After GWAS: Searching for Genetic Risk for Schizophrenia and Bipolar Disorder

Elliot S. Gershon, M.D.

Ney Alliey-Rodriguez, M.D.

Chunyu Liu, Ph.D.

Ten years ago it was widely expected that the genetic basis of common disease would be resolved by genome-wide association studies (GWAS), large-scale studies in which the entire genome is covered by genetic markers. However, the bulk of heritable variance remains unexplained. The authors consider several alternative research strategies. For instance, whereas it has been hypothesized that a common disease is associated primarily with common genetic variants, it is now plausible that multiple rare variants each have a potent effect on disease risk and that they could accumulate to become a substantial component of common disease risk. This idea has become more appealing since the discovery that copy number variants (CNVs) are a substantial source of human mutations and are associated with multiple common diseases. CNVs are structural genomic variants consisting of microinsertions, microdeletions, and transpositions in the human genome. It has been argued that numerous rare CNVs are plausible causes of a substantial proportion of common disease, and rare CNVs have been found to be potent risk factors in schizophrenia and autism. Another approach is to "parse the genome," i.e., reanalyze subsets of current GWAS data, since the noise inherent in genome-wide approaches may be hiding valid associations. Lastly, technological advances and declining costs may allow large-scale genome-wide sequencing that would comprehensively identify all genetic variants. Study groups even larger than the 10,000 subjects in current meta-analyses would be required, but the outcomes may lead to resolution of our current dilemma in common diseases: Where is the missing heritability?

(Am J Psychiatry 2011; 168:253–256)

Twin and adoption studies during the 20th century firmly established a genetic basis for the major mental illnesses and numerous other common diseases. Heritability based on twin studies appeared to account for at least 60% of disease risk for bipolar disorder and schizophrenia (heritability is variance in illness in the population due to additive genetic causes). Ten years ago it was widely expected that the genetic basis of common disease would be resolved by genome-wide association studies (GWAS), large-scale studies in which the entire genome is covered by genetic markers. As it evolved, the GWAS strategy became identified with the "common disease, common variant" hypothesis of common disease, a conjecture that has since been found to be valid only to a limited extent. Many of the alternative strategies can be grouped into the rubric of the "common disease, multiple rare variant" hypothesis, which has become more attractive after the relatively sparse findings of GWAS-based associations, particularly in bipolar disorder and schizophrenia, as discussed in the following.

Meta-analysis of GWAS data from many thousands of schizophrenia and bipolar disorder patients and comparison subjects has revealed a few weak-effect associations, which account for only a small part of the genetic risk (1–3). (Effect strength is measured as the odds ratio of an allele or genotype frequency in patients relative to comparison subjects.) An additional but still modest proportion of disease variance appears to be accounted for in the same GWAS studies by polygenic inheritance (4). In polygenic inheritance, there are a large number of markers that, collectively, account for risk of disease, but the risk of each one is so small it cannot be detected independently. Apart from these findings, previous associations based on candidate genes have not been replicated in these very large-scale GWAS analyses. The bulk of heritable variance remains unaccounted for.

The problem of sparse results after very large-scale studies applies generally to the genetics of common diseases in the era of GWAS, and not only to neuropsychiatric diseases. We do not consider here whether there are special aspects of the psychiatric disorders that would make them less amenable to genetic analysis, because they do not appear to be less amenable. Despite the importance of progress so far, the "missing heritability" problem has attracted attention in the media (5, 6) and among scientists in the field; there is a sense of disappointment in the air. There have been several illuminating reviews of this topic (1–3, 7, 8).

What could cause current GWAS methods to fail in detection of true associations for heritable common diseases? What alternative research strategies might we consider?

Questioning the "Common Disease, Common Variant" Hypothesis

On the basis of the infrequent human DNA mutation rate (10^{-8} per DNA base pair per generation) and the relatively short evolutionary history of the human population expansion from approximately 10,000 to billions of individuals, it was reasonable to conclude that the genetic architecture of common disease in humans would generally be only one predominant and common disease allele (gene variant) for each causative gene in a common disease prior to the population expansion. Because the expansion was so rapid, it would remain the predominant allele afterward (9). These alleles could be detected by genotyping markers that, if they were chosen to effectively tag the nearby common genetic variants, would "cover" the entire human genome. All GWAS chips so far developed have predominantly common genetic variants, based on "common disease, common variant" as a working hypothesis. There are hundreds of successes of this strategy in common disease, and they have had a major impact on understanding the role of specific genes in many diseases (10) and on drug development. But generally these associations do not explain the bulk of common disease inheritance, and most of them have limited value in a clinical context.

Other Hypotheses

An alternative and plausible hypothesis (11) is that multiple rare variants each have a potent effect on the risk of a disease and that these effects accumulate to make the disease common—the "common disease, multiple rare variants" hypothesis. This would include hypotheses of extreme heterogeneity of causative variants in a single gene. However, the statistical power of current GWAS chips to detect association with rare single-nucleotide polymorphism (SNP) variants is significantly constrained by having predominantly common alleles (only about 12% of usable SNPs in the most commonly used GWAS system would detect infrequent alleles, that is, with a frequency less than 1%). Markers with frequencies similar to those of the presumed causative variants, and very large study groups, would be required to reliably detect associations with rare SNPs.

