Molecular Psychiatry (2009) 14, 416–428

www.nature.com/mp

ORIGINAL ARTICLE

Gene discovery through imaging genetics: identification of two novel genes associated with schizophrenia

SG Potkin[1], JA Turner[1], JA Fallon[1], A Lakatos[1], DB Keator[1], G Guffanti[2] and F Macciardi[2], FBIRN[1]

[1]*Department of Psychiatry and Human Behavior, Brain Imaging Center, University of California, Irvine, CA, USA and* [2]*Department of Science and Biomedical Technology, University of Milan, Milan, Italy*

We have discovered two genes, *RSRC1* and *ARHGAP18*, associated with schizophrenia and in an independent study provided additional support for this association. We have both discovered and verified the association of two genes, *RSRC1* and *ARHGAP18*, with schizophrenia. We combined a genome-wide screening strategy with neuroimaging measures as the quantitative phenotype and identified the single nucleotide polymorphisms (SNPs) related to these genes as consistently associated with the phenotypic variation. To control for the risk of false positives, the empirical *P*-value for association significance was calculated using permutation testing. The quantitative phenotype was Blood-Oxygen-Level Dependent (BOLD) Contrast activation in the left dorsal lateral prefrontal cortex measured during a working memory task. The differential distribution of SNPs associated with these two genes in cases and controls was then corroborated in a larger, independent sample of patients with schizophrenia ($n=82$) and healthy controls ($n=91$), thus suggesting a putative etiological function for both genes in schizophrenia. Up until now these genes have not been linked to any neuropsychiatric illness, although both genes have a function in prenatal brain development. We introduce the use of functional magnetic resonance imaging activation as a quantitative phenotype in conjunction with genome-wide association as a gene discovery tool.
Molecular Psychiatry (2009) **14,** 416–428; doi:10.1038/mp.2008.127; published online 9 December 2008

Keywords: cognition; fMRI; GWAS; case–control study; quantitative trait loci; prefrontal cortex

Introduction

The genetic aspect of neuropsychiatric disorders, such as schizophrenia, bipolar disorder and autism, is often initially determined through linkage studies which can point to areas of the genome linked to the disorder (for example, references[1–5]). Combining this approach with known molecular functions of the genes in the linkage areas identifies likely genes that can then be specifically tested in case–control designs. Although useful, this method depends on current knowledge regarding genes found in the linkage area; it greatly limits the potential genes studied, both in number and mechanism.

Genome-wide association studies (GWAS) allow the identification of genes whose relationship with the disease phenotype has not even been hypothesized. Genome-wide association techniques have developed exponentially in the past few years; current chips allow 100 000–1 000 000 single nucleotide polymorphisms (SNPs) to be assessed in an individual. GWAS offer enormous promise in identifying genetic variation involved with illness and its response to treatment (for example, Ozaki *et al.*;[6] Klein *et al.*[7]) by allowing all areas of the genome to be considered. However, as the number of potential genetic variations under study increases, making it more likely to find the important variations, so does the likelihood of spurious findings. Solutions to this statistical problem have been to increase the sample size to tens of thousands, or more; to increase the significance threshold astronomically; or to limit the number of SNPs considered to a handful of *a priori* determined candidates.[8,9] We suggest that use of a quantitative trait (QT) as the phenotype, rather than a categorical case–control design, has considerably greater statistical power, thereby reducing the sample size needs in a GWAS by many fold.[10]

Functional MRI reflects neural activation while performing tasks and has been clearly shown to reflect neuropsychiatric dysfunction.[11–15] Imaging Genetics is the emerging field that integrates genetic and neuroimaging data, usually through candidate gene approaches.[16–20] The integration of the quantitative imaging phenotype analysis with GWAS is the basis of our approach to Imaging Genetics. This approach both directly addresses the candidate gene limitations with a genome-wide scan, and has the increased statistical power of QT analysis coupled

Correspondence: Professor SG Potkin, Department of Psychiatry and Human Behavior, Brain Imaging Center, University of California, Irvine, 5251 California Ave, Suite 240, Irvine, CA 92617, USA.
E-mail: sgpotkin@uci.edu
Received 18 February 2008; revised 6 October 2008; accepted 9 October 2008; published online 9 December 2008

with the sensitivity of imaging to reflect brain function.

We use differential brain imaging activation patterns as the starting point in Imaging Genetics analyses, based on the assumption that brain imaging will reveal important pathophysiological differences. We then determine the impact of genetic variation on these brain activation phenotypic patterns, to identify genetic influences potentially key to understanding the pathophysiology. In this way, we use a QT as a marker for the neuropsychiatric dysfunction, and gain the statistical power of using a QT while targeting a phenotype that separates the population of interest from the control population. The differential brain imaging patterns can be either be chosen from the literature or empirically determined within the study. We chose activation in the dorsolateral prefrontal cortex (DLPFC) for this study because of the extensive literature implicating it in schizophrenia.[21–30]

The use of QTs with a comprehensive genome-wide scan has not been commonly applied to neuropsychiatric disorders, perhaps because of difficulty in determining the QT. A notable exception was the discovery of Kidney and Brain Protein (KIBRA) using memory performance as the QT, based on quartile ranking in verbal episodic memory, in a genomics scan of pooled DNA.[31] Recently, Almasy *et al*.[32] used cognitive function as a QT in a family study of schizophrenia with 386 microsatellite markers. Neuroimaging has been used to reveal the function of candidate genes, for example COMT,[33,34] using studies designed to begin with a specific gene and explore its effects on various phenotypes. Brain imaging has been used to study the function of a number of other genes such as 5HTTR transporter, DRD4, DRD1, HTR3A, TPH2 and MAOA[20,35–38] and genes associated with schizophrenia including NRG1, RGS4, COMT, GRM3, G72, DISC1 and BDNF.[16,18,19,39] In functional neuroimaging studies of neuropsychiatric patients and healthy controls, differential activation in regions of interest (ROI) or putative circuits can be identified. We limit our imaging phenotypes to these areas—in this case, left DLPFC (BA46)—and then examine the function of individual genetic variation on these phenotypes at an individual level, that is, how each SNPs predicts activation in the DLPFC. Our approach reverses the candidate gene strategy: Rather than beginning with a specific candidate gene as a grouping factor and searching for differences in neuroimaging results within groups, we begin with brain imaging as a phenotype and determine the SNPs that influence that phenotype.

We implemented this Imaging Genetics approach in a small, discovery sample. This combination of genome-wide exploration with a quantitative imaging phenotype led to the identification of two genes, *RSRC1* and *ARHGAP18*, not previously associated with schizophrenia or other brain disorders. To support our findings and to validate our method of gene discovery we conducted a case–control association study in a new sample of schizophrenia patients and healthy controls. SNPs representing these two genes were verified in this sample.

Materials and methods

Subjects

The discovery sample consisted of 28 chronic schizophrenic patients diagnosed according to Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) criteria by a Structured Clinical Interview for Diagnosis (SCID).[40,41] A total of 75% of the patients were right handed. The average age was 43 years (range 27–60 years old). The mean duration of illness was 13.6 years (range 1–32 years). All were treated with stable doses of atypical antipsychotic drugs, all except two with conventional antipsychotic agents. See the Supplementary Materials for additional patient characteristics.

The corroborative sample consisted of 82 subjects with chronic schizophrenia and 91 controls overall matched for gender and age, also diagnosed according to DSM-IV criteria with a SCID interview. This was part of a larger cohort collected by the Functional Imaging Biomedical Informatics Research consortium (FBIRN). A total of 89% of both groups were right handed. The average age was 38 (range 18–61 years) and 36.2 (range 18–65 years) years of age for the subjects with schizophrenia and the controls, respectively. The mean duration of illness was 14.3 years (range 2–43 years). All were treated with stable doses of antipsychotic drugs. This sample is typical of chronic schizophrenic patients in treatment with a moderate degree of stable symptoms (Supplementary Materials).

fMRI methods (also see Supplementary Materials)

fMRI scanning. All subjects in the discovery sample were scanned on a 1.5 T Phillips (Picker Eclipse) scanner using a T2*-weighted gradient echo sequence, during which the subjects performed three runs of a Serial Item Recognition Paradigm (SIRP).[27] The SIRP task included two memory loads (2 digits and 5 digits to remember) and a control condition (left and right pointing arrows, to control for movement activations), each in a blocked design. In the baseline condition blocks, subjects were presented a series of arrows and asked to indicate the direction in which the arrow pointed (left or right). In both memory load conditions blocks, subjects were presented with a set of numbers (presented simultaneously for 5 s) then presented with a series of 10 probe trials each consisting of a single number presented for 2 s. Subjects indicated whether the probe was in the memory set of numbers or not. The memory sets were different in every block and every run. Each run was 240 s long and consisted of nine blocks, beginning and ending with a baseline condition block.

fMRI analyses
The first two images of each scan were deleted to allow for saturation effects. The remaining functional magnetic resonance imaging (fMRI) data for each scan were motion-corrected, normalized to a standard space, smoothed using an 8-mm FWHM Gaussian filter, and analyzed using SPM2 (http://www.fil.ion.ucl.ac.uk/spm/), using the canonical HRF. The General Linear Model (GLM) modeled the effects of the low and high memory load relative to the control condition. The contrast of interest compared the high memory load against the low memory load. The voxel-wise analysis was performed on each subject to provide a subject-specific measure of BOLD signal change.

