From Candidate Genes to Genome-wide Association: The Challenges and Promise of Posttraumatic Stress Disorder Genetic Studies

Karestan C. Koenen, **Laramie E. Duncan**, **Israel Liberzon**, and **Kerry J. Ressler**
Department of Epidemiology (KCK), Mailman School of Public Health, Columbia University, New York, New York; Department of Social and Behavioral Sciences (KCK), Harvard School of Public Health; Analytic and Translational Genetics Unit (LED), Massachusetts General Hospital; Department of Medicine (LED), Harvard Medical School, Boston; Broad Institute of MIT and Harvard (LED), Cambridge, Massachusetts; Department of Psychiatry (IL), University of Michigan Medical School; Ann Arbor Veterans Health System (IL), Ann Arbor, Michigan; Department of Psychiatry and Behavioral Sciences (KJR), Emory University, Atlanta, Georgia; and the Howard Hughes Medical Institute (KJR), Chevy Chase, Maryland.

Although most persons are exposed to traumatic events throughout their lives, many to multiple traumatic events, only a minority develop posttraumatic stress disorder (PTSD) (1), suggesting that individual vulnerability and resilience factors are important in PTSD pathophysiology. Posttraumatic stress disorder is heritable (2), suggesting that these individual differences might, in part, be explained by genetic factors. Two articles in this issue, Wilker *et al.* (3) and Xie *et al.* (4), use different study designs and methods to reach the same goal: to understand the genetic basis of PTSD.

Candidate gene studies, such as that of Wilker *et al.*, are predicated on the argument that the use of prior biological knowledge will lead to the identification of robust genetic risk variants. This study builds upon one of the central features of PTSD, memory disturbance, characterized by re-experiencing symptoms. Wilker *et al.* present a careful justification for studying KIBRA, a memory-related protein. The authors report their finding that the minor alleles of two intronic single nucleotide polymorphisms (SNPs) within *WWC1*—the gene encoding the KIBRA protein—rs10038727 and rs4576167, were associated with reduced risk of lifetime PTSD in Rwandan genocide survivors. They found a dose–response relationship with these alleles and replicated the finding in a highly-traumatized Ugandan sample. This is the first report of an association between KIBRA and PTSD.

In contrast to candidate gene studies, genome-wide association studies (GWAS), such as described by Xie *et al.*, take an agnostic approach to identifying genetic variants for disease. Xie *et al.* conducted a GWAS in 1578 European Americans and 2766 African Americans who had previously participated in genetic studies of substance dependence. The authors

found a single genome-wide significant hit, rs406001, which is "intergenic" (i.e., see Table 1, a glossary of common genetic terms). The second strongest association mapped to the first intron of the Tolloid-like 1 gene (*TLL1*) in the European American sample but not the African-American sample. Two SNPs (rs6812849, rs7691872) in *TLL1* replicated in an independent European-American sample. These results provide the first evidence for a role for *TLL1* in PTSD.

Wilker *et al.*, with its strong a priori hypothesis, good SNP coverage across the gene, and replication sample, represents the best of modern candidate gene studies. However, such studies have fallen out of favor in many areas of medical genetics, due to the high rates of false positive findings and low rates of replication (5). Many journals, including *Behavior Genetics* and *Neuropsychiatric Genetics*, have established policies requiring stricter standards for such studies. An additional criticism of candidate genes studies is that current biological knowledge is often insufficient to correctly specify candidate gene hypotheses. The GWAS studies have revealed that the majority of true risk variants identified by GWAS are found in noncoding regions rather than being located in the protein-coding portions or exomes of genes (note that the SNPs associated with PTSD in both Wilker *et al.* and Xie *et al.* were intronic and intergenic). In total, only approximately 1.5% of the genome codes for proteins is collectively known as the "exome." Given that the human genome is vast (containing 3 billion base pairs of DNA; 1 million common, independent genetic variants; and approximately 20,000 protein-coding genes), the ability of GWAS to assess millions of genetic variants across the genome is viewed by many as preferable for discovering robust pathways, compared with the candidate approach, which focuses on a single genetic location (Figure 1).

Candidate gene research continues, despite these concerns. Strengths of such studies include dense coverage of targeted genes and the ability to test mechanistic hypotheses. We believe that the best approach is to proceed with appropriate caution with regard to candidate gene findings (given their poor track record in many fields), rather than dismiss candidate gene studies from the outset, but to be open to the possibility that they might yield important new leads. Replication through many studies, meta-analyses, and ultimately large GWAS approaches will lend credibility (or not) to PTSD candidate gene findings. Furthermore, it is important to note that samples sizes under 5000 or even 10,000 are now considered to be relatively "small" by modern genetics standards (6). Convincing demonstrations of association, underscored by the discovery of hundreds of loci for individual traits and disorders, now come from GWAS of tens or even hundreds of thousands of individuals (7).

Along with our colleagues in the field, we established the PTSD working group within the Psychiatric Genomics Consortium to accomplish the critical next step in PTSD genetics: to conduct very large meta- and mega-analyses of candidate and GWAS studies of PTSD—work that can only be accomplished by large "team science." Taken together, Wilker *et al.* and Xie *et al.* highlight five challenges to consider as the field moves forward.

