological bases of disorders of spoken and written language (DSWL). I comment on the changing landscape of psychological and educational research prompted by this infusion and its recursive impact on studies in genetics and genomics.

[1]Yale University
[2]Moscow State University

Address correspondence to Elena L. Grigorenko, Child Study Center, Yale University, 230 South Frontage Road, New Haven, CT 06510; e-mail: elena.grigorenko@yale.edu

GENERAL CONSIDERATIONS

Behaviors and behavioral disorders (i.e., manifested traits typically referred to as phenotypes) of interest to psychologists are complex, and complex phenotypes have complex etiologies. For the past 30 years or so, it has been appreciated that these complex phenotypes are affected by both genetic and environmental factors through main effects and interactions (gene–gene, environment–environment, and gene–environment). Recently, however, this appreciation has gained significant empirical support through the identification of measurable connections among specific genes and environments and specific behaviors or disorders.

Estimates suggest that there are ~24,500 genes in the human genome. The number of environments in which humans develop and live is almost infinite. And unlike genes, human environments are not discrete but rather create a continuum of conditions in which we exist. Even putting environments aside, imagining the number of possible combinations among genes is mind boggling. When these combinations are considered at different levels of continuous environments, the

number becomes unimaginable. Studies of patterns of familial transmission of complex behaviors and behavioral disorders also suggest that many genes are involved in these transmissions (e.g., estimates for phonological processing—the capacity to process meaningful speech sounds or phonemes—suggest the influence of 2–5 genes and for autism up to 100 genes). Moreover, each of these genes can contribute both risk-related and protective variants to phenotypes of interest. Finally, even if reliably identified, individual alleles are likely neither necessary nor sufficient for the manifestation of complex phenotypes. In sum, the phenotypes of interest to psychologists and educators are most likely governed by many genes of small effect whose contributions are modified by other genes and environments.

We now understand many of the risk, preventive, and mixed impacts of different environments on phenotypes of interest to psychologists and educators. However, we have a long road ahead to identify specific genes affecting these phenotypes. And although we have stumbled across a number of combined gene–environment high-order impacts (e.g., the interaction effect of maltreatment in childhood and specific genetic risk factors), there is much to do to replicate and understand the reported findings and establish new ones.

GENERAL OR SPECIFIC?

Everything in cognition correlates to a certain degree; the same is true of indicators of spoken and written language functioning. A review of psychological processes referred to as "deficient" in DSWL (i.e., speech and sound disorder [SSD], specific language impairment [SLI], developmental dyslexia [DD]) generates a relatively limited list of components that are repeatedly featured in these disorders. These processes include, but are not limited to, phonology, orthography, morphology, and semantics. For example, specific psychological functions such as phonological memory appear to be deficient in all these disorders. Thus, questions have arisen about whether these are different disorders or variable manifestations of the same disorder and whether these disorders have distinct or fully or partially overlapping psychological architectures. In the context of this review, the most interesting issue is whether studies of the genetic bases of these disorders and their components can clarify their behavioral and psychological architecture.