The primary ROI, the left DLPFC, and 11 areas with neuroanatomical connections to the DLPFC and 2 control areas were defined by sampling the digital Talairach atlas developed by Lancaster *et al*.[42] in coordination with the NIH/NIMH Human Brain Project. A summary statistic for each atlas determined region was calculated for each subject (a mean β-value across all the voxels in each ROI for the high memory load > low memory load contrast). These summary statistics, reflecting differential imaging signals, were used as the initial imaging phenotypes. The left DLPFC was defined *a priori* as the primary dependent variable given the extensive literature implicating it in schizophrenia.[21–29] The 11 other ROI and 2 control ROIs were secondary dependent variables chosen to address the biological plausibility of the DLPFC findings.

Genotyping methods

Discovery sample. Genotyping of the discovery sample was performed by the Broad Institute (http://www.broad.mit.edu/gen_analysis/genotyping/) with the Illumina Infinium Human1 chip. This provided a total of 105 950 autosomic SNPs. Samples successfully genotyped in less than 90% of markers on either array were excluded from analysis. We removed 5297 SNPs with missing genotypes >10% across subjects and 2117 SNPs with minor allele frequency (MAF) <1% that partially overlapped with the previous category. After removal of SNPs that did not pass the quality assurance measures we had 98 648 SNPs to analyze, with a global mean call rate of 98.2%.

This GWAS analysis identified two genes described in 'Results'. To further characterize these two genes additional SNPs were genotyping with the HumanCNV370-Duo Bead Chip performed by the Laboratory for Genetic Epidemiology of Complex Traits of the University of Milan. The choice of this platform was based on cost-effective considerations.

Corroborative sample. Corroborative samples for *RSRC1* and *ARHGAP18* genes were assayed with the Illumina Infinium HumanHap300 Bead Arrays for cost effectiveness and Taqman SNP Genotyping Assay (Applied Biosystems, Foster City, CA, USA) for those *RSRC1* and *ARHGAP18* markers selected for confirmation that were not present on the HumanHap300. Quality assurance criteria were the same as applied to the discovery sample WGS. Six *RSRC1* SNPs and one *ARHGAP18* SNP were genotyped by Custom Taqman SNP Genotyping Assay (Applied Biosystems; see Supplementary Materials for details).

Stratification. To correct for possible population stratification in our sample(s) we used the program EIGENSTRAT[43] that controls for the risk of stratification by performing a principal component analysis with the highest possible number of available SNPs. We thus used the entire set of SNPs (described above), as suggested by the method, and additionally performed a parallel analysis using a subset of SNPs 18 036 that (1) were not in linkage disequilibrium (LD) to each other, (2) showed a MAF >0.30 and (3) were not in chromosomal regions previously known to be related to schizophrenia. With both approaches, our samples did not show evidence of stratification.

QT statistical analyses on the discovery sample
For each of the SNPs of the Human1 Bead Array and the additional *RSRC1* and *ARHGAP18* SNPs on the HumanCNV370-Duo Bead Chip, we performed a QT regression on the imaging phenotype in the discovery sample. Using the PLINK permutation procedure for QTs we identified statistically significant results. PLINK (http://pngu.mgh.harvard.edu/purcell/plink/)[44] performs a regression-like approach for QTs, using either an asymptotic (likelihood ratio or Wald tests) or an empirical significance test for each SNP analyzed. To control for the risk of false positives, we calculated the empirical *P*-value for association significance with permutations; for each individual SNP we exchanged labels across the different quantitative values, and using the adaptive permutation algorithm of PLINK, we determined the empirical significance for each SNP in the genotype analyses. With this procedure, the number of permutations performed is different for each SNP and is constrained to the observed significance 'raw' value for each SNP. The empirical permuted *P*-value refers to the proportional number of times that the permuted test passed the original significance value. We also permuted genotype assignments 550 000 times for all the SNPs that were included in the discovery sample analysis as a check on the individual SNP results, and observed that the smallest *P*-value by chance was $P = 10^{-5}$, which is consistent with the individual SNP permutation results.

Case–control association analyses on the corroborative sample
To determine the support of the results of the discovery sample, we focused only on the SNPs mapping the *RSRC1* and *ARHGAP18* genes, with 35 and 61 SNPs, respectively. A logistic regression

analysis for small case–control samples, BINREG,[45] was used. BINREG incorporates a permutation procedure to control for false positives. Tables 1 and 2 provide the case–control results for comparison with the discovery sample.

Results

QT analyses on the discovery sample

Using the left hemisphere DLPFC summary statistic as the imaging phenotype, one gene, *RSRC1* (arginine/serine-rich coiled-coil 1), was identified as having at least one SNP whose QT analysis was significant at $P<10^{-7}$, with an empirical *P*-value of 10^{-6} by permutation. We found that SNPs rs12696067, rs6803630 and rs1915935, all mapping within the *RSRC1* gene on chromosome 3, significantly affected activation of DLPFC (BA46), with *P*-values of 10^{-7} (empirical by permutation 10^{-6}; Table 1). Of the 19 SNPs associated with *RSRC1* in the Human 1 Array based on Build 35, 14 SNPs were significant by permutation testing of each SNP. An additional 31 SNPs for *RSRC1* were evaluated using the HumanCNV370-Duo; 13 of those were also statistically significant. Thus, for the studied SNP data set, 27 of the 48 *RSRC1* SNPs tested were significant by permutation testing of each SNP.

This BA46 imaging phenotype was also affected by SNP rs11154490 at a permutation significance level of $P<10^{-5}$. rs11154490 is one of nine SNPs on the Human 1 array that map the gene *ARHGAP18* on chromosome 6 (Rho GTPase activating protein 18). Of the nine *ARHGAP18* SNPs in the Human 1 Array (Build 35), four were significant by permutation testing. An additional 49 SNPs for *ARHGAP18* were evaluated using the Human370CNV; 15 of those were also statistically significant by permutation testing (Table 2).

Thus, 27 SNPs in *RSRC1* and 19 in *ARHGAP18* were significantly associated with activation in the DLPFC. Tables 1 and 2 show the results of the QT analysis using PLINK, as well as the corroborative results (see below). See Supplementary Tables S1 and S2 for full results and Figures 1 and 2 for physical maps of the two genes.

Circuitry exploration

Our analysis based on *a priori* assumptions focused on measures of the left DLPFC activation as the primary imaging phenotype, and then considered imaging measures in other regions of interest in exploring biological plausibility. Examining the effect of the significant genetic locus (using permuted *P*-values) across other brain regions determines if the effect of that locus follows the pattern of known brain circuitry, or if it appears randomly across the brain. We measured the same summary statistic, a change in BOLD signal between the harder and easier memory load in the SIRP task, over each of an additional 11 ROIs for each subject, as for the DLPFC (BA46), and two control areas. The 11 ROIs were chosen based on the literature documenting their relationship to working memory and executive and cognitive functions, as well as anatomically connected circuitry.[15,24,46,47] The hypothalamus and right uncus served as the control areas.

The brain areas significantly related to the polymorphic variation in SNPs that represent both genes appear to be involved in circuitry previously implied in schizophrenia, and not in the control areas. The circuitry areas include the left cortical areas BA46, 9, 8, 6, 7, the limbic lobe cortex, amygdala, as well as the right cerebellar cortex and dentate nucleus and, subcortically, the left thalamus, putamen and globus pallidus. Although BA7 is not part of the prefrontal system formally, it has strong direct interconnections with the dorsal prefrontal areas BA46, 9, 8. These brain areas are involved in several circuits relevant to schizophrenia including the 'prefrontal system' linking prefrontal and limbic cortices with associated crossed cerebellar, and thalamic and basal ganglia loops.[46,48,49] Thus, there is a consistent association between both genes and the prefrontal and dorsal neocortical circuits relevant to both the memory task and schizophrenia deficits.

In Figure 3 the distribution of permuted *P*-values is shown across a single portion of chromosome 3 and chromosome 6, by brain area. This figure shows multiple imaging phenotype results simultaneously. The pattern of peaks (low *P*-values) is localized to one area of chromosome 3 and one area of chromosome 6, and appears strongly in BA46 and functionally related brain areas, but not in the two control areas (right uncus and hypothalamus). The number of statistically significant SNPs in the depicted regions of approximately 7 and 2 million base pairs, respectively, is generally limited to the *ARHGAP18* and *RSRC1* genes, rather than randomly distributed.

Gene annotation

Gene ontology, canonical pathway and functional network analyses were executed using Ingenuity Pathways Analysis (Ingenuity Systems, Mountain View, CA, USA), Ensemble and SWISSPROT (http://ca.expasy.org/sprot/) tools; these indicated a potential DNA binding property and physical interaction with growth differentiation factor 9 precursor (GDF9) of the *RSRC1* gene product. The dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/index.html) revealed intronic localization in the gene for the SNPs, respectively.

