Review

Biological Psychiatry

Genetic Epidemiology and Insights into Interactive Genetic and Environmental Effects in Autism Spectrum Disorders

Young Shin Kim and Bennett L. Leventhal

ABSTRACT

Understanding the pathogenesis of neurodevelopmental disorders has proven to be challenging. Using autism spectrum disorder (ASD) as a paradigmatic neurodevelopmental disorder, this article reviews the existing literature on the etiological substrates of ASD and explores how genetic epidemiology approaches including gene-environment interactions (G × E) can play a role in identifying factors associated with ASD etiology. New genetic and bioinformatics strategies have yielded important clues to ASD genetic substrates. The next steps for understanding ASD pathogenesis require significant effort to focus on how genes and environment interact with one another in typical development and its perturbations. Along with larger sample sizes, future study designs should include sample ascertainment that is epidemiologic and population-based to capture the entire ASD spectrum with both categorical and dimensional phenotypic characterization; environmental measurements with accuracy, validity, and biomarkers; statistical methods to address population stratification, multiple comparisons, and G × E of rare variants; animal models to test hypotheses; and new methods to broaden the capacity to search for G × E, including genome-wide and environment-wide association studies, precise estimation of heritability using dense genetic markers, and consideration of G × E both as the disease cause and a disease course modifier. Although examination of G × E appears to be a daunting task, tremendous recent progress in gene discovery has opened new horizons for advancing our understanding of the role of G × E in the pathogenesis of ASD and ultimately identifying the causes, treatments, and even preventive measures for ASD and other neurodevelopmental disorders.

Keywords: Autism spectrum disorders, Environment, Genes, Genetic epidemiology, Interactions, Neurodevelopmental disorders

http://dx.doi.org/10.1016/j.biopsych.2014.11.001

Pathogenesis of psychiatric disorders is complex, but gene discovery has provided insight into biological mechanisms underlying neurodevelopmental disorders (NDDs), including autism spectrum disorder (ASD), intellectual disabilities, and schizophrenia. Gene discovery has led to only a modest understanding of ASD biological pathways; however, recent work with rare, de novo mutations is pointing the field in new directions, including converging actions of ASD-associated mutations and the rapidly evolving progress in the use of translational strategies (in animals and humans) to understand the effects of replicated ASD-associated mutations (1). Understanding roles of rare variants in ASD has highlighted the multifactorial etiology of NDDs characterized by pleiotropy (diverse phenotypes from identical genetic factors), genetic heterogeneity (different genes causing same phenotypes), and interactions: epistasis (between genes) and gene-environment interactions (G × E) (2–10). Evidence also suggests that environmental factors lead to diverse phenotypes, depending on the developmental timing of exposure (11–13). Research to disentangle this complexity requires strategies that specifically incorporate both genetic and environmental factors. This article provides 1) an overview of the role of genetic epidemiology in identifying both genetic and environmental factors and their joint roles in NDD etiology, 2) evidence regarding genetic and environmental influences on NDDs, and 3) research strategies to advance our understanding of NDD etiology. We use ASD as an exemplar of the broader group of NDDs.

GENETIC EPIDEMIOLOGY IN ASD

Genetic epidemiology uses disparate data from bioinformatics, population genetics, epidemiology, and molecular genetics to elucidate roles for genes and their interactions with environment in the occurrence of disease in populations (14). Genetic epidemiology 1) focuses on systematic sampling to enhance generalizability of research findings, 2) studies joint effects of genes and environment, and 3) incorporates disease biology into conceptual models (15). Twin, family, linkage, and association studies are among study designs that allow examination of genetic or environmental factors in diseases. With increasing evidence for NDD heterogeneity, genetic epidemiologic studies must attend to ASD phenotypic variability in both sampling and phenotype definition.

SEE COMMENTARY ON PAGE 3

 ISSN: 0006-3223

GENETICS

Twin studies with sample sizes of 11–67 monozygotic (MZ) and 9–210 dizygotic (DZ) twin pairs yield 47%–96% ASD concordance rates in MZ twins and 0–36% in DZ twins for autism and broader ASD phenotypes, suggesting strong heritability associated with ASD (16–21). Sibling relative risk (λ_s) (ratio of ASD prevalence among siblings of individuals with ASD to general population) ranges from 1.5–19.4 (22–26).

Initial genome-wide linkage studies were underpowered for detecting genes of small effect, leading to inconsistent findings across the samples and resistance to replication. Linkage analyses using larger samples, endophenotypes, and quantitative traits yielded positive findings in specific chromosomal regions: 2q, 5, 7q, 15q, and 16p (27,28).

Candidate gene association studies are used for ASD because they are more powerful than linkage, at a given locus, allowing detection of genes of weaker effect. Because of limited knowledge about ASD pathophysiology and gene functions, only a few genes, including *SLC6A4, GABR, RELN, NLGN, MET,* or *EN2,* have been examined with infrequent and inconsistent replications (27,29). Meta-analysis of 14 family-based association studies of *5HTT,* reported findings from each study as inconsistent, with main analyses showing no association (30).

