What is a functional genetic polymorphism? Defining classes of functionality

Paul R. Albert, PhD

Ottawa Hospital Research Institute (Neuroscience), University of Ottawa, Ottawa, Ont.

In the human genome there are at least 3.1 million single nucleotide polymorphisms (SNPs), or about 1 SNP per kilobase of sequence.[1] There have been thousands of studies in mental health, including genome-wide association studies, attempting to associate mental illness or various behavioural, imaging or biochemical endophenotypes with genetic polymorphisms or sequence variants (following recommendations of the Human Genome Variation Society, "sequence variant" is a more inclusive term than "polymorphism." However, a common problem has been a lack of consistent association with a given polymorphism. For example, 5-HTTLPR is the most studied functional polymorphism in psychiatry, yet controversy remains as to the strength of association, with both positive and negative findings having been reported.[2] In psychiatry in particular, there remains a mismatch between association of SNPs with disease and their functional role in disease. To begin to link SNP associations with potential functional roles, I propose a classification system that ranks the extent to which a given SNP has been demonstrated to have a functional role, with the highest rank being its role in the behaviours that result in mental illness. However, the proposed classification does not rank the magnitude of functional impact of a polymorphism, and hence cannot predict the likelihood of association of a given SNP with mental illness.

To understand the mechanistic basis by which a polymorphism is associated with a particular phenotype or behavioural outcome, it is necessary to know whether that polymorphism is functional (i.e., whether it alters the function of a gene or set of genes). In most cases, the function of an associated polymorphism is not defined and must be surmised or extrapolated as an effect on the gene that contains this polymorphism. In rare cases, a polymorphism may be a nonsynonymous coding region variation that alters the gene product protein structure. Most common polymorphisms are potential regulatory polymorphisms located in noncoding regions, including promoter/upstream, downstream and intron regions, that may affect transcription;[3] in intron and untranslated regions transcribed as RNA that may affect

transcription, RNA splicing, stability or translation;[4] or in intergenic regions of unknown function.[5] For example, 2 SNPs in the dopamine D2 receptor gene introns 5–6 alter its splicing to favour the generation of the D2 long over short receptor isoform and are associated with reductions in working memory and reduced frontostriatal activation in people with schizophrenia and people who abuse cocaine.[6–8] Individual SNPs may have minimal functional impact but may be in linkage disequilibrium with a set of polymorphisms that form a haplotype associated with a functional outcome on gene expression or function.[9] Whereas a polymorphism may have a demonstrated effect on the function of a gene, there are varying degrees of demonstrating the function of a genetic variant. These range from in vitro studies to studies that seek to determine the functional impact of a given polymorphism on gene expression in humans as the clearest indication of a "functional polymorphism." Hence, for the purposes of generating discussion on this issue, a potential classification system to rank the function of a polymorphism or sequence variant is presented.

I propose 4 classes of genetic polymorphisms:
- Class 0: Function not determined. Either (A) no function is known, or (B) theoretical function is predicted but has not been experimentally demonstrated.
- Class 1: Functional in vitro. The functional effect of the polymorphism on a target DNA element or regulatory mechanism has been demonstrated using in vitro assays (e.g., gel shift, reporter assay, ligand binding); however, the function of the polymorphism on endogenous gene expression or in vivo is unknown.
- Class 2: Functional in vivo. In addition to class 1 requirements, (A) function effect of the polymorphism on the endogenous gene has been tested in model cellular systems (e.g., human transformed cell lines, human B lymphoblasts, primary cell cultures) using methods such as relative allelic expression and chromatin immunoprecipitation, and (B) in vitro function is correlated with a functional change in human tissue.

- Class 3: Functional phenocopy. In addition to class 1 requirements, (A) function has been demonstrated in vivo using model organisms such as knockin mice, and (B) function is correlated with a functional change in human tissue.

Note that within each class, the degree of functional impact may vary. Hence, the class ranking does not indicate the magnitude of functional impact of the polymorphism; rather, it indicates only the degree to which the impact has been investigated. Furthermore, the class of a polymorphism does not imply its importance in a given population: a polymorphism may have a large impact on function in vivo, but be exceedingly rare and not useful as a marker in the general population, as is the case for TPH2 G1463A.[10,11]

Class 0 (no function demonstrated)

For most polymorphisms, the specific functional effect of the polymorphism on gene expression has not been determined. For example, if the SNP is not located in a region of predicted or known functional interest, "function" can only be inferred from the gene location. In some cases (0B), the polymorphism may lie in a region of functional interest, such as a coding region or promoter region. It may even be a region with strong interspecies homology or known function. However, if no data are presented supporting the proposed function, the polymorphism is still nonfunctional. For example, a polymorphism located in a consensus sequence for a DNA binding protein may be predicted to disrupt that site, but often consensus sites lack demonstrated function. It has been found that many polymorphisms in promoter regions that may "disrupt" a putative consensus DNA element do not actually influence gene function. In genes differentially expressed in people with schizophrenia, 25%–35% of "promoter" polymorphisms demonstrated in vitro "functionality" of a greater than 50% effect on transcription, as detected using a reporter assay.[12,13] However, as discussed below, this level of functionality remains insufficient to ascertain a functional polymorphism.