Exploring the same genetic databases revealed poor annotation for the three most significant SNPs that related to BA46 within *ARHGAP18*, rs11154490, rs9372944 and rs9388724. We found all to be intronic. Given that *ARHGAP18* belongs to the RhoGAP family, members of this family may control aspects of synapse function. The *ARHGAP18* gene products such as Rho GTPases are linked to RAS and EGFR-mediated proliferation, migration and differentiation of forebrain progenitors (EGFR, epidermal and transforming growth factor-α growth factor receptors

Table 1 QT analysis on the discovery sample, and case–control association on the corroborative sample, for SNPs related to the *RSRC1* gene in the Human 1 and HumanHap300 BeadArrays

Chromosome 3		*QT (left Brodmann area 46) analysis—28 cases*		*Case–control analysis 82 cases and 91 controls*				
Marker	*Position*	*Uncorrected* P-*value*	*Empirical* P-*value*	*Minor allele*	*MAF (case)*	*MAF (ctrl)*	*OR (95% CI)*	P-*value (permuted)*
rs11713363	159311016	0.377c	0.297	C	0.32	0.31	0.94 (0.79–1.11)	0.45
rs12696067	159315676	1.52E-07	1.00E-06	NA	NA	NA	NA	NA
rs6803630	159319767	1.52E-07	1.00E-06	T	0.27	0.33	0.76 (0.46–1.26)	0.287
rs4542974	159328111	0.421	0.347	C	0.12	0.17	1.39 (1.16–1.66)	4.00E-04
rs7624303	**159333917**	**2.31E-04**	**2.30E-04**	**G**	**0.18**	**0.25**	**0.23 (0.10–0.57)**	**1.00E-03**
rs4318510	**159339123**	**6.96E-04**	**7.83E-04**	**A**	**0.36**	**0.44**	**0.76 (0.61–0.94)**	**0.01**
rs4680416	159341624	0.079	0.087	T	0.05	0.05	0.93 (0.73–1.19)	0.565
rs4271876	**159350337**	**6.96E-04**	**7.83E-04**	**C**	**0.36**	**0.44**	**0.79 (0.65–0.95)**	**0.013**
rs9857883	159369966	0.028	0.029	C	0.47	0.36	0.96 (0.56–1.65)	0.89
rs1915935	**159387102**	**1.52E-07**	**1.00E-06**	**A**	**0.28**	**0.33**	**5.59 (3.98–7.84)**	**<1.00E-04**
rs1624094	**159404804**	**7.28E-05**	**6.85E-05**	**T**	**0.37**	**0.41**	**5.59 (3.98–7.84)**	**<1.00E-04**
rs1729997	**159410601**	**8.31E-05**	**6.54E-05**	**T**	**0.38**	**0.42**	**5.59 (3.98–7.84)**	**<1.00E-04**
rs10513524	159418320	0.665	0.65	G	0.09	0.08	0.53 (0.27–1.03)	0.062
rs827123	**159433575**	**2.80E-04**	**2.09E-04**	**T**	**0.33**	**0.4**	**0.20 (0.11–0.37)**	**<1.00E-04**
rs1213048	159472681	0.01	0.012	NA	NA	NA	NA	NA
rs1851062	159475057	0.005	0.006	NA	NA	NA	NA	NA
rs7643792	159475508	0.009	0.011	NA	NA	NA	NA	NA
rs2682405	159476532	0.058	0.061	NA	NA	NA	NA	NA
rs1210359	159490501	0.013	0.011	T	0.33	0.28	1.03 (0.83–1.27)	0.805
rs1730007	159499699	0.01	0.012	NA	NA	NA	NA	NA
rs2693542	159509681	0.01	0.012	NA	NA	NA	NA	NA
rs10513526	159513076	0.457	0.359	C	0.11	0.17	0.07 (0.03–0.15)	<1.00E-04
rs9827781	159532890	0.081	0.088	T	0.54	0.45	0.07 (0.01–0.39)	0.002
rs6441188	159551801	0.063	0.093	A	0.41	0.51	0.93 (0.86–1.01)	0.066
rs827134	159551899	0.01	0.012	NA	NA	NA	NA	NA
rs7610713	159554959	0.076	0.08	NA	NA	NA	NA	NA
rs6441190	159558925	NA	NA	NA	NA	NA	NA	NA
rs6778370	159577116	0.085	0.091	C	0.55	0.45	1.08 (1.00–1.17)	0.056
rs7639867	159581870	0.098	0.096	C	0.55	0.45	1.08 (1.00–1.17)	0.057
rs939114	159597947	0.062	0.086	T	0.42	0.51	0.93 (0.86–1.01)	0.082
rs9843252	159608485	NA	NA	NA	NA	NA	NA	NA
rs4402974	159601362	4.27E-04	4.18E-04	NA	NA	NA	NA	NA
rs7632059	159616001	0.105	0.109	T	0.53	0.44	6.98 (3.50–13.91)	<1.00E-04
rs16828998	159652271	0.665	0.65	G	0.09	0.08	1.11 (0.74–1.66)	0.615
rs10936142	159655099	0.043	0.037	C	0.29	0.39	0.70 (0.44–1.11)	0.127
rs4324501	159658691	0.038	0.035	NA	NA	NA	NA	NA
rs6806967	**159660947**	**5.96E-04**	**5.82E-04**	**T**	**0.15**	**0.23**	**8.42 (3.31–21.40)**	**<1.00E-04**
rs10936143	**159662666**	**6.65E-04**	**9.30E-04**	**C**	**0.13**	**0.19**	**3.24 (1.42–7.40)**	**0.005**
rs2885663	**159666957**	**0.043**	**0.037**	**A**	**0.34**	**0.43**	**0.09 (0.02–0.38)**	**0.001**
rs10936145	**159675830**	**0.001**	**0.001**	**G**	**0.13**	**0.22**	**2.29 (1.36–3.83)**	**0.002**
rs6769314	159681416	0.457	0.359	G	0.15	0.2	0.64 (0.37–1.13)	0.125
rs1978781	159689485	0.049	0.055	NA	NA	NA	NA	NA
rs1369562	159691262	0.049	0.055	NA	NA	NA	NA	NA
rs1714518	159696117	0.213	0.252	T	0.54	0.43	0.96 (0.72–1.28)	0.794
rs2117764	**159719554**	**0.04**	**0.039**	**C**	**0.28**	**0.33**	**0.01 (0.00–0.05)**	**<1.00E-04**
rs1714505	159719745	0.163	0.141	T	0.07	0.05	1.42 (1.18–1.71)	2.00E-04
rs1714509	159735652	0.054	0.057	A	0.35	0.5	1.71 (1.13–2.59)	0.011
rs11708784	**159738389**	0.005	0.005	NA	NA	NA	NA	NA
rs16829102	159738822	0.903	1	C	0.16	0.09	1.87 (0.69–5.04)	0.215
rs1630524	**159752993**	**0.044**	**0.049**	**G**	**0.35**	**0.5**	**0.63 (0.45–0.88)**	**0.006**
rs1714524	159755798	NA	NA	C	0.48	0.4	0.96 (0.69–1.34)	0.806

Abbreviations: NA, not genotyped; QT, quantitative trait; MAF, minor allele frequency; OR, odds ratio; CI, confidence interval. The four columns on the left represent the findings from the discovery sample using a quantitative trait analysis and the five columns on the right are the finding from the corroborative sample using a case–control analysis for the identical SNPs. Uncorrected *P*-value represents the probability of the likelihood ratio test between the models with and without the SNP. Empirical *P*-value represents the permutation-determined *P*-value (analyses performed using PLINK); OR with 95% confidence interval; *P*-value (permuted) represents the BINREG case–control logistic regression. Tinted areas indicate statistical significance and bold text indicates SNPs positive in both QT and case–control analyses.