Other genetic hypotheses, such as two-hit hypotheses, are plausible and have yet to be tested (12). The most tantalizing two-hit hypothesis is that there is gene-gene interaction of common SNPs, such that association is not detected by testing SNPs or genes one at a time. The challenge in performing an analysis of this hypothesis on a genome-wide basis is that there are 23,000 genes, each with multiple SNPs, that can interact with each other. The statistical analysis of such data and the large study groups that would be required to have enough power to detect association in such an analysis have so far deterred genome-wide exploration of association with gene-gene interactions. Hypotheses on specific gene-gene interactions, of course, can be tested.

Possible Error

Another possibility for the missing heritability is that the heritability estimates may not be accurate. True heritable variance may be smaller than observed in twin studies, and true variance of known associations may be higher (1). Manolio et al. considered this possibility (2) but noted that heritability estimates from pedigree studies in animals agree well with heritability estimated from response to artificial selection, and in humans heritability of height estimated from genetic marker sharing by siblings agrees well with traditional heritability estimates, suggesting that estimates from family studies are not necessarily inflated.

CNVs as Rare Variants With Strong Effects on Disease Risk

The hypothesis of a common disease with multiple rare variants has become more appealing since the discovery that copy number variants (CNVs) are a substantial source of human mutations (13, 14) and are associated with multiple common diseases. Several of the associations were discovered by reanalysis of GWAS data for the intensity of light signals in chromosomal regions (since all GWAS chips generate such signals for SNP detection). This is a remarkable and fortunate serendipity, because so many more individuals were studied by GWAS chips than by analogous chips designed to capture CNVs.

CNVs are structural genomic variants, stretches of DNA several hundred to several million base pairs in size, consisting of microinsertions, microdeletions, and transpositions in the human genome. CNVs generally occur only in certain specific segments of the genome. Rare CNVs have been found to be potent causes of schizophrenia (15–17). Several of the rare CNVs associated with schizophrenia are also associated with autism (17). Also, in several mental and neuropsychiatric disorders, including schizophrenia, autism, and possibly bipolar disorder, there is a net increase of very rare CNVs throughout the genome.

It has been argued that a large number of rare CNVs are plausible causes of a substantial proportion of common disease for several reasons. First, the mutation rate for generation of new CNVs ranges from 100 to 10,000 times the rate of nucleotide substitutions (DNA base-pair changes) in the human genome (8). Second, all of the currently detected associations of recurrent rare CNVs with schizophrenia and autism have high odds ratios. As Vassos et al. have stated (18), these "represent genetic variants that bridge the gap between highly penetrant mutations in Mendelian, single-gene diseases and the common low-risk genetic variants typically associated with complex genetic disorders." Third, the rare CNV associations discovered so far have been found by reanalysis of GWAS marker data and by DNA hybridization with genome-wide

tiling arrays, which have yielded acceptable evidence only for fairly large CNVs (more than 100,000 DNA base pairs in length). It has recently been shown that such large CNVs constitute only about 5% of all CNVs, so it is reasonable to expect that more associations will be discovered as more sensitive CNV detection methods develop (19).

Parsing the Genome

The noise inherent in genome-wide approaches may be hiding valid associations in currently feasible numbers of subjects. GWAS analyses based on subsets may be more fortunate, despite the introduction by such analyses of multiple-testing issues. Subset analyses of GWAS data could be based on functional characterization of specific SNPs, such as their effects on gene expression (20) or DNA methylation, evolutionarily conserved regions of the genome, and pathway analyses that yield groups of genes with functional interactions. One appealing strategy based on linkage data (2) is up-weighting variants according to regional linkage scores in pedigrees; this could be done on a pedigree-by-pedigree basis.

Beyond GWAS

Technological advances in GWAS can be expected; one can envisage GWAS chips with many millions of SNP markers, which would encompass a large proportion of known rare alleles, although necessarily would miss "private" mutations. Even larger study groups than the approximately 10,000 case and comparison groups in current meta-analyses could be tested. Undoubtedly, these strategies would detect some additional associations, but they have pronounced statistical limits for rare variants (1).

An additional factor is the rapid and exponential decline in sequencing costs since the first human genome sequences appeared. Sequencing of several thousand individuals in a single project can be expected to be financially feasible within the next few years. Sequence data would enable detection of all structural variants (CNVs) and all SNP variants. There are, of course, numerous challenges in generating these enormous quantities of data, in managing them, and in statistical analyses of association. Nonetheless, these are not impossible with existing analytic approaches. Next-generation sequencing appears to be the next step in disease association analysis and would at least inform us if the missing heritability is due to missed genetic associations.