Three issues bear mention. First, a number of publications indicate componential overlap among various developmental disorders and hypothesize that some genetic regions harbor genes that contribute to multiple disorders. Specifically, researchers have considered phenotypic and genetic overlaps between SLI and autism spectrum disorders (ASD), SSD and DD, and attention deficit and hyperactivity disorder (ADHD) and DD. Here I present only select examples to illustrate the logical underpinning of this approach. The first example pertains to the comparison of holistic diagnoses, such as SLI and ASD, that share atypicality of language development although in very different manifestations. It has been noted in the literature that, when certain language-related aspects of the autism phenotype are considered (e.g., delay in the development of phrase speech, language impairment in parents of probands, or using a quantitative trait of language functioning), the same chromosomal regions (2q, 7q, and 13q) appear to be linked to both disorders in different samples (Bartlett et al., 2004; Spence et al., 2006). The second example pertains to the investigation of so-called componential processes of developmental disorders. Specifically, both patients with SSD and patients with DD demonstrate deficits on tasks related to phonological processing (e.g., nonword repetition and nonword decoding). When these phenotypes are considered in a sample of probands with SSD and their relatives, four previously identified susceptibility regions for DD (3p12–q13, 6p22, 15q21, and 1p36) also demonstrated linkage with SSD (Smith, Pennington, Boada, & Shriberg, 2005; Stein et al., 2004), indicating multiple, or pleiotropic, influences of the same genes on multiple phenotypes. The third example pertains to cross-referenced phenotypes—for example, reading performance in a sample of ADHD probands and their family members and performance on attention tasks in a sample of DD probands and their family members. When these strategies are implemented, there is also evidence of overlap in genetic susceptibility regions (e.g., 6p21–22; Willcutt et al., 2002). Fourth, there is an intriguing tendency for a number of genetic regions to crop up in investigations of different phenotypes. A relevant illustration is region 7q31–35, reported by a number of research groups to harbor genes for various forms of speech and language disorders (e.g., *FOXP2*[1]) and to be associated with mental retardation and autism (Alarcon, Yonan, Gilliam, Cantor, & Geschwind, 2005). Fifth, there are many other case examples where a chromosomal abnormality (e.g., deletion, addition, or relocation of chromosomal material) results in a number of psychological challenges, including mental retardation and language delay (e.g., 7q11, 15q11–13; Somerville et al., 2005; Veenstra-Vanderweele, Christian, & Cook, 2004). To summarize, when different language-related phenotypes are considered within distinct developmental disorders, evidence suggests that particular regions of the human genome appear to be linked to these phenotypes no matter how the disorder is defined or through which disorders a given sample was ascertained. In other words, there might be genes involved in, for example, language deficit no matter what particular disorder (e.g., SLI or autism) such a deficit characterizes; thus, genetic deficits might cross diagnostic categories, forming the basis for comorbidities. Correspondingly, psychological models depicting the process behind these disorders, connectionist models, for example,

might explain the observed overlap in genetic etiologies of multiple developmental disorders by featuring specific componential processes deficient in more than one disorder.

The second issue to consider comes from studies of DD (for details, see Haworth, Meaburn, Harlaar, & Plomin, 2007; Schulte-Körne et al., 2007). Today, there are four interesting candidate genes for DD and reading-related processes (a 15q gene *DYX1C1*,[2] two 6p genes, *KIAA0319* and *DCDC2*,[3] and a 3p gene *ROBO1*[4]; Fisher & Francks, 2006), all of which are active candidates, but none of which has been either fully accepted or rejected as such. The consensus is that any association of all or some of these four genes with DD is probably of small magnitude. Yet, despite an inconsistent pattern of replications and nonreplications of these genes' involvement with DD, one remarkable feature of these four genes is a certain degree of communality, at least in broad strokes, of their functions. Three of the four genes (the function of *DYX1C1* has not been established yet) appear to be involved in early stages of brain development. Correspondingly, it is unlikely that these genes are reading specific; however, there are not much data on what other behavioral phenotypes these genes might be involved with (there is a published report of the lack of association between *DYX1C1* and ASD). To summarize, the four DD-associated candidate genes appear to be involved, at least to a minor degree, with the formation of the anatomical foundation for higher cognitive functions such as reading (possibly among others).

Third, it is important to recognize the *huge* degree of variability in both linkage and association studies (see Table 1) with regard to what particular phenotypes are considered (Grigorenko, 2005). Genetic studies of DD have benefited tremendously from cognitive models of reading and reading difficulties. These models are complex: They reference multiple psychological processes and consider dynamics of development and learning in the establishment of the brain-based systems and representations supporting reading. Not surprisingly, the pattern of results emerging from behavior and molecular genetic studies of these models is also complex. Different componential processes show different estimates of heritability (Grigorenko, 2004). And when considered as phenotypes in linkage and association studies, they do not show identical or even similar patterns of results. In fact, although poorly understood at this point, this pattern of results cannot be explained either by the presence of general behavioral deficits for DD or by the existence of generalist genes. Although it is appealing to assume the existence of main genetic effects for holistic disorders, results of quantitative and molecular genetic studies of psychological components of these disorders do not permit us to default to parsimonious models of major main effects. Results repeatedly suggest both overlapping (i.e., underlying comoribities) and unique (i.e., generating specificities) genetic influences on componential phenotypes.