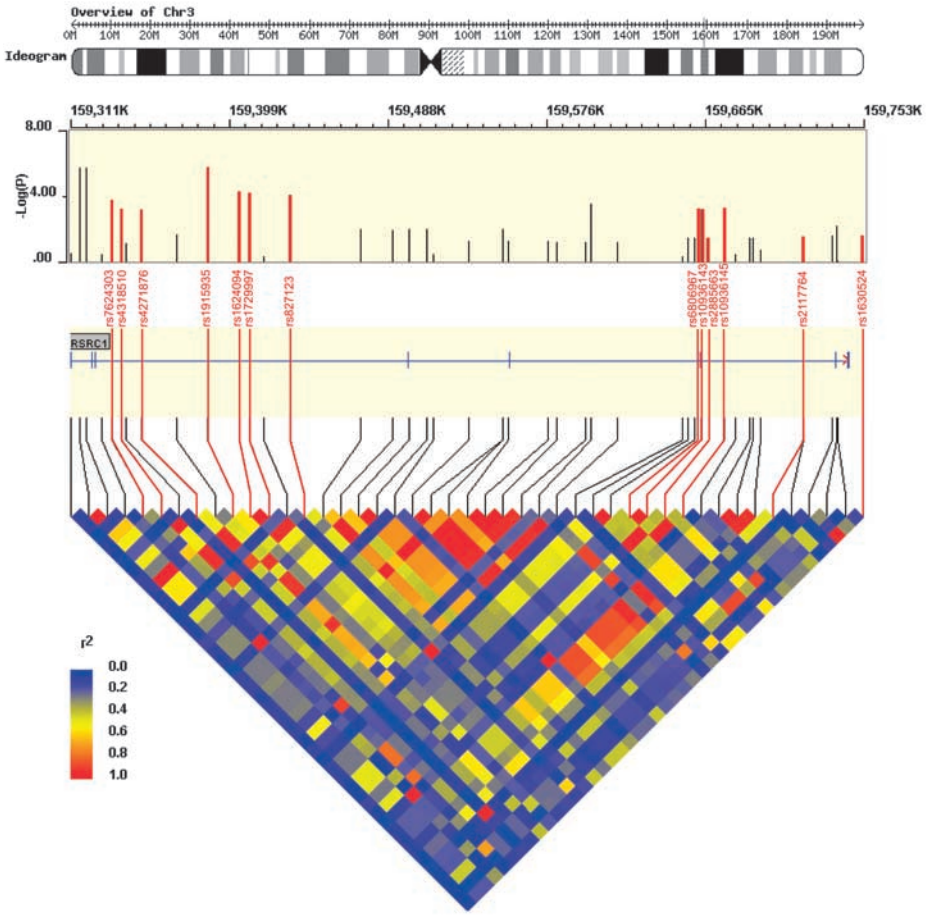

Figure 1 Physical map of the single nucleotide polymorphisms (SNPs) associated with *RSRC1* in the Discovery sample produced by WGAViewer.[72] The topmost sector is the ideogram of Chromosome 3; the vertical line shows the relative of location of *RSRC1*. Below this is the graph showing the –log(*P*) significance values of the individual SNPs on the imaging phenotype. The vertical lines in red indicate the SNPs that are significant in both the Discovery and Corroborative samples. The small blue lines below this indicate the location of the exons in the transcripts annotated (translated region of the DNA). The vertical lines above the accompanying triangular matrix indicate the SNP locations, and demonstrate the linkage disequilibrium (LD) pattern between SNPs (r^2). The warmer colors on the flame scale indicate greater LD whereas the blue indicates absence of LD. These patterns in this figure suggest minimal LD between the studied SNPs in the Discovery sample.

(erb B)[50]). In addition, based on *in silico* analyses both genes have a function in dopamine, glutamate and fibroblast growth factor receptor signaling.

Corroboration in an independent case–control sample
To establish if our method of gene discovery using brain imaging as a quantitative phenotype could successfully identify genes related to schizophrenia, we tested the most promising genes discovered (*RSRC1* and *ARHGAP18*) in an independent case–control study collected by the FBIRN consortium (www.nbirn.net). The statistically significant *RSRC1* and *ARHGAP18* SNPs from the discovery sample were analyzed in a case–control design, using a logistic regression analysis for small samples that included a permutation procedure to determine significance threshold *P*-value for each SNP (GLM BINREG[45]). Genotyping data were available for 17 of the 27 positive discovery sample SNPs for *RSRC1* and for 18 of the 19 positive SNPs for *ARHGAP18*. Of the 17, 13 *RSRC1* SNPs tested and 6 of the 18 *ARHGAP18* SNPs tested were significantly associated with a diagnosis of schizophrenia in the corroborative sample, substantiating a function of *RSRC1* and *ARHGAP18* in the genetics of schizophrenia (see Figure 4 for summary; Tables 1 and 2). These significant SNPs in the logistic regression (*P*-values shown in the table) demonstrate the value of the imaging genetics approach in identifying genes associated with the disorder. These findings highlight the ability of our method using a quantitative brain imaging approach to identify genes related to schizophrenia.

Discussion

Imaging genetics provides a powerful method for identifying genetic variations that affect a QT of interest in neuropsychiatric populations. Admittedly, this method excludes genes or polymorphisms that do

Table 2 QT analysis on the discovery sample and case–control association on the corroborative sample for SNPs related to the *ARHGAP18* gene in the Human 1 and HumanHap300 BeadArrays

Chromosome 6		*QT (left Brodmann area 46) analysis—28 cases*		*Case–control analysis 82 cases and 91 controls*				
Marker	*Position*	*Uncorrected* P-*value*	*Empirical* P-*value*	*Minor allele*	*MAF (case)*	*MAF (ctrl)*	*OR (95%CI)*	P-*value (permuted)*
rs7760943	129910960	0.728	0.625	G	0.28	0.22	0.78 (0.52–1.16)	0.227
rs2571588	129911201	0.287	0.324	G	0.34	0.41	0.64 (0.46–0.87)	0.005
rs10457525	129914659	0.62	0.463	T	0.22	0.2	1.15 (0.77–1.73)	0.481
rs9321172	129919692	0.145	0.179	T	0.27	0.3	1.40 (1.13–1.73)	0.002
rs17810534	129928992	0.551	0.684	C	0.13	0.15	1.40 (0.98–1.99)	0.062
rs9385500	129930382	0.585	0.706	T	0.36	0.33	1.01 (0.67–1.52)	0.956
rs13214956	129930727	0.337	0.353	T	0.07	0.05	0.62 (0.40–0.95)	0.03
rs10457526	129938194	0.019	0.021	T	0.22	0.22	0.83 (0.66–1.02)	0.082
rs12197456	129946942	0.11	0.132	G	0.04	0.04	0.61 (0.43–0.88)	0.008
rs9492347	129948346	0.145	0.179	C	0.24	0.28	0.94 (0.73–1.21)	0.641
rs6935162	129949280	0.675	1	T	0.31	0.27	0.97 (0.72–1.31)	0.87
rs4504494	129954780	0.124	0.151	A	0.12	0.1	0.97 (0.75–1.24)	0.797
rs7758025	129959932	0.83	0.857	T	0.16	0.17	0.49 (0.31–0.79)	0.003
rs10499162	129960703	0.399	0.354	C	0.14	0.13	2.09 (1.31–3.35)	0.002
rs9321174	129962909	0.91	1	A	0.35	0.37	0.81 (0.46–1.41)	0.461
rs12530181	129966696	0.41	0.625	G	0.18	0.17	1.28 (0.68–2.42)	0.442
rs9388717	129978524	0.022	0.024	NA	NA	NA	NA	NA
rs9375638	129984510	0.015	0.017	C	0.23	0.23	0.72 (0.46–1.13)	0.155
rs9375641	129989891	0.018	0.02	C	0.44	0.42	1.15 (0.83–1.59)	0.399
rs9375644	129993694	0.105	0.116	NA	NA	NA	NA	NA
rs12664247	**129997020**	**9.44E-04**	**0.001**	**C**	**0.29**	**0.3**	**2.27 (1.24–4.16)**	**0.008**
rs4509146	**129997180**	**9.44E-04**	**0.001**	**T**	**0.29**	**0.3**	**2.27 (1.23–4.16)**	**0.008**
rs9402155	130000790	9.44E-04	0.001	T	0.17	0.23	0.59 (0.27–1.28)	0.186
rs9483050	130002942	0.2776	0.255	G	0.25	0.22	1.154 (0.77–1.73)	0.485
rs10499163	130004326	0.034	0.038	C	0.27	0.28	0.87 (0.52–1.43)	0.589
rs9372944	130007047	6.58E-05	4.18E-05	A	0.19	0.23	0.95 (0.67–1.34)	0.782
rs9398913	130012655	NA	NA	T	0.4	0.37	0.65 (0.29–1.45)	0.3
rs4479964	130013237	0.093	0.119	A	0.33	0.3	1.92 (0.93–3.95)	0.075
rs9388721	130013622	0.013	0.018	T	0.17	0.19	0.54 (0.28–1.02)	0.058
rs4499954	130020089	0.003	0.003	C	0.23	0.26	1.98 (0.98–3.99)	0.056
rs6922711	130028485	0.002	0.002	C	0.49	0.43	1.21 (0.98–1.49)	0.072
rs9398913	130012655	NA	NA	A	0.03	0.05	0.54 (0.36–0.79)	0.002
rs13203608	130035588	0.386	0.338	T	0.09	0.08	0.64 (0.32–1.27)	0.203
rs11154490	**130036348**	**3.80E-05**	**3.68E-05**	**T**	**0.27**	**0.3**	**0.58 (0.35–0.97)**	**0.04**
rs9388724	130041921	6.98E-05	6.38E-05	C	0.17	0.22	1.25 (0.83–1.87)	0.28
rs2051632	**130050625**	**0.037**	**0.039**	**C**	**0.42**	**0.46**	**0.68 (0.50–0.93)**	**0.015**
rs11154491	130053359	0.008	0.008	G	0.11	0.12	0.85 (0.55–1.34)	0.495
rs1041915	130055946	0.116	0.092	C	0.1	0.12	0.83 (0.49–1.40)	0.485
rs13208724	130057842	0.195	0.262	C	0.23	0.29	1.09 (0.81–1.49)	0.554
rs1476042	130058786	0.091	0.106	A	0.3	0.24	1.36 (0.65–2.82)	0.407
rs1894641	130062643	0.871	1	A	0.18	0.14	0.50 (2.41–1.04)	0.065
rs1023480	130065893	0.23	0.29	C	0.55	0.44	0.85 (0.58–1.27)	0.44
rs2032533	130069013	0.911	1	C	0.32	0.25	1.56 (1.00–2.44)	0.051
rs3752536	130072908	0.916	0.857	A	0.16	0.13	0.99 (0.75–1.31)	0.964
rs12213388	130074438	0.329	0.274	C	0.17	0.11	0.91 (0.69–1.19)	0.504
rs9385512	130075559	0.122	0.14	G	0.25	0.3	0.58 (0.42–0.79)	0.001
rs1984408	130079698	0.8	0.857	G	0.32	0.34	1.04 (0.79–1.36)	0.79
rs9398918	130087370	0.537	0.857	A	0.21	0.14	1.02 (0.84–1.23)	0.848
rs7753327	130092303	0.656	0.857	C	0.35	0.43	1.04 (0.86–1.26)	0.665
rs17469847	**130098274**	**0.054**	**0.037**	**C**	**0.2**	**0.17**	**1.31 (1.11–1.54)**	**0.001**
rs376632	130100763	0.252	0.377	A	0.13	0.13	1.03 (0.79–1.35)	0.808
rs208869	130104410	0.058	0.055	C	0.48	0.44	0.86 (0.63–1.18)	0.351
rs10484284	**130105402**	**0.013**	**0.022**	**G**	**0.15**	**0.16**	**1.48 (1.01–2.17)**	**0.044**
rs208870	130105554	0.925	0.857	G	0.07	0.07	0.87 (0.52–1.45)	0.593
rs10484283	130106363	0.565	0.625	C	0.07	0.05	0.88 (0.60–1.29)	0.518
rs10484282	130107771	0.252	0.207	T	0.23	0.28	0.72 (0.52–0.99)	0.041
rs9402168	130108429	0.419	0.487	G	0.37	0.39	0.74 (0.53–1.03)	0.072
rs17057773	130112009	NA	NA	T	0.05	0.06	1.07 (0.71–1.63)	0.737
rs208865	130114249	NA	NA	A	0.06	0.07	0.65 (0.43–0.99)	0.043
rs9388732	130122966	0.057	0.041	C	0.24	0.27	1.18 (0.94–1.48)	0.159
rs17057830	130123366	0.106	0.086	G	0.1	0.11	0.81 (0.56–1.15)	0.243
rs4897336	130123754	0.832	0.857	A	0.1	0.14	0.76 (0.59–0.99)	0.04
rs208872	130126061	NA	NA	G	0.08	0.07	1.41 (1.10–1.81)	0.006

