Published in final edited form as:

Biol Psychiatry. 2013 March 15; 73(6): 499–500. doi:10.1016/j.biopsych.2013.01.028.

Complex Rare Variation and Its Role in Endophenotypic Variation in Schizophrenia

John Blangero

Department of Genetics, Texas Biomedical Research Institute, San Antonio, Texas

In this issue of *Biological Psychiatry*, Yeo *et al.* (1) present some tantalizing (albeit preliminary) evidence that the cumulative number of complex rare sequence variants, specifically copy number variants (CNVs) representing deletions, correlates with postulated endophenotypes for schizophrenia such as general cognitive ability and various structural magnetic resonance imaging measures. I will use the term endophenotype for such quantitative traits because they have heritable components and they show some evidence for genetic correlation with disease liability (2). This evidence for genetic correlation may be indirect as reflected by the observed differences in means between cases, their first-degree relatives, and controls or measured directly by formal quantitative genetic analysis in families. A review of such studies by Burdick *et al.* (3) concludes that general cognitive ability is an endophenotype for schizophrenia. Multiple studies [e.g., (4)] also support schizophrenia-related endophenotypic status for a number of neuroimaging derived measures. Thus, the focal phenotypes in this study are also likely endophenotypes for schizophrenia.

The sequence variants being utilized in this study, CNVs, are deviations (in this case, a loss of sequence) from the expected human diploid state for a particular contiguous section of sequence. Focusing on this still relatively novel class of sequence variation, the investigators find a modest association between these endophenotypes and total CNV deletion burden in schizophrenics, but not in cases, and posit the existence of genotype $\times$ environment or genotype $\times$ genotype (i.e., epistasis) interactions in which there is a differential response to schizophrenia liability.

The authors write from an epidemiologic perspective, whereas in my brief comments I will provide a perspective from complex disease genetics. From a geneticist's view, there are some potentially troubling aspects of this study that should be mentioned. The study (as are many that involve expensive neuroimaging phenotyping) is very small (79 schizophrenia cases and 110 healthy controls) for detecting genetic associations. With such a small sample size, only relatively large effects can be detected. Given this low power, the great potential for biased estimates of effects expected in the winner's curse problem is substantially exacerbated (5). Additionally, the study sample subjects represent a mixture of ethnic groups that self-identify as white. These individuals are further broken into approximately 80%

Anglo and 20% Hispanic. Most studies of complex disease genetics actively avoid such ethnic mixtures or use advanced methods to deal with the potential heterogeneity to reduce the potential of type 1 errors of genetic association. This study does not use the standard approach of including principal components of the genome-wide single-nucleotide polymorphism variation as nuisance covariates to absorb hidden heterogeneity. In fact, even within studies uniquely focusing on Hispanics, it is standard practice to always use such statistical corrections for latent population stratification because of Amerindian admixture.

Ignoring potential latent population structure may or may not have affected the observed results, depending on whether there are also differences in endophenotypic distributions. However, even with these weaknesses (that are common to many studies not specifically designed around testing genetic hypotheses), the study merits careful consideration.

Based on the now universal observation that specific common single-nucleotide polymorphism variants fail to account for the genetic component of schizophrenia or cognitive ability, the authors rightfully focus on the rarer part of the allelic frequency spectrum that has been largely ignored. Large-scale whole genome sequencing now unequivocally supports the fact that, by far, most human sequence variation is rare (6). However, the dominant genome-wide association paradigm, which has largely failed to deliver causal gene identification in psychiatry, has focused exclusively on the effects of sequence variants with minor allele frequencies (MAFs) generally greater than .05 (or more typically .10).

It is both timely and wise to now focus on the rare part of the frequency spectrum because of the substantial disease heritability still unaccounted for and the expected enrichment of functional relevance resulting from natural selection. To further increase their likelihood of success, Yeo *et al.* (1) focus on a class of sequence variants (CNV deletions greater than 500 bp in size) that are more likely to be intrinsically functional because of their clear potential to alter gene structure or gene regulation. The overall benefit of focusing on such rarer CNV deletions is that the study should be biased toward detecting larger genetic effects.

