OPEN

Molecular Psychiatry (2013), 1–3

www.nature.com/mp

LETTER TO THE EDITOR

Response to 'Predicting the diagnosis of autism spectrum disorder using gene pathway analysis'

Molecular Psychiatry advance online publication, 22 October 2013; doi:10.1038/mp.2013.125

In a recent paper published online in *Molecular Psychiatry*, Skafidas *et al.*[1] report a classifier for identifying individuals at risk for autism spectrum disorders (ASDs). Their classifier is based on 267 single-nucleotide polymorphisms (SNPs) that were selected from the results of a pathway analysis using cases from the Autism Genetic Resource Exchange (AGRE).[1] Using within-sample cross-validation, the authors claim a classification accuracy for ASDs of 85.6%. They subsequently applied their classifier to ASD cases from the Simons Foundation Autism Research Initiative (SFARI) and controls from the Wellcome Trust Birth Cohort (WTBC) and report ASD classification accuracy of 71.7%.

We believe that the claims made by Skafidas *et al.*[1] are inconsistent with current knowledge of the genetics of ASDs,[2] and inconsistent with the expected precision of risk predictions for complex psychiatric disorders. Further, as classification accuracy depends on the size of the discovery sample, the results are also inconsistent with the size of the sample they employed (only 123 controls were included in the discovery set).

To examine the validity of Skafidas *et al.*'s claims, we pursued a range of analyses to assess the evidence for association between ASDs and (1) the individual SNPs named in their paper as most predictive, (2) their genetic classifier, to the extent it was described and (3) the pathways identified in the report, from which the predictive SNPs were selected. For each analysis, where possible, we attempted to replicate the analytic approach of Skafidas *et al.*[1] using data from the Psychiatric Genomics Consortium (PGC) autism group, which includes ~5400 cases, more than three times the number used in the original report. The methodology of these analyses is described in detail in Supplementary Information.

First, we found no evidence for single SNP associations between any of the 30 most contributory SNPs listed by Skafidas *et al.*[1] in their Table 2 and ASDs in the PGC (Table 1). In the current PGC meta-analysis, the mean *P*-value for these SNPs was 0.47 with a minimum 0.007, and none are notable or survive a 30 SNP correction for multiple testing. Further information on these associations can be found in Supplementary Information.

Second, we examined the classification ability of the 30 SNPs disclosed in Skafidas *et al.*[1] (their Table 2) for ASDs in the PGC. We wrote to the authors, asking for the complete list of 237 SNPs and weights, but they declined to provide the complete list. We accordingly built a classifier using the data for 30 SNPs disclosed in Skafidas *et al.*,[1] which the authors identify as the most influential (explaining approximately 58% of the total predictive power of the classifier). We constructed the classifier using two approaches. We initially used the weights provided by Skafidas *et al.*[1] and examined the predictive ability of the 30 SNP classifier in the full PGC autism sample. As described in detail in Supplementary Information, the classifier did not differ from chance in its ability to predict ASDs (AUC = 0.505, $P = 0.22$).

We then built the score using the SNP weights estimated from the PGC data. We randomly selected a set of 732 trios to build a classifier and then tested the predictive ability of the classifier in a distinct set of 243 trios (these number mirror those used by Skafidas *et al.*[1]). For all trios, we created case pseudo–control pairs to perform model building and validation, but otherwise followed the methods proposed in Skafidas *et al.*[1] (for example, using 0, 1, 3 scoring against minor allele count). We repeated this procedure across 100 random samples of the same size from the PGC autism data. Across these replicates, we tested for a difference between case and control risk scores using a *t*-test (mean risk score of cases—mean risk score of controls) and found an average t-statistic of 0.492 with an average *P*-value of 0.50 for the validation samples. We conclude that the classifier presented by Skafidas *et al.*,[1] at least as constructed using the 30 top SNPs

Table 1. Meta-analytic results for the 30 most predictive SNPs in the Skafidas classifier

SNP	*Chr*	*BP*	*A1*	*A2*	*ln(OR)*	P-*value*
rs260808	11	103 909 166	A	C	−0.024	0.510
rs769052	5	138 944 433	T	C	−0.042	0.422
rs876619	16	56 283 534	A	C	0.044	0.398
rs905646	11	88 353 802	A	G	0.062	0.167
rs968122	12	70 791 615	T	C	0.001	0.974
rs984371	11	55 577 698	T	C	0.018	0.594
rs1243679	14	21 093 733	A	G	0.027	0.710
rs1818106	11	103 913 376	A	C	0.009	0.736
rs2239118	12	2 660 753	T	C	0.054	0.097
rs2240228	19	15 852 872	A	G	0.083	0.007
rs2300497	14	90 865 283	T	C	0.034	0.408
rs2384061	2	25 135 620	A	G	0.052	0.058
rs3773540	3	55 096 928	A	G	−0.085	0.273
rs4128941	17	63 531 331	A	G	−0.123	0.085
rs4308342	4	71 884 205	T	G	−0.107	0.142
rs4648135	4	103 536 670	A	G	0.008	0.894
rs6483362	11	88 412 451	A	G	−0.0335	0.513
rs7313997	12	71 265 958	A	C	0.035	0.450
rs7562445	2	213 192 048	T	G	0.042	0.279
rs7842798	8	131 890 170	A	G	0.033	0.241
rs8053370	16	56 262 906	T	C	−0.042	0.415
rs9288685	2	233 987 114	T	C	−0.007	0.804
rs10193128	2	233 987 722	T	C	−0.015	0.581
rs10409541	19	13 433 127	T	C	0.087	0.048
rs11020772	12	70 792 582	T	G	0.001	0.966
rs11145506	9	80 264 584	T	C	−0.117	0.282
rs12317962	12	70 792 582	T	G	0.001	0.966
rs12582971	12	18 459 387	T	C	−0.001	0.981
rs17629494	10	53 560 898	T	C	−0.060	0.217
rs17643974	10	126 792 798	T	C	0.002	0.964