Abbreviations: NA, not genotyped; QT, quantitative trait; MAF, minor allele frequency; OR, odds ratio; CI, confidence interval. The four columns on the left represent the findings from the discovery sample using a quantitative trait analysis and the five columns on the right are the finding from the corroborative sample using a case–control analysis for the identical SNPs. Uncorrected *P*-value represents the probability of the likelihood ratio test between the models with and without the SNP. Empirical *P*-value represents the permutation-determined *P*-value (analyses performed using PLINK); OR with 95% confidence interval; *P*-value (permuted) represents the BINREG case–control logistic regression. Tinted areas indicate statistical significance and bold text indicates SNPs positive in both QT and case–control analyses.

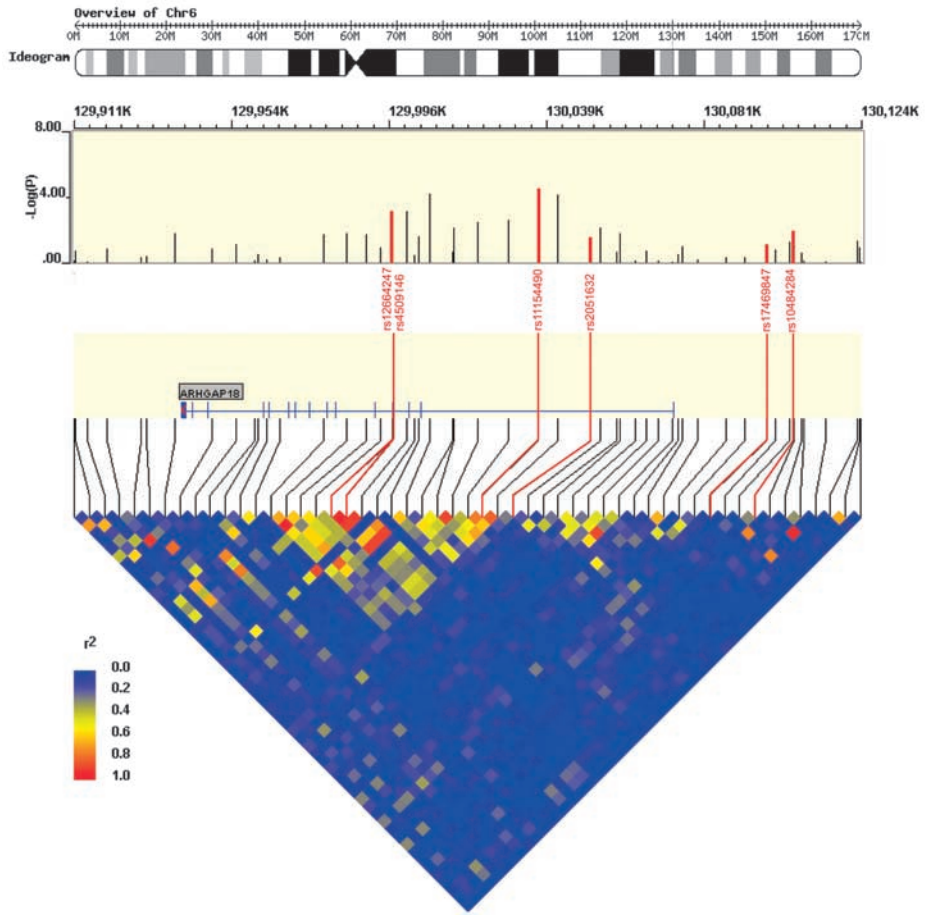

Figure 2 Physical map of the single nucleotide polymorphisms (SNPs) associated with *ARHGAP18* in the discovery sample produced by WGA Viewer.[72] See Figure 3 legend for description. The clustering pattern of primarily blue and yellow squares indicate minimal linkage disequilibrium (LD) between the SNPs studied. This indicates relative statistical independence between the SNPs analyzed in the discovery sample. The association of these SNPs with the quantitative phenotype strengthening the proposed relationship of the *ARHGAP18* gene with schizophrenia.

not influence differences in brain area activation, or the particular imaging phenotypes chosen. On the other hand, structural and functional brain imaging provide sensitive quantitative and objective measures of brain function in neuropsychiatric illness. Brain development, structure and function are heritable and have strong genetic influences. Using an imaging phenotype has face validity and biological relevance as it constrains the GWAS analyses.

Studies of schizophrenia must consider potential medication effects. Manoach[21] in reviewing fMRI studies in schizophrenia pointed out that working memory deficits are observed in medicated, unmedicated and medication–naïve patients as well as their siblings, concluding that such deficits are present regardless of medication[22,23] and have at least, in part, a genetic basis. Using the Sternberg working memory task as activation in an fMRI twin study of schizophrenia and healthy controls, evidence was found for both the behavioral and fMRI measures being heritable traits.[24] Using the same task in schizophrenia, BOLD activation patterns involving the DLPFC differed by DRD1 genotype despite equivalent memory performance.[25] The degree of heritability has yet to be defined although working memory performance decreased with increasing genetic load among schizophrenic twins discordant for schizophrenia and control twins.[26] The Sternberg task-related abnormalities are found in the relatives of persons with schizophrenia as well as in the patients both on medication, and with no medication, which is also consistent with a heritable trait.[21,27]

We address the issue of false positives, always a concern in such high-dimensional data, from both statistical and biological viewpoints. The imaging data are summarized by averaging BOLD activation over an atlas-determined ROI. The primary ROI of interest, the DLPFC was chosen *a priori*. The hundreds of thousands of imaging-genetic analyses are corrected for chance findings by permutation tests.[51] The anatomically and/or functionally connected regions in the brain should show a similar pattern of genotype influence. This is not a statistical consideration, but a biologically driven one. Finally,