Class 1 (functional in vitro)

Many so-called functional polymorphisms fall into this class. For polymorphisms in coding regions, a change in a single amino acid is not sufficient to determine functionality. Large deletions or premature termination of the coding sequence are admittedly very likely to be functional, but it remains critical to determine whether the truncated fragment is itself functional (e.g., as a dominant negative). Hence, without determination of the protein function, even coding polymorphisms should be designated as class 0. For example, a 4-bp deletion in DISC1 that induces a frame-shift mutation and was linked with schizophrenia in a family should be functional, yet it has been subsequently found to be present only in control individuals in a subsequent association study.[14]

For most polymorphisms located in noncoding regions, a variety of approaches, including bioinformatic search for location in consensus DNA elements, can be used to elucidate their function in regulation of transcription. However, many consensus elements are degenerate, and most sequence variants fall within a "consensus" DNA element that has not been validated for its function. Hence, the function of a specific transcription factor at the "element" and the effect of the polymorphism on this function must be experimentally demonstrated. A variety of in vitro methods can be used for known DNA elements, including electrophoretic mobility shift and super-shift assay, chromatin immunoprecipitation assays to demonstrate binding of the transcription factor, and reporter assays or reverse transcription polymerase chain reaction to demonstrate transcription regulation and allele-specific differences in transcription factor binding or activity in cell lines.[3,15] For unknown DNA elements, screening approaches, such as yeast one-hybrid or expression cloning, can be used to identify the transcription factor involved and show the effect of the polymorphism.[15] Without identification of the specific transcription factor involved, simply showing a small 50% effect on transcription by reporter assay is insufficient to define a functional polymorphism. To achieve class 1, the transcription factor affected needs to be identified, and the effect of the polymorphism on this transcription factor's actions, as well as on basal transcription of the target gene, should be demonstrated. Furthermore, demonstration that alteration of the transcription factor phenocopies the effect of the polymorphism in vivo would place this in a higher class: class 3.

Class 2 (functional in vivo)

Class 2 requires that, in addition to in vitro studies (class 1), the expression of the endogenous gene is shown to be affected by the functional polymorphism. For example, the 5-HTTLPR polymorphism was originally shown to alter serotonin transporter expression in transformed B lymphoblasts from humans in an allele-specific manner.[16] However, studies of serotonin transporter expression in the brain do not reveal allele-specific alterations in its apparent level of expression. Hence, a more stringent criterion is the demonstration of altered expression in brain tissue. For example, the *HTR1A* C(-1019)G allele has been shown to have an association with higher levels of 5-HT1A autoreceptors, consistent with its disruption of repressor function in raphe cells.[17] However, positron emission tomography studies are limited in their ability to determine precise receptor expression levels, and allele-specific expression in human cells remains to be demonstrated.

To attain class 2 designation, demonstration of in vivo activity is critical since a polymorphism may show activity in in vitro assays but may not have a great impact on gene expression in vivo. For example, we found a polymorphism in the dopamine D2 receptor gene that attenuates the binding and repression of the D2 receptor gene by transcription factor CC2D1A/Freud-1, which recognizes a site incorporating this polymorphism. Although we found that Freud-1 strongly regulates the expression of endogenous D2 receptors in cell lines,[18] preliminary results suggest that the polymorphism itself may not have a great impact on D2 receptor gene

expression because of the presence of multiple CC2D1A sites in *D2DR*. The loss of one of the sites due to the "functional" polymorphism may not greatly affect expression since the other sites compensate. Thus, I propose that in vitro demonstration of function of a polymorphism be classified as class 1.

Class 3 (functional phenocopy)

The most rigorous demonstration of functionality requires in vivo analysis of the phenotype of the polymorphism, or phenocopy of its activity.[4] A good example is the brain-derived neurotrophic factor (BDNF) Val66Met polymorphism that was modelled in mice and was shown to increase anxiety and depression behaviours and reduce BDNF protein levels in vivo.[19,20] However, despite the evidence in animal models, this polymorphism is not strongly associated with depression in multiple genome-wide association studies, and there is still no evidence that the polymorphism alters BDNF levels in humans.