Abbreviations: BP, base pair in HG19; Chr, chromosome; OR, odds ratio; SNP, single-nucleotide polymorphism. The SNP name, chromosome, base pair, reference allele, alternate allele, natural log of the odds ratio and *P*-value are presented from the meta-analysis of autism spectrum disorders from the Psychiatric Genomics Consortium. This meta-analytic strategy reflects the weighted combination of the contributing cohorts reflective of power to detect association. None of the SNPs meet a multiple testing significance threshold, let alone the genome-wide association threshold of 5×10^{-8}.

Table 2. Pathway results from the PGC meta-analysis of ASDs

KEGG pathway name	*FORGE*	*INRICH*	*MAGENTA*	*SS*	*ALIGATOR*
Purine metabolism	0.715	0.012	0.140	0.477	0.255
Calcium signaling	0.907	0.719	0.828	0.782	0.987
Chemokine signaling pathway	0.060	0.870	0.614	0.418	0.879
Phosphotidylinositol signaling	0.256	0.734	0.317	0.480	0.632
Oocyte meiosis	0.986	0.522	0.743	0.771	0.301
Ubiquitin-mediated proteolysis	0.658	0.429	0.741	0.451	0.943
Wnt signaling	0.863	0.480	0.626	0.408	0.552
Axon guidance	0.611	0.502	0.289	0.083	0.654
Focal adhesion	0.837	0.435	NA	0.685	0.374
Cell adhesion molecules	0.278	0.472	0.963	0.054	0.255
Gap junction	0.786	0.768	0.780	0.676	0.926
LTM	0.006	0.011	0.078	0.066	0.014
Long-term potentiation	0.937	0.883	0.961	0.742	0.969
Long-term depression	0.727	0.450	0.643	0.230	0.422
Taste transduction	0.510	1.000	0.900	0.670	0.692
Insulin signaling pathway	0.455	0.318	0.013	0.693	0.187
GnRH signaling	0.357	0.589	0.658	0.575	0.927
Melanogenesis	0.520	0.496	0.509	0.444	0.660

Abbreviations: ASD, autism spectrum disorder; GWAS, genome-wide association study; LTM, leukocyte transendothelial migration; NA, not applicable. Pathway results from the PGC Network and Pathway Analysis (PGC-NPA) group as applied to the meta-analysis results from PGC Autism. Five different methods are presented: FORGE, INRICH, MAGENTA, Set Screen (SS) and ALIGATOR. These methods have been documented elsewhere[6–10] and represent some of the leading methods for pathway analysis using GWAS data. None of the pathways identified in the Skafidas paper survive a multiple-testing correction based on the PGC ASD meta-analysis.

named in their report, does not generalize to predict ASDs in other samples. This result strongly suggests that the Skafidas *et al.*[1] results cannot be used to predict ASDs.

We repeated the set of analyses above using a case–control design, to mirror the approach employed by Skafidas *et al.*[1] We used 732 cases matched with 732 population controls for discovery, and 243 cases matched with 243 population controls for validation, much as the authors initially reported. In these comparisons, when principal components were included in the analysis to control for population ancestry, we observed nearly identical results to what we found in the family-based study described above (see Supplementary Information). However, without controlling for population ancestry, we observed a bias in estimates of the AUC for the curve, suggesting that such bias may have contributed to the results reported by Skafidas *et al.*, as has already been suggested.[3]

Finally, we evaluated the significance of the pathways identified by Skafidas *et al.*[1] (their Table 1), the analysis which provided the basis for their SNP selection. We did not observe significant evidence for a relationship between any of these pathways and ASDs using five different pathway analysis tools in the combined PGC ASD sample set (Table 2). This result strongly suggests that the pathway analyses do not generalize to external samples and therefore cannot be validly used in the development of a classifier.