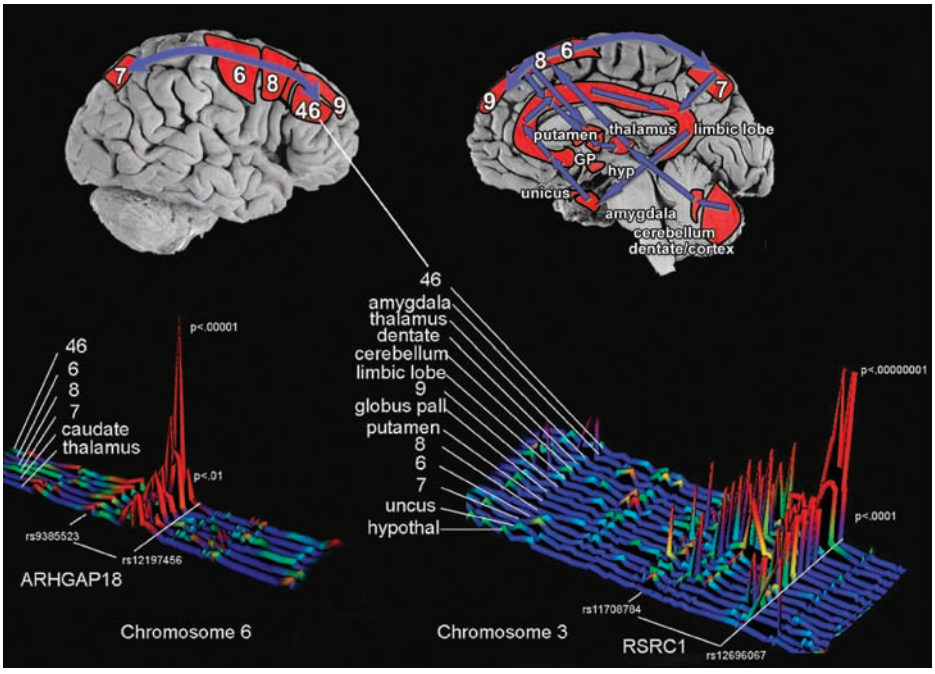

Figure 3 Imaging Genophenotypes in the discovery sample. The relationship between activation in brain areas and single nucleotide polymorphisms (SNPs) on chromosomes 6 and 3 from the quantitative trait analysis are depicted. Each tube represents activation in a different brain area. The brain areas are grouped to explore similarities within the significant genes (SNPs) in related areas forming anatomical circuits. Peaks in the tubes represent permuted *P*-values (plotted as –log(*P*)) for all SNPs represented over an approximately 7 million base pair region of chromosome 6 around *ARHGAP18* (lower left) and 2 million base pairs on chromosome 3 around *RSRC1* (lower right). The height of each tube depicts the significance level for those SNPs related to brain activation in each brain area. The brain areas are represented by their Brodmann area designation or their commonly used anatomical labels. The specific RS numbers for SNPs bracketing the main peaks are listed in their approximate locations and mark the boundaries of *RSRC1* and *ARHGAP18*. The directional arrows on the anatomical template demonstrate the implied anatomical circuitry for brain areas represented in figure. GP, globus pallidus; hyp, hypothalamus; BA46, dorsolateral prefrontal cortex.

these SNPs identified in an Imaging Genetics analysis become candidates which must be replicated in an independent sample.

We have provided a proof of concept of this novel approach using activation in the left DLPFC as a quantitative phenotype. The Imaging Genetics analyses included massively parallel analyses of all 105 950 SNPs in conjunction with summary imaging results. We picked an ROI based on known differences between patients and controls from the published literature.[22,23,28,29,33,48] This ROI, used as a quantitative phenotype, identified genes which were previously unlinked to schizophrenia through molecular hypotheses—and these genes showed case–control differences in an independent sample, where they would never have been considered using a standard candidate gene approach.

The QT can be determined based on the literature as in this study or can be empirically derived. In the latter application, the brain imaging patterns between the patient population and normal healthy controls in the data set are contrasted to allow a choice of ROIs that best distinguish the groups. We then generate summary measures on activation patterns in those ROIs for each patient and each control subject. The SNPs that influence these brain activation quantitative phenotypes are identified through a series of GLM calculations. The statistical approach is built upon a GLM that combines imaging phenotypes, disease diagnosis and genetic data in a single model:

Imaging phenotype = genotype effect + diagnosis effect + genotype–diagnosis interaction effect

The value of this latter application is that it includes the diagnosis-by-genotype interaction, as well as the ability to add additional terms for gene–gene interactions. The GLM models the effect of each SNP with the brain activation measure as the dependent variable. Significance of the interaction effect is determined through permutation testing or other appropriate methods. The genotype–diagnosis interaction represents disease effects on the neuroimaging phenotype that are influenced by normal genetic variation, thus, the conceptual fusion of the interaction between imaging and genetics by diagnosis. It is more than the additive effects of these domains, but rather their interaction, modeled explicitly in the analysis.

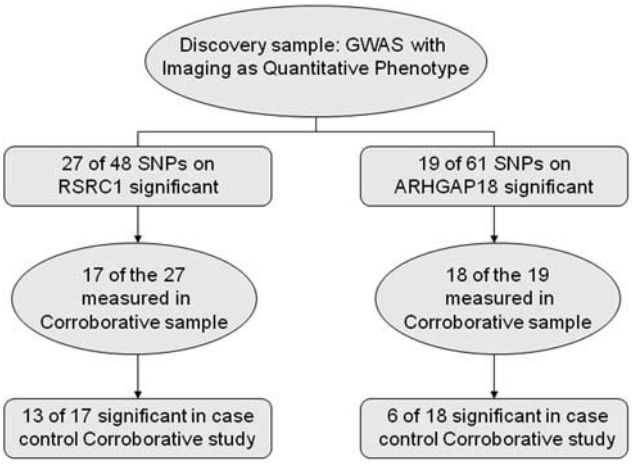

Figure 4 Summary of the statistically significant single nucleotide polymorphisms (SNPs) associated with *RSRC1* and *ARHGAP18* in the discovery and corroborative samples. The top rectangles indicate the number of statistically significant SNPs from those that were tested in the discovery sample using brain imaging as a quantitative phenotype. Seventeen of the positive *RSRC1* SNPs and eighteen of the positive *ARHGAP18* SNPs from the discovery sample were measured in the Corroborative sample. The lower rectangles indicate the number of these SNPs that were also statistically significant in the case–control Corroborative study. The results for all SNPs in *RSRC1* and *ARHGAP18* are in Tables 1 and 2.

The correction for multiple testing is an ongoing point of research. The BOLD activation is a summary score of the activation of all imaging voxels in the designated ROI chosen for the dependent variable or QT. The most conservative Bonferroni correction is not appropriate to SNP data as they typically do not meet the assumption of independence of multiple tests. Other methods used to correct for the risk of falsely concluding for a positive association, that is increasing the risk of the frequency of false positives, range from the Benjamini–Hochberg proposal[52] adapted for genome analyses by Storey and Tibshirani[53] with their FDR 'correction', to the Nyholt[54] and Meng *et al*.[55] methods that consider the dependency across SNPs. Some methods however, like Nyholt's and Meng's, are well-suited for a 'small' SNP set, for example as SNPs across a gene or in a chromosomal region, but are not easily generalizable to whole genome association studies. Other methods establish a sample-based significance threshold by a permutation approach.[8,56] Any correction for type I errors should be balanced with the risk of increasing the false-negative results. We used permutation testing, as it makes no assumptions regarding independence of tests and uses the sample being tested to empirically determine appropriate statistical thresholds.[8,29] Permutation is a preferred method for determining the likelihood of chance findings; the advantages are discussed in Hirschhorn and Daly,[8] Balding,[57] Dudbridge and Koeleman.[58]

There are six published GWAS in schizophrenia.[32,59–63] For the most part these studies have not studied the SNP panels that we studied and were not sufficiently powered to confirm our findings. Only the study by Kirov *et al*. used the same SNPs profile and of the other studies only two of our positive *ARHGAP18* SNPs were overlapping. Each of these GWAS has identified 1 or 2 genes (SNPs) that have passed genome-wide significance. None of the studies provide data or *P*-values below the genome-wide significance levels for the overlapping *RSRC1* and *ARHGAP18* SNPs that were positive in our study. The actual *P*-values are necessary to confirm our finding for these two genes that do not require the genome-wide threshold of significance for testing the entire genome.

Up to half of all genes are expressed in the brain (assuming 22–25K genes). Any or all of the SNPs tested in this GWAS could correlate with brain activation. We expected to find some clearly spurious results identifying genes that are not expressed in the brain. Empirically, we have determined that this is a relatively rare event in this data set. The top eight most significant results, implicate the same two genes. Most of the analyzed SNPs were not related to task-specific brain activation used in this study. Further, investigation of gene annotation shows all the significant SNPs represent genes that are expressed in the brain, providing additional face validity.

The sign of the regression coefficient is positive for 17 of the 19 β-coefficients for the 19 SNPs in *RSRC1* or *AHRGAP18* that are significant for both the discovery and the case–control samples. A positive coefficient means that the minor allele increases the phenotype mean, in the QT analysis. An increase in the phenotype mean implies that those subjects with the minor allele activate their DLPFC more than those without the minor allele. Schizophrenia subjects are expected to have a greater deviant activation, that is, less activation with a comparatively high memory load than controls; therefore having less of the minor allele than healthy controls. This is precisely what was observed in the case–control analyses. Plenge *et al*.[64] interpreted similar observations as 'the minor allele was associated with protection', but this interpretation is premature in our study. It is important to emphasize that SNPs on the Illumina chips used in our study were primarily chosen because they are tagging SNPs and not necessarily causal or coding SNPs. More detailed investigations of the areas identified by these tagging SNPs are needed to investigate which SNPs are causally linked to schizophrenia and to determine their functionality.

The results of the discovery and corroborative sample analyses are intriguing in several ways. Brain regions connected to left DLPFC (BA46) also showed a significant influence of *RSRC1* and *ARHGAP18* SNPs on brain activation measures, as shown in the figures, whereas control brain areas do not. These anatomically connected areas have several interesting

features in common; all are neocortical regions that receive a dense dopamine innervation, all are highly interconnected with each other, and participate in a dorsal cortical circuitry that is consistently implicated in the etiology of schizophrenia, especially the DLPFC.[48] Interestingly, these areas are associated with dopamine function especially of the D1 receptor class. We have not, however, tested the effects of these genotypes on brain activation in healthy controls; that will have to be the topic of future studies.