There is a novelty to this study. It is the first that I know of to examine potential disease liability-mediated heterogeneity in the effects of genetic variants on schizophrenia endophenotypes. The most compelling result of the study by Yeo *et al.* (1) is their observation that a negative relationship (as measured by a partial correlation coefficient of $-.36$) between total CNV deletion burden and cognitive ability is observed only in schizophrenics. Similarly for ventricle size, a positive relationship with CNV burden was observed in schizophrenics (partial correlation coefficient $.30$), whereas there was no such correlation in controls (partial correlation coefficient $.00$). In both cases, the absolute correlation is larger in schizophrenics, suggesting that individuals with the disease are more susceptible to the effects of these CNV variants. How can such complexity occur and what does it mean?

Consider the problem rephrased as the examination of two traits, CNV deletion burden (d) and general cognitive ability (c). Conditional on the covariate effects (age, sex, and ethnicity), the authors have examined the phenotypic correlation between these two traits. In

the classical quantitative genetic model including only random genetic and random environmental effects (i.e., the classically additive polygenic model), this phenotypic correlation can be decomposed as follows:

$$\rho_{P(d,c)} = \sqrt{h_d^2 h_c^2} \rho_{G(d,c)} + \sqrt{(1-h_d^2)(1-h_c^2)} \rho_{E(d,c)},$$

where h^2 symbolizes the additive genetic heritability of a given indexed trait, $\rho_{G(d,c)}$ is the genetic correlation between the two traits, and $\rho_{E(d,c)}$ is the random environmental correlation between the traits. The first term on the right-hand side of this equation,

$\sqrt{h_d^2 h_c^2} \rho_{G(d,c)}$, represents the standardized genetic covariance between d and c . In our case, one (CNV deletion burden or d) of the two traits is theoretically completely genetically determined (i.e., its heritability is expected to be 1). This reduces the relationship to:

$$\rho_{P(d,c)} = \sqrt{h_c^2} \rho_{G(d,c)}.$$

Furthermore, because the causal direction is known (CNV variants may influence our quantitative phenotype; it is unlikely that our quantitative phenotype influences CNV burden), the square of this term ($\rho_{P(d,c)}^2 = h_c^2 \rho_{G(d,c)}^2$) provides us with an estimate of the heritability in cognitive ability that is due specifically to CNV deletion burden,

$$h_{c|d}^2 = h_c^2 \rho_{G(d,c)}^2.$$

Yeo *et al.* (1) implicitly allow this CNV burden-specific heritability to be a function of schizophrenia status. Using their partial correlation results and the above covariance mathematics, we can rephrase their core result. For schizophrenics, their data yield a CNV burden-specific heritability for general cognitive ability of .06, with an approximate 95% confidence region of .02 and .30. For controls, this CNV burden-specific heritability is estimated at .00, with an approximate 95% confidence interval of 0 and .04. The wide confidence intervals reflect the imprecision inherent in such a small sample.

The results for ventricle size are qualitatively similar. These heritabilities (or measures of genetic covariance in our reformulated case) are significantly different (as was shown by the authors' classical test of statistical interaction between group and CNV burden). Thus, from a genetic perspective, Yeo *et al.* (1) have observed evidence for a differential endophenotypic response to CNV burden as a function of the disease environment. A genetic covariance that is a function of the environment is direct evidence of genotype $\times$ environment interaction (7). Disease is an endogenous environment that has both genetic and environmental determinants. Hence, some of the observed interaction may in reality be due to genotype $\times$ genotype interaction or epistasis as the authors duly note. Although we typically search for the signature of the effects of functional sequence effects on means, here we are observing their influence on genetic covariance between two traits. This represents a

higher-order functional effect of substantial complexity that merits further examination in larger samples.

