frontiers in GENETICS

ORIGINAL RESEARCH ARTICLE
published: 18 February 2014
doi: 10.3389/fgene.2014.00033

A scoring strategy combining statistics and functional genomics supports a possible role for common polygenic variation in autism

Jérôme Carayol[1]*, Gerard D. Schellenberg[2], Beth Dombroski[2], Claire Amiet[1], Bérengère Génin[1], Karine Fontaine[1], Francis Rousseau[1], Céline Vazart[1], David Cohen[3], Thomas W. Frazier[4], Antonio Y. Hardan[5], Geraldine Dawson[6] and Thomas Rio Frio[1]

[1] *IntegraGen, Evry, France*
[2] *Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, USA*
[3] *Groupe Hospitalier Pitié-Salpêtrière, Department of Child and Adolescent Psychiatry, AP-HP, Université Pierre et Marie Curie, Paris, France*
[4] *Center for Pediatric Behavioral Health and Center for Autism, Cleveland Clinic, Cleveland, OH, USA*
[5] *Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA*
[6] *Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC, USA*

Edited by:
Ravinesh A. Kumar, University of Chicago, USA

Reviewed by:
Judith A. Badner, University of Chicago, USA
Yong-Kyu Kim, Howard Hughes Medical Institute, USA
David Duffy, Queensland Institute of Medical Research, Australia
Benjamin Neale, Massachusetts General Hospital, USA

****Correspondence:***
Jerôme Carayol, IntegraGen, 5 rue Henri Desbrueres, 91000 Evry, France
e-mail: jerome.carayol@integragen.com

Autism spectrum disorders (ASD) are highly heritable complex neurodevelopmental disorders with a 4:1 male: female ratio. Common genetic variation could explain 40–60% of the variance in liability to autism. Because of their small effect, genome-wide association studies (GWASs) have only identified a small number of individual single-nucleotide polymorphisms (SNPs). To increase the power of GWASs in complex disorders, methods like convergent functional genomics (CFG) have emerged to extract true association signals from noise and to identify and prioritize genes from SNPs using a scoring strategy combining statistics and functional genomics. We adapted and applied this approach to analyze data from a GWAS performed on families with multiple children affected with autism from Autism Speaks Autism Genetic Resource Exchange (AGRE). We identified a set of 133 candidate markers that were localized in or close to genes with functional relevance in ASD from a discovery population (545 multiplex families); a gender specific genetic score (GS) based on these common variants explained 1% ($P = 0.01$ in males) and 5% ($P = 8.7 \times 10^{-7}$ in females) of genetic variance in an independent sample of multiplex families. Overall, our work demonstrates that prioritization of GWAS data based on functional genomics identified common variants associated with autism and provided additional support for a common polygenic background in autism.

Keywords: autism, genetic variance, polygenic model, common variants, genetic score, functional genomics

INTRODUCTION

Autism Spectrum Disorders (ASDs) are characterized by impairments in social interaction and communication, restricted interests, and repetitive behaviors with a 4:1 male to female ratio (Johnson and Myers, 2007). A recent systematic review estimated a median ASD prevalence of 62/10,000 globally and 65.5/10,000 in the US and Canada (Elsabbagh et al., 2012). Based on a standardized assessment of description of behaviors from administrative or health records, not on standardized diagnostic interview in the general population, the Centers for Disease Control estimated in 2012 that up to 1 in 88 children have an ASD (Wingate et al., 2012).

The recurrence risk in siblings of children with ASD is estimated to be 18.7% (Ozonoff et al., 2011), which is 16 times higher than in the general population. Earlier twin studies have reported pairwise concordance rates for monozygotic (MZ) twins in the range of 36–96% while in dizygotic (DZ) twins this rate was lower than 30% resulting in heritability estimates higher than 90% (Folstein and Rutter, 1977; Steffenburg et al., 1989; Bailey et al., 1995; Farley et al., 2009). The latest twin study to date confirms this high heritability with 95.2% concordance rate in MZ twins and 4.3% in DZ twins (Nordenbaek et al., 2013). To contrast these results, one recent study provided a lower estimate of heritability (37%) but with wide confidence intervals (CI) (8–84%) (Hallmayer et al., 2011). As many as 15% of cases may be attributable to rare genetic factors like *de novo* mutations, rare copy number variations (CNVs) or chromosomal abnormalities but several common variants including CNVs and single nucleotide polymorphisms (SNPs) have also been strongly linked to autism (Cook and Scherer, 2008; Freitag et al., 2010; Devlin and Scherer, 2012). The contribution of rare and common variants to autism and their possible interactions remain to be determined, but available evidence suggests that common variants, despite each not being causal, increase the susceptibility to the disorder (Abrahams and Geschwind, 2008; Klei et al., 2012; Stein et al., 2013). There is an increasing interest in making predictions of complex traits phenotypes from genetic information (Manolio, 2013). With the availability of genome-wide data, many common variants have been identified in complex diseases like cancers. Based on these results, common polygenic models are tested for

their discriminatory power and potential application in screening programs (Pharoah et al., 2008; So et al., 2011). To date, much of the ASD risk information is complex and ethical concerns have been recently discussed (Rossi et al., 2013). However, prediction may positively impact patient outcome by contributing to an earlier diagnosis of autism, thereby providing earlier access to services. Based on a small number of common variants, polygenic models has recently been proposed to estimate autism risk (Carayol et al., 2011). Assuming that thousands of common variants may explain more than 40% of autism genetic variation, one could expect a high predictive power of a polygenic model synthesizing their information (Klei et al., 2012).

Several genome-wide association studies (GWASs) have been performed to decipher the genetic etiology of autism that is attributable to common variants (i.e., SNPs) with only a few variants having shown significant associations and replicated in an independent population or in endophenotypes (Wang et al., 2009; Weiss et al., 2009; Anney et al., 2010; Connolly et al., 2012). Considering their small effect, these SNPs represent only a small proportion of the large number of common variants required to reach the reported high heritability estimates (Klei et al., 2012). By virtue of their joint effect, SNPs that do not reach the stringent genome-wide significance level in GWASs may be biologically important nonetheless (Meuwissen et al., 2001; Wray et al., 2007), and their information, if combined in genetic risk models, could be used to estimate the genetic susceptibility to complex diseases (Purcell et al., 2009; Bush et al., 2010; Levinson et al., 2012; Stahl et al., 2012). To overcome the lack of power inherent in GWASs and to decipher the common polygenic background of complex disease, novel methods (e.g., statistical noise reduction or gene ontology enrichment) have emerged that make it possible to prioritize results that did not reach statistical significance in autism studies (Anney et al., 2011; Hussman et al., 2011). In autism, a pathway-based approach was proposed to prioritize SNPs from GWAS data and combine them in a genetic classifier (Skafidas et al., 2012). Convergent functional genomic (CFG) analysis is an alternative gene-level analysis method that is used to prioritize common variants that did not reach the significance level in GWAS, based on the function of the genes in which they stand and the potential impact on the disease, through the integration of multiple lines of biological and genetic evidence (Le-Niculescu et al., 2009; Patel et al., 2010; Ayalew et al., 2012). This strategy was demonstrated to be successful in studies of bipolar disorder and schizophrenia in which SNPs, despite displaying weak statistical signals, were selected on the basis of gene-based analyses. Combined in a genetic risk prediction score, these prioritized common variants were demonstrated to discriminate affected subjects from healthy controls (Patel et al., 2010; Ayalew et al., 2012), while polygenic models based on SNPs selected according to simple *P*-value criteria from GWAS provide evidence about the role of common variants in the disorders (up to 3% of genetic variance explained) but have little value for risk prediction (Purcell et al., 2009).