How can these issues be interpreted? At this point, our knowledge of the genetic mechanisms involved in typical and atypical spoken and written language is far from advanced, but neither are we entirely ignorant. We know there are likely genetic factors of general and specific impact that follow a certain hierarchy: (a) disruptions in genes of major impact are rare, lead to severe phenotypic deficiencies that are not specific, and do not seem to account for individual differences in spoken and written language in the general population; (b) disruptions in

Table 1
Molecular Genetic Designs Used in Genetic Studies of Psychological Traits

Study type	*Function*	*Design (units of analyses)*	*Outcome*
Linkage study	Permits geneticists to track patterns of inheritance of specific genetic variants or larger chunks of genetic materials (e.g., chromosomal pieces or regions) within families	Uses related people only, that is, family members: extended, nuclear, or pairs of any degree of relatedness (parents and children, siblings, cousins, etc.)	Establishes a link between a disorder or trait (i.e., a phenotype) and a particular location in the genome that can be subsequently investigated for an association with specific genes harbored in this location
Association study	Allows geneticists to investigate connections in the general population between a particular variant in a particular gene (e.g., a variant that alters a production of a particular protein) and the disorder or trait of interest by detecting a statistical correlation between the two	Uses related and unrelated people; related people: family members (see above); unrelated people: cases (people with the phenotype of interest) and controls (people who matched to the cases on a number of important parameters, e.g., ethnicity, gender, age, exposure to a particular type of environment but do not have the phenotype of interest)	Identifies a specific genetic factor (a genetic variant) associated with the manifestation of the phenotype of interest

genes expressed early in development are infrequent, result in a jeopardized or atypically structured anatomical foundation and brain function, which are recruited in the development of higher order psychological processes, and might account for common disorders in combination with specific environmental exposures; and (c) variations in genes expressed later in development and throughout the life span are common, characterized by small or very small effects in isolation, and contribute to orchestras of genetic and environmental effects forming the texture of individual differences in the general population. Finally, given the mosaic nature of our knowledge, we can only hypothesize the mechanisms by which these rather diverse types of genetic impacts might occur. Figure1 and its legend illustrate some of these hypotheses.

CAUSATION AND CORRELATION

So far, this discussion has considered only the establishment of a connection or association between a particular gene and a particular phenotype. However, the nature of this association is statistical, and its meaning is in capturing the presence of a correlational, not causational, link. Thus, the next—huge—step is to establish causation from association. And far from every correlation has causation underlying it.

There is debate in the field with regard to guidelines for translating correlation into causation. The primary possibilities are the discovery of biological pathways and development and verification of animal models of behavior. Given the nature of this brief contribution, I comment only selectively here. First, it is important to understand the limitations of research. The majority of behavior/molecular genetic labs can go only as far as establishing correlation. Verifying causation requires collaborative efforts among colleagues in a variety of disciplines. Thus, it is important to tread carefully when interpreting results of studies correlating behaviors and genes.

Second, it is possible that certain correlations between genes and behaviors will never be leveraged into causal links. For example, even if the pathway of genetic events underlying reading acquisition is established, the unpredictability