To put the results reported in Skafidas *et al.*[1] into perspective, consider the magnitude of effects implied by the results of the classifier. From the external validation experiment, the authors report an area under the receiver operating characteristic curve 0.747 (Skafidas *et al.*, Supplementary Figure S2). This result implies that their SNP-set explains ~11% of variation in liability to ASDs (assuming a prevalence of 1% and a liability threshold model).[4] For complex traits, in particular psychiatric disorders, explaining so much variation with so few SNPs and such a small discovery sample size (732 cases and 123 controls) is unprecedented, and inconsistent with results from genome-wide association studies. For example, to achieve similar levels of variance explained in human height, sample sizes of ~180 000 individuals were required.[5]

We find no evidence that the implicated SNPs, the classifier or the pathways named in Skafidas *et al.*[1] are associated with ASDs. We therefore conclude that the classifier, as presented, cannot be used in a general way to predict ASDs, and consequently is unlikely to have any translational value.

The differences between the report of Skafidas *et al.*[1] and our analyses are striking. We suspect that our failures to replicate their claims originate from several issues with the original analyses and data. In particular, the failure to control for potential population stratification in Skafidas *et al.*[1] has likely led to biased estimates of allelic effects, as suggested in a recent letter.[3] We detail other technical issues in Supplementary Information, which may also explain the differences in the results.

There are a great many challenges to the accurate interpretation of genomic data and multiple false-positive associations from technical or study design biases have been identified in the literature. We conclude that the classifier presented in Skafidas *et al.*[1] will not usefully identify individuals at risk for ASDs in the population. Nevertheless, there are increasing numbers of robust and replicable finding emerging in psychiatric genetics. These findings hold great promise for understanding the biological basis of psychiatric disorders and for translation.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

EB Robinson[1,2,3], D Howrigan[1,2,3], J Yang[4,5], S Ripke[1,2,3,6], V Anttila[1,2,3,6], LE Duncan[3,7,8,9], L Jostins[10], JC Barrett[10], SE Medland[11], DG MacArthur[1,2,3], G Breen[12], MC O'Donovan[13], NR Wray[4,5], B Devlin[14], MJ Daly[1,2,3,6], PM Visscher[4,5], PF Sullivan[15] and BM Neale[1,2,3,6]

[1]*Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA;*
[2]*Department of Medicine, Harvard Medical School, Boston, MA, USA;*
[3]*Medical and Population Genetics Program, Broad Institute for Harvard and MIT, Cambridge, MA, USA;*
[4]*The University of Queensland, Queensland Brain Institute, Brisbane, QLD, Australia;*
[5]*The Queensland Brain Institute, The University of Queensland, Brisbane, QLD, Australia;*
[6]*Stanley Center for Psychiatric Research, Broad Institute for Harvard and MIT, Cambridge, MA, USA;*
[7]*Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA;*
[8]*Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts, General Hospital, Boston, MA, USA;*
[9]*Department of Psychiatry, Harvard Medical School, Boston, MA, USA;*
[10]*Wellcome Trust Sanger Institute, Cambridge, UK;*
[11]*Queensland Institute of Medical Research, Brisbane, QLD, Australia;*
[12]*Social Genetic and Developmental Psychiatry Center, Institute of Psychiatry, King's College London, London, UK;*
[13]*MRC Centre for Neuropsychiatric Genetics & Genomics, Cardiff University School of Medicine, Cardiff, UK;*
[14]*Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA and*
[15]*Department of Genetics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, USA*
E-mail: bneale@broadinstitute.org

REFERENCES

1 Skafidas E, Testa R, Zantomio D, Chana G, Everall IP, Pantelis C. *Mol Psychiatry* advance online publication, 11 September 2012; doi:10.1038/mp.2012.126 (e-pub ahead of print).
2 Anney R, Klei L, Pinto D, Almeida J, Bacchelli E, Baird G *et al. Hum Mol Genet* 2012; **21**: 4781–4792.
3 Belgard TG, Jankovic I, Lowe JK, Geschwind DH. *Mol Psychiatry* advance online publication, 2 April 2013; doi:10.1038/mp.2013.34 (e-pub ahead of print).
4 Wray NR, Yang J, Goddard ME, Visscher PM. *PLoS Genet* 2010; **6**: e1000864.
5 Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F *et al. Nature* 2010; **467**: 832–838.
6 Moskvina V, O'Dushlaine C, Purcell S, Craddock N, Holmans P, O'Donovan MC. *Genet Epidemiol* 2011; **35**: 861–866.
7 Pedroso I, Lourdusamy A, Rietschel M, Nöthen MM, Cichon S, McGuffin P *et al. Biol Psychiatry* 2012; **72**: 311–317.
8 Holmans P, Green EK, Pahwa JS, Ferreira MA, Purcell SM, Sklar P *et al. Am J Hum Genet* 2009; **85**: 13–24.
9 Segrè AV, DIAGRAM Consortium, MAGIC investigators, Groop L, Mootha VK, Daly MJ, Altshuler D. *PLoS Genet* 2010; **6**: e1001058.
10 Lee PH, O'Dushlaine C, Thomas B, Purcell SM. *Bioinformatics* 2012; **28**: 1797–1799.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)