Using a discovery sample of multiplex families from the Autism Genetic Resource Exchange (AGRE) (Lajonchere, 2010), we aimed to identify SNPs that were overtransmitted from parents to affected or unaffected children and which also discriminate children with autism from their unaffected siblings (Sacco et al., 2007). For this purpose, a GWAS that combined association information from parent-offspring transmission and a comparison of affected-unaffected siblings was performed. Because there is significant sexual dimorphism of the genetic risk for autism, ignoring sex effects lessens the statistical power and could lead to the failure to identify a significant proportion of the genes that contribute to risk (Ober et al., 2008; Lu and Cantor, 2012). Through adding evidence of sex-specific risk alleles, genetic models, and variable penetrance of alleles in ASD, we performed the GWAS with and without sex stratification (Stone et al., 2004; Schellenberg et al., 2006; Weiss et al., 2006; Sato et al., 2012). A scoring approach similar to CFG was then applied to extract genome-wide association signals from "noise." The scoring algorithm of the prioritization method integrates statistical characteristics from the GWAS with functional genomic evidence.

Using an independent validation population of multiplex families, SNPs that were detected within the discovery population were selected and combined in a sex specific genetic score (GS). The GS was tested for association in this validation sample and the proportion of genetic variance explained evaluated.

MATERIALS AND METHODS

PARTICIPANTS AND GENOTYPING

Two independent sets of DNA samples from autism multiplex families were used in this study (**Figure 1**). The discovery population consisted of 545 multiplex families from AGRE repository and included 964 affected siblings (773 males, 191 females; 4.1:1 male-to-female sex ratio) and 317 unaffected siblings (144 males, 173 females). The validation population consisted of 288 multiplex families from a collection at the University of Pennsylvania (originally collected at the University of Washington) that was enriched with a complementary set of 339 families from AGRE (independent from the discovery sample). It was composed of 1,000 affected siblings (812 males, 188 females; 4.3:1 male-to-female sex ratio) and 288 unaffected siblings (141 males, 147 females). The diagnosis of autism was made using the standard Autism Diagnostic Interview-Revised (ADI-R) algorithm (Lord et al., 1994). Only individuals with a "strict" definition of autism (93 and 86% of children with ASD in the discovery and validation populations, respectively), including individuals who exceeded the ADI-R cut-off for autism in all domains as defined in the AGRE repository (http://www.agre.org/), were selected to improve the power of the GWAS by homogenizing the phenotype (Shao et al., 2002; McCarthy et al., 2008). Members of the AGRE families were genotyped using the Infinium II HumanHap550 BeadChip at the Center for Applied Genomics at The Children's Hospital of Philadelphia (CHOP), as previously described (Wang et al., 2009). SNPs that failed the Hardy Weinberg Equilibrium Test ($P < 10^{-3}$) or that had a call rate less than 90% or a minor allele frequency less than 5% were removed. Mendelian transmission of alleles was checked for every SNP; genotypes that were inconsistent with Mendelian inheritance in one or several families were considered as unknown in all members of the families showing the error. SNPs identified in the discovery population were genotyped in the University of Pennsylvania collection, as previously described (Carayol et al., 2011).

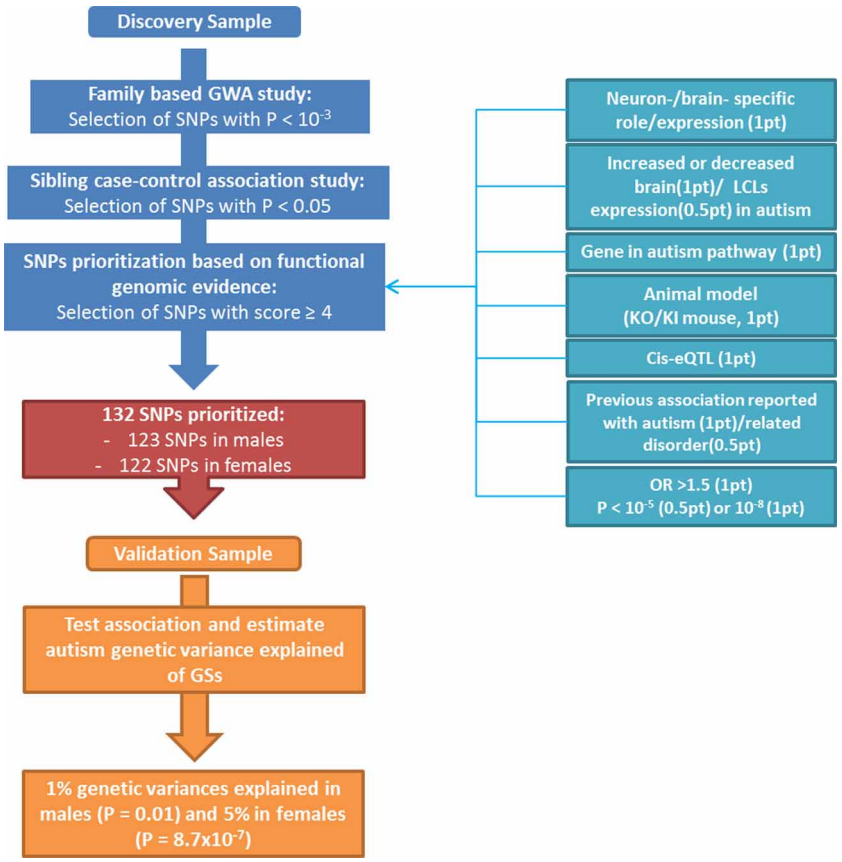

FIGURE 1 | Flow chart displaying the different steps for SNPs prioritization with details of the scoring strategy.