In contrast, genome-wide association studies (GWAS) exploit strengths of association studies without guessing the identity of causal genes a priori (hypothesis-free approach); this leads to unbiased, comprehensive searches for susceptibility alleles (31). Although GWAS have yielded successes in medicine (e.g., age-related macular degeneration, obesity, hypertension, diabetes) (32–35), GWAS with samples >2000 failed to identify replicable common variants for ASD (36–39). Scarce replications in searches for ASD risk alleles result from 1) sample sizes insufficient to detect modest effect sizes; 2) poor control for population stratification in case-control studies; 3) overly permissive approaches in multiple comparison corrections, especially in early candidate gene studies; 4) varied ASD phenotype definitions; and 5) diverse samples primarily selected from nonrepresentative sources (40,41).

Significant increases in copy number variations (CNVs), which are submicroscopic variations in chromosomal structure, especially de novo CNVs, in simplex ASD families (i.e., only one affected individual) have been identified (42). Several investigators reported structural variations on the short arm of chromosome 16 associated with idiopathic ASD (42–45). This 16p11.2 CNV includes ~600 kilobases and ~29 genes, 22 of which are expressed in human fetal brain (46). Studies have demonstrated 16p11.2 deletions in ~.1%–.7% of ASD and 16p11.2 duplications in ~.1%–.5%, a rate 10 times greater than base rates for this CNV in the general population (7,47–49).

The 16p11.2 CNV is associated with the following phenotypes: intellectual disabilities, developmental delay, speech problems, schizophrenia, seizures, increased body weight or obesity, and increased head circumference (4–9,48–54). Similar phenomena have been reported for a large number of ASD-associated CNVs, including CNVs on 1q21.2, 3q29, 7q11.23, 7q36.3, 15q11.2, 15q13.3, 16p13.11, 17p12, 17q12, and 22q11.21 (3).

Excess de novo single nucleotide variant burden has been observed only for loss of function (i.e., nonsense, canonical splice site, and frameshift mutations) (55–60). Multiple de novo single nucleotide variants at the same locus, compared with the null distribution in control subjects, allowed identification of genome-wide significant loci, including loss of function mutations in *SCN2A, CHD8, DYRK1A, GRIN2B, KATNAL2, POGZ, CUL3,* and *TBR1* (55–58,61). Similar to CNVs, some single nucleotide variants initially associated with single disorders are now associated with other disorders (e.g., *SCN2A* in ASD and epilepsy) (55,57,62).

More recent studies confirm ASD-related genetic heterogeneity found in earlier twin, family, and linkage studies. The ASD-related genes seem to converge on a few pathophysiologic pathways related to synaptic function and plasticity, GTPase and Ras signaling, and neurogenesis (45,63–67). Phenotype pleiotropy suggests interaction with additional genetic and nongenetic factors, acting at various developmental time points and resulting in divergent phenotype manifestations of a single genetic variant (68).

ENVIRONMENTAL FACTORS

Twin studies provide strong evidence equally for genetics and environmental factors in ASD risk. High levels of heritability (phenotypic variance owing to genetic factors), in the range of ~90%, were reported in early twin studies; a more recent twin study found larger environmental influences on ASD risk—37% heritability and 55% shared environmental liability (20,69,70). These findings have been replicated in a large independent population-based Swedish National Registry study of 2,049,973 siblings including DZ and MZ twins, yielding 50% heritability and 50% nonshared environmental influence for ASD (26). Diagnostic disparities in some MZ twin pairs also suggest that environmental factors contribute to both liability for and expression of autism-related traits (71).

Progressively higher ASD prevalence estimates (.07%–2.6%) suggest that most of the increase is attributable to greater public awareness, better case ascertainment, broadening of ASD diagnostic construct, and diagnostic substitution (72–74). If increasing ASD prevalence is even partly caused by increasing incidence, environmental factors or their interactions with as-yet-unknown genetic vulnerability may represent other ASD risk mechanisms.

Nutrients, smoking, alcohol, medications, and pesticides are the most commonly examined exposures during pregnancy because of their known neurotoxicity or specific adverse or protective impacts on developing brains (75). Epigenetics (long-term or heritable changes in function of a locus or chromosome without alteration of underlying DNA) may represent one pathway for G×E (76). There is an association of ASD with fragile X, Rett, and Angelman syndromes, each of which involves epigenetic mechanisms (77–79). Associations have been also reported between ASD and parent-of-origin syndromes (e.g., 15q11.3, Turner syndrome) (10,75,80). Two recent studies report differences in DNA methylation profiles between individuals with and without ASD; one group studied 20 postmortem brains of individuals with ASD and 21 of control subjects, finding differentially