

ORIGINAL ARTICLE

DRD2 co-expression network and a related polygenic index predict imaging, behavioral and clinical phenotypes linked to schizophrenia

G Pergola¹, P Di Carlo¹, E D'Ambrosio^{1,2}, B Gelao¹, L Fazio¹, M Papalino¹, A Monda¹, G Scozia¹, B Pietrangelo¹, M Attrotto¹, JA Apud³, Q Chen², VS Mattay^{2,4}, A Rampino^{1,5}, G Caforio⁵, DR Weinberger^{2,6}, G Blasi⁵ and A Bertolino^{1,5}

Genetic risk for schizophrenia (SCZ) is determined by many genetic loci whose compound biological effects are difficult to determine. We hypothesized that co-expression pathways of SCZ risk genes are associated with system-level brain function and clinical phenotypes of SCZ. We examined genetic variants related to the dopamine D2 receptor gene *DRD2* co-expression pathway and associated them with working memory (WM) behavior, the related brain activity and treatment response. Using two independent post-mortem prefrontal messenger RNA (mRNA) data sets (total $N = 249$), we identified a *DRD2* co-expression pathway enriched for SCZ risk genes. Next, we identified non-coding single-nucleotide polymorphisms (SNPs) associated with co-expression of this pathway. These SNPs were associated with regulatory genetic loci in the dorsolateral prefrontal cortex ($P < 0.05$). We summarized their compound effect on co-expression into a Polygenic Co-expression Index (PCI), which predicted *DRD2* pathway co-expression in both mRNA data sets (all $P < 0.05$). We associated the PCI with brain activity during WM performance in two independent samples of healthy individuals (total $N = 368$) and 29 patients with SCZ who performed the n-back task. Greater predicted *DRD2* pathway prefrontal co-expression was associated with greater prefrontal activity and longer WM reaction times (all corrected $P < 0.05$), thus indicating inefficient WM processing. Blind prediction of treatment response to antipsychotics in two independent samples of patients with SCZ suggested better clinical course of patients with greater PCI (total $N = 87$; $P < 0.05$). The findings on this *DRD2* co-expression pathway are a proof of concept that gene co-expression can parse SCZ risk genes into biological pathways associated with intermediate phenotypes as well as with clinically meaningful information.

Translational Psychiatry (2017) **7**, e1006; doi:10.1038/tp.2016.253; published online 17 January 2017

INTRODUCTION

Many genome-wide studies have demonstrated that complex heritable diseases such as schizophrenia (SCZ) are associated with numerous common genetic variants with small effects on susceptibility across heterogeneous populations. Risk variants are not randomly interspersed in the genome, but cluster in selective molecular pathways.^{1,2} Therefore, molecular pathways may be more informative than any individual molecule or molecular event *per se*.³ As genetic variation in risk genes translates into biological risk associated with measurable phenotypes,⁴ identifying common molecular pathways that predict the phenotypes of interest may help elucidating the basis of genetic liability.

Many risk loci for SCZ are associated with the modulation of gene expression.^{3,5} A cardinal principle of the organization of molecular pathways is that gene expression is co-regulated and pathways are likely co-expressed.⁶ This may be also the case of SCZ genes.⁷ Therefore, risk genes for SCZ may be linked through co-expression pathways.^{8,9}

Here we studied brain-specific gene co-expression as a principle to identify molecular pathways of risk genes and detect further genes related with SCZ as 'guilty by association'. With respect to

SCZ risk, the *DRD2* gene coding for the D2 dopamine receptor is an optimal candidate to investigate the genetic architecture of molecular pathways affected in patients with SCZ. A long-standing hypothesis holds that dopaminergic dysfunctional regulation in subcortical areas and in the prefrontal cortex (PFC) is a key pathophysiological mechanism of SCZ.^{10,11} For example, working memory (WM) deficits, among the core symptoms in SCZ, are related with alterations of PFC activity.^{12–19} Moreover, PFC activity during WM is predicted by midbrain dopamine,²⁰ striatal dopamine²¹ and by reductions of amphetamine-induced release of prefrontal dopamine binding to D2 receptors in SCZ.²² Critically, research by the Psychiatric Genomic Consortium (PGC) supported the notion that the *DRD2* gene coding for the D2 dopaminergic receptor is associated with risk for SCZ.²³ *DRD2* is not only related to SCZ risk, but genetic variation within this gene has been linked with phenotypes related to SCZ and its treatment.^{24–27} Consistent with studies showing that D2 dopaminergic receptors are involved in WM,²⁸ functional genetic variants in *DRD2* have also been associated with WM performance and related cortical activity in healthy subjects and in patients with SCZ.^{21,27} For example, a non-coding single-nucleotide polymorphism (SNP) of *DRD2*, rs1076560, modulates alternative splicing of the D2 dopaminergic receptor

¹Department of Basic Medical Science, Neuroscience and Sense Organs, University of Bari Aldo Moro, Bari, Italy; ²Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, MD, USA; ³National Institutes of Health, National Institute of Mental Health, Clinical and Translational Neuroscience Branch, NIMH, Bethesda, MD, USA; ⁴Departments of Neurology and Radiology, Johns Hopkins School of Medicine, Baltimore, MD, USA; ⁵Institute of Psychiatry, Department of Neuroscience, Sense Organs and Locomotive System, Bari University Hospital, Bari, Italy and ⁶Departments of Psychiatry, Neurology, Neuroscience and The Mckusick-Nathans Institute of Genomic Medicine, Johns Hopkins School of Medicine, Baltimore, MD, USA. Correspondence: Professor A Bertolino, Department of Basic Medical Science, Neuroscience and Sense Organs, University of Bari Aldo Moro, Piazza Giulio Cesare 11, Bari 70120, Italy. E-mail: alessandro.bertolino@uniba.it