The first challenge is how to assess, quantify, and account for trauma exposure in the analysis. Wilker *et al.* considered a cumulative measure of lifetime trauma (trauma load) exposure in their analysis. The authors also tested whether trauma load modified genetic risk

but found no evidence for interaction. Xie *et al.* did not consider trauma exposure except in a sensitivity analysis, whereby they repeated their analyses limiting the sample to trauma-exposed persons. Although the single genome-wide genotype × environment interaction study on a psychiatric phenotype (8), did not yield significant findings, such genome-wide gene × environment studies have been successful (9). Genome-wide association studies of PTSD will need to consider trauma load, type, timing, and severity, given that trauma exposure is a necessary condition for the PTSD diagnosis.

The second challenge is the definition of the PTSD phenotype. The diagnosis of PTSD is often described as a "Chinese menu" because case status depends on the person meeting a specified number of criteria within three (in DSM-IV) or four (in the new DSM-V) symptom clusters. Posttraumatic stress disorder is heterogeneous in that two people can be diagnosed with PTSD and yet have different subsets of symptoms in each of the clusters. Also, whether there are unique genetic influences on each cluster (or even each individual symptom type) is yet unknown. Wilker *et al.* tested the association of *WWC1* SNPs and PTSD symptom clusters; an implicit hypothesis was that a stronger association with *WWC1* SNPs would be observed with the re-experiencing cluster; however, no strong evidence of differential association was found. The authors acknowledge that the power for these analyses was limited.

The third challenge is comorbidity. Posttraumatic stress disorder is highly comorbid with other mental disorders, even in nonclinical samples (1). Twin studies suggest that the high comorbidity of PTSD with other disorders—including major depression; panic disorder; generalized anxiety disorder; and nicotine, alcohol, and drug dependence—is due to shared genetic factors (2). Cross-disorder analyses of genome-wide data reveal common genetic risk variants across disorders (10), and we expect such variants will also be associated with PTSD.

The fourth challenge is population stratification. Genotype distributions at many loci vary by ancestry. For example, the minor allele frequency of the genome-wide significant SNP in Xie *et al.* (rs406001) is approximately 6% in Europeans and approximately 30% in Sub-Saharan Africans. If case status also varies by ancestry, and this is not taken into account, a spurious genotype– phenotype association might be observed. Concerns about population stratification have prompted the use of predominately racially and ethnically homogenous samples. The Psychiatric Genomics Consortium (PGC) has, thus far, only published a mega-analysis on samples of European descent. In contrast, the PGC PTSD working group includes almost 10,000 samples from African and African-American participants. Thus, large scale meta- and mega-analysis of candidate gene and GWAS of PTSD will have to grapple with data from multiple populations.

The fifth challenge is sampling. Like most genetic studies in psychiatry, neither Wilker *et al.* nor Xie *et al.* use epidemiologic methods to identify population-representative samples. Thus, persons with PTSD in these studies, as in most other studies published to date, are not representative of persons with the disorder in the general population. The implication is that findings from these studies might not generalize to samples with different genetic backgrounds, trauma exposure histories, and comorbidity profiles. The lack of systematic

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

sampling is also a barrier to the use of genetic findings for risk prediction. A major focus of current biologically driven PTSD research is to identify biomarkers of the disorder that distinguish between persons at high and low risk of developing PTSD after trauma exposure to better target evidence-based interventions to high-risk groups. For genetic variants to be useful in this broader context, studies of representative samples are needed.

Wilker *et al.* and Xie *et al.* studies represent the best of the new generation of PTSD genetic research. Nevertheless, experience with other complex genetic disorders suggests reason for cautious adoption of findings from candidate gene studies and relatively "small" GWAS. Fortunately, a roadmap for successfully identifying reliable biomarkers for other complex genetic diseases such as diabetes and schizophrenia has been drawn, with hundreds of genetic loci being identified for many diseases. The international PGC-PTSD Consortium aims to achieve the same success for PTSD by bringing together researchers around the globe to conduct mega- and meta-analyses of PTSD GWAS.

Acknowledgments

References

1. Kessler RC, Sonnega A, Bromet E, Hughes M, Nelson CB. Posttraumatic stress disorder in the National Comorbidity Survey. Arch Gen Psychiatry. 1995; 52:1048–1060. [PubMed: 7492257]

2. Afifi TO, Asmundson GJ, Taylor S, Jang KL. The role of genes and environment on trauma exposure and posttraumatic stress disorder symptoms: A review of twin studies. Clin Psychol Rev. 2010; 30:101–112. [PubMed: 19892451]

3. Wilker S, Kolassa S, Vogler C, Lingenfelder B, Elbert T, Papassotiropoulos A, et al. The role of memory-related gene *WWC1* (*KIBRA*) in lifetime posttraumatic stress disorder: Evidence from two independent samples from African conflict regions. Biol Psychiatry. 2013; 74:664–671. [PubMed: 23582269]

