



Published in final edited form as:

Biol Psychiatry. 2017 November 15; 82(10): 696–697. doi:10.1016/j.biopsych.2017.09.004.

How good were candidate gene guesses in schizophrenia genetics?

Patrick F Sullivan^{1,2,3}

¹Department of Genetics, University of North Carolina at Chapel Hill, NC, USA

²Department of Psychiatry, University of North Carolina at Chapel Hill, NC, USA

³Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

“If it disagrees with experiment it is wrong. In that simple statement is the key to science. It does not make any difference how beautiful your guess is. It does not make any difference how smart you are, who made the guess, or what his name is – if it disagrees with experiment it is wrong. That is all there is to it.”

Richard Feynman (1, page 156-7)

Picking a candidate gene for an association study of schizophrenia is a guess. Paraphrasing Feynman (quantum physicist, Nobelist, and legendarily incisive thought experimentalist), the guess may fit beautifully into the core of an elegant neurobiological process. The guess might have sprung from the mind of a dauntingly brilliant researcher of sterling repute. But, even then, and perhaps especially then, a candidate gene is merely a guess. The proof of the guess is the “experiment”, the hard-nosed statistical evidence: is the association evidence with schizophrenia extremely strong?

As a field, we have been guessing at candidate genes for schizophrenia for over 40 years. If these papers were combined into a single academic career, if considered as the output of one researcher, they have had an enviable career: 1,064 papers, *h*-index 98, and 47.2 citations/paper (PubMed IDs from SZGene (2), 1965-2006, citations from Web of Science 9/2017). The top 10 are familiar: *BDNF*, *COMT*, neuregulin 1, dysbindin, *AKT1*, *DRD2*, and *DISC1*. The top six papers have been cited 564-2007 times. Guessing at genes for schizophrenia has been an important stratagem.

In an article on page XX (3), Matt Keller and colleagues ask an important question: how good was our guesswork? They posed a clever contrast: do lists of the top 25 or the top 86

Correspond with Dr Sullivan: Department of Genetics, CB#7264, 5097 Genomic Medicine, University of North Carolina, Chapel Hill, NC, 27599-7264, USA, pfsulliv@med.unc.edu.

URLs

Psychiatric Genomics Consortium, <https://pgc.unc.edu>

Functional Mapping and Annotation of Genome-Wide Association Studies, <http://fuma.ctglab.nl>

EGA, <https://www.ebi.ac.uk/ega>

dbGaP, <https://www.ncbi.nlm.nih.gov/gap>

NIMH Genomics, <https://www.nimhgenetics.org>

Conflicts of Interest

PFS reports the following potentially competing financial interests: Lundbeck (advisory committee, grant recipient), Pfizer (Scientific Advisory Board), Element Genomics (consultation fee), and Roche (speaker reimbursement).

historical candidate genes for schizophrenia (2) have, as a set, better evidence of association with schizophrenia? In effect, they ask, “how good was the field at guessing?”

This important question was (in my opinion) fairly, thoughtfully, and comprehensively evaluated using state-of-the-art methods and the best publically-available results for schizophrenia (4). The conclusion was clear: the field was pretty bad at guessing. In the authors’ words, “we found little evidence that common SNPs within these genes are any more relevant to schizophrenia than SNPs within control sets of non-candidate genes”, and note that ~\$US 250 million was spent on candidate gene studies (3).

Readers may raise a few obvious questions. Q: are gene set methods ever informative? A: these methods are commonly applied and often informative (e.g., reference (5)). Q: even if the full list strikes out perhaps one or two genes were correctly identified? A: inconsistent with the data (see (3, 6) for detail). Q: this study evaluated common variation; what about rare exonic variation? A: implausible, given that most of the original studies explicitly evaluated common variants (6). We now know that rare exon variants are very difficult to find, and the genes identified to date were not on anyone’s guess list (*SETD1A* and *RBM12*). Q: what about epistatic interactions? (Restated: the explicit initial candidate gene guesses didn’t pan out, so double-down on a more exotic mechanism?). A: implausible, not parsimonious, unlikely except under fairly weird circumstances (most interactions are detectable under additivity), and inconsistent with empirical data (Extended Data Figure 7 in reference (4)).

But, these objections miss the point. The track record of candidate gene guessers was no different from picking genes at random. Application of the candidate gene approach is predicated on the assumption of reasonably good guessing, and we cannot convincingly reject the null (H_0 : candidate gene guessing is indistinguishable from picking genes at random).

Johnson et al. (3) add importantly to the literature on this topic, in aggregate and for specific genes, and for schizophrenia as well as other psychiatric disorders. For example, one of the most highly cited papers in psychiatry (>4300 citations) was in *Science* in 2003 by Caspi et al. who reported a gene-environment association of HTTLPR and early stress on risk for major depressive disorder (7). It seems pretty clear that this study is wrong given lack of replication in a meta-analysis (N=38,802) (8) and in an exceptionally similar study (9).

Implications? Following on from Johnson et al. (3), if candidate gene guessing doesn’t work, how can we make progress? The strongest and most consistent clue that we have into the etiology of schizophrenia is its marked twin/pedigree heritability. How can we move from this broad clue to specific, reproducible, and actionable hypotheses about the etiology of schizophrenia? We recently put forth an agenda (10). I suggest the following:

(1) Historical candidate gene studies didn’t work, and can’t work (following from elementary school math and the now extensive knowledge of the genetic “architecture” of schizophrenia (10); caveat, there may be a few edge-case exceptions). There is little evidence to support almost all of the historical candidate genes for schizophrenia (including impact factor heavy-hitters like *COMT*, *BDNF*, *DISC1*, and *dysbindin*).