Received 24 June 2016; revised 28 September 2016; accepted 13 October 2016

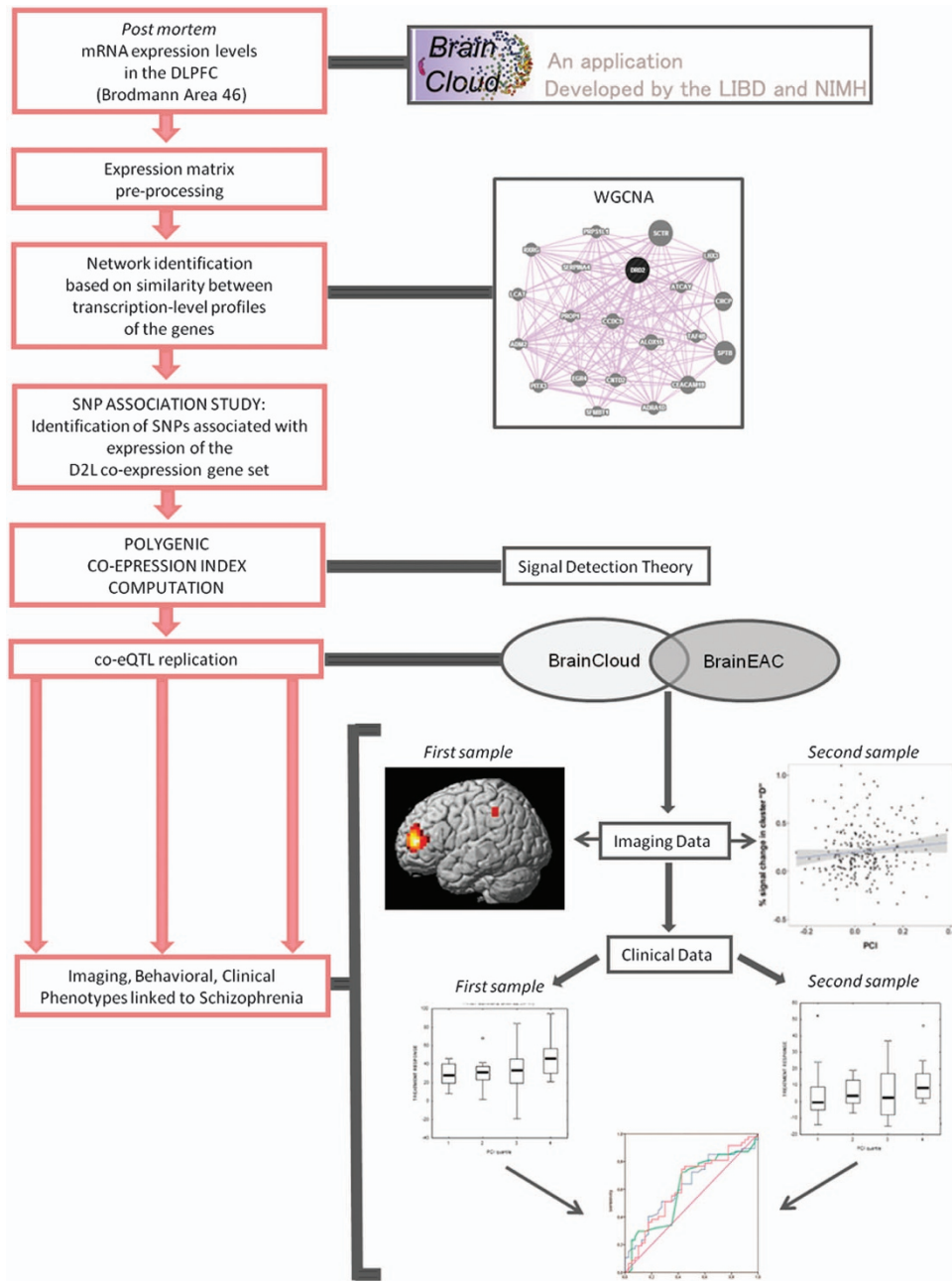


Figure 1. Concept of the study. DLPFC, dorsolateral prefrontal cortex; SNP, single-nucleotide polymorphism.

transcript into two major isoforms (D2 long, D2L, and D2 short, D2S), likely affecting neuronal activity.²⁷ However, *DRD2* single-SNP associations explain only a small fraction of risk for SCZ and related phenotypes.^{21,27}

We investigated the compound association between SNPs approximating *in vivo* the transcription levels of a gene set co-expressed with D2L and phenotypes of relevance to SCZ. We focused on D2L in the PFC because it is mainly found at the post-synaptic terminal,²⁹ its expression is altered in the PFC of patients with SCZ,¹³ and we were interested in the modulation of activity of neuronal populations receiving dopaminergic afferents during WM, a mechanism that is altered in SCZ.²² Furthermore, D2L is targeted by antipsychotic medications.^{29,30} To achieve our aim, we performed four consecutive steps (Figure 1). First, we identified the co-expression gene set of D2L from human post-mortem

PFC.³¹ Second, we identified the association of independent genetic variants of the genes included in the gene set with expression of the whole gene set. Third, we combined these genetic variants into a Polygenic Co-expression Index (PCI) designed to index the genetic component of gene transcription co-regulation and validated this index in an independent post-mortem data set. Fourth, we associated our PCI with brain activity and behavioral performance during WM, that is, D2L-dependent and key intermediate phenotypes for SCZ,³¹ and response to treatment with antipsychotics which target D2 signaling in SCZ. Based on prior evidence,²² we hypothesized that genetic variants associated with greater co-expression of D2L and its gene set would also be associated with greater BOLD signal in the PFC during WM and poorer behavioral performance. In patients with SCZ, we investigated the preservation of the PCI-BOLD signal

Table 1. Demographic data in all studies reported

Sample name	Sample size	Female (male) [ratio]	Age mean \pm s.d. years	Age range (years)
Braincloud	199	60 (139) [0.43]	32.3 \pm 20	0–78
BrainEAC	50	15 (35) [0.43]	57 \pm 19	20–91
First healthy fMRI study	124	54 (70) [0.77]	27.5 \pm 6.6	19–48
Second healthy fMRI study	244	132 (112) [1.20]	29.4 \pm 8.4	18–55
fMRI patients study	29	3 (26) [0.12]	29.3 \pm 7.0	15–42
First clinical study	47	8 (39) [0.21]	28.7 \pm 7.0	16–42
Second clinical study	40	11 (29) [0.38]	27.4 \pm 6.5	18–40

Abbreviation: fMRI, functional magnetic resonance imaging.

relationships and assessed the sensitivity and specificity of the PCI as predictor of treatment outcome. Previous findings reported that the T allele of rs1076560, which is associated with a greater D2L/D2S ratio in the PFC,² is also associated with greater symptom improvement in patients with SCZ.¹⁵ Therefore, we hypothesized that greater D2L gene set co-expression in the PFC indexed by the PCI predicted greater clinical response.

MATERIALS AND METHODS

Participants

Table 1 summarizes the demographic data of the subjects included in all experiments. After receiving a complete description of the study, all participants in the clinical and the imaging studies provided written informed consent following the guidelines of the Declaration of Helsinki. Protocols and procedures were approved by the ethics committee of the University of Bari and by the institutional review board of the National Institute of Health, Bethesda, MD, USA.

