LETTER TO THE EDITOR

Copy-Number Variations and Human Disease

To the Editor: The comprehensive mapping of genomic copy-number variations (CNVs) should allow for those variants to be studied for their correlation with disease phenotypes.[1] One line of evidence that was advanced to support potential implications of CNVs for human disease was the overlap of CNVs with chromosomal loci harboring genes known to very often cause well-characterized monogenic illnesses from the OMIM database.[1] But a general concern arises from the list of CNVs that overlap rare disease genes reported in table 5 of the mapping study by Wong et al.[1] For instance, the *BSCL2* gene fell within a CNV region that had a frequency of 3 in 105 subjects. The gene was noted by the authors to be causative for spinal muscular atrophy, distal, type V (MIM 600794),[1] and indeed two missense mutations in *BSCL2* (MIM 606158.0013 and .0014) have been reported for that phenotype. However, 13 other missense mutations in that gene cause Berardinelli-Seip congenital generalized lipodystrophy (BSCL), an extremely rare autosomal recessive disorder affecting ~1 person in 10 million.[2] Thus, the observed frequency of ~3% *BSCL2*-CNV heterozygotes seems high, given the low prevalence of BSCL as ascertained clinically in the general population. Conservative assumption of a *BSCL2*-CNV frequency of 1% would predict a homozygote frequency of a major CNV rearrangement of this region of 1 person in 40,000, a frequency that is much higher than the observed prevalence of BSCL—or of any spinal muscular atrophy subtype, for that matter—in the general population. The same type of disparity between predicted and observed prevalence appears to hold true for several other genes causing very rare homozygous diseases that lie within CNV loci,[1] including *SMA3* (MIM 253400) and *SMA4* (MIM 271150), which cause spinal muscular atrophy subtypes (CNVs seen in 60 of 105 samples), and *GCK* (MIM 138979), which causes neonatal-onset diabetes (CNVs seen in 10 of 105 samples). Obviously, CNVs include both duplications and deletions, and homozygosity for a duplication-type CNV would not necessarily imply the same pathogenic consequence as a deletion-type CNV for an OMIM gene. Thus, calculations of predicted disease frequency should be based on homozygosity for the subset of deletion-type CNVs. But, because deletions would still be expected to represent a sizable subset of CNVs for at least a sizable subset of OMIM genes, the predicted disease rate would still be much higher than the observed frequency of the autosomal recessive disease phenotype in the general population.

There may be some valid biological reasons for the apparent disparities between the observed frequency of CNV heterozygotes and the reported frequencies of the related rare OMIM recessive diseases, including (1) inaccurate disease-frequency estimates in published reports that underestimate the actual frequency of the phenotype in the general population (perhaps subtle or later-onset forms of the phenotypes might be more prevalent in the general population than has been generally recognized), (2) incompatibility of homozygosity for certain completely deleted genes with fetal viability, (3) derivation of CNV-frequency estimates in nonrepresentative "normal control" samples, and (4) rescue of the lost or altered function in homozygotes for a CNV by another gene product. Alternatively, there may be systematic technical reasons for potential disparities leading to overestimation of some CNV frequencies. It would thus be important to validate those CNVs with use of alternative technologies, such as quantitative PCR,[3] and to expand the samples of normal control subjects to determine whether homozygotes for those CNVs exist among "healthy" controls. In addition, studies within families, both healthy and diseased, might help to clarify the potential pathogenicity of some of these CNVs. The findings emphasize the fact that the excitement over the biological reality of CNVs within clinical and research samples should be tempered pending the development of standards and independent wide-scale replications, possibly with use of a variety of detection methods.

ROBERT A. HEGELE

Acknowledgments

R.A.H. is supported by operating grants from the Canadian Institutes of Health Research (MT14030), the Heart and Stroke Foundation of Ontario, Genome Canada through the Ontario Genomics Institute, a Career Investigator award from the Heart and Stroke Foundation of Ontario, the Edith Schulich Vinet Canada Research Chair (Tier I) in Human Genetics, and the Jacob J. Wolfe Distinguished Medical Research Chair.

Web Resource

The URL for data presented herein is as follows:

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for spinal muscular atrophy, distal, type V; *BSCL2* mutations; *SMA3*; *SMA4*; and *GCK*)

References

1. Wong KK, deLeeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Horsman DE, MacAulay C, Ng RT, Brown CJ, Eichler EE, et al (2007) A comprehensive analysis of common copy-number variations in the human genome. Am J Hum Genet 80:91–104
2. Garg A (2004) Acquired and inherited lipodystrophies. N Engl J Med 350:1220–1234
3. Feuk L, Carson AR, Scherer SW (2006) Structural variation in the human genome. Nat Rev Genet 7:85–97

From the Schulich School of Medicine and Dentistry, University of Western Ontario, and Vascular Biology Research Group, Robarts Research Institute, London, Ontario

Address for correspondence and reprints: Dr. Robert A. Hegele, Blackburn Cardiovascular Genetics Laboratory, Robarts Research Institute, 406-100 Perth Drive, London, ON, Canada N6A 5K8. E-mail: hegele@robarts.ca

Am. J. Hum. Genet. 2007;81:414.
0002-9297/2007/8102-0022$15.00
DOI: 10.1086/519220