ASSOCIATION STUDIES

In the discovery population of 545 multiplex families from AGRE, family-based association tests were performed using the Family Based Association Test (FBAT) software (Laird et al., 2000) under additive, recessive, and dominant inheritance models to maximize power (Burton et al., 2007; Kang et al., 2010). FBAT tests for an excess of an individual SNP allele from parents to affected siblings. The interaction of sex with genotype may manifest in a genotypic effect that is apparent only in affected females, only in affected males, or in both sexes but with different magnitude or direction of effect (Strohmaier et al., 2013). Furthermore, the difference in ASD prevalence may result in male- and female- specific genes, or, more likely, from the differential penetrance of some risk alleles on the basis of the sex of the affected individual, leading to a lower power to detect existing weak effects in a particular gender stratum, compared with a higher effect in the other stratum (Hirschhorn et al., 2002; Schellenberg et al., 2006; Lu and Cantor, 2012). Consequently, affected male and female siblings were analyzed together and then independently. Because of the potential for protective alleles in autism pathogenesis, the transmission of markers from parents to unaffected siblings was also assessed (Sacco et al., 2007). SNPs associated with a *P*-value $< 10^{-3}$ were evaluated for their ability to discriminate individuals with autism from their unaffected siblings through a sibling case-control association analysis under the same gender specificity. Odds ratios (ORs) were estimated using a generalized estimating equation (GEE) model with an independence correlation matrix to account for the non-independence of individuals from the same family (Zeger and Liang, 1986; Hancock et al., 2007). Gender was introduced as an adjustment covariate when it was not used as a variable of stratification. SNPs associated at the nominal level were selected for further analysis.

SNP-GENE PAIR PRIORITIZATION

Many association tests were performed based on different genetic models and gender stratification to extract a maximum of the association signals but at the cost of an inflated number of false positive results. One way to minimize these false positive results consists of correcting *P*-values for multiple testing. At the genome-wide level, a 5×10^{-8} *P*-value threshold is generally used to define an association between a SNP and the disease as significant. Unfortunately, such criterion is too stringent in GWA study of complex disease to identify low risk common variants in samples of moderate size like in autism. To extract association signals from GWAS and minimize false-positive results, we developed a gene-based scoring method based on CFG, in which points were allocated to SNPs as indicated in **Table 1**. "Related gene" refers to the nearest gene to the SNP ($\pm$ 5 kb upstream and

downstream) or, when no gene matched this criterion, both the closest downstream and upstream genes within 50 kb of the SNP. For scoring, information was integrated from previous reports of SNP-related genes that may be linked to autism and to other neurodevelopmental disorders demonstrated to share a common genetic background [schizophrenia, bipolar disorder, and mental retardation (Ben-Shachar et al., 2009; Carroll and Owen, 2009; Berkel et al., 2010; Crespi et al., 2010)] as well as from genomic characteristics of the scored SNP (SNP location and eQTL property). Human data from autism brain gene expression and lymphoblastoid cell line expression studies were also considered in order to determine the specific role or expression of the SNP-associated gene in the central nervous system (CNS) (Purcell et al., 2001; Garbett et al., 2008). Involvement of the SNP-related gene in a pathway implicated in ASD was jointly integrated with data about impairment of the CNS or function in knockout models such as mouse. The maximum score for a gene was 9 (**Table 1**). SNPs were selected for further analyses if they displayed a score greater than or equal to 4. This cut-off was chosen based on observation of all SNPs scored and the distribution of points in the different categories. Only SNPs with a score above this cut-off displayed points allocated to statistical ("Statistical Parameters" in **Table 1**) and functional genomic evidences.

DEFINITION OF THE GENETIC MODELS AND DEVELOPMENT OF GSs

SNPs fulfilling statistical and functional genomic criteria in the discovery analysis were used to construct a GS. Sex-specific GSs of individuals were built in the discovery sample as the sum of deleterious alleles under their best fitted genetic model (Carayol et al., 2011). The "best-fitted" model was defined as the one that maximized the proportion of genetic variance in the discovery population when all SNPs were considered in a polygenic model. When only SNPs with consistent direction of allelic association were analyzed, genetic models without consistent direction of effect were excluded from the "best-fitted" model evaluation. The proportion of genetic variance explained by the GSs was estimated as in Wray et al. (2010) according to the most recent prevalence of autism in the United States, estimated to be 1 in 88 children (1 in 54 males and 1 in 252 females) and recurrence risk in siblings estimated to be 25.9 and 9.6% in male and female siblings, respectively in a large, international, multicenter,

Table 1 | Prioritization and scoring algorithm rules (details are given in Supplementary Table 2) for SNPs selected with $P < 0.001$ in the family-based GWA and $P < 0.05$ in the siblings case-control studies performed in the discovery sample.

General principle	Points	Example
Statistical parameters (max 2 points)		
P-value of the association of the SNP with autism in GWAS is $<10^{-8}$ vs. between 10^{-8} and 10^{-5}.	1 vs. 0.5	None* vs. rs7974275 (GRIN2B)
Odds ratio associated with the risk-associated allele of the SNP is $\geq$1.5 in the sibling case-control study.	1	rs4251859 (PLAUR)
Genomic characteristics (max 2 points)		
The SNP is located within the gene (including 5 kb upstream and downstream regions).	1	rs2770298 (HTR2A)
The SNP acts as an eQTL of the gene as determined by two eQTL databases, "Genevar" (Yang et al., 2010) and "eQTL resource @ Pritchard's lab" (eqtl.uchicago.edu) (Veyrieras et al., 2008; Degner et al., 2009; Pickrell et al., 2010).	1	rs2297389 (GABRR1)
Previous reporting (max 1 point)		
The gene has been associated through genome-wide or gene candidate association studies, mutations, or structural abnormalities with autism vs. with a related neurodevelopmental genetic disorder (e.g., schizophrenia, bipolar, mental retardation).	1 vs. 0.5	rs3928471 (SLC9A9) vsrs72723811 (NRG1)
Physiological properties (max 4 points)		
The expression of the gene is significantly different in patients with autism compared with controls in brain (Purcell et al., 2001; Garbett et al., 2008) or in lymphoblastoid cell lines (Gregg et al., 2008; Hu et al., 2011).	1 vs. 0.5	rs3928471 (SLC9A9) vs. rs636624 (PTPRG)
The gene has a specific role or restricted expression in the CNS.	1	rs12514116 (KCNIP1)
A mouse model exhibits either impairment of CNS development or function with or without an autism-related behavior as reported in the mouse gene informatics database from the JAX laboratory (Blake et al., 2011) and literature.	1	rs314253 (DLG4)
The gene is a part of a pathway in which other genes have been strongly associated with autism (development of the CNS, neurogenesis, neuronal migration, neuron projection, synaptogenesis, synaptic transmission) or is a part of a biochemical pathway from in which other genes have been strongly associated with autism (e.g., TSC/mTOR, MET receptor tyrosine kinase pathways).	1	rs9940922 (CDH13)

prospective study in siblings of children with ASD (Ozonoff et al., 2011; Wingate et al., 2012).

The GS model was evaluated in the validation sample testing its association to the disorder and estimating the genetic relative risk (GRR) using a GEE in males and females separately. Inflation effect on GRRs because of SNPs selection according to consistent direction of allele effect in the validation sample was evaluated based on 1,000 bootstraps resampling (Efron and Tibshirani, 1986). Mean of GRR values estimated in each bootstrap sample was then compared to the GRR observed value. Empirical 95% CIs were determined by bootstrapping 1000-times the validation sample using each family as a resampling unit.