Network identification

We used the publicly available Braincloud data set⁵ (<http://braincloud.jhmi.edu/>) for a genome-wide Weighted Genes Co-expression Network Analysis.³² The sample included 199 observations (demographics in Table 1). We preprocessed the gene expression matrix to factor out confounding variables, including demographics. The first principal component of the *DRD2* gene set (module eigengene, ME) served to track the simultaneous variation of the whole gene set. We correlated this co-expression measure with *DRD2* expression levels.³² To investigate the biological functions that may be subserved by this ensemble of co-expressed genes, we computed gene ontology enrichment analysis using AmiGO2 (<http://amigo.geneontology.org/amigo/landing>). Finally, we assessed enrichment of the gene set for the loci associated with SCZ risk by the Psychiatric Genomic Consortium (PGC²³ with a hypergeometric test (SI Materials and Methods).

SNP association study

We performed a gene set-wide association study of SNPs with the ME. The sample size of our post-mortem data set is probably small for a genetic association study, and co-expression is most likely a phenotypic trait with complex heritability, much like SCZ. Power calculations in genetic association studies have been shown to depend on many factors, including heritability of the trait, the proportion of variance explained by the genotyped SNPs, the total number of SNPs, the proportion of SNPs with no effect on the trait, the total sample size and the *P*-value threshold for SNP selection.³³ The heritability of expression quantitative trait loci (eQTLs) is extremely variable and appears to be comparable between *cis*- and *trans*-eQTLs (co-eQTLs fall in the latter category³⁴). Nevertheless, there is evidence of high replicability of eQTLs both with stringent and more lenient thresholds of significance.³⁵ These findings suggest that, beside statistical significance, genetic signals in the study of gene expression may be found beyond the threshold for corrected or nominal significance, as is also the case of complex clinical traits.³⁶

Based on these considerations, we first tested for co-eQTLs that would survive stringent Bonferroni correction; then, we employed more permissive statistics for our association to minimize false-negative findings

and performed internal cross-validations and independent replication to minimize type I errors. Through this procedure, we aimed at identifying an ensemble of SNPs that, together, predict co-expression. We used eight SNPs associated with the first principal component of gene set co-expression with *P* < 0.005 to compute the PCI. As a common method to select SNPs for subsequent inclusion in polygenic scores consists in increasing the number of SNPs until the proportion of variance plateaus,³⁷ we also tested whether such an approach selected the same set of eight SNPs with *P* < 0.005 (SI Materials and Methods). We interrogated Haploreg v4.1 (<http://archive.broadinstitute.org/mammals/haploreg/haploreg.php>) to gain information on the possible regulatory functions of these SNPs.³⁸ Then, we computed the PCI by assigning a weight to each genotype of each SNP based on the expression profile within different genotypic groups.³⁹ The greater the PCI, the greater is the messenger RNA expression level of that individual. We cross-validated the PCI and assessed ethnicity and population stratification effects, as well as age effects (SI Materials and Methods).

We used the publicly available BrainEAC data set⁴⁰ to replicate the association of the PCI with D2L co-expression in the frontal cortex. From this data set, we selected the probes of all genes included in the D2L co-expression pathway and preprocessed the data as above reported. The sample included 50 Caucasian participants.⁴⁰ We computed the PCI of each individual and associated it with the ME using Pearson's correlation.

Imaging study. We recruited 124 healthy unrelated Caucasian adults from the region of Apulia, Italy (demographics in Table 1), for a functional magnetic resonance imaging (fMRI) experiment and genotyped them for the SNPs included in the PCI. We used the n-back task to probe WM.⁴¹ Stimuli consisted of numbers (1–4) shown in random sequence and displayed at the points of a diamond-shaped box. There was a non-memory-guided control condition (0-back) that required subjects to identify the stimulus currently seen. As memory load increased, the task required the recollection of a stimulus seen one (1-back) or two (2-back) stimuli before, while keeping on encoding incoming stimuli (SI Materials and Methods). We tested the association of the PCI with brain activation using repeated measures analysis of covariance (within-subject factor: LOAD (1-back, 2-back); covariates: age, gender and handedness; whole-brain topological false discovery rate-corrected $\alpha=0.05$; extent threshold = 6, that is, > 300 mm³).

We recruited a second fMRI sample of 244 Caucasian healthy volunteers as part of the NIMH Clinical Brain Disorders Branch 'Sibling Study' (Table 1). These participants performed the same 2-back fMRI task described above and were genome-wide genotyped (SI Materials and Methods⁴²). The effect of the PCI on BOLD response was tested using robust linear models^{43–45} with age, gender and handedness as covariates of no interest. We performed one-tailed *t*-tests using as regions of interest the clusters associated with the PCI in the first healthy sample. We used MarsBar (<http://marsbar.sourceforge.net/>) to extract the percent signal change and corrected for the number of tests (false discovery rate⁴⁶).

For the fMRI study on the clinical cohort, we recruited 29 Caucasian patients with SCID diagnosis of SCZ from the region of Apulia, Italy (Table 1). Genotyping and fMRI protocols matched the procedures followed for the first fMRI healthy sample.

Behavioral analyses. In the first healthy fMRI sample, we investigated behavioral performance (accuracy and reaction time) during the WM task in the scanning session (SI Materials and Methods). We analyzed accuracy, that is, percent of correct responses, and reaction times with repeated

Table 2. SNPs associated with the first principal component of D2L gene set expression

Rank	Marker	Locus	Gene ^a	Gene name	Uncorrected P-value	MAF	PGC P-value
1	rs2486064	1q32.1	CHIT1	Chitinase 1	5.0×10^{-6}	0.22	0.088
2	rs6902039	6p22.3	GPLD1	Glycosylphosphatidylinositol-specific phospholipase D1	4.6×10^{-4}	0.23	0.48
3	rs851436	2p24.1	OSR1	Odd-skipped related 1	1.0×10^{-3}	0.48	0.77
4	rs9297283	8q22.2	POP1	Processing of precursor 1, ribonuclease P/MRP subunit	1.0×10^{-3}	0.20	0.33
5	rs12940715	17q25.1	SDK2	Sidekick cell adhesion molecule 2	1.7×10^{-3}	0.20	0.050
6	rs1805453	17p13.2	DHX33	DEAH (Asp-Glu-Ala-His) Box Polypeptide 33	2.8×10^{-3}	0.34	0.054
7	rs11213916	11q22.3	BTG4	B-cell translocation gene 4	3.0×10^{-3}	0.30	0.50
8	rs1037791	7p21.1	AGR2	Anterior gradient 2	3.2×10^{-3}	0.31	0.94

Abbreviations: MAF, minor allele frequency in the Braincloud sample; PGC, Psychiatric Genomics Consortium; SNP, single-nucleotide polymorphism. ^aThe Gene column reports the genes included in the D2L gene set. SNPs fall in a ± 100 kbp window from the set genes. P-values refer to the association with diagnosis of schizophrenia.

measures analysis of covariance (within-subject factor: LOAD (1-back, 2-back)) and PCI as a predictor. As we were testing two measures, we set the threshold for significance at $\alpha = 0.025$ (Bonferroni correction). In the clinical fMRI sample, we analyzed behavioral performance following the same procedures employed for the first fMRI sample.