Received Sept. 20, 2010; revision received Oct. 18, 2010; accepted Oct. 18, 2010 (doi: 10.1176/appi.ajp.2010.10091340). From the Department of Psychiatry and Behavioral Neuroscience, University of Chicago. Address correspondence and reprint requests to Dr. Gershon, University of Chicago Hospitals, 5841 South Maryland Ave. (MC 3077), Chicago, IL 60637; egershon@yoda.bsd.uchicago.edu (e-mail).

Funded by NIMH grants R01 MH-080425, R01 MH-078151, and R01 MH-61613.

The authors report no financial relationships with commercial interests.

Funded by NIMH grants 1R01-MH-080425, 5R01-MH-078151, and 5R01-MH-61613.

References

1. Altshuler D, Daly MJ, Lander ES: Genetic mapping in human disease. Science 2008; 322:881–888
2. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM: Finding the missing heritability of complex diseases. Nature 2009; 461:747–753
3. McClellan J, King MC: Genetic heterogeneity in human disease. Cell 2010; 141:210–217
4. International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P: Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 2009; 460:748–752
5. Wade N: A decade later, genetic map yields few new cures. New York Times, June 12, 2010. http://www.nytimes.com/2010/06/13/health/research/13genome.html
6. Maher B: The case of the missing heritability. Nature 2008; 456:18–21
7. Girirajan S, Eichler EE: Phenotypic variability and genetic susceptibility to genomic disorders. Hum Mol Genet 2010; 19(R2):R176–R187
8. Zhang F, Gu W, Hurles ME, Lupski JR: Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet 2009; 10:451–481
9. Reich DE, Lander ES: On the allelic spectrum of human disease. Trends Genet 2001; 17:502–510
10. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA: Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci USA 2009; 106:9362–9367
11. Pritchard JK: Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet 2001; 69:124–137
12. Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, Itsara A, Vives L, Walsh T, McCarthy SE, Baker C, Mefford HC, Kidd JM, Browning SR, Browning BL, Dickel DE, Levy DL, Ballif BC, Platky K, Farber DM, Gowans GC, Wetherbee JJ, Asamoah A, Weaver DD, Mark PR, Dickerson J, Garg BP, Ellingwood SA, Smith R, Banks VC, Smith W, McDonald MT, Hoo JJ, French BN, Hudson C, Johnson JP, Ozmore JR, Moeschler JB, Surti U, Escobar LF, El-Khechen D, Gorski JL, Kussmann J, Salbert B, Lacassie Y, Biser A, Donald-McGinn DM, Zackai EH, Deardorff MA, Shaikh TH, Haan E, Friend KL, Fichera M, Romano C, Gecz J, DeLisi LE, Sebat J, King MC, Shaffer LG, Eichler EE: A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. Nat Genet 2010; 42:203–209
13. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M: Large-scale copy number polymorphism in the human genome. Science 2004; 305:525–528
14. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C: Detection of large-scale variation in the human genome. Nat Genet 2004; 36:949–951
15. McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S, Perkins DO, Dickel DE, Kusenda M, Krastoshevsky O, Krause V, Kumar RA, Grozeva D, Malhotra D, Walsh T, Zackai EH, Kaplan P, Ganesh J, Krantz ID, Spinner NB, Roccanova P,

Bhandari A, Pavon K, Lakshmi B, Leotta A, Kendall J, Lee YH, Vacic V, Gary S, Iakoucheva LM, Crow TJ, Christian SL, Lieberman JA, Stroup TS, Lehtimaki T, Puura K, Haldeman-Englert C, Pearl J, Goodell M, Willour VL, Derosse P, Steele J, Kassem L, Wolff J, Chitkara N, McMahon FJ, Malhotra AK, Potash JB, Schulze TG, Nothen MM, Cichon S, Rietschel M, Leibenluft E, Kustanovich V, Lajonchere CM, Sutcliffe JS, Skuse D, Gill M, Gallagher L, Mendell NR, Wellcome Trust Case Control Consortium, Craddock N, Owen MJ, O'Donovan MC, Shaikh TH, Susser E, DeLisi LE, Sullivan PF, Deutsch CK, Rapoport J, Levy DL, King MC, Sebat J: Microduplications of 16p11.2 are associated with schizophrenia. Nat Genet 2009; 41:1223–1227

16. International Schizophrenia Consortium: Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature 2008; 455:237–241
17. Carroll LS, Owen MJ: Genetic overlap between autism, schizophrenia and bipolar disorder. Genome Med 2009; 1:102
18. Vassos E, Collier DA, Holden S, Patch C, Rujescu D, St Cloud D, Lewis CM: Penetrance for copy number variants associated with schizophrenia. Hum Mol Genet 2010; 19:3477–3481
19. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, Kristiansson K, Macarthur DG, MacDonald JR, Onyiah I, Pang AW, Robson S, Stirrups K, Valsesia A, Walter K, Wei J, Wellcome Trust Case Control Consortium, Tyler-Smith C, Carter NP, Lee C, Scherer SW, Hurles ME: Origins and functional impact of copy number variation in the human genome. Nature 2010; 464:704–712
20. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ: Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet 2010; 6(4):e1000888