ARHGAP18 gene products belong to the human RhoGAP family with approximately 80 RhoGAP proteins known to be encoded in the human genome. The RhoGAPs, GTPase-activating proteins have the ability to modulate Rho-mediated signaling pathways by controlling the balance between active and inactive Rho proteins. Rho proteins belong to the Ras superfamily that is composed of over 50 members divided into 6 families, including Ras, Sar, Rho, Ran, Rab and Arf.[65] They participate in an array of physiological processes, such as cell migration, intercellular adhesion, cytokinesis, proliferation, differentiation and apoptosis.[66] Rho GTPases are important regulators of the actin cytoskeleton and consequently influence the shape and migration of cells. GTPases of the Rho family are strong regulators of signaling pathways that link growth factors and/or their receptors to adhesions and associated structures.[67] One signaling pathway mediated by Ras is initiated by the EGF receptor (erb B epidermal growth factor receptor, EGFr), leading to cell proliferation. EGFr signaling can induce mitosis, proliferation, cell motility, differentiation and protein secretion.[68] EGFr is localized on subventricular neural progenitor (svz) cells in the fetal and adult lateral ventricles, and these progenitors give rise to forebrain neurons in development and after injury in the adult.[50] Thus, the *ARHGAP18* gene products (Rho GTPases) are linked to Ras, and thus, to EGFr-mediated proliferation, migration and differentiation of forebrain progenitors. Therefore, our finding of *ARHGAP18* SNPs related to DLPFC activation in schizophrenia is promising because schizophrenia has been linked to altered prenatal neurogenesis of cortical neurons, including those in dorsal prefrontal cortex.[30] In addition, *ARHGAP18* is precisely contained within 6q22-24, which has been shown to be linked to schizophrenia.[69]

RSRC1 is a unique marker of bone marrow-derived stem cells also expressing the cd34 marker.[70] The cd34 marker is also found in a proliferating population of the subventricular zone (svz) stem cell population in the human fetal forebrain.[71] Circulating bone marrow-derived cd34 stem cells can take up residence in subventricular niches in the developing and postnatal forebrain. These neural stem cells of the svz give rise to EGFr-responsive progenitors which can be induced to massively proliferate and migrate to lesion sites in the forebrain in the presence of transforming growth factor-α, which binds to the erb B EGFr receptor.[50] Thus, there may be an association between *RSRC1* variants, and the cd34-positive hematopoietic and neural stem cells in the progenitors that give rise to forebrain neural development.

Taken together, the present finding of a highly significant association between *ARHGAP18* and *RSRC1* SNPs with schizophrenia is intriguing as both genes have a function in prenatal brain development linking hematopoietic and neural stem cell proliferation in the svz and migration to forebrain structures such as the dorsal cortical stream and associated limbic lobe, striatal and amygdaloid circuitry.

The approach described is a screening method that makes GWAS data usable and exploratory in preparation for future studies, for example molecular studies, expression and transgenic studies, and all other functional genomic approaches. It allows for completely novel SNPs to be identified as having a function in the disease phenotype.

Acknowledgments

This research was supported by grants to the Transdisciplinary Imaging Genetics Center (TIGC-P20 RR020837-01) and to the Functional Imaging Biomedical Informatics Research Network (FBIRN-1 U24 RR021992) from the National Center for Research Resources (NCRR) at the National Institutes of Health (NIH) and by grants POCEMON (FP7-ICT-2007-216088), FIRB Italia-Israele (RBIN04SWHR) and HYPERGENES (HEALTH-F4-2007-201550) and by an anonymous foundation. The Broad Institute Center for Genotyping and Analysis is supported by grant U54 RR020278-01 from the NCRR. We acknowledge the help and support of Mita Mancini and Yann Legros from Illumina, as well as Cristina Barlassina, Chiara Dal Fiume, Alessandro Orro and Federica Torri (University of Milan) for performing the HumanHap300 Bead Array procedures, as well as Liv Trondsen and Divya Rajpoot (UCI) for editorial support. We also acknowledge the recruitment, evaluation and SCID-based diagnostic assessment of healthy controls and schizophrenic subjects by FBIRN investigators: John Lauriello and Juan Bastillo, University of New Mexico; Daniel O'Leary, University of Iowa; Kelvin Lim, University of Minnesota; Gregory McCarthy, Judith Ford, Yale University; Arthur Toga, Tyrone Cannon, UCLA; Randy Gollub, Harvard University; Aysenil Belger, University of North Carolina; Dana Nguyen, Diane Highum, University of California, Irvine. We acknowledge the helpful comments of William E Bunney and Hal Stern.

Author contributions: The fMRI task, imaging data from the discovery sample and imaging analyses for these results were programmed and implemented by Jessica Turner; the neuroanatomical and neuroscience expertize and genetic annotation was contributed by James Fallon; the genetic data analysis, PLINK and Eigenstrat analyses and genetic annotation were performed by Guia Guffanti and Fabio Macciardi; the *in silico* annotations were performed by Anita

Lakatos; the visualization and gene viewer methods were developed by David Keator; the design and oversight of the experiments and analyses were the responsibility of Steven Potkin. Article preparation was a joint effort of all authors.

References

1 McQueen MB, Devlin B, Faraone SV, Nimgaonkar VL, Sklar P, Smoller JW *et al.* Combined analysis from eleven linkage studies of bipolar disorder provides strong evidence of susceptibility loci on chromosomes 6q and 8q. *Am J Hum Genet* 2005; **77**: 582–595.
2 Vicente AM, Macciardi F, Verga M, Bassett AS, Honer WG, Bean G *et al.* NCAM and schizophrenia: genetic studies. *Mol Psychiatry* 1997; **2**: 65–69.
3 Kwasnicka-Crawford DA, Roberts W, Scherer SW. Characterization of an Autism-Associated Segmental Maternal Heterodisomy of the Chromosome 15q11-13 Region. *J Autism Dev Disord* 2006; **37**: 694–702.
4 Stein CM, Millard C, Kluge A, Miscimarra LE, Cartier KC, Freebairn LA *et al.* Speech sound disorder influenced by a locus in 15q14 region. *Behav Genet* 2006; **36**: 858–868.
5 Chagnon YC. Shared susceptibility region on chromosome 15 between autism and catatonia. *Int Rev Neurobiol* 2006; **72**: 165–178.
6 Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T *et al.* Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat Genet* 2002; **32**: 650–654.
7 Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* 2005; **308**: 385–389.
8 Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 2005; **6**: 95–108.
9 Wang WY, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet* 2005; **6**: 109–118.
10 Liu B. *Statistical Genomics: Linkage, Mapping, and QTL Analysis*. CRC Press: Boca Raton, 1997.
11 Ebmeier K, Rose E, Steele D. Cognitive impairment and fMRI in major depression. *Neurotox Res* 2006; **10**: 87–92.
12 Dickerson BC. Functional MRI in the early detection of dementias. *Rev Neurol (Paris)* 2006; **162**: 941–944.
13 Yurgelun-Todd DA, Ross AJ. Functional magnetic resonance imaging studies in bipolar disorder. *CNS Spectr* 2006; **11**: 287–297.
14 Davis CE, Jeste DV, Eyler LT. Review of longitudinal functional neuroimaging studies of drug treatments in patients with schizophrenia. *Schizophr Res* 2005; **78**: 45–60.
15 Kircher TT, Thienel R. Functional brain imaging of symptoms and cognition in schizophrenia. *Prog Brain Res* 2005; **150**: 299–308.
16 Lawrie SM, Hall J, McIntosh AM, Cunningham-Owens DG, Johnstone EC. Neuroimaging and molecular genetics of schizophrenia: pathophysiological advances and therapeutic potential. *Br J Pharmacol* 2008; **153**(Suppl 1): S120–S124.
17 Turner JA, Smyth P, Macciardi F, Fallon JH, Kennedy JL, Potkin SG. Imaging phenotypes and genotypes in schizophrenia. *Neuroinformatics* 2006; **4**: 21–49.
18 Roffman JL, Weiss AP, Goff DC, Rauch SL, Weinberger DR. Neuroimaging-genetic paradigms: a new approach to investigate the pathophysiology and treatment of cognitive deficits in schizophrenia. *Harv Rev Psychiatry* 2006; **14**: 78–91.
19 Meyer-Lindenberg A, Zink CF. Imaging genetics for neuropsychiatric disorders. *Child Adolesc Psychiatr Clin N Am* 2007; **16**: 581–597.
20 Glahn DC, Thompson PM, Blangero J. Neuroimaging endophenotypes: strategies for finding genes influencing brain structure and function. *Hum Brain Mapp* 2007; **28**: 488–501.
21 Manoach DS. Prefrontal cortex dysfunction during working memory performance in schizophrenia: reconciling discrepant findings. *Schizophr Res* 2003; **60**: 285–298.
22 Barch DM, Carter CS, Braver TS, Sabb FW, MacDonald III A, Noll DC *et al.* Selective deficits in prefrontal cortex function in medication-naive patients with schizophrenia. *Arch Gen Psychiatry* 2001; **58**: 280–288.
23 Callicott JH, Egan MF, Mattay VS, Bertolino A, Bone AD, Verchinksi B *et al.* Abnormal fMRI response of the dorsolateral prefrontal cortex in cognitively intact siblings of patients with schizophrenia. *Am J Psychiatry* 2003; **160**: 709–719.
24 Karlsgodt KH, Glahn DC, van Erp TG, Therman S, Huttunen M, Manninen M *et al.* The relationship between performance and fMRI signal during working memory in patients with schizophrenia, unaffected co-twins, and control subjects. *Schizophr Res* 2007; **89**: 191–197.
25 Tura E, Turner JA, Fallon JH, Kennedy JL, Potkin SG. Multivariate analyses suggest genetic impacts on neurocircuitry in schizophrenia. *Neuroreport* 2008; **19**: 603–607.
26 Cannon TD, Keller MC. Endophenotypes in the genetic analyses of mental disorders. *Annu Rev Clin Psychol* 2006; **2**: 267–290.
27 Manoach DS, Press DZ, Thangaraj V, Searl MM, Goff DC, Halpern E *et al.* Schizophrenic subjects activate dorsolateral prefrontal cortex during a working memory task, as measured by fMRI. *Biol Psychiatry* 1999; **45**: 1128–1137.
28 Perlstein WM, Carter CS, Noll DC, Cohen JD. Relation of prefrontal cortex dysfunction to working memory and symptoms in schizophrenia. *Am J Psychiatry* 2001; **158**: 1105–1113.
29 Manoach DS, Gollub RL, Benson ES, Searl MM, Goff DC, Halpern E *et al.* Schizophrenic subjects show aberrant fMRI activation of dorsolateral prefrontal cortex and basal ganglia during working memory performance. *Biol Psychiatry* 2000; **48**: 99–109.
30 Akbarian S, Bunney Jr WE, Potkin SG, Wigal SB, Hagman JO, Sandman CA *et al.* Altered distribution of nicotinamide-adenine dinucleotide phosphate-diaphorase cells in frontal lobe of schizophrenics implies disturbances of cortical development. *Arch Gen Psychiatry* 1993; **50**: 169–177.
31 Papassotiropoulos A, Stephan DA, Huentelman MJ, Hoerndli FJ, Craig DW, Pearson JV *et al.* Common Kibra alleles are associated with human memory performance. *Science* 2006; **314**: 475–478.
32 Almasy L, Gur RC, Haack K, Cole SA, Calkins ME, Peralta JM *et al.* A genome screen for quantitative trait loci influencing schizophrenia and neurocognitive phenotypes. *Am J Psychiatry* 2008; **165**: 1185–1192.
33 Egan MF, Goldberg TE, Kolachana BS, Callicott JH, Mazzanti CM, Straub RE *et al.* Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. *Proc Natl Acad Sci USA* 2001; **98**: 6917–6922.
34 Meyer-Lindenberg A, Nichols T, Callicott JH, Ding J, Kolachana B, Buckholtz J *et al.* Impact of complex genetic variation in COMT on human brain function. *Mol Psychiatry* 2006; **11**: 8 67–877, 797.
35 Glahn DC, Paus T, Thompson PM. Imaging genomics: mapping the influence of genetics on brain structure and function. *Hum Brain Mapp* 2007; **28**: 461–463.
36 de Geus E, Goldberg T, Boomsma DI, Posthuma D. Imaging the genetics of brain structure and function. *Biol Psychol* 2008; **79**: 1–8.
37 Aleman A, Swart M, van Rijn S. Brain imaging, genetics and emotion. *Biol Psychol* 2008; **79**: 58–69.
38 Shaw P, Gornick M, Lerch J, Addington A, Seal J, Greenstein D *et al.* Polymorphisms of the dopamine D4 receptor, clinical outcome, and cortical structure in attention-deficit/hyperactivity disorder. *Arch Gen Psychiatry* 2007; **64**: 921–931.
39 van Haren NE, Bakker SC, Kahn RS. Genes and structural brain imaging in schizophrenia. *Curr Opin Psychiatry* 2008; **21**: 161–167.
40 First MB, Spitzer RL, Gibbon M, Williams JBW. *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition. (SCID-I/P)*. Biometrics Research New York State Psychiatric Institute: New York, 2002.
41 First MB, Spitzer RL, Gibbon M, Williams JBW. *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition. (SCID-I/NP)*. Biometrics Research New York State Psychiatric Institute: New York, 2002.