Pharmacogenetics. The first clinical cohort in the pharmacogenetic study included 47 Caucasian patients with SCZ (Structured Clinical Interview for DSM (SCID) diagnosis)²⁶ recruited from the region of Apulia, Italy (Table 1). Treatment response was computed as the difference in the total Positive And Negative Syndrome Scale (PANSS) core between baseline and treatment end.

The second cohort consisted of 40 patients with SCZ with history of inadequate treatment response recruited at the Clinical Brain Disorders Branch SCZ inpatient research unit at the National Institutes of Health Clinical Center, Bethesda, MD, USA.²⁶ Treatment response was defined as the difference between symptoms severity at the end of placebo treatment and severity at the end of drug treatment.¹⁵ The clinical protocols used in the pharmacogenetic study for both samples have been described in detail previously²⁶ (see Supplementary Table 2 for clinical data).

We first performed an association study to investigate the direction of the PCI treatment response relationship using Spearman's Rho; then, we assessed the potency of the PCI as predictor of treatment response in comparison with pharmacological doses and off-medication symptom severity using Receiver Operating Characteristic curves (SI Materials and Methods).

RESULTS

Identification of the D2L gene set

We identified a co-expression gene set of 85 genes including the D2L transcript (Supplementary Table 1). Its first principal component (ME) explained 32.5% of the variance. D2L expression levels positively correlated with ME ($R^2 = 0.4$). The D2L co-expression pathway was enriched for the ontologies DNA packaging (GO:0006323, corrected P -value = 0.002), negative regulation of dopamine secretion (GO:0033602, corrected P -value = 0.004) and response to nicotine (GO:0035094, corrected P -value = 0.03); notably, some of these results may be affected by our choice of the *DRD2* module as a candidate gene set. Besides *DRD2*, this gene set included three genes associated with SCZ based on PGC2 (*GATAD2A*, *GALNT10* and *ZSCAN23*). The enrichment of the gene set for protein-coding genes located in the genome-wide association study loci associated with SCZ by PGC2²³ was significant (hypergeometric test, $P = 0.029$).

Identification of co-eQTLs associated with co-expression of the whole D2L module

Eight independent SNPs located in the genes included in the module were associated with co-expression of the whole gene set, that is, the ME (Table 2). The first SNP, rs2486064, survives even very stringent statistics using Bonferroni correction for multiple comparisons (corrected $P = 0.0033$). Table 2 shows that these SNPs are not strong predictors of SCZ status in the PGC2 work.²³ Five out of these eight non-coding SNPs modify regulatory motifs⁴⁷

and two out of eight have been previously recognized as *trans*-eQTLs.⁴⁸ As an ensemble, these SNPs are strongly associated with gene expression regulation in the dorsolateral prefrontal cortex Haploreg v4.1, $P = 0.00049$, SI Appendix). We used these SNPs to compute the PCI and to verify its correlation with expression of the whole gene set (as per definition of the PCI; $R^2 = 0.38$) and D2L transcriptional levels ($t_{198} = 5.8$, $R^2 = 0.14$, $P = 2.9 \times 10^{-8}$; Supplementary Figure 1). These effects were not affected either by population stratification or by age (Supplementary Table 3; SI Results). Multiple cross-validations performed using different procedures uniformly supported the association between co-expression of the gene set and the PCI (Supplementary Figures 2 and 3).

As in spite of all *in silico* validations it cannot be definitively ruled out that our findings were related with peculiarities of the specific sample analyzed, for example, with ethnicity, we replicated the association between the PCI and the gene set using an independent data set (BrainEAC⁴⁰). BrainEAC included 80 out of the 85 probes of the D2L gene set above identified and was not affected by ethnicity effects because it only included Caucasian subjects. The correlation between the PCI and gene set co-expression replicated in the same direction (Pearson's $R = 0.23$, one-tailed $P = 0.05$). The association was strongest for highest quality observations (for example, with RNA Integrity Number (RIN) > 6, $R = 0.38$, one-tailed $P = 0.028$; Supplementary Figure 4). This is relevant because all Braincloud data had RIN > 7, thus we found a significant association in spite of non-overlapping probes, of partly degraded messenger RNAs and of ethnicity differences.

Biological validation of the PCI by means of association with system-level phenotypes: imaging study

In the first healthy sample, activity in the fronto-parietal WM network correlated positively with the PCI surviving whole-brain peak-level correction for multiple comparisons (topological false discovery rate $q < 0.05$; Figure 2 and Table 3). Individuals with greater PCI and greater predicted D2L gene set co-expression levels had greater bilateral activation in the WM brain network during task performance, that is, they were less efficient in processing WM information. There was no significant negative correlation with the PCI or LOAD \times PCI interaction on brain activity. Further analyses supporting the robustness of these results are reported in Supplementary Tables 3–4 and Supplementary Figures 5–6.

In the second healthy sample, we used the clusters identified in the first sample to extract the signal change which we associated to the PCI. We found significant correlations between the PCI and percent signal change in the left anterior middle frontal gyrus (BA10), and in the right inferior parietal lobule (BA40; Table 3). Also in patients, brain activity correlated with the PCI (Supplementary