RESULTS

A flow chart illustrating the identification and prioritization of SNPs, as well as the validation processes followed in this study, is shown in **Figure 1**. A set of 900 SNPs (Table 1 in supplementary materials) was identified from the family-based GWAS performed in the discovery population with risk allele associated at the nominal level in the sibling case-control study. This set of SNPs was then prioritized through the selected scoring strategy. One SNP, rs11123610 in the allantoicase gene, was statistically associated at the genome-wide level ($P < 1.3 \times 10^{-8}$ assuming a 5×10^{-8} genome-wide significativity threshold and a Bonferroni correction for four different GWA studies) under a recessive model with $P = 8.8 \times 10^{-9}$. Allantoicase participates in the uric acid degradation pathway, but its activity is absent in mammals (Vigetti et al., 2002). Based on statistical evidence, this SNP should be considered for replication; however, the corresponding gene did not provide any functional evidence of relation to autism and was thus discarded by our scoring criteria as a likely false-positive association. Among genes related to SNPs identified through the GWAS, 330 were associated with some functional evidence in human and/or animal models (at least 1 point from "Physiological properties," **Table 1**), suggesting potential involvement in autism. After prioritization of SNP-gene pair scores, a subset of 133 SNPs related to 119 genes had a score greater than 4 (Supplementary Table 2). The highest score was 7 for a unique SNP-gene pair, which was observed for rs4251859 related to *PLAUR*, the gene encoding the plasminogen activator, urokinase receptor. This gene has previously been described as an autism-risk gene in a large association study and has also been linked to autism based on strong functional genomic evidence from animal models and expression studies in brains from patients with autism (Powell et al., 2003; Eagleson et al., 2005; Campbell et al., 2008; Garbett et al., 2008).

One SNP was discarded from subsequent analyses because of genotyping failure in the validation population. From the subset of 132 SNPs selected by the scoring strategy (Table 3 in Supplementary materials), 123 were associated to the disorder in males and 122 in females in the Discovery sample before selection based on consistency of direction of allele effect. They respectively explained 1 and 5% of the genetic variance in the validation sample. GSs were significantly associated with autism in males ($P = 0.01$) and in females ($P = 8.7 \times 10^{-7}$).

DISCUSSION

The scoring strategy utilized in the present study prioritized a set of 132 independent common variants with low evidence of association in a GWAS that are related to 119 genes with potential links to autism. Combined in a sex-specific GS, these common variants explained 1% of the genetic variance in males and 5% in females, producing additional support for a common polygenic background in autism and the role of common variants as a part of risk for developing autism.

The present gene-based scoring approach was inspired by CFG, which has been previously shown to identify genes and pathways implicated in bipolar disorder (Le-Niculescu et al., 2009)—some of which have been confirmed in independent studies (McGrath et al., 2009)—and schizophrenia (Ayalew et al., 2012). To determine the score of a SNP-gene pair, points were allocated to the statistical characteristics of the SNP that reflected its discriminative ability. Points were also allocated to the related gene according to its association with autism and/or its role in the CNS. For example, several SNPs received points since they were located within the locus of known autism-susceptibility genes such as *PLAUR*, *HTR2A*, *RORA*, *CADM1*, *SLC9A9*, *GRIN2A*, *GABRA4*, *GABRB1*, *RBFOX1*, and *PCDH10*. Although the vast majority of scored genes had no previous evidence (significant association, mutations, differential expression) of a link with autism, some had been previously linked to genetically-related neuropsychiatric disorders such as schizophrenia and bipolar disorder and may be potentially implicated in autism, as has been demonstrated for genes such as *HTR2A*, *GRIN2A*, and *RBFOX1* (Carroll and Owen, 2009). Additionally, in light of the abnormal brain development and functioning observed in autism, the vast majority of identified genes were prioritized since they have a predominant role in the CNS and its functioning (Courchesne et al., 2007; Polsek et al., 2011). Genes with a specific role in the major processes that are altered in autism (development of the CNS, neurogenesis/neuronal migration/neuron projection, and synaptogenesis/synaptic transmission) were allocated more points to reflect their greater potential for a role in autism etiology (Bauman and Kemper, 2005; Geschwind and Levitt, 2007; Bourgeron, 2009; Wegiel et al., 2010; Melom and Littleton, 2011). The precise function of each gene was determined through a careful review of the literature including the analysis of knockout mouse models, a crucial step in the determination of the gene's function during brain development.

An allele score approach based on thousands of SNPs selected using *P*-value criteria in a GWA study explained up to 0.78% of autism variance (Anney et al., 2012). The scoring strategy alone or combine with a selection of less than a hundred of SNPs according to consistency of their allelic effect also extract a non-negligible proportion of genetic variance. Although a different cohort from the Autism Genome Project was used in their study, our results provide some evidence that the scoring strategy allow to extract true association from noise and reduce the number of selected SNPs from thousands to hundreds of informative SNP.

A set of SNPs identified during a discovery association study may include false-positive results that will outweigh the effects from true variants, decrease the predictive accuracy, and subsequently decrease the explained genetic variance (Wray et al.,

2010). Such a SNP set could be enriched for true positives compared with false positives by increasing discovery sample size (Evans et al., 2009). However, because of the small effect size of common variants, a sufficiently large sample size may not be practical for autism at this time because of the limited number of multiplex families with available genotypic information. The scoring strategy is a powerful alternative to reduce the "noise" variants by excluding SNPs in genes without functional evidence. Moreover, the proportion of genetic variance explained by GWAS data increased when the *P*-value association threshold for autism was increased from 10^{-5} to 0.5, similar to observations in related neurodevelopmental disorders, which suggests that thresholds used in the present GWAS could be relaxed to enrich the GS models with true variants (Purcell et al., 2009; Anney et al., 2010).

A pathway-based approach was proposed in autism to prioritize SNPs from GWAS and design a polygenic score that could predict autism diagnosis with good accuracy (Skafidas et al., 2012). This method only used pathway data while our scoring strategy combined their information with other biological and genetic evidences. A set of SNPs was identified in a first case-control sample and used to predict autism diagnosis with good accuracy in a second independent sample. Despite an interesting statistical approach to generate a common polygenic model in autism, results were biased because population structure of cases and controls was not considered (Belgard et al., 2013; Robinson et al., 2013).

Population structure may confound genetic classifiers as demonstrated in autism (Belgard et al., 2013; Robinson et al., 2013). However, use of unaffected siblings in association studies (instead of independent controls) protects against false positives resulting from population stratification (Dempfle et al., 2008). So, to maintain statistical power, affected, and unaffected individuals from the different families were analyzed as a whole without considering ethnicity, but limiting false positive results coming from population stratification using unaffected siblings as controls.