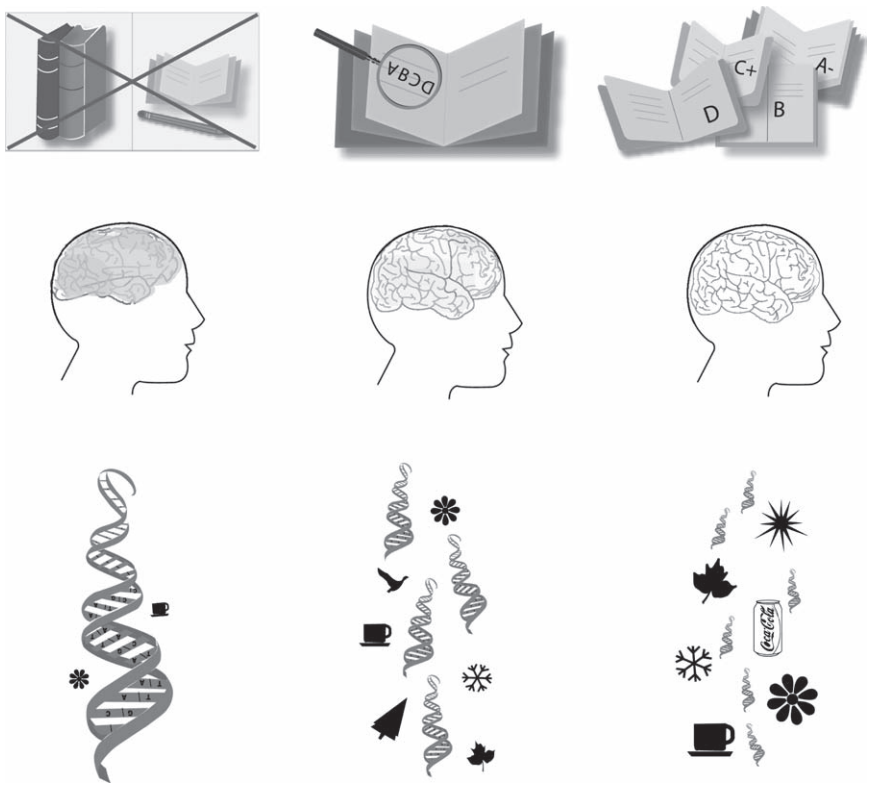

Fig. 1. Possible contributions of genetic and environmental factors in complex behavioral phenotypes. (1) For example, rare mutations in the gene *FOXP2* result in structural and/or functional brain abnormalities and a severe disorders of spoken and written language. Environment is also assumed to play a limited role in this mechanism. (2) An example might come from studies of developmental dyslexia (DD), where an impact of a genetic variant might result in the manifestation of DD if combined with the impact of risk environments. (3) An example of common variation in the function of genes expressed later in development and associated with individual difference (e.g., in school grades). One example is associated with the variants in the *COMT* gene coding for the production of the catechol-*O*-methyltransferase, a metabolic agent participating in the turnover of the neurotransmitters dopamine, epinephrine, and norepinephrine and implicated in a variety of cognitive functions. Environment (e.g., diet, weather, socioeconomic status, pedagogies) assumed to play a substantial role in trait manifestation.

introduced by various environments (e.g., type and amount of schooling) into the realization of this pathway in real life will be impossible to replicate in laboratory environments. For example, how can something like "love at first sight," which might dictate a life path for a human being, be modeled in the constrained environment of a laboratory? And, although model organisms are extremely important in understanding the genetics of behavior, some human behaviors are simply impossible to model in other organisms (e.g., although there are bookworms, there are no worms that read).

Third, without diving deeper into the philosophical and epistemological debate on causation, it is important to mention types of causes relevant to this discussion. With regard to Mendelian traits (i.e., traits whose inheritance is controlled by the simple genetic laws discovered by Mendel), the field typically considered *necessary* and *sufficient* causes, assuming that the manifestation of rare Mendelian conditions necessitates the presence of the genetic factor without which the condition cannot occur and whose presence determines the manifestation of the condition (Figure 2a). With regard to complex, non-Mendelian traits (i.e., traits whose inheritance does not obey Mendelian laws), the field typically refers to *counterfactual* and *probabilistic* causes, implying that these causes both introduce and influence the probability of the manifestation of the condition (Figure 2b). These types of causes negate the assumption of determinism and introduce room to model the etiology of complex traits, allowing for interaction and covariation of genes with other genes and environments through a distribution of probabilities.