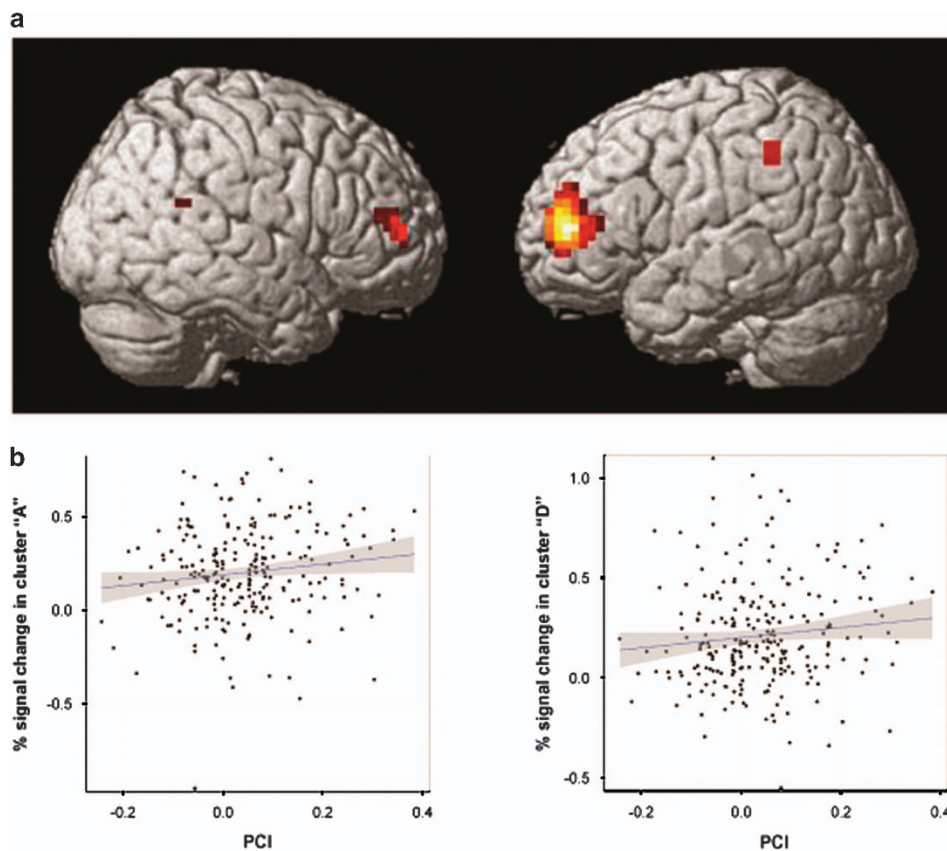


Figure 2. Discovery and replication functional magnetic resonance imaging results. (a) First sample. Significant clusters associated with the positive slope of the Polygenic Co-expression Index (PCI) at topological false discovery rate-corrected q -value < 0.05 (cluster extent = 6). (b) Second sample. Scatter plots of the % signal change in two clusters with significant positive correlation with the PCI in the second sample. The regression line is blue, 95% confidence intervals for the fit are gray. See Table 3 for the exact statistics and localization of clusters 'A' and 'D'. Left in the figure is left in the brain.

Figure 7). Hence, the PCI-BOLD association was preserved in patients with SCZ.

Behavioral results

In the first healthy fMRI sample, repeated measures analysis of covariance (within-subject factor LOAD (1-back, 2-back)) on WM accuracy and reaction times revealed a significant LOAD \times PCI interaction on reaction times surviving Bonferroni correction ($F_{1,120} = 6.9, P = 0.01$). *Post hoc* regressions were non-significant for 1-back ($t_{120} = 0.12$, adjusted $R^2 = -0.008, P = 0.91$) but yielded a significant fit for 2-back (Supplementary Figure 6; $t_{120} = 2.3$, adjusted $R^2 = 0.033, P = 0.024$). Greater PCI was related with longer reaction times at 2-back (Supplementary Figure 8), supporting the imaging findings and the interpretation of reduced efficiency in these individuals. No other significant effects or interactions involved the PCI (SI Results). The same analysis on behavioral data of the clinical fMRI sample yielded no significant effects or interactions (all $P > 0.05$).

Clinical translation of the PCI by means of association with response to treatment with antipsychotics

The correlation between clinical improvement (difference in PANSS total score between baseline and end point) and PCI was positive and significant in the first clinical sample ($N = 47, \rho = 0.39; P = 0.007$; Supplementary Figure 9). Results replicated in the second clinical sample (PANSS total score, $N = 40, \rho = 0.27$, one-tailed $P = 0.047$). Greater PCI was associated with greater clinical response. Then, we pooled the two samples ($N = 87$) and tested

the blind prediction of treatment response based on the PCI, which was significant (area under the curve (AUC) = 0.63, $P = 0.043$; Supplementary Figure 10). Similarly, off-medication symptoms (AUC = 0.62, $P = 0.048$) significantly predicted treatment response, whereas dose adjustment was a marginally significant predictor (AUC = 0.61, $P = 0.079$).

DISCUSSION

We believe the present results provide the first proof of concept that the co-expression context of SCZ risk genes, for example, *DRD2*, affects system-level and clinical phenotypes. Although previous studies indexed gene expression based on *cis*-genetic markers,^{39,49} here we detected co-expression *trans*-eQTLs (co-eQTLs) to predict gene co-expression using a network approach. This data-driven, bias-free procedure identified a molecular pathway of convergence of some genes associated with risk for SCZ. The D2L network recapitulates part of the complex neurobiology of SCZ and other systems level phenotypes, including WM performance and related brain activity.

Genes co-expressed with D2L

We found that the variability of system-level phenotypes is closely associated with a shared component between co-expressed genes. Indeed, the statistics of the association between the PCI and imaging phenotypes outperform prior reports on the effect of genetic variants associated with *DRD2*.^{26,27} Moreover, the effect of the PCI survived when we co-varied for *DRD2* rs1076560 genotype

Table 3. Statistics of the association between PCI and brain activity during working memory*

Cluster	Region/BA	First healthy fMRI sample						Second healthy fMRI sample			
		MNI (x, y, z)	Talairach (x, y, z)	Cluster extent	Z-value	Uncorrected P-value	Corrected P-value	Partial η^2	t-value	Uncorrected P-value	Corrected P-value
A	Left MIFG/10	-33, 49, 14	-33, 47, 13	156	4.8	< 0.001	0.002	0.13	2.6	0.0052	0.01
B	Right MIFG/10	31, 46, 14	30, 44, 14	26	3.44	< 0.001	0.027	0.08	0.26	0.40	0.40
C	Right STG/13	46, -44, 21	45, -42, 22	7	3.36	< 0.001	0.031	0.07	0.73	0.23	0.31
D	Left IPL/40	-59, -41, 44	-58, -38, 41	12	3.32	< 0.001	0.033	0.07	3.1	0.0011	0.0044

Abbreviations: BA, Brodmann area; FDR, false discovery rate; fMRI, functional magnetic resonance imaging; IPL, inferior parietal lobule; MIFG, middle frontal gyrus; MNI, Montreal Neurological Institute; STG, superior temporal gyrus. First study: Displayed clusters survive peak-level topological FDR-corrected threshold $q=0.05$, cluster extent = 6. Results also surviving cluster-level family-wise error (FWE) correction for multiple comparisons are in bold font. Partial η^2 refers to the effect size of the association of brain activity with the Polygenic Co-expression Index (PCI). Second study: Results surviving FDR correction for multiple comparisons (number of tests = 4, $\alpha=0.05$) are in bold font. *White fields refer to the first healthy fMRI sample, grey fields to the second healthy fMRI sample study.