4. Xie P, Kranzler HR, Yang C, Zhao H, Farrer LA, Gelernter J. Genome-wide association study identifies new susceptibility loci for posttraumatic stress disorder. Biol Psychiatry. 2013; 74:656–663. [PubMed: 23726511]

5. Ioannidis JP, Tarone R, McLaughlin JK. The false-positive to false-negative ratio in epidemiologic studies. Epidemiology. 2011; 22:450–456. [PubMed: 21490505]

6. Sullivan PF, Daly MJ, O'Donovan M. Genetic architectures of psychiatric disorders: The emerging picture and its implications. Nature Rev Genet. 2012; 13:537–551. [PubMed: 22777127]

7. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature. 2010; 467:832–838. [PubMed: 20881960]

8. Sonuga-Barke EJ, Lasky-Su J, Neale BM, Oades R, Chen W, Franke B, et al. Does parental expressed emotion moderate genetic effects in ADHD? An exploration using a genome wide association scan. Am J Med Genet B Neuropsychiatr Genet. 2008; 147B:1359–1368. [PubMed: 18846501]

9. Pare G, Cook NR, Ridker PM, Chasman DI. On the use of variance per genotype as a tool to identify quantitative trait interaction effects: A report from the Women's Genome Health Study. PLoS Genet. 2010; 6:e1000981. [PubMed: 20585554]

10. Smoller JW, Craddock N, Kendler K, Lee PH, Neale BM, Nurnberger JI, et al. Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. Lancet. 2013; 381:1371–1379. [PubMed: 23453885]
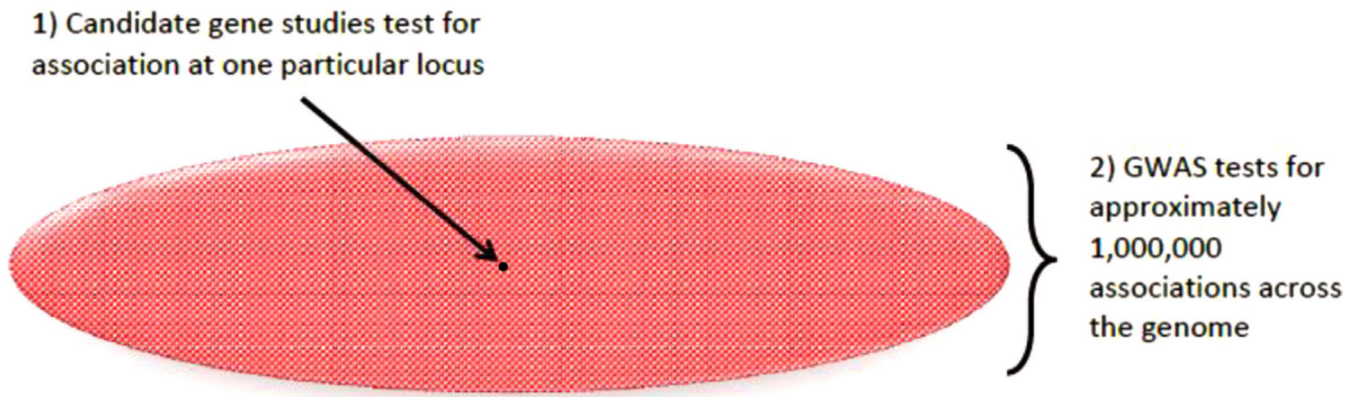
1) Candidate gene studies test for association at one particular locus

2) GWAS tests for approximately 1,000,000 associations across the genome

Figure 1.
Graphic representation of the approximately 1 million common, independent variants in genome, which provides a simplified depiction of the coverage difference between: 1) candidate gene studies; and 2) genome-wide association studies (GWAS).

Table 1

Glossary of Common Terms Used in Genetic Studies

Term	Definition
Phenotype	Outcome variable of interest, such as case/control status or a quantitative measure of symptom severity
Genome	The entirety of hereditary information of an organism. Humans have approximately 3.2 billion base pairs of primarily nuclear DNA
Locus	A location in the genome, which might be one specific position, a region such as a gene, or a segment of DNA that contains genetic variants that are correlated with one another (a "haplotype")
Allele	One of two or more alternative genetic variations at a particular locus
Genotype	For most genetic variants in the genome, humans have two copies of a genetic variant (given the 23 pairs of chromosomes of humans). The genotype of a person is the pair of alleles at a specific position in the genome
SNP	Single nucleotide polymorphism: a genetic variation indexed by a specific location in the genome and (typically) a name, which takes the form of an "rs number." For example, rs10038727 in Wilker *et al.* is an SNP with two alleles, G and A. Thus, the three genotypes that an individual might have are GG, GA, and AA
Association	Genetics term for statistical relationship, which might be assessed by logistic or linear regression or similar methods
Exome	The portion of the genome that directly codes for proteins; collectively the exome accounts for 1.5% of the genome
Intronic	From the noun "intron," intronic denotes location in an intron of a gene
Intergenic	Refers to a location in the genome that is not within known protein-coding genes