Although not in the scope of our study, it would have been of interest to better characterize the true underlying genetic model of the common variants. Reproducibility of the genetic models for each common variant was compared in the 2 populations using a bootstrap reproducibility index (RI) (Ma, 2006; Carayol et al., 2011). A subset of 57 SNPs with highly reproducible genetic models (RI > 80%) in both populations explained a large proportion of the genetic variance in the validation population (data not shown).

We acknowledge that different ways of scoring SNP-gene pairs could have given slightly different results. Nevertheless, the simple scoring system that was developed in this study provides a good prioritization of SNPs with regard to our focus of mining real statistical signals from "noise" in the GWAS. The major limitation of the proposed scoring approach is its dependence on available information about genes; however, its power will increase with the evolution of knowledge.

In conclusion, adding function evidence of genes to GWA results is a powerful alternative to allele score approach in order to identify common variants associated to autism. In a sample of moderate size, the scoring strategy like the CFG method allow to extract and prioritize biologically relevant signals when classical GWA studies based on *p*-value selection only identified a limited number of SNPs. Common variants displaying liberal association *p*-value (*p*-value > 0.05) explain a proportion of genetic variance in autism (Anney et al., 2012) like in schizophrenia and bipolar disorder (Purcell et al., 2009; Levinson et al., 2012). Apply the scoring strategy to SNPs selected with liberal *p*-value from GWAS could identify additional genes and enhance our understanding of the common genetic basis of autism.

AUTHOR CONTRIBUTIONS

Jérôme Carayol and Thomas Rio Frio participated in the design of the study, the analysis of data, the interpretation of data, and to draft the manuscript. Beth Dombroski, Bérengère Génin, Karine Fontaine, Francis Rousseau, and Céline Vazart participated to the genotyping of the samples and the interpretation of data. Claire Amiet, David Cohen, Thomas W. Frazier, Antonio Y. Hardan, Gerard D. Schellenberg, and Geraldine Dawson participated to the interpretation of data, and to review the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We gratefully acknowledge the resources provided by Autism Speaks Autism Genetic Resource Exchange (AGRE) Consortium and the participating AGRE families. The AGRE is a program of Autism Speaks and is supported, in part, by grant 1U24MH081810 from the National Institute of Mental Health to Clara M. Lajonchere (PI). The AGRE Consortium: Dan Geschwind, MD, PhD, UCLA, Los Angeles, CA; Maja Bucan, PhD, University of Pennsylvania, Philadelphia, PA; W. Ted Brown, MD, PhD, FACMG, NYS, Institute for Basic Research in Developmental Disabilities, Staten Island, NY; Rita M. Cantor, PhD, UCLA School of Medicine, Los Angeles, CA; John N. Constantino, MD, Washington University School of Medicine, St Louis, MO; T. Conrad Gilliam, PhD, University of Chicago, Chicago, IL; Martha Herbert, MD, PhD, Harvard Medical School, Boston, MA; Clara Lajonchere, PhD, Autism Speaks, Los Angeles, CA; David H. Ledbetter, PhD, Emory University, Atlanta, GA; Christa Lese-Martin, PhD, Emory University, Atlanta, GA; Janet Miller, JD, PhD, Autism Speaks, Los Angeles, CA; Stanley F. Nelson, MD, UCLA School of Medicine, Los Angeles, CA; Gerard D. Schellenberg, PhD, University of Washington, Seattle, WA; Carol A. Samango-Sprouse, EdD, George Washington University, Washington, DC; Sarah Spence, MD, PhD, UCLA, Los Angeles, CA; Matthew State, MD, PhD, Yale University, New Haven, CT; Rudolph E. Tanzi, PhD, Massachusetts General Hospital, Boston, MA. The University of Pennsylvania sample collection was funded by UW Autism Center for Excellence grant 5-P50-HD055782 to Geraldine Dawson (PI) and a grant from Autism Speaks.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fgene.2014.00033/abstract

REFERENCES

Abrahams, B. S., and Geschwind, D. H. (2008). Advances in autism genetics: on the threshold of a new neurobiology. *Nat. Rev. Genet.* 9, 341–355. doi: 10.1038/nrg2346

Anney, R. J., Kenny, E. M., O'Dushlaine, C., Yaspan, B. L., Parkhomenka, E., Buxbaum, J. D., et al. (2011). Gene-ontology enrichment analysis in two independent family-based samples highlights biologically plausible processes for autism spectrum disorders. *Eur. J. Hum. Genet.* 19, 1082–1089. doi: 10.1038/ejhg.2011.75

Anney, R., Klei, L., Pinto, D., Almeida, J., Bacchelli, E., Baird, G., et al. (2012). Individual common variants exert weak effects on the risk for autism spectrum disorder. *Hum. Mol. Genet.* 21, 4781–4792. doi: 10.1093/hmg/dds301

Anney, R., Klei, L., Pinto, D., Regan, R., Conroy, J., Magalhaes, T. R., et al. (2010). A genome-wide scan for common alleles affecting risk for autism. *Hum. Mol. Genet.* 19, 4072–4082. doi: 10.1093/hmg/ddq307

Ayalew, M., Le-Niculescu, H., Levey, D. F., Jain, N., Changala, B., Patel, S. D., et al. (2012). Convergent functional genomics of schizophrenia: from comprehensive understanding to genetic risk prediction. *Mol. Psychiatry* 17, 887–905. doi: 10.1038/mp.2012.37

Bailey, A., Le Couteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E., et al. (1995). Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol. Med.* 25, 63–77. doi: 10.1017/S0033291700028099

Bauman, M. L., and Kemper, T. L. (2005). Neuroanatomic observations of the brain in autism: a review and future directions. *Int. J. Dev. Neurosci.* 23, 183–187. doi: 10.1016/j.ijdevneu.2004.09.006

Belgard, T. G., Jankovic, I., Lowe, J. K., and Geschwind, D. H. (2013). Population structure confounds autism genetic classifier. *Mol. Psychiatry.* doi: 10.1038/mp.2013.34. [Epub ahead of print].