(SI Results). Interestingly, our gene set included genes previously associated with biochemical pathways relevant to SCZ, for example, calcium- and cannabinoid-mediated transmission like *CACNA2D4* (ref. 50) and *CNR1*.⁵¹ The gene set also includes genes not mentioned in the latest PGC2 publication, but associated with SCZ in previous genome-wide association studies, such as *CALHM3*.⁵² Interestingly, one of the set genes associated with SCZ, *GATAD2A*, ranked fourth for intramodular connectivity, hence belonging to a group of highly connected genes (hubs) within the identified network. *GATAD2A* is involved in gene silencing and is associated with histones.⁵³ Accordingly, the gene set includes further histone-related genes (*HIST1H1E*, *HIST1H3G* and *HIST2H2AC*) and several of the SNPs included in the PCI are associated with histone functions (SI Appendix; see also the SI Discussion for more information on the SNPs identified). Histone proteins have been highlighted in a recent *trans*-diagnostic gene set analysis of the genetic architecture of psychiatric disorders.²³

Because of the data-driven nature of the approach, genes associated by prior literature with D2 dopamine receptors, for example, by physical interaction, may end up in different clusters when building a network. Therefore, not all genes relevant to D2L-mediated signaling are included in the gene set studied here. However, a tenet of co-expression analyses is that co-expressing genes are co-regulated in terms of transcription, for example, they may be targeted by the same repressors/enhancers or si/miRNA. It is also important to note that the network we identified is a model of co-expression, and does not necessarily represent gene co-regulation. For example, gene expression patterns vary as a function of many variables, such as age,⁵ ethnicity and cell type. Here we clustered genes based on their expression patterns corrected for confounding variables. The SI Materials and Methods includes additional information on age and ethnicity effects (SI Results).

Because of the relatively small sample size of post-mortem data sets for a genetic association study, we used permissive procedures for co-eQTL detection. Then, we cross-validated the SNP weights and also the SNP selection within Braincloud, we replicated the effects in BrainEAC, and we also cross-validated the SNPs entering the PCI in the fMRI experiment. Together, these validation steps and the findings obtained in multiple independent data sets across multiple biological scales support the idea that the SNPs identified here are valid and modulate the D2L co-expression pathway. It is remarkable that eight common SNPs weighted for their molecular effects on gene co-expression accounted for a sizable proportion of variance in prefrontal activation and significantly predicted treatment outcome in patients with SCZ (SI Discussion).

The SNPs associated with the D2L co-expression gene set that we identified have not been previously reported for their association with *DRD2* or with psychiatric or cognitive phenotypes, except for rs1037791 that has been associated with the openness subscale of the big five personality traits.⁵⁴ However, the first-ranked SNP, which even survived Bonferroni correction for multiple comparisons, is intergenic between two paralog genes, *CHIT1* and *CH13L1*. Both genes code for stress-induced chitinases and prior evidence associated brain *CH13L1* expression with SCZ.^{55,56} Genetic variants in this intergenic region such as rs4950928 have been associated with SCZ.^{57,58} This SNP is located upstream of *CH13L1* where also rs2486064 is located (linkage disequilibrium: $r^2=0.35$, $D'=1$). This regulatory region likely interacts with specific transcription factors.^{57,58} *CH13L1* expression has been proposed to be relevant to SCZ because of its association with the *AKT1-GSK3 β* pathway,⁵⁸ which is also associated with the cyclic AMP-independent pathway of *DRD2*.⁵⁹ The present evidence suggests a further candidate molecular mechanism of action of genetic variants located in this region in SCZ, that is, regulation of a *DRD2* co-expression gene set.

The association of the D2L co-expression PCI with prefrontal activity and WM performance

Dopamine D2 receptor signaling is a critical modulator of WM. The present findings suggest that individuals bearing genotypes associated with increased simultaneous expression of the D2L gene set manifest greater PFC activity as well as longer reaction times during WM. These findings suggest that increased expression of the D2L co-expression gene set predisposes to less efficient WM processing, a well-established intermediate risk-associated phenotype for SCZ.³¹ These results are also consistent with a large body of studies demonstrating that pharmacological manipulation of dopamine D2R is associated with WM performance and with prefrontal activity measured with BOLD fMRI.^{41,60} Critically, they are consistent with a recent study reporting that greater prefrontal D2 PET binding is positively correlated with prefrontal activity during WM.²² Finally, these findings are consistent with several reports from our group suggesting that the T allele of rs1076560, associated with greater D2L/D2S ratio compared with the G allele, is also associated with inefficient prefrontal activity during WM.^{21,27,41} The present findings extend this earlier work by showing that the co-expression gene set identified by our network analysis is associated with genetic background outside the *DRD2* gene. Such genetic variation is in turn associated with intermediate phenotypes of SCZ.

The role of D2L and its co-expression gene set in treatment response to antipsychotics in patients with SCZ

D2 dopaminergic receptor is the main target of antipsychotic medications^{29,30} and previous evidence indicated that T-carriers for rs1076560 benefit more from antipsychotic treatment than other patients with SCZ.²⁶ Consistently, here we show that greater predicted expression of a D2L co-expression gene set is associated with greater clinical improvement. We found this association not only in drug-naïve/drug-free patients but also in patients with history of inadequate treatment response. The strength of the prediction based on the PCI compared favorably with clinical predictors such as pharmacological dose. Results suggest that SCZ patients with greater genetically determined availability of a main target of antipsychotics are predisposed to better treatment response.

CONCLUSIONS

The present findings suggest that a D2L co-expression gene set enriched for protein-coding genes associated with schizophrenia modulates PFC function during WM and response to D2 antagonist antipsychotic drugs. The genetic variants detected in this study were not associated with diagnosis, but were located in regulatory genetic loci. In other words, genetic variation modulating molecular pathways of SCZ risk genes may recapitulate part of the variance of SCZ-related phenotypes in healthy and clinical populations.

CONFLICT OF INTEREST

GP is the academic supervisor of a collaborative research project with Hoffman-La Roche, Ltd, which supports his and AR's salary. The remaining authors declare no conflict of interest.