The data suggest that candidate gene guessing should be retired. This is not a new statement, as candidate gene studies have been controversial for decades, but the case can now be made forcibly.

In the scrappy, vibrant, and iconoclastic free-for-all that should characterize scientific inquiry, researchers, reviewers, journal editors, and readers can of course do whatever they choose. This includes recommending, funding, publishing, and reading/citing poor-quality candidate gene studies that do not meet the mature and widely accepted quality standards of human complex trait genomics (i.e., professional consideration of sample size, false findings due to poor control of multiple comparisons, power, population stratification). In my opinion, ignoring the body of work that has been amassed about the genetic basis of schizophrenia is wasteful and unscientific. It might yield a paper somewhere but it won't contribute to true progress. (The free-for-all sword cuts both ways, and the genomic twitterverse delights in refuting shoddy candidate gene guesses hours after appearing on-line.)

(2) Perhaps a reader disagrees deeply with these conclusions, and has some novel candidate gene guess. Most who care deeply about schizophrenia genomics are highly pragmatic, and would be pleased to be proven incorrect. But, note the emphasis on “proven” and not “opined”: if you want your guess to be believed, the burden of proof is appropriately very high and requires meeting the standards now applied in human complex trait genomics. “Suggestive” findings aren't enough. Mimicking an approach that yielded a high-profile paper in the early 2000's won't work now.

(3) How do we progress? This requires a longer answer (10). Briefly, we now know what to do, and we are making real progress. Nature has designed the genetic architectures of basically all common human diseases, disorders, and traits in a complex way. (For the present audience, this includes structural brain imaging phenotypes whose architectures are similar to other complex traits.) Schizophrenia is truly complex, and simple approaches, models, and guesswork have consistently failed. We should use approaches that have yielded evidence – but, we now know that the required sample sizes are huge. Therefore, progress requires collaboration and open-science approaches, and psychiatry is among the leaders in medicine (URLs) (10). If you have an idea, test it out using (a) on-line resources like FUMA (URLs), (b) freely available summary statistics (URLs), or (c) using individual-level data obtainable by application to genomic repositories (URLs).

(4) If progress requires meta-analysis and consortia (e.g., the next PGC schizophrenia paper has over 60,000 cases), what is one researcher to do? (a) If you have an idea, test it out (#3 above) and, if meritorious, figure out an effective way to collaborate with groups that can put your idea to a stronger test. (b) Instead of doing candidate gene genotyping, genotype with a SNP array. SNP array prices are historically low (~\$45/subject for 700K markers) to get a large amount of useful information. These can be used to identify ancestry, large copy number variants, and to generate genetic risk scores that summarize the inherited liability to schizophrenia. These are surely far more useful than genotyping *BDNF* val/met, *COMT* val/met, or *HTTLPR*. As with every technology, and although the methods are standard, there

are many ways to make a complete hash of the data, and this not for the unwise, incautious, or inexperienced.

Scientific inquiry should be self-correcting. I strongly suggest that we abandon candidate gene guesswork (as historically applied) as they have only provided false directions and wasted effort. Better approaches are of proven value. Circling back to Feynman (1): “if it disagrees with experiment it is wrong. That is all there is to it.”

Acknowledgements

PFS gratefully acknowledges support from the Swedish Research Council (Vetenskapsrådet, award D0886501), NIMH U01 MH109528 and NIMH R01 MH077139.

References

1. Feynman R (1985): The Character of Physical Law. Cambridge, MA: MIT Press.
2. Allen N, Bagade S, McQueen M, Ioannidis J, Kavvoura F, Khoury M, et al. (2008): Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: The SzGene Database. *Nature Genetics*. 40:827–834. [PubMed: 18583979]
3. Johnson EC, Border R, Melroy-Greif WE, de Leeuw C, Ehringer MA, Keller MC (2017): No evidence that schizophrenia candidate genes are more associated with schizophrenia than non-candidate genes. *Biol Psychiatry*.
4. Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014): Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 511:421–427. [PubMed: 25056061]
5. Pathway Analysis Subgroup of the Psychiatric Genomics Consortium (2015): Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat Neurosci*. 18:199–209. [PubMed: 25599223]
6. Farrell MS, Werge T, Sklar P, Owen MJ, Ophoff RA, O'Donovan MC, et al. (2015): Evaluating historical candidate genes for schizophrenia. *Molecular Psychiatry*. 20:555–562. [PubMed: 25754081]
7. Caspi A, Sugden K, Moffitt TE, Taylor A, Craig IW, Harrington H, et al. (2003): Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science*. 301:386–389. [PubMed: 12869766]
8. Culverhouse RC, Saccone NL, Horton AC, Ma Y, Anstey KJ, Banaschewski T, et al. (2017): Collaborative meta-analysis finds no evidence of a strong interaction between stress and 5-HTTLPR genotype contributing to the development of depression. *Mol Psychiatry*.
9. Fergusson DM, Horwood LJ, Miller AL, Kennedy MA (2011): Life stress, 5-HTTLPR and mental disorder: findings from a 30-year longitudinal study. *The British journal of psychiatry : the journal of mental science*. 198:129–135. [PubMed: 21282783]
10. Sullivan PF, Agrawal A, Bulik CM, Andreassen OA, Børglum AD, Breen G, et al. (In press): Psychiatric Genomics: An Update and an Agenda. *Am J Psychiatry*.