Ben-Shachar, S., Lanpher, B., German, J. R., Qasaymeh, M., Potocki, L., Nagamani, S. C., et al. (2009). Microdeletion 15q13.3: a locus with incomplete penetrance for autism, mental retardation, and psychiatric disorders. *J. Med. Genet.* 46, 382–388. doi: 10.1136/jmg.2008.064378

Berkel, S., Marshall, C. R., Weiss, B., Howe, J., Roeth, R., Moog, U., et al. (2010). Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. *Nat. Genet.* 42, 489–491. doi: 10.1038/ng.589

Blake, J. A., Bult, C. J., Kadin, J. A., Richardson, J. E., and Eppig, J. T. (2011). The Mouse Genome Database (MGD): premier model organism resource for mammalian genomics and genetics. *Nucleic Acids Res.* 39, D842–D848. doi: 10.1093/nar/gkq1008

Bourgeron, T. (2009). A synaptic trek to autism. *Curr. Opin. Neurobiol.* 19, 231–234. doi: 10.1016/j.conb.2009.06.003

Burton, P. R., Clayton, D. G., Cardon, L. R., Craddock, N., Deloukas, P., Duncanson, A., et al. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661–678. doi: 10.1038/nature05911

Bush, W. S., Sawcer, S. J., De Jager, P. L., Oksenberg, J. R., McCauley, J. L., Pericak-Vance, M. A., et al. (2010). Evidence for polygenic susceptibility to multiple sclerosis–the shape of things to come. *Am. J. Hum. Genet.* 86, 621–625. doi: 10.1016/j.ajhg.2010.02.027

Campbell, D. B., Li, C., Sutcliffe, J. S., Persico, A. M., and Levitt, P. (2008). Genetic evidence implicating multiple genes in the MET receptor tyrosine kinase pathway in autism spectrum disorder. *Autism Res.* 1, 159–168. doi: 10.1002/aur.27

Carayol, J., Schellenberg, G. D., Dombroski, B., Genin, E., Rousseau, F., and Dawson, G. (2011). Autism risk assessment in siblings of affected children using sex-specific genetic scores. *Mol. Autism* 2, 17. doi: 10.1186/2040-2392-2-17

Carroll, L. S., and Owen, M. J. (2009). Genetic overlap between autism, schizophrenia and bipolar disorder. *Genome Med.* 1, 102. doi: 10.1186/gm102

Connolly, J. J., Glessner, J. T., and Hakonarson, H. (2012). A genome-wide association study of autism incorporating autism diagnostic interview-revised, autism diagnostic observation schedule, and social responsiveness scale. *Child Dev.* 84, 17–33. doi: 10.1111/j.1467-8624.2012.01838.x

Cook, E. H. Jr., and Scherer, S. W. (2008). Copy-number variations associated with neuropsychiatric conditions. *Nature* 455, 919–923. doi: 10.1038/nature07458

Courchesne, E., Pierce, K., Schumann, C. M., Redcay, E., Buckwalter, J. A., Kennedy, D. P., et al. (2007). Mapping early brain development in autism. *Neuron* 56, 399–413. doi: 10.1016/j.neuron.2007.10.016

Crespi, B., Stead, P., and Elliot, M. (2010). Evolution in health and medicine Sackler colloquium: Comparative genomics of autism and schizophrenia. *Proc. Natl. Acad. Sci. U.S.A.* 107(Suppl. 1), 1736–1741. doi: 10.1073/pnas.0906080106

Degner, J. F., Marioni, J. C., Pai, A. A., Pickrell, J. K., Nkadori, E., Gilad, Y., et al. (2009). Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics* 25, 3207–3212. doi: 10.1093/bioinformatics/btp579

Dempfle, A., Scherag, A., Hein, R., Beckmann, L., Chang-Claude, J., and Schafer, H. (2008). Gene-environment interactions for complex traits: definitions, methodological requirements and challenges. *Eur. J. Hum. Genet.* 16, 1164–1172. doi: 10.1038/ejhg.2008.106

Devlin, B., and Scherer, S. W. (2012). Genetic architecture in autism spectrum disorder. *Curr. Opin. Genet. Dev.* 22, 229–237. doi: 10.1016/j.gde.2012.03.002

Eagleson, K. L., Bonnin, A., and Levitt, P. (2005). Region- and age-specific deficits in gamma-aminobutyric acidergic neuron development in the telencephalon of the uPAR(-/-) mouse. *J. Comp. Neurol.* 489, 449–466. doi: 10.1002/cne.20647

Efron, B., and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.* 1, 54–75. doi: 10.1214/ss/1177013815

Elsabbagh, M., Divan, G., Koh, Y. J., Kim, Y. S., Kauchali, S., Marcin, C., et al. (2012). Global prevalence of autism and other pervasive developmental disorders. *Autism Res.* 5, 160–179. doi: 10.1002/aur.239

Evans, D. M., Visscher, P. M., and Wray, N. R. (2009). Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Hum. Mol. Genet.* 18, 3525–3531. doi: 10.1093/hmg/ddp295

Farley, M. A., McMahon, W. M., Fombonne, E., Jenson, W. R., Miller, J., Gardner, M., et al. (2009). Twenty-year outcome for individuals with autism and average or near-average cognitive abilities. *Autism Res.* 2, 109–118. doi: 10.1002/aur.69

Folstein, S., and Rutter, M. (1977). Infantile autism: a genetic study of 21 twin pairs. *J. Child Psychol. Psychiatry* 18, 297–321. doi: 10.1111/j.1469-7610.1977.tb00443.x

Freitag, C. M., Staal, W., Klauck, S. M., Duketis, E., and Waltes, R. (2010). Genetics of autistic disorders: review and clinical implications. *Eur. Child Adolesc. Psychiatry* 19, 169–178. doi: 10.1007/s00787-009-0076-x

Garbett, K., Ebert, P. J., Mitchell, A., Lintas, C., Manzi, B., Mirnics, K., et al. (2008). Immune transcriptome alterations in the temporal cortex of subjects with autism. *Neurobiol. Dis.* 30, 303–311. doi: 10.1016/j.nbd.2008.01.012

Geschwind, D. H., and Levitt, P. (2007). Autism spectrum disorders: developmental disconnection syndromes. *Curr. Opin. Neurobiol.* 17, 103–111. doi: 10.1016/j.conb.2007.01.009

Gregg, J. P., Lit, L., Baron, C. A., Hertz-Picciotto, I., Walker, W., Davis, R. A., et al. (2008). Gene expression changes in children with autism. *Genomics* 91, 22–29. doi: 10.1016/j.ygeno.2007.09.003

Hallmayer, J., Cleveland, S., Torres, A., Phillips, J., Cohen, B., Torigoe, T., et al. (2011). Genetic heritability and shared environmental factors among twin pairs with autism. *Arch. Gen. Psychiatry* 68, 1095–1102. doi: 10.1001/archgenpsychiatry.2011.76

Hancock, D. B., Martin, E. R., Li, Y. J., and Scott, W. K. (2007). Methods for interaction analyses using family-based case-control data: conditional logistic regression versus generalized estimating equations. *Genet. Epidemiol.* 31, 883–893. doi: 10.1002/gepi.20249

Hirschhorn, J. N., Lohmueller, K., Byrne, E., and Hirschhorn, K. (2002). A comprehensive review of genetic association studies. *Genet. Med.* 4, 45–61. doi: 10.1097/00125817-200203000-00002

Hussman, J. P., Chung, R. H., Griswold, A. J., Jaworski, J. M., Salyakina, D., Ma, D., et al. (2011). A noise-reduction GWAS analysis implicates altered regulation of neurite outgrowth and guidance in autism. *Mol. Autism* 2, 1. doi: 10.1186/2040-2392-2-1