ACKNOWLEDGMENTS

We are grateful to Dr Andrew Jaffe, Dr Richard Straub, Dr Gianluca Ursini (Lieber Institute for Brain Development), Dr Nicola Amoroso (Department of Physics, University of Bari Aldo Moro) and Dr Silvia Torretta (Department of Basic Medical Science, Neuroscience and Sense Organs, University of Bari Aldo Moro) for helpful methodological insights. Data acquisition was made possible by Rita Maselli, Dr Paolo Taurisano, Tiziana Quarto, Dr Raffaella Romano, Dr Annamaria Porcelli, Dr Marina Mancini, Linda Antonucci and Dr Silvia Torretta (Department of Basic Medical

Science, Neuroscience, and Sense Organs, University of Bari Aldo Moro). Finally, we gratefully acknowledge the work by Dr Pierluigi Selvaggi, Alessandra Raio, Roberta Passiatore and Dalila Albergo (Department of Basic Medical Science, Neuroscience, and Sense Organs – University of Bari Aldo Moro), who contributed to data analysis. This research has been funded by a 'Capitale Umamo ad Alta Qualificazione' grant by Fondazione Con Il Sud and by the NARSAD grant (number: 28935) awarded to Alessandro Bertolino, by the NIMH Clinical Brain Disorders Branch 'Sibling Study' protocol 95-M-0150 awarded to Daniel R. Weinberger, and by the Lieber Institute for Brain Development. The funding sources had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; preparation, review or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

- Kendler KS. What psychiatric genetics has taught us about the nature of psychiatric illness and what is left to learn. *Mol Psychiatry* 2013; **18**: 1058–1066.
- Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat Neuroscience* 2015; **18**: 199–209.
- Harrison PJ, Weinberger DR. Schizophrenia genes, gene expression, and neuro-pathology: on the matter of their convergence. *Mol Psychiatry* 2005; **10**: 40–68.
- Insel TR. The NIMH Research Domain Criteria (RDoC) Project: precision medicine for psychiatry. *Am J Psychiatry* 2014; **171**: 395–397.
- Colantuoni C, Lipska BK, Ye T, Hyde TM, Tao R, Leek JT et al. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* 2011; **478**: 519–523.
- Gaiteri C, Ding Y, French B, Tseng GC, Sibille E. Beyond modules and hubs: the potential of gene coexpression networks for investigating molecular mechanisms of complex brain disorders. *Genes Brain Behav* 2014; **13**: 13–24.
- Haugberg ME, Roussos P, Grove J, Borglum AD, Mattheisen M. Analyzing the role of microRNAs in schizophrenia in the context of common genetic risk variants. *JAMA Psychiatry* 2016; **73**: 369–377.
- Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998; **95**: 14863–14868.
- Ramani AK, Li Z, Hart GT, Carlson MW, Boutz DR, Marcotte EM. A map of human protein interactions derived from co-expression of human mRNAs and their orthologs. *Mol Syst Biol* 2008; **4**: 180.
- Weinberger DR, DeLisi LE, Perman GP, Targum S, Wyatt RJ. Computed tomography in schizophreniform disorder and other acute psychiatric disorders. *Arch Gen Psychiatry* 1982; **39**: 778–783.
- Abi-Dargham A. Schizophrenia: overview and dopamine dysfunction. *J Clin Psychiatry* 2014; **75**: e31.
- Bertolino A, Caforio G, Petruzzella V, Latorre V, Rubino V, Dimalta S et al. Prefrontal dysfunction in schizophrenia controlling for COMT Val158Met genotype and working memory performance. *Psychiatry Res* 2006; **147**: 221–226.
- Bertolino A, Esposito G, Callicott JH, Mattay VS, Van Horn JD, Frank JA et al. Specific relationship between prefrontal neuronal N-acetylaspartate and activation of the working memory cortical network in schizophrenia. *Am J Psychiatry* 2000; **157**: 26–33.
- Callicott JH, Ramsey NF, Tallent K, Bertolino A, Knable MB, Coppola R et al. Functional magnetic resonance imaging brain mapping in psychiatry: methodological issues illustrated in a study of working memory in schizophrenia. *Neuropsychopharmacology* 1998; **18**: 186–196.
- Goldman-Rakic PS. Working memory dysfunction in schizophrenia. *J Neuro-psychiatry Clin Neurosci* 1994; **6**: 348–357.
- Lee J, Park S. Working memory impairments in schizophrenia: a meta-analysis. *J Abnorm Psychol* 2005; **114**: 599–611.
- Manoach DS. Prefrontal cortex dysfunction during working memory performance in schizophrenia: reconciling discrepant findings. *Schizophr Res* 2003; **60**: 285–298.
- Schwarz E, Tost H, Meyer-Lindenberg A. Working memory genetics in schizophrenia and related disorders: an RDoC perspective. *Am J Med Genet B Neuropsychiatr Genet* 2016; **171**: 121–131.
- Weinberger DR, Egan MF, Bertolino A, Callicott JH, Mattay VS, Lipska BK et al. Prefrontal neurons and the genetics of schizophrenia. *Biol Psychiatry* 2001; **50**: 825–844.
- Meyer-Lindenberg A, Kohn PD, Kolachana B, Kippenhan S, McInerney-Leo A, Nussbaum R et al. Midbrain dopamine and prefrontal function in humans: interaction and modulation by COMT genotype. *Nat Neurosci* 2005; **8**: 594–596.
- Bertolino A, Taurisano P, Pisciotto NM, Blasi G, Fazio L, Romano R et al. Genetically determined measures of striatal D2 signaling predict prefrontal activity during working memory performance. *PLoS One* 2010; **5**: e9348.
- Slifstein M, van de Giessen E, Van Snellenberg J, Thompson JL, Narendran R, Gil R et al. Deficits in prefrontal cortical and extrastriatal dopamine release in