Thus, we must consider the possibility that proof of the causal nature of genetic contributions to behavior and behavior disorders may remain elusive: People cannot be randomly assigned to genotypes and/or environments. Yet, a causal link between a gene or gene variant and behavior can be substantiated by repeated, systematic efforts at replication in various samples and environments.

NONREPLICATIONS: HOW WORRISOME ARE THEY?

Popper's (1961) remark that the strongest basis for our belief in a hypothesis is in our repeated unsuccessful attempts to refute it unequivocally stresses the importance of replication. The problem of multiple testing is especially central to the merger of the fields of genomics, psychology, and education, where an almost infinite number of tests can be carried out for different behavioral phenotypes with hundreds of thousands of genetic markers. Thus, it is very worrisome that only a fifth to a quarter of initially reported positive associations between genetic markers and complex diseases or traits tend to be replicated (Ioannidis, Trikalinos, Ntzani, & Contopoulos-Ioannidis, 2003).

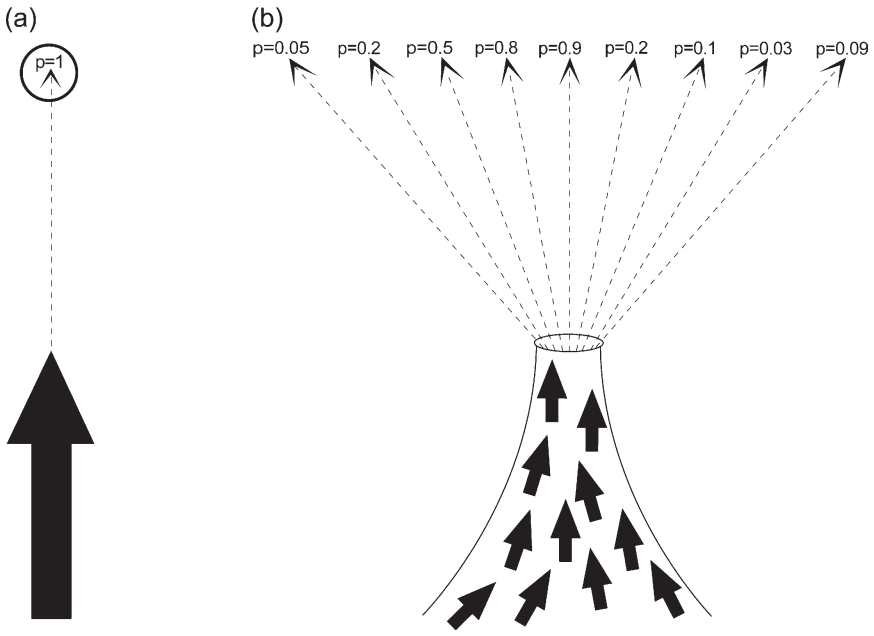

Fig. 2. Illustrations of different causal mechanisms that might be involved in the emergence of complex behavioral phenotypes. (a) Necessary and sufficient causes result in certain, determined outcomes. This type of causation is typically associated with traits that are considered to be influenced by a single genetic mechanism whose inheritance follows Mendel's laws. A deterministic genetic cause (shown in a solid arrow) leads to an inevitable outcome ($p = 1$). (b) Counterfactual and probabilistic causes: These causes introduce and influence but do not determine outcomes. This type of causation is typically associated with traits whose inheritance is governed by mechanisms other than those captured by Mendel, that is, complex behavioral traits. The impact of genetic forces (shown in solid arrows) is "funneled" by environment (or any combination of forces acting interactively!) to result in different probabilistic outcomes (shown in dotted arrows); variation in environment (i.e., different type of funneling) leads to an array of possible probabilistic outcomes.