Other complications arise in the study. The standard fixed-effect regression model utilized by Yeo *et al.* (1) implicitly assumes that all CNV deletion effects are equal and in the same direction. This is very unlikely to be true. As is the case with most current studies of rare sequence variants in which we must aggregate variants to enable statistical testing, it may be more useful to employ variance component-based random-effects models (such as reference 8), which do not require such assumptions and in which the relative variance (e.g., the CNV burden-specific heritability or gene-specific heritability) is directly estimated.

Although the investigators did not observe a correlation between total CNV deletion burden and schizophrenia disease liability, such correlations between CNVs in specific genomic regions have been reliably detected in larger samples (9,10). The observed association between the CNV burden and the schizophrenia-related endophenotypes in schizophrenics alone may have been due to the unique set of CNVs captured in the schizophrenic subjects but not observed in the controls. Because of the rarity of the CNV deletions, it is highly likely that the composition of the variants included in the total CNV burden varies substantially from cases and controls. If this is the case, then the use of a global metric such as total CNV deletion burden may be masking more specific causal genetic relationships. Larger samples would ameliorate this problem.

Composite measures of genetic risk have limited utility with regard to the enhancement of translational biologic knowledge. We may be better served in the future focusing on specific CNVs or cumulative CNV variation within specific genomic regions, a strategy that has previously been shown to be valuable for schizophrenia (9,10). Although understanding the aggregate effects of sequence variants helps us to understand the overall importance of genetic factors in disease, the identification of specific functional variants directly endorses causal genes that can provide useful biologic anchors for drug target discovery and a window into the understanding of the underlying pathologic pathways.

This study has succeeded in providing us with a set of novel hypotheses to test regarding the role of genetic variation in schizophrenia liability and schizophrenia endophenotypes. The observation of context-dependent expression of genetic variance influencing endophenotypes strongly suggests that we may need to utilize more complex (but available) statistical models to enhance the potential for causal gene discovery.

Acknowledgments

Dr. Blangero reports having received research funds from Eli Lilly & Co.

References

1. Yeo RA, Gangestad SW, Liu J, Ehrlich S, Thoma RJ, Pommy J, et al. The impact of copy number deletions on general cognitive ability and ventricle size in patients with schizophrenia and healthy control subjects. *Biol Psychiatry*. 2013; 73:540–545. [PubMed: 23237311]

2. Glahn DC, Curran JE, Winkler AM, Carless MA, Kent JW Jr, Charlesworth JC, et al. High dimensional endophenotype ranking in the search for major depression risk genes. *Biol Psychiatry*. 2012; 7:6–14. [PubMed: 21982424]
3. Burdick KE, Gunawardane N, Woodberry K, Malhotra AK. The role of general intelligence as an intermediate phenotype for neuropsychiatric disorders. *Cogn Neuropsychiatry*. 2009; 14:299–311. [PubMed: 19634032]
4. Turner JA, Calhoun VD, Michael A, van Erp TG, Ehrlich S, Segall JM, et al. Heritability of multivariate gray matter measures in schizophrenia. *Twin Res Hum Genet*. 2012; 15:324–335. [PubMed: 22856368]
5. Göring HH, Terwilliger JD, Blangero J. Large upward bias in estimation of locus-specific effects from genomewide scans. *Am J Hum Genet*. 2001; 69:1357–1369. [PubMed: 11593451]
6. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; 491:56–65. [PubMed: 23128226]
7. Blangero J. Statistical genetic approaches to human adaptability. *Hum Biol*. 1993; 65:941–966. [PubMed: 8300087]
8. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet*. 2011; 89:82–93. [PubMed: 21737059]
9. Van Den Bossche MJ, Johnstone M, Strazisar M, Pickard BS, Goossens D, Lenaerts AS, et al. Rare copy number variants in neuropsychiatric disorders: Specific phenotype or not? *Am J Med Genet B Neuropsychiatr Genet*. 2012; 159B:812–822. [PubMed: 22911887]
10. Grayton HM, Fernandes C, Rujescu D, Collier DA. Copy number variations in neurodevelopmental disorders. *Prog Neurobiol*. 2012; 99:81–91. [PubMed: 22813947]