- schizophrenia: a positron emission tomographic functional magnetic resonance imaging study. *JAMA Psychiatry* 2015; **72**: 316–324.
- 23 Schizophrenia Working Group of the Psychiatric Genomics Consortium Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**: 421–427.
 - 24 Kaalund SS, Newburn EN, Ye T, Tao R, Li C, Deep-Soboslay A et al. Contrasting changes in DRD1 and DRD2 splice variant expression in schizophrenia and affective disorders, and associations with SNPs in postmortem brain. *Mol Psychiatry* 2014; **19**: 1258–1266.
 - 25 Zhang JP, Lencz T, Malhotra AK. D2 receptor genetic variation and clinical response to antipsychotic drug treatment: a meta-analysis. *Am J Psychiatry* 2010; **167**: 763–772.
 - 26 Blasi G, Selvaggi P, Fazio L, Antonucci LA, Taurisano P, Masellis R et al. Variation in dopamine D2 and serotonin 5-HT2A receptor genes is associated with working memory processing and response to treatment with antipsychotics. *Neuropsychopharmacology* 2015; **40**: 1600–1608.
 - 27 Zhang Y, Bertolino A, Fazio L, Blasi G, Rampino A, Romano R et al. Polymorphisms in human dopamine D2 receptor gene affect gene expression, splicing, and neuronal activity during working memory. *Proc Natl Acad Sci USA* 2007; **104**: 20552–20557.
 - 28 Kellendonk C, Simpson EH, Polan HJ, Malleret G, Vronskaya S, Winiger V et al. Transient and selective overexpression of dopamine D2 receptors in the striatum causes persistent abnormalities in prefrontal cortex functioning. *Neuron* 2006; **49**: 603–615.
 - 29 Usiello A, Baik JH, Rouge-Pont F, Picetti R, Dierich A, LeMeur M et al. Distinct functions of the two isoforms of dopamine D2 receptors. *Nature* 2000; **408**: 199–203.
 - 30 Richtand NM, Welge JA, Logue AD, Keck PE Jr, Strakowski SM, McNamara RK. Dopamine and serotonin receptor binding and antipsychotic efficacy. *Neuropsychopharmacology* 2007; **32**: 1715–1726.
 - 31 Callicott JH, Egan MF, Mattay VS, Bertolino A, Bone AD, Verchinski 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.
 - 32 Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 2005; **4**: Article17.
 - 33 Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet* 2013; **9**: e1003348.
 - 34 Yang S, Liu Y, Jiang N, Chen J, Leach L, Luo Z et al. Genome-wide eQTLs and heritability for gene expression traits in unrelated individuals. *BMC Genomics* 2014; **15**: 13.
 - 35 Wright FA, Sullivan PF, Brooks AI, Zou F, Sun W, Xia K et al. Heritability and genomics of gene expression in peripheral blood. *Nat Genet* 2014; **46**: 430–437.
 - 36 Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460**: 748–752.
 - 37 Maher BS. Polygenic scores in epidemiology: risk prediction, etiology, and clinical utility. *Curr Epidemiol Rep* 2015; **2**: 239–244.
 - 38 Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res* 2016; **44**: D877–D881.
 - 39 Pergola G, Di Carlo P, Andriola I, Gelao B, Torretta S, Attrotto MT et al. Combined effect of genetic variants in the GluN2B coding gene (GRIN2B) on prefrontal function during working memory performance. *Psychol Med* 2015; 1–16.
 - 40 Trabzuni D, Rytten R, Walker R, Smith C, Imran S, Ramasamy A et al. Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies. *J Neurochem* 2011; **119**: 275–282.
 - 41 Gelao B, Fazio L, Selvaggi P, Di Giorgio A, Taurisano P, Quarto T et al. DRD2 genotype predicts prefrontal activity during working memory after stimulation of D2 receptors with bromocriptine. *Psychopharmacology* 2014; **231**: 2361–2370.
 - 42 Rasetti R, Sambataro F, Chen Q, Callicott JH, Mattay VS, Weinberger DR. Altered cortical network dynamics: a potential intermediate phenotype for schizophrenia and association with ZNF804A. *Arch Gen Psychiatry* 2011; **68**: 1207–1217.
 - 43 Holland PaW R. Robust regression using iteratively reweighted least-squares. *Commun Stat Theory Methods* 1977; **6**: 813–827.
 - 44 Huber PJ. Robust regression: asymptotics, conjectures and Monte Carlo. *Ann Stat* 1973; **1**: 799–821.
 - 45 Huber PJ. *Robust Statistics*. John Wiley & Sons: New York, 1981.
 - 46 Benjamini Y, Hochberg Y. On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J Educ Behav Stat* 2000; **25**: 60–83.
 - 47 Kheradpour P, Kellis M. Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Res* 2014; **42**: 2976–2987.
 - 48 Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* 2013; **45**: 1238–1243.
 - 49 Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* 2015; **47**: 1091–1098.
 - 50 Hertzberg L, Katsel P, Roussos P, Haroutunian V, Domany E. Integration of gene expression and GWAS results supports involvement of calcium signaling in Schizophrenia. *Schizophr Res* 2015; **164**: 92–99.
 - 51 Colizzi M, Iyegbe C, Powell J, Ursini G, Porcelli A, Bonvino A et al. Interaction between functional genetic variation of DRD2 and Cannabis use on risk of psychosis. *Schizophr Bull* 2015; **41**: 1171–1182.
 - 52 Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kahler AK, Akterin S et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Gen* 2013; **45**: 1150–1159.
 - 53 Walavalkar NM, Gordon N, Williams DC Jr.. Unique features of the anti-parallel, heterodimeric coiled-coil interaction between methyl-cytosine binding domain 2 (MBD2) homologues and GATA zinc finger domain containing 2 A (GATAD2A/p66alpha). *J Biol Chem* 2013; **288**: 3419–3427.
 - 54 Terracciano A, Sanna S, Uda M, Deiana B, Usala G, Busonero F et al. Genome-wide association scan for five major dimensions of personality. *Mol Psychiatry* 2010; **15**: 647–656.
 - 55 Chung C, Talerico T, Seeman P. Schizophrenia hippocampus has elevated expression of chondrex glycoprotein gene. *Synapse* 2003; **50**: 29–34.
 - 56 Arion D, Unger T, Lewis DA, Levitt P, Mirmics K. Molecular evidence for increased expression of genes related to immune and chaperone function in the prefrontal cortex in schizophrenia. *Biol Psychiatry* 2007; **62**: 711–721.
 - 57 Zhao X, Tang R, Gao B, Shi Y, Zhou J, Guo S et al. Functional variants in the promoter region of Chitinase 3-like 1 (CHI3L1) and susceptibility to schizophrenia. *Am J Hum Genet* 2007; **80**: 12–18.
 - 58 Yang MS, Morris DW, Donohoe G, Kenny E, O'Dushlaine CT, Schwaiger S et al. Chitinase-3-like 1 (CHI3L1) gene and schizophrenia: genetic association and a potential functional mechanism. *Biol Psychiatry* 2008; **64**: 98–103.
 - 59 Blasi G, Napolitano F, Ursini G, Di Giorgio A, Caforio G, Taurisano P et al. Association of GSK-3beta genetic variation with GSK-3beta expression, prefrontal cortical thickness, prefrontal physiology, and schizophrenia. *Am J Psychiatry* 2013; **170**: 868–876.
 - 60 Kahnt T, Weber SC, Haker H, Robbins TW, Tobler PN. Dopamine D2-receptor blockade enhances decoding of prefrontal signals in humans. *J Neuroscience* 2015; **35**: 4104–4111.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017

Supplementary Information accompanies the paper on the Translational Psychiatry website (<http://www.nature.com/tp>)