Despite the excitement associated with its frontiers, this field is challenged with nonreplications. For example, the report of the first DD candidate gene, *DYX1C1*, was followed by numerous nonreplications. Similarly, out of many regions reportedly linked to ASD through nine genome-wide scans, only one for autism proper and none for ASD survived the rigors of meta-analysis (Trikalinos et al., 2006). Possible reasons for these nonreplications include the likely heterogeneity of genetic mechanisms underlying complex traits, large groups of genes of small effect contributing to complex traits, and numerous biases in the field. There is no overt consensus about what constitutes a true finding and how many replications are necessary to accept it or nonreplications to dismiss it. Yet, because random assignments are impossible and modeling, whether computer or model organism based, is capable of reproducing only a particular aspect of complex systems supporting the development of complex human traits, replications are an absolutely essential means of verifying or substantiating a result. It is also critical to account for nonreplications and provide researchers access to reports on failures to replicate.

I want to note the distinction between "operational" and "constructive" replications (Lykken, 1968); what the field needs is constructive attempts at replication that challenge and extend original findings, not merely operational reproduction of the initial study. The field needs to know about all, not only successful, attempts to replicate a reported gene–behavior association.

SUMMARY

Psychological and educational sciences will continue to be saturated with studies correlating genes and typical and atypical behaviors and skills. The infusion of genetics and genomics into psychology and education results in a number of innovations. The first impact is the sheer number of psychological and educational studies that include a genetic component; more and more investigations consider the genome (i.e., measured genetic variants) as a source of group differences between disordered and typical persons and as a source of individual differences in the general population. This infusion might require new training for psychologists and educators in genetics and genomics. The second outcome of this infusion is methodological, related primarily to the development of statistical methods that permit proper inclusion of genetic data in cross-sectional and longitudinal models of psychological traits. The multivariate presentation of complex human traits and the apparent pleiotropic influence of genes necessitate the development of complex models for analyzing these data. The third impact relates to the realization that a simple association between a variant and a behavior does not indicate a causal link between the two. The common belief that everything genetic is causal is being revised as genetics penetrates further into psychology. Fourth, the strict requirements for replicability in genetics influence the perception of reported associations between genes and behaviors; the initial excitement of the first reports of these associations has waned; what is expected is a convincing ratio of replications and nonreplications. In fact, only replicated findings can be considered indicative of reliable genetic linkages and associations. These and other outcomes of the infusion of genetics into psychology and education may change the profile of current research, opening new directions for research aimed at understanding the etiology of human behavior.

Finally, the strengthening bridge between psychology and genetics affects not only psychological but also genetic research as well. One major outcome is a deepened appreciation of complex phenotypes and the necessity for their careful characterization and modeling prior to conducting genetic studies. The early genetic investigations of complex human disorders used a simple dichotomous affected/unaffected differentiation of individuals; now, psychological models of complex human behavior such as reading or depression are used in genetic research. Another outcome is a growing appreciation of the role of environment; similarly, if early genetic studies of complex traits intentionally ignored the role of environment, genetic studies now willingly incorporate into their methodologies environmental manipulations or variability in an attempt to decipher genetic mechanisms of behavior. Finally, like psychology, genetics is loosening up its views of causal roles of genes, having substituted its early deterministic with probabilistic schemas for complex behaviors.

Thus, the infusion of genetics into psychology and education has already resulted in gains for these fields. And there is every reason to believe that these gains will only grow in magnitude.

Acknowledgments—Preparation of this article was supported by Grants R21 TW006764-02 from the Fogarty Program as administered by the National Institutes of Health, Department of Health and Human Services, and R01 DC007665 as administered by the National Institute of Deafness and Communication Disorders (Principal Investigator: Grigorenko). Grantees undertaking such projects are encouraged to express their professional judgment freely. Therefore, this article does not necessarily reflect the position or policies of the National Institutes of Health, and no official endorsement should be inferred. I am thankful to Robyn Rissman for her editorial assistance and Beata Moryl for preparing the figures.

NOTES

1 Forkhead box P2.
2 Dyslexia susceptibility-1, candidate 1.
3 Gene encoding Kazusa cDNA 0319 and doublecortin domain containing 2, respectively.
4 Roundabout 1.