Apart from yielding mechanistic information on the underlying basis of disease for the design of more effective treatment strategies, functional polymorphisms may provide information for a more function-targeted approach that can be used to refine pathway analysis of genome-wide association studies, although preliminary data suggest that this approach may need further refinement.[21] For example, a combined genotype association study of class 2 or above polymorphisms may be more productive than a study of individual polymorphisms or genome-wide associations of polymorphisms of unknown function. Conversely, a strong association identified and replicated in genome-wide association studies would provide an important target for studies to elucidate their functional importance and would possibly provide mechanistic insights into associated mental illness. However, the challenge will be to scale the analysis of function so that it could be more systematically applied to a larger number of polymorphisms. This can be done readily for reporter assays, but for identification of specific transcription factor binding or function the development of bioinformatic, array or proteomic technologies will be required to assess this on a large scale.[22]

Competing interests: None declared.

References

1. Frazer KA, Ballinger DG, Cox DR, et al.; International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007;449:851-61.
2. Karg K, Burmeister M, Shedden K, et al. The serotonin transporter promoter variant (5-HTTLPR), stress, and depression meta-analysis revisited: evidence of genetic moderation. *Arch Gen Psychiatry* 2011;68:444-54.
3. Chorley BN, Wang X, Campbell MR, et al. Discovery and verification of functional single nucleotide polymorphisms in regulatory genomic regions: current and developing technologies. *Mutat Res* 2008;659:147-57.
4. Sadee W, Wang D, Papp AC, et al. Pharmacogenomics of the RNA world: structural RNA polymorphisms in drug therapy. *Clin Pharmacol Ther* 2011;89:355-65.
5. Liao P-Y, Lee KH. From SNPs to functional polymorphism: the in-
sight into biotechnology applications. *Biochem Eng J* 2010;49:149-58.
6. Zhang Y, Bertolino A, Fazio L, et al. Polymorphisms in human dopamine D2 receptor gene affect gene expression, splicing, and neuronal activity during working memory. *Proc Natl Acad Sci U S A* 2007;104:20552-7.
7. Bertolino A, Fazio L, Caforio G, et al. Functional variants of the dopamine receptor D2 gene modulate prefronto-striatal phenotypes in schizophrenia. *Brain* 2009;132:417-25.
8. Moyer RA, Wang D, Papp AC, et al. Intronic polymorphisms affecting alternative splicing of human dopamine D2 receptor are associated with cocaine abuse. *Neuropsychopharmacology* 2011;36:753-62.
9. Drysdale CM, McGraw DW, Stack CB, et al. Complex promoter and coding region beta 2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. *Proc Natl Acad Sci U S A* 2000;97:10483-8.
10. Blakely RD. Overview: a rare opportunity or just one less reason to be depressed. *Neuron* 2005;48:701-2, author reply 705-6.
11. Zhang X, Gainetdinov RR, Beaulieu JM, et al. Loss-of-function mutation in tryptophan hydroxylase-2 identified in unipolar major depression. *Neuron* 2005;45:11-6.
12. Hoogendoorn B, Coleman SL, Guy CA, et al. Functional analysis of human promoter polymorphisms. *Hum Mol Genet* 2003;12:2249-54.
13. Buckland PR, Hoogendoorn B, Guy CA, et al. A high proportion of polymorphisms in the promoters of brain expressed genes influences transcriptional activity. *Biochim Biophys Acta* 2004;1690:238-49.
14. Green EK, Norton N, Peirce T, et al. Evidence that a Drosoph Inf ServC1 frame-shift deletion associated with psychosis in a single family may not be a pathogenic mutation. *Mol Psychiatry* 2006;11:798-9.
15. Remes-Lenicov F, Jacobsen KX, Rogaeva A, et al. Identification of novel transcriptional regulators in the nervous system. In: Chattopadhyay A, editor. *Serotonin receptors in neurobiology*. Boca Raton (FL): CRC Press; 2007. p. 81-103.
16. Lesch KP, Bengel D, Heils A, et al. Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science* 1996;274:1527-31.
17. Parsey RV, Ogden RT, Miller JM, et al. Higher serotonin 1A binding in a second major depression cohort: modeling and reference region considerations. *Biol Psychiatry* 2010;68:170-8.
18. Rogaeva A, Ou XM, Jafar-Nejad H, et al. Differential repression by Freud-1/CC2D1A at a polymorphic site in the dopamine-D2 receptor gene. *J Biol Chem* 2007;282:20897-905.
19. Chen ZY, Jing D, Bath KG, et al. Genetic variant BDNF (Val66Met) polymorphism alters anxiety-related behavior. *Science* 2006;314:140-3.
20. Chiaruttini C, Vicario A, Li Z, et al. Dendritic trafficking of BDNF mRNA is mediated by translin and blocked by the G196A (Val66Met) mutation. *Proc Natl Acad Sci U S A* 2009;106:16481-6.
21. Collins AL, Kim Y, Sklar P, et al. Hypothesis-driven candidate genes for schizophrenia compared to genome-wide association results. *Psychol Med* 2011 Aug. 19:1-10. [Epub ahead of print]
22. Macintyre G, Bailey J, Haviv I, et al. is-rSNP: a novel technique for in silico regulatory SNP detection. *Bioinformatics* 2010;26:i524-30.