42 Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Freitas CS, Rainey L *et al.* Automated Talairach atlas labels for functional brain mapping. *Hum Brain Mapp* 2000; **10**: 120–131.
43 Price A, Patterson N, Plenge R, Weinblatt M, Shadick N, DA R. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38**: 904–909.
44 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D *et al.* PLINK: a tool set for whole-genome association and population-based linkage analysis. *Am J Hum Genet* 2007; **81**: 559–575.
45 StataCorp. *Stata Statistical Software Release 10*. Stata Corporation: College Station, Texas, 2007.
46 Heimer L. Basal forebrain in the context of schizophrenia. *Brain Res Brain Res Rev* 2000; **31**: 205–235.
47 Petrides M. Lateral prefrontal cortex: architectonic and functional organization. *Philos Trans R Soc Lond B Biol Sci* 2005; **360**: 781–795.
48 Fallon JH, Opole IO, Potkin SG. Neuroanatomy of schizophrenia: circuitry and neurotransmitter systems. *Clin Neurosci Res* 2003; **3**: 77–107.
49 Potkin SG, Alva G, Fleming K, Anand R, Keator D, Carreon D *et al.* A PET study of the pathophysiology of negative symptoms in schizophrenia. Positron emission tomography. *Am J Psychiatry* 2002; **159**: 227–237.
50 Fallon J, Reid S, Kinyamu R, Opole I, Opole R, Baratta J *et al. In vivo* induction of massive proliferation, directed migration, and differentiation of neural cells in the adult mammalian brain. *Proc Natl Acad Sci USA* 2000; **97**: 14686–14691.
51 Nichols T, Hayasaka S. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Stat Methods Med Res* 2003; **12**: 419–446.
52 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B* 1995; **57**: 289–300.
53 Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 2003; **100**: 9440–9445.
54 Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* 2004; **74**: 765–769.
55 Meng Z, Zaykin DV, Xu CF, Wagner M, Ehm MG. Selection of genetic markers for association analyses, using linkage disequilibrium and haplotypes. *Am J Hum Genet* 2003; **73**: 115–130.
56 Dudbridge F, Gusnanto A, Koeleman BP. Detecting multiple associations in genome-wide studies. *Hum Genomics* 2006; **2**: 310–317.
57 Balding DJ. A tutorial on statistical methods for population association studies. *Nat Rev Genet* 2006; **7**: 781–791.
58 Dudbridge F, Koeleman BP. Efficient computation of significance levels for multiple associations in large studies of correlated data, including genomewide association studies. *Am J Hum Genet* 2004; **75**: 424–435.
59 Kirov G, Zaharieva I, Georgieva L, Moskvina V, Nikolov I, Cichon S *et al.* A genome-wide association study in 574 schizophrenia trios using DNA pooling. *Mol Psychiatry* 2008; DOI:10.1038/mp.2008.33.
60 Lencz T, Morgan TV, Athanasiou M, Dain B, Reed CR, Kane JM *et al.* Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. *Mol Psychiatry* 2007; **12**: 572–580.
61 Mah S, Nelson MR, Delisi LE, Reneland RH, Markward N, James MR *et al.* Identification of the semaphorin receptor PLXNA2 as a candidate for susceptibility to schizophrenia. *Mol Psychiatry* 2006; **11**: 471–478.
62 Shifman S, Johannesson M, Bronstein M, Chen SX, Collier DA, Craddock NJ *et al.* Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. *PLoS Genet* 2008; **4**: e28.
63 Sullivan PF, Lin D, Tzeng JY, van den Oord E, Perkins D, Stroup TS *et al.* Genomewide association for schizophrenia in the CATIE study: results of stage 1. *Mol Psychiatry* 2008; **13**: 570–584.
64 Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, Maller J *et al.* Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat Genet* 2007; **39**: 1477–1482.
65 Takai Y, Sasaki T, Matozaki T. Small GTP-binding proteins. *Physiol Rev* 2001; **81**: 153–208.
66 Symons M. Rho family GTPases: the cytoskeleton and beyond. *Trends Biochem Sci* 1996; **21**: 178–181.
67 Kozma R, Ahmed S, Best A, Lim L. The Ras-related protein Cdc42Hs and bradykinin promote formation of peripheral actin microspikes and filopodia in Swiss 3T3 fibroblasts. *Mol Cell Biol* 1995; **15**: 1942–1952.
68 Wells A. EGF receptor. *Int J Biochem Cell Biol* 1999; **31**: 637–643.
69 Cao Q, Martinez M, Zhang J, Sanders AR, Badner JA, Cravchik A *et al.* Suggestive evidence for a schizophrenia susceptibility locus on chromosome 6q and a confirmation in an independent series of pedigrees. *Genomics* 1997; **43**: 1–8.
70 Zhang QH, Ye M, Wu XY, Ren SX, Zhao M, Zhao CJ *et al.* Cloning and functional analysis of cDNAs with open reading frames for 300 previously undefined genes expressed in CD34+ hematopoietic stem/progenitor cells. *Genome Res* 2000; **10**: 1546–1560.
71 Rakic S, Zecevic N. Early oligodendrocyte progenitor cells in the human fetal telencephalon. *Glia* 2003; **41**: 117–127.
72 Ge D, Zhang K, Need AC, Martin O, Fellay J, Urban TJ *et al.* WGAViewer: software for genomic annotation of whole genome association studies. *Genome Res* 2008; **18**: 640–643.

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