Hu, V. W., Addington, A., and Hyman, A. (2011). Novel autism subtype-dependent genetic variants are revealed by quantitative trait and subphenotype association analyses of published GWAS data. *PLoS ONE* 6:e19067. doi: 10.1371/journal.pone.0019067

Johnson, C. P., and Myers, S. M. (2007). Identification and evaluation of children with autism spectrum disorders. *Pediatrics* 120, 1183–1215. doi: 10.1542/peds.2007-2361

Kang, J., Cho, J., and Zhao, H. (2010). Practical issues in building risk-predicting models for complex diseases. *J. Biopharm. Stat.* 20, 415–440. doi: 10.1080/10543400903572829

Klei, L., Sanders, S. J., Murtha, M. T., Hus, V., Lowe, J. K., Willsey, A. J., et al. (2012). Common genetic variants, acting additively, are a major source of risk for autism. *Mol. Autism* 3, 9. doi: 10.1186/2040-2392-3-9

Laird, N. M., Horvath, S., and Xu, X. (2000). Implementing a unified approach to family-based tests of association. *Genet. Epidemiol.* 19(Suppl. 1), S36–S42. doi: 10.1002/1098-2272(2000)19:1+<::AID-GEPI6>3.0.CO;2-M

Lajonchere, C. M. (2010). Changing the landscape of autism research: the autism genetic resource exchange. *Neuron* 68, 187–191. doi: 10.1016/j.neuron.2010.10.009

Levinson, D. F., Shi, J., Wang, K., Oh, S., Riley, B., Pulver, A. E., et al. (2012). Genome-Wide Association Study of Multiplex Schizophrenia Pedigrees. *Am. J. Psychiatry* 169, 963–973. doi: 10.1176/appi.ajp.2012.11091423

Le-Niculescu, H., Patel, S. D., Bhat, M., Kuczenski, R., Faraone, S. V., Tsuang, M. T., et al. 3rd (2009). Convergent functional genomics of genome-wide association data for bipolar disorder: comprehensive identification of candidate genes, pathways and mechanisms. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 150B, 155–181. doi: 10.1002/ajmg.b.30887

Lord, C., Rutter, M., and Le Couteur, A. (1994). Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism Dev. Disord.* 24, 659–685. doi: 10.1007/BF02172145

Lu, A. T., and Cantor, R. M. (2012). Allowing for sex differences increases power in a GWAS of multiplex Autism families. *Mol. Psychiatry* 17, 215–222. doi: 10.1038/mp.2010.127

Manolio, T. A. (2013). Bringing genome-wide association findings into clinical use. *Nat. Rev. Genet.* 14, 549–558. doi: 10.1038/nrg3523

Ma, S. (2006). Empirical study of supervised gene screening. *BMC Bioinformatics* 7:537. doi: 10.1186/1471-2105-7-537

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P., et al. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* 9, 356–369. doi: 10.1038/nrg2344

McGrath, C. L., Glatt, S. J., Sklar, P., Le-Niculescu, H., Kuczenski, R., Doyle, A. E., et al. (2009). Evidence for genetic association of RORB with bipolar disorder. *BMC Psychiatry* 9:70. doi: 10.1186/1471-244X-9-70

Melom, J. E., and Littleton, J. T. (2011). Synapse development in health and disease. *Curr. Opin. Genet. Dev.* 21, 256–261. doi: 10.1016/j.gde.2011.01.002

Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.

Nordenbaek, C., Jorgensen, M., Kyvik, K. O., and Bilenberg, N. (2013). A Danish population-based twin study on autism spectrum disorders. *Eur. Child Adolesc. Psychiatry* 23, 35–43. doi: 10.1007/s00787-013-0419-5

Ober, C., Loisel, D. A., and Gilad, Y. (2008). Sex-specific genetic architecture of human disease. *Nat. Rev. Genet.* 9, 911–922. doi: 10.1038/nrg2415

Ozonoff, S., Young, G. S., Carter, A., Messinger, D., Yirmiya, N., Zwaigenbaum, L., et al. (2011). Recurrence risk for autism spectrum disorders: a baby siblings research consortium study. *Pediatrics* 128, e488–e495. doi: 10.1542/peds.2010-2825

Patel, S. D., Le-Niculescu, H., Koller, D. L., Green, S. D., Lahiri, D. K., McMahon, F. J., et al. (2010). Coming to grips with complex disorders: genetic risk prediction in bipolar disorder using panels of genes identified through convergent functional genomics. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 153B, 850–877. doi: 10.1002/ajmg.b.31087

Pharoah, P. D., Antoniou, A. C., Easton, D. F., and Ponder, B. A. (2008). Polygenes, risk prediction, and targeted prevention of breast cancer. *N. Engl. J. Med.* 358, 2796–2803. doi: 10.1056/NEJMsa0708739

Pickrell, J. K., Marioni, J. C., Pai, A. A., Degner, J. F., Engelhardt, B. E., Nkadori, E., et al. (2010). Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* 464, 768–772. doi: 10.1038/nature08872

Polsek, D., Jagatic, T., Cepanec, M., Hof, P. R., and Simic, G. (2011). Recent Developments in neuropathology of autism spectrum disorders. *Transl. Neurosci.* 2, 256–264. doi: 10.2478/s13380-011-0024-3

Powell, E. M., Campbell, D. B., Stanwood, G. D., Davis, C., Noebels, J. L., and Levitt, P. (2003). Genetic disruption of cortical interneuron development causes region- and GABA cell type-specific deficits, epilepsy, and behavioral dysfunction. *J. Neurosci.* 23, 622–631.

Purcell, A. E., Jeon, O. H., Zimmerman, A. W., Blue, M. E., and Pevsner, J. (2001). Postmortem brain abnormalities of the glutamate neurotransmitter system in autism. *Neurology* 57, 1618–1628. doi: 10.1212/WNL.57.9.1618

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., et al. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460, 748–752. doi: 10.1038/nature08185

Robinson, E. B., Howrigan, D., Yang, J., Ripke, S., Anttila, V., Duncan, L. E., et al. (2013). Response to "Predicting the diagnosis of autism spectrum disorder using gene pathway analysis". *Mol. Psychiatry*. doi: 10.1038/mp.2013.125. [Epub ahead of print].