REFERENCES

Alarcon, M., Yonan, A. L., Gilliam, T. C., Cantor, R. M., & Geschwind, D. H. (2005). Quantitative genome scan and ordered-subsets analysis of autism endophenotypes support language QTLs. *Molecular Psychiatry*, *10*, 747–757.

Bartlett, C. W., Flax, J. F., Logue, M. W., Smith, B. J., Vieland, V. J., Tallal, P., et al. (2004). Examination of potential overlap in autism and language loci on chromosomes 2, 7, and 13 in two independent samples ascertained for specific language impairment. *Human Heredity*, *57*, 10–20.

Fisher, S. E., & Francks, C. (2006). Genes, cognition and dyslexia: Learning to read the genome. *Trends in Cognitive Sciences*, *10*, 250–257.

Grigorenko, E. L. (2004). Genetic bases of developmental dyslexia: A capsule review of heritability estimates. *Enfance*, *3*, 273–287.

Grigorenko, E. L. (2005). A conservative meta-analysis of linkage and linkage-association studies of developmental dyslexia. *Scientific Studies of Reading*, *9*, 285–316.

Haworth, C. M. A., Meaburn, E. L., Harlaar, N., & Plomin, R. (2007). Reading and generalist genes. *Mind, Brain, and Education*, *1*, 173–180.

Ioannidis, J. P., Trikalinos, T. A., Ntzani, E. E., & Contopoulos-Ioannidis, D. G. (2003). Genetic associations in large versus small studies: An empirical assessment. *Lancet*, *361*, 567–571.

Lykken, D. (1968). Statistical significance in psychological research. *Psychological Bulletin*, *70*, 151–159.

Popper, K. (1961). *The logic of scientific discovery.* New York: Science Editions.

Schulte-Körne, G., Ludwig, K. U., el Sharkawy, J., Nöthen, M. M., Müller-Myhsok, B., & Hoffmann, P. (2007). Genetics and neuroscience in dyslexia: Perspectives for education and remediation. *Mind, Brain, and Education*, *1*, 162–172.

Smith, S. D., Pennington, B. F., Boada, R., & Shriberg, L. D. (2005). Linkage of speech sound disorder to reading disability loci. *Journal of Child Psychology & Psychiatry & Allied Disciplines*, *46*, 1057–1066.

Somerville, M. J., Mervis, C. B., Young, E. J., Seo, E. J., del Campo, M., Bamforth, S., et al. (2005). Severe expressive-language delay related to duplication of the Williams-Beuren locus. *New England Journal of Medicine*, *353*, 1694–1701.

Spence, S. J., Cantor, R. M., Chung, L., Kim, S., Geschwind, D. H., & Alarcon, M. (2006). Stratification based on language-related endophenotypes in autism: Attempt to replicate reported linkage. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, *141*, 591–598.

Stein, C. M., Schick, J. H., Taylor, G. H., Shriberg, L. D., Millard, C., Kundtz-Kluge, A., et al. (2004). Pleiotropic effects of a chromosome 3 locus on speech-sound disorder and reading. *American Journal of Human Genetics*, *74*, 283–297.

Trikalinos, T. A., Karvouni, A., Zintzaras, E., Ylisaukko-oja, T., Peltonen, L., Jarvela, I., et al. (2006). A heterogeneity-based genome search meta-analysis for autism-spectrum disorders. *Molecular Psychiatry*, *11*, 29–36.

Veenstra-Vanderweele, J., Christian, S. L., & Cook, E. H. J. (2004). Autism as a paradigmatic complex genetic disorder. *Annual Review of Genomics & Human Genetics*, *5*, 379–405.

Willcutt, E. G., Pennington, B. F., Smith, S. D., Cardon, L. R., Gayan, J., Knopik, V. S., et al. (2002). Quantitative trait locus for reading disability on chromosome 6p is pleiotropic for attention-deficit/hyperactivity disorder. *American Journal of Medical Genetics*, *114*, 260–268.