Rossi, J., Newschaffer, C., and Yudell, M. (2013). Autism spectrum disorders, risk communication, and the problem of inadvertent harm. *Kennedy Inst. Ethics J.* 23, 105–138. doi: 10.1353/ken.2013.0006

Sacco, R., Papaleo, V., Hager, J., Rousseau, F., Moessner, R., Militerni, R., et al. (2007). Case-control and family-based association studies of candidate genes in autistic disorder and its endophenotypes: TPH2 and GLO1. *BMC Med. Genet.* 8:11. doi: 10.1186/1471-2350-8-11

Sato, D., Lionel, A. C., Leblond, C. S., Prasad, A., Pinto, D., Walker, S., et al. (2012). SHANK1 deletions in males with autism spectrum disorder. *Am. J. Hum. Genet.* 90, 879–887. doi: 10.1016/j.ajhg.2012.03.017

Schellenberg, G. D., Dawson, G., Sung, Y. J., Estes, A., Munson, J., Rosenthal, E., et al. (2006). Evidence for multiple loci from a genome scan of autism kindreds. *Mol. Psychiatry* 11, 1049–1060. doi: 10.1038/sj.mp.4001874

Shao, Y., Raiford, K. L., Wolpert, C. M., Cope, H. A., Ravan, S. A., Ashley-Koch, A. A., et al. (2002). Phenotypic homogeneity provides increased support for linkage on chromosome 2 in autistic disorder. *Am. J. Hum. Genet.* 70, 1058–1061. doi: 10.1086/339765

Skafidas, E., Testa, R., Zantomio, D., Chana, G., Everall, I. P., and Pantelis, C. (2012). Predicting the diagnosis of autism spectrum disorder using gene pathway analysis. *Mol. Psychiatry*. doi: 10.1038/mp.2012.126. [Epub ahead of print].

So, H. C., Kwan, J. S., Cherny, S. S., and Sham, P. C. (2011). Risk prediction of complex diseases from family history and known susceptibility loci, with applications for cancer screening. *Am. J. Hum. Genet.* 88, 548–565. doi: 10.1016/j.ajhg.2011.04.001

Stahl, E. A., Wegmann, D., Trynka, G., Gutierrez-Achury, J., Do, R., Voight, B. F., et al. (2012). Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat. Genet.* 44, 483–489. doi: 10.1038/ng.2232

Steffenburg, S., Gillberg, C., Hellgren, L., Andersson, L., Gillberg, I. C., Jakobsson, G., et al. (1989). A twin study of autism in Denmark, Finland, Iceland, Norway and Sweden. *J. Child Psychol. Psychiatry* 30, 405–416. doi: 10.1111/j.1469-7610.1989.tb00254.x

Stein, J. L., Parikshak, N. N., and Geschwind, D. H. (2013). Rare inherited variation in autism: beginning to see the forest and a few trees. *Neuron* 77, 209–211. doi: 10.1016/j.neuron.2013.01.010

Stone, J. L., Merriman, B., Cantor, R. M., Yonan, A. L., Gilliam, T. C., Geschwind, D. H., et al. (2004). Evidence for sex-specific risk alleles in autism spectrum disorder. *Am. J. Hum. Genet.* 75, 1117–1123. doi: 10.1086/426034

Strohmaier, J., Amelang, M., Hothorn, L. A., Witt, S. H., Nieratschker, V., Gerhard, D., et al. (2013). The psychiatric vulnerability gene CACNA1C and its sex-specific relationship with personality traits, resilience factors and depressive symptoms in the general population. *Mol. Psychiatry* 18, 607–613. doi: 10.1038/mp.2012.53

Veyrieras, J. B., Kudaravalli, S., Kim, S. Y., Dermitzakis, E. T., Gilad, Y., Stephens, M., et al. (2008). High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet.* 4:e1000214. doi: 10.1371/journal.pgen.1000214

Vigetti, D., Monetti, C., Prati, M., Gornati, R., and Bernardini, G. (2002). Genomic organization and chromosome localization of the murine and human allantoicase gene. *Gene* 289, 13–17. doi: 10.1016/S0378-1119(02)00541-3

Wang, K., Zhang, H., Ma, D., Bucan, M., Glessner, J. T., Abrahams, B. S., et al. (2009). Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* 459, 528–533. doi: 10.1038/nature07999

Wegiel, J., Kuchna, I., Nowicki, K., Imaki, H., Marchi, E., Ma, S. Y., et al. (2010). The neuropathology of autism: defects of neurogenesis and neuronal migration, and dysplastic changes. *Acta Neuropathol.* 119, 755–770. doi: 10.1007/s00401-010-0655-4

Weiss, L. A., Arking, D. E., Daly, M. J., and Chakravarti, A. (2009). A genome-wide linkage and association scan reveals novel loci for autism. *Nature* 461, 802–808. doi: 10.1038/nature08490

Weiss, L. A., Kosova, G., Delahanty, R. J., Jiang, L., Cook, E. H., Ober, C., et al. (2006). Variation in ITGB3 is associated with whole-blood serotonin

level and autism susceptibility. *Eur. J. Hum. Genet.* 14, 923–931. doi: 10.1038/sj.ejhg.5201644

Wingate, M., Mulvihill, B., Kirby, R. S., Pettygrove, S., Cunniff, C., Meaney, F., et al. (2012). Prevalence of autism spectrum disorders–autism and developmental disabilities monitoring network, 14 sites, United States, 2008. *MMWR. Surveill. Summ.* 61, 1–19.

Wray, N. R., Goddard, M. E., and Visscher, P. M. (2007). Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res.* 17, 1520–1528. doi: 10.1101/gr.6665407

Wray, N. R., Yang, J., Goddard, M. E., and Visscher, P. M. (2010). The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS Genet.* 6:e1000864. doi: 10.1371/journal.pgen.1000864

Yang, T. P., Beazley, C., Montgomery, S. B., Dimas, A. S., Gutierrez-Arcelus, M., Stranger, B. E., et al. (2010). Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics* 26, 2474–2476. doi: 10.1093/bioinformatics/btq452

Zeger, S. L., and Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 42, 121–130. doi: 10.2307/2531248

Conflict of Interest Statement: Jérôme Carayol, Claire Amiet, Karine Fontaine, Bérengère Génin, Francis Rousseau, Céline Vazart, and Thomas Rio Frio arecurrently salaried employees of IntegraGen. Jérôme Carayol, Francis Rousseau, and Thomas Rio Frio have patent applications with IntegraGen. Gerard D. Schellenberg and Beth Dombroski declare that they have no competing interests; Antonio Y. Hardan, David Cohen, and Geraldine Dawson are compensated consultants for IntegraGen. Thomas W. Frazier has received a clinical research grant from IntegraGen.

*Received: 16 July 2013; accepted: 29 January 2014; published online: 18 February 2014. Citation: Carayol J, Schellenberg GD, Dombroski B, Amiet C, Génin B, Fontaine K, Rousseau F, Vazart C, Cohen D, Frazier TW, Hardan AY, Dawson G and Rio Frio T (2014) A scoring strategy combining statistics and functional genomics supports a possible role for common polygenic variation in autism. Front. Genet. **5**:33. doi: 10.3389/fgene.2014.00033*

This article was submitted to Behavioral and Psychiatric Genetics, a section of the journal Frontiers in Genetics.

Copyright © 2014 Carayol, Schellenberg, Dombroski, Amiet, Génin, Fontaine, Rousseau, Vazart, Cohen, Frazier, Hardan, Dawson and Rio Frio. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.